Title
A Convenient Partnership: The Ribosome and the Nascent Chain Interact to Modulate Protein Synthesis and Folding

Permalink
https://escholarship.org/uc/item/0rh988bb

Author
Goldman, Daniel Hersel

Publication Date
2015

Peer reviewed|Thesis/dissertation
A Convenient Partnership: The Ribosome and the Nascent Chain Interact to Modulate Protein Synthesis and Folding

by

Daniel Hershel Goldman

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

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos Bustamante, Chair
Professor Jamie Doudna Cate
Associate Professor Andreas Martin

Spring 2015
A Convenient Partnership: The Ribosome and the Nascent Chain Interact to Modulate Protein Synthesis and Folding

Copyright 2015
by
Daniel Hershel Goldman
Abstract

A Convenient Partnership: The Ribosome and the Nascent Chain Interact to Modulate Protein Synthesis and Folding

by

Daniel Hershel Goldman

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Carlos Bustamante, Chair

During translation, the ribosome reads the genetic code of the messenger RNA, adding one amino acid at a time to the nascent protein. The sequence of the polypeptide determines the three dimensional structure of the natively folded protein, and thus encodes its biological activity. Because folding rates are often fast compared to translation, many proteins likely undergo folding transitions during synthesis, with folding potentially modulated by the sequential appearance of the polypeptide and the chemical environment of the ribosome. Traditional experimental approaches to study protein folding employ chemical, temperature or pH-induced denaturation and would irreversibly destroy the ribosome; thus, such techniques cannot be used to probe folding on the ribosome. In this work, we implement a novel single-molecule optical tweezers assay to probe folding transitions of the nascent polypeptide as it emerges from the ribosome. We demonstrate that the ribosome can modulate the kinetics of folding through interactions between the nascent chain and the charged ribosomal surface. Additionally, the ribosome can prevent misfolding of incompletely synthesized protein fragments. These observations point to a chaperone-like role for the ribosome in guiding the nascent protein to its native state.

In addition to interacting with the exterior of the ribosome, some nascent chain sequences can form specific contacts with the ribosome exit tunnel. These contacts lead to conformational changes of the ribosome, and reduced translation rates. The Secretion Monitor protein stalls the ribosome upon translation of a 17 amino acid motif. Arrest release requires targeting of the stalled ribosome-nascent chain complex to the translocon; thus, it has been hypothesized that arrest is released by a mechanical pulling force generated as the polypeptide is translocated across the membrane. By applying force to the nascent polypeptide of stalled ribosomes, we demonstrate that translation arrest at SecM is released by mechanical force. Additionally, we show that the force needed to release stalling can be generated by a protein folding in close proximity to the ribosome tunnel exit. Our results demonstrate the feasibility of a feedback mechanism, whereby a folding protein can modulate its synthesis through the generation of force. More generally, since the nascent polypeptide in the cell
can undergo a number of potentially force-generating events—chaperone binding, protein or ligand binding, translocation and membrane insertion—force applied to the nascent chain may be an important modulator of protein synthesis.
This work is dedicated to my grandparents, Alta, Phil, Joan and Herb, my parents, Marilyn and Robert, and my sister Rebecca, all of whom have loved and supported me unconditionally.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Translating the Genetic Code: Structure and Function of the Ribosome</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Stop and Go: Kinetic Control Over Protein Synthesis</td>
<td>5</td>
</tr>
<tr>
<td>1.3 From Sequence to Structure: Folding of the Nascent Polypeptide</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Probing Biological Processes at the Single-Molecule Level Using Optical Tweezers</td>
<td>9</td>
</tr>
<tr>
<td>2 The Ribosome Modulates Nascent Protein Folding</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Summary</td>
<td>12</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>12</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>20</td>
</tr>
<tr>
<td>2.5 Methods</td>
<td>21</td>
</tr>
<tr>
<td>3 Mechanical Force Releases Nascent Chain-Mediated Ribosome Arrest In Vitro and In Vivo</td>
<td>40</td>
</tr>
<tr>
<td>3.1 Summary</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>40</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>41</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>45</td>
</tr>
<tr>
<td>3.5 Methods</td>
<td>48</td>
</tr>
<tr>
<td>4 Development of an Optimized In Vitro Translation System for Single-Molecule Experiments</td>
<td>66</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>66</td>
</tr>
<tr>
<td>4.2 Results</td>
<td>67</td>
</tr>
<tr>
<td>4.3 Discussion</td>
<td>78</td>
</tr>
<tr>
<td>5 Conclusion</td>
<td>80</td>
</tr>
</tbody>
</table>
List of Figures

1.1 The central dogma of molecular biology .................................................. 1
1.2 The crystal structure of the 70S ribosome in complex with tRNAs and mRNA . 2
1.3 The translation cycle, showing the ribosome, mRNA, tRNAs and auxiliary protein factors ................................................................. 3
1.4 The translation elongation cycle ................................................................. 4
1.5 The chemical mechanism of peptidyl transfer on the ribosome .................. 4
1.6 Schematic of SecM-mediated ribosome stalling based on cryo-EM structure ... 7
1.7 SecM-mediated ribosome stalling regulates expression of SecA .................. 8
1.8 Proteins fold along a “funnel-shaped” energy landscape .............................. 9
1.9 De novo folding is distinct from in vitro refolding of full-length polypeptides . 10
1.10 A micron-sized particle can be stably trapped by a focused laser beam ....... 11

2.1 Mechanical unfolding of ribosome-nascent chain complexes using optical tweezers 14
2.2 Ribosome-bound T4 lysozyme folds more slowly than the free protein ........ 16
2.3 The ribosome decelerates the final, irreversible transition of the nascent polypeptide from ‘I’ to “N” ................................................................. 18
2.4 The ribosome prevents misfolding of incompletely-synthesized polypeptides ... 19
2.5 Modification of ribosomal protein L17 for tethering in optical tweezers experiments 29
2.6 Attachment of incomplete polypeptides to DNA “handles” for tethering in optical tweezers experiments ............................................................... 30
2.7 Force-extension curves of DNA “handles” ................................................... 31
2.8 Single-molecule puromycin release experiment ......................................... 32
2.9 Force-dependent lifetimes of free and ribosome-bound T4 lysozyme .......... 33
2.10 Force-extension curves for free and ribosome-bound T4 lysozyme ............... 34
2.11 Force-extension curves for free and ribosome-bound T4 lysozyme ............... 35
2.12 Extension histograms for force-clamp measurements ............................... 36
2.13 Extension changes from BHMM analysis ................................................. 37
2.14 Force-extension curve of a ribosome-bound incomplete T4 lysozyme polypeptide 38
2.15 Solubility of incomplete T4 lysozyme polypeptides ................................ 39

3.1 A direct applied force catalyzes release of SecM-mediated arrest ................ 42
3.2 Nascent protein folding near the ribosome tunnel exit can rescue SecM-mediated stalling .......................... 44
3.3 Top7 refolds against an applied mechanical load .......................... 46
3.4 Kinetic model for folding-induced release of stalled ribosomes ....................... 47
3.5 Primary sequence diagram for optical tweezers experiments ....................... 54
3.6 Preparation of SecM-stalled ribosomes ........................................... 55
3.7 Arrest release of SecM-stalled ribosomes in bulk after prolonged time .......... 56
3.8 Schematic for the optical tweezers stall-release experiment showing A, P and E sites of the ribosome ........................................ 57
3.9 Control stall-release experiments in the absence of puromycin, EF-G and GTP .... 58
3.10 Alternative method for calculating force-catalyzed arrest release rates takes into account non-specific tether rupture ............................................ 60
3.11 Assembly of oligonucleotides encoding linker sequences ........................ 61
3.12 The ribosome does not influence folding of nascent Top7 ........................ 62
3.13 Colonies sequenced irrespective of fluorescence recovered the entire linker length distribution ........................................ 63
3.14 Experimental geometry for pulling experiments with single Top7 molecules ... 63
3.15 The force-dependent rates used in the folding-SecM release kinetic scheme are plotted together ........................................ 64
3.16 The effective folding-SecM release rate as a function of force .................... 65

4.1 Translation rates in single-molecule experiments are slow compared to in vivo rates ....... 68
4.2 Translation rates in the PURE system are comparable to those obtained in single-molecule experiments ........................................ 69
4.3 The translation substeps starting with initial selection and ending with peptidyl transfer ........................................ 70
4.4 A schematic for quench flow experiments to monitor the progress of a translation reaction ........................................ 70
4.5 Rates of dipeptide formation display a Michaelis-Menten dependence on ternary complex concentration ........................................ 71
4.6 A scheme describing a tripeptide formation assay to measure translocation rates ....... 72
4.7 Results of the tripeptide assay to measure translocation rates ....................... 73
4.8 Fitting the tripeptide assay data to extract kinetic rates ................................ 73
4.9 The magnesium-dependence of tripeptide formation ................................ 74
4.10 The magnesium-dependence of dipeptide formation ................................ 74
4.11 Experimental geometry for optical tweezers experiments to measure translation rates through an mRNA hairpin ........................................ 75
4.12 Example traces of single ribosomes translating through an mRNA hairpin at 1.3 mM free magnesium ........................................ 76
4.13 Coding sequence and kinetic scheme for the heptapeptide experiment .............. 77
4.14 Products from the heptapeptide experiment are separated into distinct peaks by HPLC ........................................ 77
4.15 Data and fits for the time-dependent concentrations of products in the heptapeptide experiment .................................................. 78
4.16 Rate variation in the heptapeptide experiment is not explained by differential tRNA concentration or wobble base-pairing ............................ 79
Acknowledgments

I would like to acknowledge all of the people who have supported me on the path towards my Ph.D. I have been fortunate to interact with many great scientists and friends, making for a rich experience. Firstly I would like to thank my advisor, Dr. Carlos Bustamante, for his unwavering support throughout my time at Berkeley and for fostering an unparalleled environment for doing science. Our discussions left me feeling encouraged and excited about my work, and were often culturally enriching as well. I am grateful for the numerous opportunities he afforded me, through projects in the laboratory, travel, conferences, introductions and social engagements.

Dr. Christian Kaiser—a postdoc in the lab—was my mentor during the first few years of my Ph.D. I could not have asked for a better mentor and colleague. Christian is an exceptional scientist, and I am grateful to him for teaching me a wide variety of lab techniques, advising me on my presentations and writing, and for setting an example of curiosity, hard work and scholarship. His approach to science left a lasting impression that will remain with me throughout my career.

Dr. Ignacio Tinoco, Jr. has been a supportive presence throughout my Ph.D. He has kept me on track by making sure that I ask good questions. During our work together, I have been inspired by his dedication to science, and challenged by our engaging discussions.

I would like to acknowledge all of the others with whom I collaborated directly on scientific projects in the lab: postdoc Dr. Maurizio Righini, graduate students Lisa Alexander and Varsha Desai and undergraduates Apurva Shah, Anthony Milin and Karim Merchant for affording me the chance to work together in a supportive, pleasant atmosphere. I am grateful to all of my colleagues in the Bustamante and Tinoco labs for fostering an intellectually challenging and warm environment.

I was fortunate to spend several months working in the lab of Dr. Måns Ehrenberg in Uppsala, Sweden. This was a satisfying and memorable experience, and I am grateful to Dr. Ehrenberg for welcoming me into his lab for these months. My work with him endowed me with a fresh perspective on my research and enabled me to learn new skills, to make new personal connections and to immerse myself in a different culture. In the lab, Anneli Borg was extremely generous with her time and patient in teaching me the techniques of the lab.

I thank my qualifying exam committee members—Dr. Jamie Cate, Dr. Andreas Martin, Dr. David Wemmer and Dr. Phillip Geissler—for guiding me during my qualifying exam and appreciate the input of the former two as members of my dissertation committee. In particular, I thank Dr. Cate for occasional discussions and for the opportunity to assist in teaching his biophysical chemistry course. In addition, I am indebted to Dr. Geissler for imparting me with a profound appreciation of statistical mechanics through his course during my first year.

Finally, I thank all of my friends and family for their support during my studies. My classmates Evan Wang, Claudia Avalos, Maya Sen and Manchuta Dangkulwanich, have been great friends, as have Laura Grau and Maurizio Righini, who have enhanced my life in Berkeley with adventure and spontaneity. My parents—Robert and Marilyn—have been
a source of continuous strength and love throughout my life, have supported me through difficult times and have instilled within me a passion for learning. They are the best parents I could possibly hope for. My sister Rebecca has been my best friend and confidant, and I am thankful to have someone on this earth with whom I share such high DNA sequence similarity. All of my grandparents—Joan, Herb, Alta and Philip—have unconditionally supported and loved me, and fueled my ambitions in life.
Chapter 1

Introduction

1.1 Translating the Genetic Code: Structure and Function of the Ribosome

The ribosome is a large and complex molecular machine found in the cells of all living organisms. In the final step of the central dogma of molecular biology, the ribosome reads the genetic code of the messenger RNA (mRNA) template and “translates” the code to synthesize proteins [1](Figure 1.1). The result of this process is a linear chain of amino acids which folds to a specific three dimensional structure to carry out its biological function [2].

Figure 1.1: The central dogma of molecular biology. The ribosome translates the genetic code of the mRNA and catalyzes peptide bond formation to synthesize proteins.

The ribosome is a ribonucleoprotein complex comprising two subunits. In bacteria, the subunits sediment as 30S (small subunit) and 50S (large subunit), and together form a 70S complex of roughly $2.5 \times 10^6$ Daltons [3](Figure 1.2). While the eukaryotic ribosome is considerably larger and more complex, many core components of the structure are conserved, indicating that the basic mechanics of translation are similar [4]. The overall architecture of the ribosome resembles an RNA scaffold around which folded proteins dock [5]. At the interface between the large and small subunits there are three binding sites for tRNAs: the aminoacyl-tRNA (A) site, the peptidyl-tRNA (P) site, and the exit (E) site. Interestingly, structural studies have demonstrated that the ribosome is a ribozyme; its core catalytic function is carried out by RNA [6].
Translation can be divided into three stages: initiation, elongation and termination [3](Figure 1.3). In the initiation stage, the 30S subunit binds to a conserved motif on the mRNA—called the Shine-Delgarno sequence—located just upstream of the start codon. Ribosome binding to the Shine-Delgarno sequence is mediated by a complementary sequence located at the 3’-end of the 16S ribosomal RNA (rRNA) [7]. With the help of initiation factors 1, 2 and 3 (IF1-3), the initiator fMet-tRNA
\[fMet\] is positioned at the start codon in the P site, and release of IF3 allows for 50S subunit joining, resulting in an elongation-competent 70S complex [3].

During the elongation cycle, the ribosome decodes successive three-nucleotide repeats, or codons, of the mRNA and catalyzes peptide bond formation, adding the specified amino acids to the nascent polypeptide chain [8](Figure 1.4). The cycle begins when Elongation Factor Thermo-unstable (EF-Tu), a GTPase, delivers an aminocyl-tRNA (aa-tRNA) to the
A site of the ribosome. In initial selection, the ternary complex of aa-tRNA/EF-Tu/GTP binds to the ribosome. Although there are 41 distinct anti-codons among the tRNA species in *E. coli* [9], the correct (or cognate) tRNA is stabilized by complementary base pairing between the codon of the mRNA and the anti-codon of the tRNA. Hydrolysis of GTP by EF-Tu provides the thermodynamic driving force for the proofreading step, in which the aa-tRNA can be irreversibly discarded from the ribosome [10]. Together, initial selection and proofreading ensure that the error frequency of amino acid incorporation is $\approx 10^{-5}$ [11].

Once the proper tRNA is accommodated into the A site, the ribosome catalyzes peptidyl transfer, leading to nucleophilic attack of the $\alpha$-amino group of the aminoacyl-tRNA on the carbonyl carbon of the peptidyl-tRNA [12](Figure 1.5). Although the ribosome enhances the rate of this reaction approximately $10^5$ to $10^7$-fold [3, 12], catalysis is thought to occur primarily by entropic means, through positioning of the reactants at the active site [13].

Following peptidyl transfer, Elongation Factor G (EF-G) binds to the ribosome. Hydrolysis of GTP by EF-G leads to translocation—the movement of the tRNAs and mRNA with respect to the ribosome by one codon. After EF-G is released, the elongation cycle is reset and is repeated until the ribosome reaches a stop codon.
**Figure 1.4:** The translation elongation cycle [8].

**Figure 1.5:** The chemical mechanism of peptidyl transfer on the ribosome [12].
When the ribosome recognizes a stop codon, the polypeptide is cleaved from the tRNA by either Release Factor 1 or Release Factor 2 (RF1 or RF2). Release Factor 3 (RF3) catalyzes dissociation of RF1 or RF2 from the ribosome. Ribosome Recycling Factor (RRF) together with EF-G dissociate the 70S complex into its constituent 30S and 50S subunits, liberating them for the next round of translation [3].

Because many clinical antibiotics target the bacterial ribosome [14], the mechanism of translation and structure of the translation machinery are of biomedical interest. Examples of ribosome targeting drugs include aminoglycosides, which cause misreading of the genetic code, macrolides, which block the progression of the nascent polypeptide through the ribosome exit tunnel, and tetracyclines, which prevent accommodation of aminoacyl-tRNA. Structures of the ribosome in complex with antibiotics have elucidated binding sites and mechanisms of action of these drugs [15]. Emerging antibiotic resistance has fueled a renewed interest in understanding the mechanisms of protein synthesis and identifying new therapeutic strategies involving the translation machinery.

1.2 Stop and Go: Kinetic Control Over Protein Synthesis

The translation cycle comprises a series of chemical reactions, as shown in figure 1.3. While \textit{in vivo} protein synthesis rates in bacteria are \( \sim 10 \) amino acids per second [16], the rate of some substeps \textit{in vitro} depends on the solution conditions. For example, the kinetics of peptidyl transfer vary with pH [17], while that of tRNA selection and EF-G-catalyzed translocation depends on \([\text{Mg}^{2+}]\) [18]. The \([\text{Mg}^{2+}]\)-dependence of tRNA selection has been exploited to study the accuracy-efficiency trade-off of translation [19]. Increasing \([\text{Mg}^{2+}]\) results in higher efficiency (as measured by \( \frac{k_{\text{cat}}}{K_m} \)) at the expense of accuracy in cognate tRNA selection.

One major challenge of single-molecule translation experiments is that rates are slow compared to those measured \textit{in vivo} and \textit{in vitro} ensemble experiments [16, 20, 21]. This issue raises questions about the relevance of some experiments, and introduces technical difficulties related to the detection of small changes in extension per unit time in optical tweezers experiments. Development of a speed-optimized \textit{in vitro} translation system for single-molecule experiments is the focus of chapter 4.

The speed that the ribosome moves along the mRNA also varies in a context-specific manner [22]. Kinetic variations arise from differences in the abundance of tRNA isoacceptors and the nature of codon-anticodon base pairing [23], and mRNA secondary structure [20]. Kinetic control over translation is important for gene expression and can affect the folding and activity of the nascent protein [24]. For example, synonymous mutations that do not change the amino acid sequence of a protein can affect the cadence of ribosome transit and thus change the folding pathway of a protein [25, 26]. The relationship between elongation kinetics and folding of the nascent polypeptide will be discussed in more detail in section 1.3.
CHAPTER 1. INTRODUCTION

The ribosome exit tunnel can also interact with specific nascent polypeptide sequences—termed ribosome arrest peptides (RAPs)—to reduce translation rates [27]. In these cases, contacts between amino acid residues in the nascent chain and the tunnel wall lead to allosteric changes, which reduce the activity of the ribosome. The *E. coli* Secretion Monitor protein (SecM) contains a 17-amino acid arrest motif (SecM17) which stalls ribosomes upon its synthesis [28]. A recent cryo-EM structure of the SecM-stalled ribosome has documented unique contacts between the nascent chain and the exit tunnel, as well as conformational changes of the P site tRNA, elucidating how peptidyl transfer is inhibited at a structural level [29](Figure 1.6).

In the cell, SecM regulates expression levels of SecA, the motor component of the bacterial translocon [27]. Stalling at the arrest motif disrupts a hairpin loop in the downstream mRNA, releasing the Shine-Delgarno (SD) site for the SecA message and allowing a second ribosome to bind and translate the message (Figure 1.7). Thus, robust stalling upregulates SecA expression, while weak stalling downregulates its expression. Arrest is released by targeting of the stalled RNC to the translocon. Hence, this system senses the availability of the secretion machinery, feeding back this information to regulate its expression. Because SecA interacts with the nascent polypeptide to thread it through the translocon pore [30], it has been proposed that stalling is relieved by a mechanical “pulling” of the nascent chain by the translocon [31].

Given that a number of additional RAPs have recently been discovered [27, 33, 34], some of which are connected to potentially force-generating processes [35], we wondered whether force on the nascent polypeptide might have a general role in regulating translation. We set out to test the hypothesis that mechanical force applied to the nascent polypeptide can release SecM-mediated ribosome arrest. This work is the subject of chapter 3.

1.3 From Sequence to Structure: Folding of the Nascent Polypeptide

Most proteins must fold to a defined three-dimensional structure, termed the “native state”, in order to carry out their specified biological functions. Because proteins carry out many of the essential functions of a living organism, and folding defects can result in a variety of human diseases—including some neurodegenerative disorders, cystic fibrosis and some cancers [36]—protein folding is an important area of biomedical research. From experimental and theoretical work focusing on fully synthesized proteins, many principles guiding folding have emerged. One of the central questions in the field has been: Among the astronomically large number of possible conformations, how does a polypeptide chain reach its native state on a physiologically relevant timescale? The emerging consensus is that proteins fold along “funnel-shaped” energy landscapes (Figure 1.8). On-pathway folding intermediates are energetically downhill from the unfolded state, while the native state is a global energy minimum. The folded state is stabilized hydrogen bonding, van der Walls interactions, pre-
Figure 1.6: Schematic of SecM-mediated ribosome stalling based on cryo-EM structure. [29]. The aminoacyl-tRNA is shown in orange, peptidyl-tRNA in green (SecM-stalled position) and grey (canonical position), and rRNA bases in blue (SecM-stalled position) and grey (canonical position). When the ribosome is in the arrested state, a critical arginine in the nascent chain interacts with base A2602 of the 23S rRNA, leading to a cascade of rRNA conformational changes and resulting ultimately in the repositioning of the P site tRNA.
Figure 1.7: SecM-mediated ribosome stalling regulates expression of SecA. [32]. The bicistronic SecM-SecA message is shown. When there is no ribosome on the message, the Shine-Delgarno (SD) sequence for SecA translation is sequestered in a hairpin and cannot be accessed. Thus there is no translation of SecA. When a ribosome translates the SecM message and stalls at the arrest site, the hairpin is disrupted and the SD sequence is liberated, allowing translation of SecA. Thus, robust SecM stalling upregulates SecA.

ferred peptide backbone angles, electrostatic interactions and hydrophobic interactions [37]. In this way, the polypeptide chain can reach the functional native state in a timely fashion.

The majority of work on protein folding is based on the study of full-length polypeptides in solution. However, in the cell, proteins are synthesized by the ribosome and may begin to fold during translation. This is supported by work showing that a nascent protein domain that has emerged from the ribosome exit tunnel—but which has not yet been released—can fold to its native structure [38]. Additionally, while protein folding rates can range from microseconds to seconds [39, 40], translation rates in bacteria are \(\sim 10\) amino acids per second [16]. These observations indicate that many proteins likely begin to fold before synthesis is complete.

The conditions for de novo folding—the folding of newly-synthesized nascent polypeptides—are thus distinct from the conditions of most folding experiments, primarily in three respects: 1) The protein may begin to fold before all of its sequence is available, 2) Folding will take place in close proximity to the ribosome surface, and 3) The ribosome-nascent chain complex is a platform for binding of molecular chaperones, which promote proper protein folding (Figure 1.9). These differences may lead to changes in the kinetics or thermodynamics of folding, altered folding trajectories or alternative conformations of nascent polypeptides, relative to their full-length counterparts.

Due to experimental limitations, relatively little is known about how proteins fold during synthesis by the ribosome. Traditional folding experiments rely on chemical denaturants, or temperature or pH jumps to probe folding. These approaches, however, are not amenable to de novo folding studies, as such harsh conditions would irreversibly destroy the ribosome. We developed a single-molecule optical tweezers assay to probe folding on the ribosome. Using mechanical force as a denaturant, we are able to selectively unfold the nascent protein, while leaving the ribosome intact. We show that the ribosome can modulate the kinetics of folding and prevent misfolding of incompletely synthesized polypeptides, pointing to a role
CHAPTER 1. INTRODUCTION

Figure 1.8: Proteins fold along a “funnel-shaped” energy landscape. [37]. Unfolded conformations are relatively high-energy, while the native state represents a global energy minimum. The polypeptide chain is guided towards the native state by following an energetically downhill trajectory.

for the ribosome in promoting proper folding to the native state. This work is the subject of chapter 2.

1.4 Probing Biological Processes at the Single-Molecule Level Using Optical Tweezers

In 1970, Arthur Ashkin demonstrated that micron-sized particles can be stably trapped in a focused beam of laser light [41]. Photons incident on a dielectric particle which is larger than the wavelength of incident light experience a change in momentum and impart a force on the particle [42], according to Newton’s Third Law. The resulting force can be divided into a scattering force and gradient force (Figure 1.10). Scattering of photons by the bead results in a net “pushing” force in the direction of propagation of the light. If the laser beam is tightly focused, the particle, slightly displaced from the focus of the beam by scattering, will refract photons enough to generate a restoring force in the opposite direction, counteracting
CHAPTER 1. INTRODUCTION

Figure 1.9: **De novo folding is distinct from in vitro refolding of full-length polypeptides.** [32]. Left: The nascent polypeptide can fold while still bound to the ribosome. Folding occurs before the entire sequence has emerged from the confines of the exit tunnel, and in close proximity to the ribosome surface. Right: The ribosome-nascent chain complex is a platform for binding of chaperones, which promote proper protein folding.

the “pushing” by the scattered photons. The spatial gradient exists in three dimensions, such that a similar restoring force will act on the bead if it is displaced from the equilibrium position along either of the axes orthogonal to the axis of propagation.

Optical tweezers instruments often rely on trapping of microspheres. By tethering the molecule(s) of interest between two microspheres using specific biochemical attachments, it is possible to apply and measure pico-Newton-scale forces and to measure nanometer-scale displacements [42]. This approach has been used to probe fundamental problems in biology at a mechanistic level, including the mechanical properties of nucleic acids, folding and unfolding of nucleic acids and proteins, and the operating principles of molecular motors and enzymes [43]. The application of force allows one to dissect how molecular machines convert chemical energy into mechanical work [44]. Single-molecule approaches are often advantageous over ensemble experiments because they allow observation of rare or alternate pathways, which may otherwise be obscured by averaging. For these reasons, optical tweezers are an excellent experimental tool and have yielded many fundamental biological insights. In this work, we use optical tweezers to probe the mechanics of nascent chain-mediated ribosome stalling, as well as the folding of nascent polypeptides on the ribosome.
CHAPTER 1. INTRODUCTION

For a nisotropic scatter, the resulting forces cancel in all but the forward direction, and an effective scattering cross section can be calculated for the object. For most conventional situations, the scattering force dominates. However, near the focus, the second component of the optical force, the gradient force, must be considered. The gradient force, as the name suggests, arises from the fact that a dipole in an inhomogeneous electric field experiences a force in the direction.

In an optical trap, the laser induces a dipole in the dielectric particle, and it is the interaction of these dipoles with the inhomogeneous electric at the focus that gives rise to the gradient trapping force. The gradient force is proportional both to the polarizability of the particle and to the gradient of the electric field intensity.

Figure 1.10: **A micron-sized particle can be stably trapped by a focused laser beam.** [42]. a) A particle illuminated with an intensity gradient experiences a force in the direction of increasing intensity. Black lines represent the paths of incident photons. Grey lines represent the resultant force on the particle. In this case, a greater intensity of incident photons on the right side of the particle results in a force in that direction. b) When the particle is “pushed” out of the focus of the beam by scattered photons, it is “pulled” back to the focus by the gradient force.
Chapter 2

The Ribosome Modulates Nascent Protein Folding


2.1 Summary

Proteins are synthesized by the ribosome and generally must fold to become functionally active. Although it is commonly assumed that the ribosome affects the folding process, this idea has been extremely difficult to demonstrate. We have developed an experimental system to investigate the folding of single ribosome-bound stalled nascent polypeptides with optical tweezers. In T4 lysozyme, synthesized in a reconstituted in vitro translation system, the ribosome slows the formation of stable tertiary interactions and the attainment of the native state relative to the free protein. Incomplete T4 lysozyme polypeptides misfold and aggregate when free in solution, but they remain folding-competent near the ribosomal surface. Altogether, our results suggest that the ribosome not only decodes the genetic information and synthesizes polypeptides, but also promotes efficient de novo attainment of the native state.

2.2 Introduction

Proteins can spontaneously fold into their native structures under appropriate conditions [2]. In vitro, some small proteins and single domains attain their native structures within microseconds [39], whereas topologically complex and larger proteins may require many seconds to fold [40] and often populate folding intermediates along the way [45, 46]. In vivo, however, folding is not necessarily limited to full-length proteins or domains. Proteins can begin to
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

fold before they are fully synthesized and while still bound to the ribosome [38, 47–52]. Moreover, during protein synthesis on the ribosome [53], elongation rates are regulated by factors including tRNA abundance [54], codon order [55], and mRNA secondary structure [20]. Decreasing these rates locally [24, 56] or globally [57] can affect the folding efficiency of newly synthesized proteins. The complete synthesis of even small proteins (100 amino acids or less) requires at least several seconds at a maximum rate of ~20 amino acids per second in Escherichia coli [58], giving nascent-chain segments sufficient time to conformationally equilibrate in the environment of the ribosome, perhaps adopting structures that are distinctly different from the native protein fold [48] [59]. However, the observation of folding transitions in ribosome-bound nascent proteins has not been possible, and a detailed analysis has been limited to computational approaches [60, 61].

We have developed an experimental system to directly probe the folding of single ribosome-bound nascent chains [62, 63] by subjecting them to force using optical tweezers (Figure 2.1, Figure 2.5, Figure 2.6, Figure 2.7). The force is applied between the nascent chain and the large ribosomal subunit. Because force acts locally [64, 65], we can selectively perturb the stability of ribosome-bound nascent polypeptides without disrupting the structural integrity of the ribosome. We studied a cysteine-free version of T4 lysozyme [66], a monomeric cytosolic protein composed of two globular regions, or subdomains (Figure 2.1, C and D). T4 lysozyme folding has been studied in ensemble [67–69] and single-molecule [65] experiments. The native fold requires interactions between the N- and C-terminal sequences whose synthesis is separated in time during vectorial translation by the ribosome.

2.3 Results

Using a reconstituted in vitro translation system supplemented with E. coli ribosomes, we first translated the protein with an unstructured C-terminal extension of 41 amino acids such that the entire T4 lysozyme sequence emerges from the narrow ribosomal exit tunnel [70]. This experimental design allows us to study the folding dynamics of the full-length protein on the ribosome. When we stretched the molecule by continuously increasing the tension applied across the nascent chain (force ramp), we observed single rips in the resulting force-extension traces, representing cooperative unfolding events (Figure 2.1 E). Puromycin-release experiments confirmed that these signals originated from ribosome-bound nascent proteins (Figure 2.8).

The mechanical unfolding pathways of free and ribosome-bound full-length T4 lysozyme are very similar: The latter unfolds at a mean force \( F_{Unf} \) of 17.0 ± 2.0 pN (N = 125 unfolding events) at a pulling speed of 100 nm/s (Figure 2.1 G). Using the wormlike chain (WLC) model [71], we calculated a contour length increase upon unfolding (\( \Delta L_C \)) of 59.9±2.1 nm, consistent with the value expected for full-length T4 lysozyme (164 amino acids × 0.36 nm per amino acid  0.9 nm corresponding to the folded end-to-end distance = 58.1 nm). Experiments with the free protein in the absence of the ribosome (Figure 2.1, F and H) revealed similar unfolding characteristics (\( F_{Unf} = 17.2 \pm 1.8 \) pN, \( \Delta L_C = 60.1 \pm 0.9 \) nm,
Figure 2.1: (A) Schematic of the molecular assembly for optical-tweezers experiments. Force can be applied to the nascent polypeptide by moving the optical trap relative to the pipette. (B) Surface representation of the ribosome [Protein Data Bank identification numbers (PDB IDs): 2aw4 and 2avy] showing the opening of the ribosomal exit tunnel in the large subunit and the location of ribosomal protein L17, serving as the attachment site in the optical-tweezers experiments. Ribosomal RNA, pink; ribosomal proteins, white; L17, blue. The small subunit is shown in a semitransparent rendering. (C) Cartoon diagram of T4 lysozyme (PDB ID: 4lzm). The N-terminal subdomain (light orange) is composed of residues 13 to 59; the C-terminal subdomain (dark orange) comprises residues 1 to 12 and 60 to 164. (D) Primary structure diagram of the protein construct translated for optical-tweezers experiments (32). Red spheres indicate stalling positions along the sequence. f.l., full length. (E and F) Force-extension traces of T4 lysozyme unfolding (red) and refolding (blue) on the ribosome (41 amino acid linker) (E) and free in solution (F). The protein unfolds in one cooperative transition near 17 pN. (G and H) Rupture-force histograms (gray bars) for unfolding of the ribosome-bound (G) and free (H) protein. Red lines, rupture force distribution reconstructed from the force-dependent lifetimes (fig. S5).
N=453), confirming that the entire protein is able to emerge from the ribosomal tunnel by means of the 41 amino acid linker [70]. Analysis of the unfolding force distributions [72] of free and ribosome-bound polypeptides (Figure 2.1, G and H) also reveals similar distances to the transition state ($\Delta f^t_{\text{free}} = 2.3 \pm 0.5 \text{ nm}$, $\Delta f^t_{\text{ribosome-bound}} = 2.0 \pm 0.2 \text{ nm}$) and native-state lifetimes (Figure 2.9). Thus, the ribosome-bound protein folds to the same native structure as the free protein and unfolds through the same pathway in the pulling experiments.

Whereas the unfolding transitions of free and ribosome-bound T4 lysozyme are indistinguishable in our experiments, the refolding exhibits marked differences: During repeated cycles of pulling and relaxation, the free protein virtually always refolds, but the ribosome-bound protein refolds only 28% of the time and lacks a well-defined refolding transition in the force-ramp measurements (Figure 2.10). We carried out “force-clamp” experiments (Figure 2.11) to better resolve the folding transition. From measurements of the refolding time, $\tau_{\text{refold}}$ (Figure 2.6 A), we estimated apparent refolding rates at 3.6 pN (Figure 2.6 B). The free protein folds rapidly with an apparent rate ($k_{\text{app}}$) of 5.4 s$^{-1}$ [95% confidence interval (CI): 4.4 s$^{-1}$, 6.5 s$^{-1}$]. Remarkably, the folding rate of the ribosome-bound protein (with a 41 amino acid C-terminal linker) is 0.012 s$^{-1}$ (95% CI: 0.004 s$^{-1}$, 0.024 s$^{-1}$), more than two orders of magnitude slower than the free protein.

The 41 amino acid linker is long enough to allow the entire protein to emerge from the exit tunnel but restricts the folding environment to the immediate vicinity of the ribosomal polypeptide exit site. Increasing the linker length by 19 amino acids provides $\sim 2.1 \text{ nm}$ at 3.6 pN (calculated based on a WLC model) of additional separation from the ribosomal surface and increases the folding rate: This construct, harboring a 60 amino acid linker, folds with $k_{\text{app}} = 0.24 \text{ s}^{-1}$ (95% CI: 0.21 s$^{-1}$, 0.28 s$^{-1}$) (Figure 2.2 B), significantly faster than the rate observed with the shorter linker, but still more slowly than the free protein (Figure 2.2 B, also compare Figure 2.2 C with Figure 2.2 A). We used this construct to characterize the folding of the ribosome-bound protein in more detail.

Given the high negative charge density of the ribosomal RNA (rRNA) moiety, the observed deceleration in folding is likely mediated by electrostatic interactions of the ribosomal surface with charged residues in the nascent chain [73]. Increasing the potassium chloride concentration from 150 to 500 mM, which results in more effective screening of electrostatic interactions, increased the folding rate of the ribosome-bound protein, whereas the folding rates of the free protein were less sensitive to salt (Figure 2.2 D). Thus, the effect of the ribosome on the folding of the nascent polypeptide is mediated, at least in part, by electrostatic interactions.

Close inspection of the force clamp traces revealed that, before folding to the native state (N), the protein transiently and reversibly samples an intermediate (I) from the unfolded state (U) (Figure 2.2, A and C), exhibiting bistability, or “hopping.” In almost all traces examined, folding to the native state then proceeds from the intermediate state (Figure 2.2 A, arrowhead), indicating that the latter is an on-pathway state and, therefore, obligatory for productive attainment of the native state. The I-to-N transition is practically irreversible at the refolding force; thus, the folding pathway can be written as
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

Figure 2.2: (A) Example extension-versus-time trace of T4 lysozyme refolding at constant force (5 pN). Before transitioning to the native state (N) at ~0.5 s, the protein visits a folding intermediate (I). Enlargement of the first second of the trace (red dashed box) reveals that the intermediate is visited immediately before folding to the native state (arrowhead), demonstrating that the intermediate is on-pathway. U, unfolded state. (B) Apparent refolding rates for ribosome-bound T4 lysozyme with 41 amino acid (+41) and 60 amino acid (+60) linkers and for the free protein (free). Error bars: 95% CIs. (C) Example trace of ribosome-bound protein (60 amino acid linker) refolding at 5 pN. (D) Apparent refolding rates of free protein (blue circles) and +60 linkers (red triangles) at 150 mM (filled symbols) and 500 mM (open symbols) KCl. Increasing the salt concentration mitigates the effect of the ribosome on the refolding rate. (E) Population of the folding intermediate before refolding. The unfolded and intermediate states are equally populated (P = 0.5) at a force of ~3.6 pN, both in the free and the ribosome-bound protein.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

\[ U \leftrightarrow I \rightarrow N \]

From the extension changes in the force-clamp traces, we estimate that between 96 and 108 residues participate in the formation of the folding intermediate. These dimensions are consistent with a folding intermediate that is formed largely by the C-terminal T4 lysozyme subdomain and might be related to a previously described “hidden intermediate” [74].

The probability of occupying the state I (relative to U) during hopping (Figure 2.12) is similar in the presence and absence of the ribosome (Figure 2.2E). From the force at which U and I are equally populated \((F_{1/2} \approx 3.6 \text{ pN})\) and the observed change in extension between U and I at that force \((\Delta x_{U-I} \approx 10 \text{ nm})\), we estimate that I is stabilized by a Gibbs free energy of unfolding \([\Delta G_{\text{unfolding}}^0 \text{ (intermediate)}] = 3.0 \text{ kcal/mol} (5.1 k_B T, \text{ where } k_B \text{ is the Boltzmann constant and } T \text{ is temperature})\) relative to U [43] for both the free and the ribosome-bound protein. This value is small compared with the stability of the folded protein \([\Delta G_{\text{unfolding}}^0 \text{ (native)}] = 14.1 \text{ kcal/mol} [68]\), indicating that the intermediate is lacking many of the interactions that stabilize the natively folded protein. It is also smaller than the reported stability of the “hidden” intermediate, perhaps because the N-terminal A helix is not part of the intermediate observed in our experimental geometry.

To explore the ribosomal effect within the folding pathway, we used a Bayesian Hidden Markov model (BHMM) approach to conduct a kinetic analysis of the force-clamp data (Figure 2.3A). The values for the rates of the U-to-I and I-to-U transitions \((k_{U\rightarrow I} \text{ and } k_{I\rightarrow U}, \text{ respectively})\) are essentially the same for the free and the ribosome-bound protein (60 amino acid linker). In contrast, \(k_{I\rightarrow N}\) of the ribosome-bound protein is smaller by at least one order of magnitude compared with that of the free protein (Figure 2.3A). The BHMM analysis also yields estimates of the extensions of each species (Figure 2.13). Interestingly, the unfolded protein, both free and ribosome-bound, is more compact than predicted from a WLC model [71] at forces below 4 pN (Figure 2.3B). This deviation from the model is consistent with a compaction of the unfolded protein, perhaps reflecting transient local secondary structure formation at those low forces before cooperative folding transitions. For the free protein, the unfolded-state extensions gradually increase to the values predicted by the WLC model when the force is raised to 5 pN. Notably, the unfolded state remains more compact in the ribosome-bound protein (over the limited force range accessible in these measurements; Figure 2.3B). Thus, the ribosomal interactions appear to have a dual effect: deceleration of native tertiary structure formation and stabilization of a compacted state before folding (Figure 2.3C), presumably mediated at least in part by electrostatic interactions.

How might the ribosome contribute to efficient de novo protein folding? So far, we have described folding of the entire protein. To study folding before completion of synthesis, we translated the N-terminal 149 amino acids of T4 lysozyme, so that \(\sim 110 \text{ to 120 residues of the full T4 lysozyme sequence (164 residues) exited the ribosome (70\% of the full protein), whereas 30 to 40 amino acids remained within the exit tunnel. Upon stretching the ribosome-bound polypeptide, the force-extension curves did not reveal any cooperative folding or}
Figure 2.3: (A) Kinetic rates along the T4 lysozyme folding pathway. The rate of the final, irreversible step ($k_{I-N}$) along the folding pathway is significantly slower for the ribosome-bound protein (60 amino acid linker). (B) Distance changes upon unfolding ($N \Rightarrow U$, open symbols) and refolding ($U \Rightarrow N$, filled symbols) at various forces. At low forces, the distance is shorter than expected from a WLC model (gray line), indicating a partial compaction of the polypeptide that does not resist forces above 4 pN. The compact structure is stabilized in the ribosome-bound protein. Error bars: 95% CIs. (C) Schematic folding energy landscapes of free and ribosome-bound T4 lysozyme. The height of the barrier from I to N is affected by the ribosome, resulting in a decrease in $k_{I-N}$. 
unfolding transitions (Figure 2.14). Even when we extended the T4 lysozyme sequence by a 20 amino acid linker (so that ~144 to 154 out of the 164 amino acids of T4 lysozyme are outside the tunnel), we did not detect folding (Figure 2.4 A), presumably because the C-terminal residues, which interact with the N-terminal A helix in the native structure, are still sequestered within the ribosomal exit tunnel.

Figure 2.4: (A) Force-extension curves of T4 lysozyme (164 amino acids) with a 20 amino acid C-terminal extension bound to the ribosome (150 amino acids outside the ribosomal tunnel), exhibiting no defined unfolding transitions. (B) A free protein fragment of 149 amino acids misfolds into a heterogeneous ensemble of structures that unfold over a wide range of forces. (C) Comparison of unfolding events recorded for the free protein and the 149 amino acid fragment. The full-length protein unfolds within a narrow, stochastic range of forces and extension changes, whereas the unfolding transitions of the isolated fragment are highly heterogeneous.
Attempts to express soluble fragments in E. coli were unsuccessful (Figure 2.15). Thus, we used puromycin-modified DNA oligonucleotides [75] to release the in vitro translated 149 amino acid fragment from the ribosome. In contrast to the lack of transitions observed on the ribosome, the released fragment unfolded at a range of forces and contour length changes (Figure 2.4 B), indicating that it adopts a highly heterogeneous ensemble of structures, some of which exhibit considerable (mechanical) stability. Given the homogeneous unfolding of the complete protein, it is unlikely that all of these structures represent productive, on-pathway species. Rather, the diverse behavior of this fragment is probably due to misfolding, aggregation, entanglement of several polypeptides immobilized in close vicinity, or interactions of the misfolded protein with the bead surface, none of which are observed for the ribosome-bound nascent protein. Thus, the ribosome appears to prevent misfolding of the incomplete protein through a kinetic mechanism, effectively acting as a molecular chaperone for nascent polypeptides. This chaperone activity is probably mediated by the surface surrounding the polypeptide exit tunnel and is distinct from the previously described protein folding activity of the ribosome mediated by the 26S rRNA [76]. Mechanisms that keep proteins in a folding-competent conformation may be particularly important if C-terminal residues, which are synthesized last, are required for productive folding, as in T4 lysozyme and other proteins [59, 77–79]

2.4 Discussion

In the cell, molecular chaperones interact with nascent polypeptides during their synthesis [80]. The in vitro experiments described here suggest that the ribosome contributes to efficient de novo folding in several ways. Polypeptide compaction, an early event during protein folding [81], is promoted by the ribosome and may, in conjunction with the spatial arrangement within polysomes [82], limit aberrant interactions among nascent chains. Rather than acting as an inert wall, which would be expected to increase folding rates entropically [83, 84], the ribosome slows folding of T4 lysozyme, presumably by attracting positively charged residues and repelling negatively charged ones within the same nascent polypeptide chain. Thus, it can bias the conformational search of the protein and its folding rate. This mechanism may complement the mode of action of other molecular chaperones that bind their substrates through hydrophobic interactions and may be particularly pronounced for T4 lysozyme and other basic proteins that represent a large group in most proteomes [85]. However, such a mechanism should be operative even for proteins harboring a net negative charge and may apply at least to cytosolic proteins, which are held in close proximity to the ribosomal surface during synthesis. Our findings may represent a paradigm for how the ribosome can, in principle, affect nascent-chain folding. The system is easily adaptable for investigating proteins other than T4 lysozyme and should be amenable to observing nascent-chain elongation in real time. These future experiments will shed more light onto how protein folding is tuned to synthesis.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

2.5 Methods

Preparation of derivatized ribosomes

To generate a unique attachment site within the 50S ribosomal subunit, we targeted the C-terminus of the large ribosomal subunit protein L17. L17 was selected for its location away from the ribosomal subunit interface and from the immediate vicinity of the ribosomal exit tunnel. We chromosomally replaced the open reading frame of rplQ, the single gene encoding L17 in Escherichia coli, with the coding sequence of C-terminally ybbR tagged [86] L17 in *E. coli* strain MC4100 using the phage λ Red recombinase system as described [87]. The plasmids and strains for the Red recombinase system were generously provided by Prof. Barry Wanner (Purdue University). A tetracycline resistance gene was used as the selection marker. The resulting strain produces ribosomal protein L17 with a C-terminal ybbR-tag. Engineered ribosomes were isolated from the recombinant strain as described previously [88] and stored frozen.

The serine residue at position 2 in the 11 amino acid (aa) long ybbR tag sequence [86] serves as an acceptor of the phosphopantethein moiety of Coenzyme A (CoA) in an enzymatic reaction catalyzed by the phosphopantethein transferase Sfp from Bacillus subtilis [86]. It has been shown that the thiol group of CoA can be derivatized with chemical moieties such as fluorophores without interfering with the specificity or efficiency of the Sfp-mediated reaction [86]. The expression plasmid for the Sfp enzyme was generously provided by Prof. Christopher Walsh (Harvard Medical School). The protein was expressed and purified as described [89].

To demonstrate the specificity of the reaction in the chemically complex environment of the ribosome, we derivatized purified engineered ribosomes with a fluorescent CoA-adduct (F-CoA) prepared by reacting reduced CoA (Fisher Scientific) with fluorescein-5-maleimide (Invitrogen). The covalent attachment of the fluorescent phosphopantethein-fluorescein moiety allows in-gel detection of the labeled L17 protein after separation by SDS-PAGE (Supplementary Figure S1). As expected, a single fluorescent band was detected after incubating engineered ribosomes (5 µM) with F-CoA (10 µM) and purified Sfp enzyme (1.5 µM). Labeling was dependent on the presence of the ybbR tag sequence on L17, the CoA moiety and the Sfp enzyme, confirming the specificity of the reaction in this system.

To covalently couple a DNA handle to L17, we repeated the reaction replacing F-CoA with a 25 nucleotide (nt) long oligonucleotide that was chemically modified at its 5-end with CoA and hybridized to a shorter, complementary oligonucleotide prior to the reaction, resulting in a double-stranded (ds) oligonucleotide with a 4 nt long single stranded overhang. Unreacted oligonucleotide was removed by sedimenting the derivatized ribosomes through a 1M sucrose cushion. To assess the efficiency of the reaction, an aliquot of the derivatized ribosomes was subsequently reacted with F-CoA. From quantitation of the in-gel fluorescence and the amount of ribosomes, we estimated that derivatization with the DNA handle was ≥50% efficient.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

Generation of ribosome-nascent chain complexes (RNCs)

We used messenger RNAs (mRNAs) lacking stop codons as templates for in vitro translation reactions. If the ribosome does not encounter an in-frame stop codon, termination of translation and nascent chain release do not occur. The ribosome proceeds to the 3-end of the mRNA and stalls there. As a consequence, the translation product accumulates as a ribosome-bound peptidyl-tRNA species. Thus, we can precisely determine where the ribosome stalls by choosing the appropriate mRNA template. To produce specific templates, we first amplified specific regions of the T4 lysozyme construct shown in Figure 2.1 D using the polymerase chain reaction (PCR) with specific primers. PCR products were gel-purified and served as templates in in vitro transcription reactions yielding mRNA (T7 MegaScript Kit, Ambion). The mRNA was isolated (MegaClear Kit, Ambion) and used as a template in in vitro translation reactions.

We generated T4 lysozyme constructs that harbored an N-terminal Avi tag sequence by subcloning. The T4 lysozyme source plasmid was a gift from Prof. Susan Marqusee (University of California, Berkeley). The Avi tag is recognized by the E. coli biotin ligase BirA with high specificity [90]. BirA covalently modifies Lys10 in the 15 residue Avi tag sequence with a biotin moiety. The extraordinarily strong biotin-streptavidin interaction has previously been utilized to immobilize macromolecules for optical tweezers experiments [91] and provides an attachment point near the N-terminus of the T4 lysozyme constructs in our experiments. We cloned the birA gene into an expression vector and purified the protein by means of an engineered His6 tag.

Translation reactions were carried out using the PURE in vitro protein synthesis system [92]. We utilized a ribosome-free version of this Kit (PURExpress ∆ Ribosome Kit, New England Biolabs) supplemented with engineered ribosomes (∼1 µM) and purified mRNA template (∼5 µM). We chose to use a molar excess of template over ribosomes to minimize the initiation of more than one ribosome per template molecule which would result in polysome formation and product heterogeneity. In vitro translation reactions also contained 50 µM D-biotin and 0.5 1 µM purified BirA enzyme. Translation was carried out for 30 min at 37. Reactions were chilled on ice and briefly clarified by centrifugation (10 min, 16,000 g, 4° C). Ribosome-nascent chain complexes were sedimented through a 1M sucrose cushion in Buffer HKM (25 mM HEPES-KOH pH 7.4, 150 mM KCl, 5 mM Mg-acetate) by centrifugation for 40 min at 200,000 g, 4° C. The supernatant was discarded and the pelleted ribosome-nascent chain complexes were resuspended in a small volume of Buffer HKM.

Translation products were analyzed in bulk by SDS-PAGE followed by detection with Streptavidin-HRP (Southern Biotech) after electro-blotting onto a nitrocellulose membrane. As expected, the ribosome-bound nascent polypeptides accumulated as an RNase A-sensitive peptidyl-tRNA species with an apparent molecular weight ∼15–20 kDa larger than expected for the free protein (Figure 2.6 A). Treatment of RNCs with puromycin also led to peptidyl-tRNA hydrolysis (Figure 2.6 A). Since the ribosomal peptidyl-transferase activity is required for puromycin to replace the tRNA, this finding confirms that the C-terminus of the nascent polypeptide is stably anchored in the peptidyl-transferase center, as predicted based on our
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

experimental design.

**Immobilization of RNCs for optical tweezers experiments**

To immobilize ribosome-nascent chain complexes on the surface of a polystyrene microsphere for manipulation with optical tweezers, we functionalized the surface of 2.1 µm carboxyl polystyrene beads (Spherotech) with double-stranded DNA oligonucleotides. The double-stranded oligonucleotides contained an amino-modifier on one of the 5-ends and a 4 nt single-stranded overhang on the other. The DNA was covalently coupled to the bead surface by chemical crosslinking with 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC; Pierce Protein Research Products) as described [93]. The reaction was quenched with 50 mM Tris-HCl, pH 7.6, and the beads were washed extensively to remove free oligonucleotides.

The single-stranded overhang at the distal end of the oligonucleotide was designed to be complementary to that of the oligonucleotide attached to ribosomal protein L17 (see above). It thus provides a “sticky end” and allows a directed, covalent coupling of the RNCs and the functionalized beads via DNA ligation. After ligation, the RNCs are separated from the bead surface by 51 bp of double-stranded DNA. For ligation, beads were equilibrated in T4 DNA ligase Buffer (New England Biolabs), mixed with RNC and T4 DNA ligase (New England Biolabs), and incubated at 16°C for 30 min. The resulting “sample beads” were stored on ice until they were injected into the microchamber of the optical tweezers instrument.

**Immobilization of free T4 lysozyme for optical tweezers experiments**

The enzymatic reactions described above for RNCs were also used to immobilize free T4 lysozyme polypeptides on polystyrene beads. The T4 lysozyme open reading frame was subcloned into an expression vector such that it is flanked by sequences encoding an N-terminal Avi-tag and a C-terminal ybbR tag. After expression in *E. coli* and biotinylation in vitro, the protein was purified by affinity chromatography on monomeric avidin agarose (Pierce Protein Research Products) and size exclusion chromatography. A DNA handle was covalently linked to the C-terminal ybbR tag in an Sfp-mediated reaction as described above. To remove unreacted handle, the protein was again subjected to chromatography on monomeric avidin agarose. The derivatized protein was ligated to functionalized polystyrene beads as described above. The sample was stored on ice until injection into the microchamber of the optical tweezers instrument.

**Puromycin-mediated handle attachment**

We utilized oligonucleotides with a puromycin (Pm)-moiety at the 3-terminus (“Pm oligonucleotides”) to release nascent chains from stalled RNCs and simultaneously conjugate a DNA
handle to the C-terminus of the released nascent chain (Figure 2.6 B). The ability of Pm-oligonucleotides to function as Pm analogs in nascent chain release has been demonstrated previously with RNA conjugates [75] and results in covalent attachment of the oligonucleotide to the C-terminal amino acid of the nascent polypeptide. We incubated purified RNCs with Pm-oligonucleotides that had previously been hybridized to yield a dsDNA oligonucleotide with a “sticky end” (see above). The resulting product was analyzed by SDS-PAGE, electroblotting and detection with Streptavidin-HRP (Figure 2.6). The released product (which we term “peptidyl-DNA”) is RNase A-resistant and has an electrophoretic mobility between the free polypeptide and the peptidyl-tRNA. We ligated this product to functionalized polystyrene beads as described above.

Synthesis of immobilized DNA handles

In optical tweezers experiments, we grabbed the N-terminal end of nascent chains or the free polypeptides through a ~2kbp double-stranded DNA handle. We developed an approach to synthesize this handle directly on the polystyrene bead in a PCR-like reaction. Carboxyl-polystyrene beads (2.1 µm diameter) were functionalized by crosslinking a single-stranded DNA oligonucleotide to the bead surface with EDC as described above. This cross-linked oligonucleotide functioned as an immobilized primer in a subsequent PCR-like reaction containing, in addition, Taq DNA polymerase (New England Biolabs), dNTPs, a linear template and a biotinylated reverse primer. The template was comprised of 1789 bp of double-stranded DNA. The ends of the template were complementary in sequence to the bead-immobilized and the free, biotinylated primer, respectively. We subjected the reactions to 35 cycles of denaturation (45 s at 94°C), annealing (45 s at 45°C) and extension (2 min at 72°C). During the reaction, the immobilized primer anneals to the template and is extended by the DNA polymerase, which also synthesizes the complementary strand after annealing of the biotinylated primer. The product is thus a double-stranded DNA fragment of 1789 bp in length. One of the 5-ends is covalently linked to the bead surface, the other harbors a biotin moiety. Subsequently, these “handle beads” were washed extensively with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and stored at 4°C.

We verified successful DNA synthesis on the bead surface by injecting an aliquot of these beads into a microchamber and captured a single bead in the optical trap (see below). We readily obtained very stable tethers to a 2.1 µm streptavidin bead (Spherotech) held in a micropipette. When we stretched these tethers, we obtained the force-extension curves (Figure 2.7) expected for a ~2kbp dsDNA molecule, with an overstretching plateau at ~65 pN [91]. For experiments involving RNCs or the free protein, the biotin moiety of the bead-coupled DNA handle was derivatized with streptavidin. To this end, an aliquot of the “handle beads” was incubated with streptavidin (~2 mg/ml) for 10 min at room temperature. Subsequently, the beads were diluted 1000-fold and injected into the microchamber.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

Optical tweezers measurements

We conducted our force-spectroscopy measurements on an instrument with a counterpropagating dual laser beam single trap design that measures force directly via the change in light momentum [94]. Experiments were carried out in Buffer HKM using a microchamber with a micropipette. The “handle bead” is held in the optical trap, while the “sample bead” is held on the tip of the micropipette by suction. We obtained tethers by moving the optical trap relative to the chamber, bringing the beads into close proximity. To exert force, we moved the trap in the opposite direction. Force and trap-position were recorded at a sampling frequencies of 1000 Hz (force ramp measurements) or 3125 Hz (force clamp measurements). Rupture forces and extension changes from force extension curves (from force ramp measurements) were determined as described [95]. Constant force experiments, in which the force feedback of the instrument operates at 2000 Hz, were analyzed using a Bayesian Hidden Markov Model (BHMM) approach (see below).

Several findings confirm that the rupture events observed in these measurements represent unfolding events within the nascent polypeptide: No transitions (other than overstretching of DNA at high force) are observed when the DNA handle alone is stretched (Figure 2.7). The unfolding events observed for the ribosome-bound full-length T4 lysozyme are similar to those of the free protein (Figure 2.1 E-H). Ribosome-bound T4 lysozyme fragments that do not acquire stable structures do not exhibit rupture events (Figure 2.4 A). Thus, the unfolding events observed originate from the nascent polypeptide, not from structural transitions within the ribosome.

Tethers containing ribosome-nascent chain complexes are stable up to \( \sim 50 \) pN of force without exhibiting unfolding transitions other than those resulting from nascent chain unfolding. At forces above 50 pN, the tether is lost due to breakage, presumably due to extraction of protein L17 (the attachment point in our geometry) from the large ribosomal subunit. Extraction of the nascent chain through the exit tunnel might also be possible, but would require unfolding of the P-site tRNA-moiety to which the stalled nascent chain remains attached in our samples.

Bayesian Hidden Markov Model (BHMM) analysis

We used a Bayesian hidden Markov model (BHMM) approach [96] to analyze the optical tweezers data collected in constant force mode. Hidden Markov models (HMM) have been used previously to analyze single molecule trajectories [97]. The Bayesian extension of the standard approach of HMM analysis allows the experimental uncertainties to be directly quantified by sampling from the posterior of the model parameters (transition probabilities and Gaussian state observable distributions) given the data, rather than simply identifying the maximum-likelihood model parameters as in the traditional HMM approach. To do this, we augment the standard HMM likelihood function with a prior that enforces the physical detailed balance constraint in the transition matrix. Sampling from the posterior proceeds by a Gibbs sampling approach, alternating updates of the reversible transition matrix [98],
the hidden state sequence, and the Gaussian state observable distributions using standard techniques.

The BHMM code we used here sampled models over the measurement histories of trap position at constant force feedback, producing estimates of average extension characterizing each state and the transition probabilities among the states, as well as confidence intervals that characterize the uncertainty in these values due to finite-sample statistics. After subsampling the extension data to 100 Hz to produce Markovian statistics, the method samples models consistent with the data using a Gibbs sampling strategy that assumes the extension measurements from each state (including measurement error) are normally distributed about the average extension for that state. Here, the number of states was fixed to three after verifying the three-state nature of the data by inspection of the recorded traces. The first 50 BHMM samples after starting from the maximum likelihood estimate were discarded, and 1000 samples were subsequently generated to collect statistics on average extensions and transition probabilities, as well as generate the 95% confidence intervals reported here. Matrices of rate constants $K$ were computed from transition matrices $T(\tau)$ using the standard relationship $T(\tau) = e^{K \tau}$ through use of the matrix logarithm.

Calculations of Debye lengths at 150 mM vs 500 mM KCl

The Debye length $\kappa^{-1}$ is a measure of how the electrostatic potential around a charge decays due to charge screening in a given medium. The potential around a charge drops to $\frac{1}{\kappa}$ over a length of $\kappa^{-1}$. $\kappa^{-1}$ in our aqueous buffer was estimated according to

$$\kappa^2 = \frac{\varepsilon^2}{\varepsilon_0 k_B T} \sum_i n_i z_i^2$$

with $\varepsilon = 1.6 \times 10^{-19}$ C (elementary charge), $\varepsilon = 78.5$ (relative dielectric constant for water at 25°C), $\varepsilon_0 = 8.854 \times 10^{-12} C V^{-1} m^{-1}$ (permittivity of vacuum), $k_B = 1.38 \times 10^{-23} J K^{-1}$ (Boltzmann constant), and $T = 298$ K. $n_i$ is the ion concentration in ions per $m^3$ and $z_i$ the ion valence for component i. For our standard buffer (Buffer HKM; 25 mM HEPES·KOH, pH 7.4, 150 mM KCl, 5 mM Mg-acetate), we obtain a Debye length of $\kappa^{-1} = 0.72$ nm (assuming that half of the HEPES molecules are deprotonated at this pH). Increasing the KCl concentration to 500 mM yields $\kappa^{-1} = 0.42$ nm, a Debye length 42% shorter than the one at 150 mM KCl.

Estimation of extension fluctuations

The time resolution in our optical tweezers measurements is fundamentally limited by the relaxation time of the polystyrene beads. The thermal fluctuations of the single molecule tether that are faster than this response time (determined by the corner frequency of the bead) are thus not resolved in our measurements. We estimated the mean square fluctuations in the extension of the N-terminus of the unfolded protein in our single-molecule experiments.
using the equipartition theorem:

\[
\langle x^2 \rangle = \frac{k_B T}{k}
\]

with \( k_B = 1.38 \times 10^{-23} \) pN nm (Boltzmann constant), and \( T = 298 \) K. \( k \) is the combined stiffness of the unfolded protein and the DNA handles. To estimate an upper bound for the stiffness, we treat the unfolded protein (164 aa) and the 1789 bp DNA handle as springs on either side of the N-terminal residue. In this case, the combined stiffness is the sum of the individual stiffnesses. We do not include any other components, such as the ribosome, in these estimations. The bead relaxes much more slowly than the DNA/protein tether (we determined a corner frequency of \( \sim 1200 \) Hz for the 2.1 \( \mu \)m diameter beads used in our experiments). We can distinguish two extreme cases: Very fast fluctuations on the timescale of \( \mu s \) and slower fluctuations on the timescale of ms. The fluctuations of the bead will contribute only to the slower fluctuations.

In the case of very fast fluctuations, we neglect contributions of the bead. We calculated the stiffnesses of the individual components (dsDNA handles and the unfolded polypeptide) at \( F = 3.6 \) pN, assuming a WLC model with \( P_{DNA} = 53 \) nm, \( P_{protein} = 0.65 \) nm, and contour length increments of \( L_C(DNA) = 0.34 \) nm/bp and \( L_C(protein) = 0.36 \) nm/aa. The DNA handle in our assembly is 1789 bp long, yielding \( k_{dsDNA,1789bp} = 0.158 \) pN/nm. The stiffness of the protein is \( k_{polypeptide,164aa} = 0.268 \) pN/nm. Thus, the combined stiffness at 3.6 pN is \( k = 0.426 \) pN/nm, and the tether fluctuates with a root mean square (RMS) deviation of \( \sqrt{\langle x^2 \rangle} = 3.1 \) nm. To account for the contributions of the bead that is held in the optical trap in our geometry, we combined the trap stiffness with the DNA handle stiffness as springs connected in series:

\[
\frac{1}{k_{DNA+trap}} = \frac{1}{k_{dsDNA,1789bp}} + \frac{1}{k_{trap}}
\]

The stiffness of our optical trap is \( k_{trap} = 0.1 \) pN/nm. We obtain \( k_{DNA+trap} = 0.061 \) pN/nm, and, combined with the unchanged polypeptide stiffness, \( \sqrt{\langle x^2 \rangle} = 3.5 \) nm.

These estimates indicate that the polypeptide spends a significant fraction of the time in closer proximity to the ribosome than indicated by the measured extensions, even if held at a force of 3.6 pN. Thus, even at a Debye length of less than 1 nm, charged residues within the polypeptide chain can come close enough to the ribosome to be subject to the electrostatic potential surrounding its surface. We only account for the DNA handle and the polypeptide in the above stiffness calculations. Therefore, the values obtained for the stiffness represent upper bounds, since any additional components, such as the ribosome and the linkages will have a finite stiffness that can only lower the total stiffness of the tether. As a consequence, the fluctuations calculated from this stiffness represent a lower bound and are likely to be higher than estimated here. We assume, however, that the contributions of the ribosome to the overall stiffness are negligible in the refolding force regime, because we expect this highly ordered, folded structure to be significantly less flexible under these conditions than the unfolded polypeptide and the DNA handle.
Dimensions and stability of the T4 lysozyme folding intermediate

We observed a T4 lysozyme folding intermediate immediately prior to folding to the native state in most recordings when the C-terminal segment is allowed to exit from the ribosomal tunnel. The intermediate appears to be obligatory: Consistent with the expected exponential distribution of state lifetimes, progression through the intermediate happens too quickly to be observed at the sampling resolution in a small fraction of traces. Consequently, the transition through the intermediate during folding to the native state is observed in most, but not all of the recorded traces. We estimated the dimensions of the intermediate by analyzing extension traces recorded in force clamp mode (Figure 2.2 A). We limited this analysis to data obtained with the free protein, because the ribosome affects the extension of the nascent polypeptide (Figure 2.3), complicating the analysis. Choosing the relatively rigid native state [43] as a point of reference, we find that the end-to-end extension of I is 10.1 nm longer than that of N at 5 pN (Figure 2.2 A), which corresponds to a difference in contour length, $\Delta L_C(I - N)$, of 26.4 nm, compared to a difference between U and N, $\Delta L_C(U - N)$, of 60.1 nm, according to the WLC model (see above). Thus, the unfolded molecule compacts by $\Delta L_C(U - I) = 33.7$ nm upon forming the intermediate. Assuming an end-to-end distance of the intermediate in the range from 0.9 nm (the folded state end-to-end distance) to 5.0 nm (the longest C- distance in the folded protein), between 96 to 108 aa participate in the intermediate formation.

We used our force-clamp data to estimate the stability of the intermediate. The unfolded and intermediate states (U and I, respectively) are equally populated at a force of $F_{1/2} \approx 3.6$ pN (Figure 2.2 E and Figure 2.12), with an extension change of $\Delta x_{(U-I)} \approx 10$ nm separating U and I. According to the WLC model, a change in extension of 10 nm at 3.6 pN corresponds to a change in contour length of $\Delta L_C \approx 30$ nm (assuming a persistence length for unfolded polypeptides of $P = 0.65$ nm). We calculated a force extension curve with these parameters ($\Delta L_C = 30$ nm, $P = 0.65$ nm) and determined $\Delta G$(stretch) by integrating the WLC equation with boundaries $F = 0$ pN to $F = F_{1/2} = 3.6$ pN. It yields $\Delta G$(stretch)$= 15$ pN nm. Thus, $\Delta G_0 = (36 - 15)$ pN nm $= 21$ pN nm $= 12.7$ kJ/mol $= 3$ kcal/mol.

An active-like “hidden” intermediate, composed largely of the C-terminal T4 lysozyme subdomain, including the N-terminal 12 aa that comprise the “A-helix” [74, 99], has previously been observed. The thermodynamic stability of this hidden intermediate is 7.9 kcal/mol [67], significantly more stable than the intermediate we observe (3.0 kcal/mol at zero force). Presumably, the N-terminal A helix does not participate in the formation of the intermediate under our experimental conditions, because it is held at a distance from the C-terminal part of the protein by force. As a consequence, the intermediate seen here contains fewer stabilizing contacts and is thermodynamically weaker. Similarly, spatially biasing the A-helix away from the remainder of the protein may allow us to observe the intermediate in our experiments, in contrast to other geometries [65].
Figure 2.5: Specificity of L17 modification in tagged ribosomes. Isolated ybbR-tagged (lanes 1-3) or WT (lanes 4 and 5) ribosomes were incubated with Sfp enzyme in the presence of a fluorescent CoA adduct (F; lanes 3 and 5) or a control substrate (c; lanes 2 and 4) or in the absence of a substrate (lane 1). After separation of samples on SDS-PAGE gels, the fluorescence was imaged (right panel) and proteins were subsequently visualized by Coomassie staining. Coomassie staining revealed the typical ribosomal protein pattern. Fluorescence imaging revealed that a single protein band (arrowhead) is labeled in the ybbR-tagged ribosome sample, demonstrating the specificity of the Sfp-mediated reaction. No labeling is detected in the control reactions with either untagged ribosomes (lanes 4 and 5) or a control substrate (lane 2).
Figure 2.6: **(A)** Western blot of stalled in vitro translation product. After translation of stop codon-free mRNA, RNCs were sedimented through 1M sucrose cushions. Samples were treated with RNase A ("RNase", 100 g/ml), puromycin ("Pm", 1 mM) or puromycin-DNA oligonucleotides (Pm-DNA, 10 M) as indicated and separated on SDS-PAGE gels. Proteins were electro-blotted onto nitrocellulose membranes and detected with streptavidin-HRP by means of their biotin moiety. The electrophoretic mobility indicates that the translation product accumulates as a peptidyl-tRNA species (Lane 1; calculated Mw of the protein: 20.5 kDa; observed apparent Mw: ~38 kDa). RNase A treatment reverts the electrophoretic mobility to the expected value of ~20 kDa (Lane 2). Puromycin treatment also reverts the apparent Mw to 20 kDa (Lane 3). Since puromycin is incorporated into the nascent chain by the ribosome, this observation indicates that the peptidyl-tRNA is still anchored in the peptidyl transferase center. Incubation with a puromycin-derivatized DNA oligonucleotide results in the (partial) formation of a peptidyl-DNA species with an apparent Mw of ~30 kDa. **(B)** Schematic illustration of DNA handle attachment through puromycin-DNA oligonucleotides. The derivatized oligonucleotide is incubated with ribosome-bound nascent chains. Ribosome-mediated incorporation of the puromycin moiety results in covalent attachment of the oligonucleotide to the C-terminus and release of the nascent chain.
Figure 2.7: Force-extension curves of DNA handles. A 1789 bp long fragment of DNA was synthesized directly on the surface of the bead in a bulk PCR-like reaction. A single molecule of this DNA was stretched (red curves) and relaxed (blue curves) repeatedly in the optical tweezers instrument. The curvature at forces below $\sim 10$ pN, due to the entropic elasticity of the molecule, and the overstretching plateau at $65$ pN are characteristic for double-stranded DNA. The molecule persists over several rounds of stretching to high forces ($\sim 70$ pN), demonstrating stable attachment, and does not exhibit hysteresis upon relaxation, indicating that it is free of nicks.
Figure 2.8: Single-molecule puromycin release experiment. (A) Schematic of the experimental setup. A tether is formed with a RNC and held at a constant low force ($F_{\text{Pull}} \approx 5$ pN). A solution of 1 mM puromycin is flowed into the chamber, resulting in a drag force on the bead ($F_{\text{Flow}}$) perpendicular to $F_{\text{Pull}}$. During the experiment, the forces $F_{\text{Pull}}$ and $F_{\text{Flow}}$ are recorded over time. (B) Puromycin release experiment (red trace: $F_{\text{Pull}}$, grey trace: $F_{\text{Flow}}$). A RNC tether is held at 5 pN constant force. At $t \approx 20$ s, the flow of the puromycin into the microchamber is initiated (grey arrowhead). At $t \approx 220$ s (when the puromycin that is being flowed into the chamber reaches the area around the trapped bead), the tether ruptures, indicated by a drop in $F_{\text{Pull}}$ to 0 pN (red arrowhead). The flow of puromycin is stopped at $t \approx 260$ s, and $F_{\text{Pull}}$ also returns to 0 pN. (C) Buffer only control. Instead of puromycin, buffer is flowed into the chamber under conditions similar to those described in (A). The tether does not rupture over the course of the experiment ($\sim 500$ s). Note that the tether is still intact when the recording stops at $t \approx 550$ s.
Figure 2.9: Force-dependent lifetimes of free (circles) and ribosome-bound (triangles) T4 lysozyme. Very similar lifetimes were obtained from rupture-force distributions shown in figure 2.1 G and H for the ribosome-bound and free proteins, respectively. The lines represent fits based on Kramers theory with a scaling factor of \( \nu = 1 \), which reduces to Bells formula. Varying the scaling factor over a range of \( \nu = 1/2 \) to \( \nu = 1 \) changed the values obtained from the fitting but did not affect the similitude of the values for free and ribosome-bound T4 lysozyme. The transitions state distances are very similar in both cases (\( \Delta^\dagger x_{\text{free}} = 2.3 \pm 0.5 \) nm, \( \Delta^\dagger x_{\text{ribosome-bound}} = 2.0 \pm 0.2 \) nm). The native state lifetimes at \( F = 0 \) pN of \( \tau_0,\text{free} = 7503 \) s (95% CI: 1053 s, 53637 s) and \( \tau_0,\text{ribosome-bound} = 1826 \) s (95% CI: 912 s, 3652 s) are afflicted with large errors because their determination is based on extrapolation, since the low force regime close to \( F = 0 \) pN is not accessible in these force ramp experiments.
Figure 2.10: Force-extension curves for free T4 lysozyme (A) and the “+41” RNC (B). Consecutive cycles of pulling (red) and relaxing (blue) the molecule are shown, offset along the x-axis for visual inspection. The traces recorded for the free protein exhibit unfolding transition at 14-19 pN during pulling, and a refolding transition is discernable at \( \sim 5 \) pN during relaxation. The refolding transition is not apparent in the RNC recordings, and unfolding is observed in only 2 of the 8 traces in this representative example.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

Figure 2.11: Representative example of force clamp measurements of T4 lysozyme refolding (free protein). The force (A) and the position of the optical trap (B) are recorded over time, while the force is cycled between 5 pN and 15 pN. At 5 pN, the molecule refolds, indicated by a sudden compaction (red arrowhead). The molecule remains stably in the native state at this force, and no unfolding is observed over the course of 60 s. Raising the force to 15 pN destabilizes the folded protein and the molecule unfolds, indicated by a sudden increase in the extension (open arrowhead). The extension change upon unfolding at 15 pN (open arrowhead) is larger than the change upon refolding at 5 pN (solid arrowhead), in agreement with the entropic elasticity of the unfolded polypeptide which results in the unfolded state being more extended at 15 pN than at 5 pN.
Figure 2.12: Example histograms of state extensions of the unfolded and intermediate states. (A) Several refolding traces were recorded for a single free T4 lysozyme molecule during refolding at 3.8, 4.0 and 4.2 pN. The hopping between the unfolded and the intermediate state was analyzed. Extensions are relative to the native state (at 0 nm, not shown in the histograms). The extension data were fitted to a mixed Gaussian model with two components (solid red line). The individual components are plotted as dashed lines. The mixing proportions of the two components reflect the relative population of the unfolded and intermediate states prior to refolding. Increasing the force during refolding biases the molecule towards increasingly populating the unfolded state. (B) Same as in (A), recorded for a single “+60” RNC molecule. The numbers above the peaks represent the probabilities of the respective state as determined by fitting to the mixed Gaussian model. The shape of the histograms in (B) better approximates the mixed Gaussian model than the histograms in (A), because the reduced folding rates on the ribosome allow us to collect more data points. However, the relative probabilities for U and I at a given force are similar for the free and the ribosome-bound protein.
Figure 2.13: Extension changes from BHMM analysis of force-clamp refolding experiments (open circles: $\Delta x_{I-N}$, filled circles: $\Delta x_{U-N}$) for the free (blue) and the ribosome-bound protein (“+60” construct, red).
Figure 2.14: Force extension curves recorded with ribosome-bound nascent chains comprising 149 N-terminal residues of T4 lysozyme (full-length T4 lysozyme: 164 amino acids). The traces do not exhibit any unfolding transitions, indicating that the nascent polypeptide does not acquire any stable structures.
CHAPTER 2. THE RIBOSOME MODULATES NASCENT PROTEIN FOLDING

Figure 2.15: Expression of T4 lysozyme fragments in *E. coli*. Fragments of T4 lysozyme composed of 25, 69, 99 and 149 N-terminal amino acids were expressed in *E. coli* BL21(DE3) cells from pET-based plasmids by induction with 1 mM IPTG for 2 h. All constructs contained an N-terminal Avi-tag and a C-terminal ybbR-tag. The lengths of the fragments are indicated by red dots in the diagram of the T4 lysozyme primary structure (top panel). Total cell extracts (T) were prepared by lysis under non-denaturing conditions and separated into soluble (S) and pellet (P) fractions by centrifugation at 20,000 g. T, S and P fractions were separated by SDS-PAGE. Proteins were stained with Coomassie Brilliant Blue. The positions of bands representing the overexpressed proteins are indicated by red dots (bottom panel). The 25 (lanes 2-4), 99 (lanes 8-10) and 149 (lanes 11-13) amino acid fragments are largely insoluble, indicating misfolding and aggregation of the translation products. The 69 amino acid fragment (lanes 5-7) comprising the N-terminal subdomain is largely soluble, similar to the full length protein (“f.l.”; lanes 14-16) shown for reference. control: uninduced sample (no overexpression; lanes 17-19)
Chapter 3

Mechanical Force Releases Nascent Chain-Mediated Ribosome Arrest *In Vitro* and *In Vivo*


3.1 Summary

Protein synthesis rates can affect gene expression and the folding and activity of the translation product. Interactions between the nascent polypeptide and the ribosome exit tunnel represent one mode of regulating synthesis rates. The SecM protein arrests its own translation, and release of arrest at the translocon has been proposed to occur by mechanical force. Using optical tweezers, we demonstrate that arrest of SecM-stalled ribosomes can indeed be rescued by force alone, and that the force needed to release stalling can be generated in vivo by a nascent chain folding near the ribosome tunnel exit. We formulate a kinetic model describing how a protein can regulate its own synthesis by the force generated during folding, tuning ribosome activity to structure acquisition by a nascent polypeptide.

3.2 Introduction

The ribosome translates messenger RNA (mRNA) into amino acid sequences that contain the information needed for the polypeptide to attain its native structure. Differential usage of synonymous codons and structural elements in the mRNA modulate polypeptide elongation rates [22]. Such rate variations may be required for proper folding and processing of nascent proteins. Moreover, interactions of specific nascent chain sequences [27, 100] with
the ribosome exit tunnel [32] result in reduced rates of elongation. The bacterial SecM protein represents an example of a stalling sequence that interacts with the ribosome exit tunnel and allosterically represses the peptidyl transferase activity of the ribosome [32, 101–103]. Translation of SecM regulates expression of SecA, the motor component of the bacterial Sec translocon [27]. Release of stalling in vivo requires interactions between nascent SecM and the translocon machinery [104, 105]. It has been suggested that mechanical force exerted by the translocon relieves elongation arrest and leads to translation restart [31].

3.3 Results

To investigate the effect of force on the release of SecM-stalled ribosome-nascent chains (RNCs), we adapted a single-molecule optical tweezers assay [106] ((Figure 3.1 A), enabling the application of defined forces to single ribosome-associated nascent polypeptides. We generated stalled RNCs that contained the C-terminal domain of human calmodulin (CaM) (Figure 3.5 and 3.6). CaM provides a mechanical fingerprint [107] in our experiments by exhibiting equilibrium folding and unfolding (“hopping”) at ~7 pN (Figure 3.1 B). To detect release of stalled ribosomes, we utilized the antibiotic puromycin. Puromycin binds to the ribosomal A site and is incorporated into the nascent polypeptide, leading to its release from the ribosome [108]. SecM-arrested ribosomes are refractory to treatment with puromycin because they contain a prolyl-tRNApro stably bound in the A site, but become sensitive after arrest release, proline incorporation, and translocation [109](Figure 3.7 and 3.8). In the presence of puromycin and EF-G, arrest release will become apparent as a rupture of the tether (Figure 3.1 B and 3.8).

We applied a defined, constant force to the molecule in the range of 10-30 pN and measured the time required to restart translation, as measured by tether rupture. The mean restart times decreased with increasing force (Figure ??). We calculated the rate of stalling rescue as a function of the applied force (Figure 3.1 C and D, Figure 3.9 and 3.10). By fitting the force-dependent rates to Bells model [110], we estimated a distance to the transition state ($\Delta x^t$) of 0.4 nm (95% CI: 0.1 nm, 0.8 nm) and a zero-force rupture rate ($k_0$) of $3 \times 10^{-4}$ s$^{-1}$ (95% CI: $0.5 \times 10^{-4}$ s$^{-1}$, $20 \times 10^{-4}$ s$^{-1}$). This rate is in agreement with biochemical ensemble experiments, in which no force was applied (Figure 3.1 D, blue dot and Figure 3.7). In the force range of our experiments, release of SecM-mediated arrest is accelerated by more than an order of magnitude (Figure 3.1 D), supporting the hypothesis that SecM arrest is relieved by the mechanical force generated by the SecA ATPase.

Co-translational insertion of transmembrane helices via the translocon can release SecM-mediated stalling, presumably by generating force [111]. We wondered whether folding of a nascent globular protein domain could generate a force capable of modulating elongation by acting on peptide-tunnel interactions. Such interactions could serve to tune elongation rates to folding transitions [32, 100]. The exit tunnel is too narrow to accommodate folded protein domains [70]; therefore, as a nascent polypeptide emerges from the exit tunnel and folds in close proximity to the ribosome, it will be sterically excluded from the tunnel. This
CHAPTER 3. MECHANICAL FORCE RELEASES NASCENT CHAIN-MEDIATED RIBOSOME ARREST IN VITRO AND IN VIVO

Figure 3.1: A direct applied force catalyzes release of SecM-mediated arrest. (A) Experimental setup for optical tweezers experiments. When the nascent chain is transferred to puromycin, the assembly breaks. The structure of CaM was obtained from PDB ID 1CLL. (B) Example trace for restart experiment. After the “hopping” signature of CaM is observed (inset) at 7 pN, the force is raised to 20 pN. Red arrow: The tether breaks after \( \sim 3 \text{ min} \) at 20 pN. (C) Restart lifetimes at each force. Red lines: Distributions returned by the right-censoring MLE. (D) Force-dependent rates for restart of SecM-stalled RNCs in the optical tweezers. Rates are determined as shown in (B), with error bars representing 95% confidence intervals returned by the MLE. Red dotted line: Fit of Bells model to optical tweezers data. \( \Delta x^+: 0.4 \text{ nm} \) (95% CI: 0.1 nm, 0.8 nm) and \( k_0: 3.3 \times 10^{-4} \text{ s}^{-1} \) (95% CI: 0.5 \times 10^{-4} \text{ s}^{-1}, 20 \times 10^{-4} \text{ s}^{-1}). Black points: Rates determined using a method to account for non-specific tether rupture (Figure S6). Error determined by bootstrapping. Blue dot: Lifetime obtained from bulk experiment (Figure S3).
stERIC exclusion might generate a force that pulls on the nascent chain within the exit tunnel, which could modulate ribosome activity.

Having established that force accelerates SecM arrest release, we used SecM as a sensor to detect if nascent protein folding outside the ribosome can release the arrest. We constructed a library of plasmids encoding fusion proteins in which the stalling sequence (SecM17) is separated from the de novo designed protein Top7 [112] by flexible linker sequences of various lengths, followed by a reporter GFP (Figure 3.2 A, 3.11). Top7 folds rapidly against an applied force in close proximity to the ribosome (Figure 3.12). The GFP coding sequence is translated only upon successful release of the SecM17-mediated translation arrest. Variations in the length of the linker separating Top7 and SecM17 would affect the translation outcome of these constructs (Figure 3.2 B). Short linker sequences will not allow folding of Top7 because the C-terminus of the protein will be sequestered in the exit tunnel (Figure 3.2 B, top). Intermediate length linkers will allow more of the Top7 sequence to emerge from the ribosome tunnel and for the protein to fold and produce the steric exclusion folding force (Figure 3.2 B, middle). And, while longer linker sequences would also allow Top7 to fold, increased separation between the folding domain and the ribosome should abolish the proposed force-generating steric exclusion release mechanism (Figure 3.2 B, bottom).

We transformed E. coli with the plasmid library containing linker lengths varying from 4 to 28 amino acids. When grown under inducing conditions, a fraction of the colonies exhibited green fluorescence, indicating accumulation of GFP (Figure 3.2 C), and suggesting that SecM17-mediated stalling had been rescued in some of the transformants. We isolated and sequenced plasmid DNA from 63 fluorescent colonies. Plasmids isolated from fluorescent colonies contained linker sequences between 15 and 22 amino acids in length (Figure 3.2 D, 3.13). Given that the SecM17 sequence contributes 16 amino acids to the polypeptide and the ribosome tunnel can accommodate 30-35 residues [70], a linker length of 15 to 22 amino acids corresponds to having the protein sequence barely outside the tunnel exit. These results suggest that nascent chain folding near the ribosome tunnel exit can result in release of SecM arrest by stretching the polypeptide in the tunnel.

When Top7 folds near the tunnel exit, it does so against the steric exclusion force that it generates in the process. The protein must be able to fold against this force, and remain folded for a sufficiently long period of time to release stalling by SecM. To estimate the forces generated by the protein, we performed optical tweezers force spectroscopy measurements with single Top7 molecules tethered by their termini (Figure 3.14). We measured the distributions of lifetimes of both the unfolded and folded states (Figure 3.3 A, B and C). From these distributions, we extracted the force-dependent rates of folding and unfolding events (Figure 3.3 D) [72]. Folding rates decrease with increasing force applied to the protein, and unfolding rates increase. The intersection of the folding and unfolding distributions occurs at ~12 pN, and represents the force at which the protein has equal probability of being folded or unfolded, a mean lifetime of 28 s for both states. Thus, as it emerges on the surface of the ribosome and folds, Top7 can exert a force of at least 12 pN for many seconds on the nascent chain still in the tunnel, before it unfolds. Based on our single-molecule results, we propose
**CHAPTER 3. MECHANICAL FORCE RELEASES NASCENT CHAIN-MEDIATED RIBOSOME ARREST IN VITRO AND IN VIVO**

44

Figure 3.2: 
**Nascent protein folding near the ribosome tunnel exit can rescue SecM-mediated stalling.**

(A) Primary sequence of the construct used in the GFP reporter assay. 

(B) Schematic illustrating the translation outcome for a short (top), intermediate (middle) and long (bottom) linker. 

(C) UV-illuminated image of colonies transformed with the linker library and grown under inducing conditions. 

(D) Histogram of linker lengths recovered by sequencing of fluorescent colonies. Grey shaded area: Library range.
a kinetic scheme that describes how folding can modulate arrest release rates (Figure 3.4 A): 

\[ R \leftarrow U \leftrightarrow N \Rightarrow R \]

U and N are the unfolded and natively folded states of Top7; F is the force, and R the stall-released ribosome state, which can be accessed from the folded (N) state (at a rate accelerated by the exclusion steric force, \( k_R(F) \)) or from the unfolded (U) state (at the basal, force-independent rate, \( k_0 \)). \( k_N(F) \) and \( k_U(F) \) represent the force-dependent folding and unfolding rates of Top7 obtained from single molecule experiments (Figure 3.15).

The effectiveness with which the pulling force catalyzes stall release depends on the force generated upon folding, the probability that the protein folds at that force, and the lifetime of the folded state. To determine how translation stall release rates depend on these factors, we solved the kinetic scheme above for the effective stall release rate, \( k_{Reff}(F) \) (Supplemental information). An approximate solution that assumes equilibrium between N and U (good for \( k_U \gg k_R \)) yields (See supplemental information for the exact solution):

\[ k_{Reff}(F) = f_0(k_R + \frac{k_U k_0}{k_N}) \]

where \( f_0 \) is the fraction of natively folded protein assuming equilibrium with the unfolded state. \( f_0 \) decreases with force as the equilibrium is tilted towards the unfolded state; in contrast, the term in parentheses increases with force as the release rate increases and the protein is biased towards the unfolded state (Figure 3.16). Due mainly to the difference in experimental geometry between the dual-tethered Top7 in the optical tweezers and the singly-tethered Top7 in the in vivo experiments, the force-dependent rates likely represent an underestimate to the true rates. Although the experimentally observed quantity, \( k_{Reff}(F) \), represents the composite kinetics of both the spontaneous and force-dependent release rates, we calculated the probability that release proceeds by the mechanical, as opposed to the spontaneous, process. This function provides the most likely force at which folding of the nascent protein leads to the release of the ribosome stall, and can be expressed as:

\[ P(F) = \frac{k_{Reff}(F) - k_0}{k_{Reff}(F)} \]

A plot of \( P(F) \) vs. force shows that the probability of force-catalyzed stall release is a maximum near 10 pN (Figure 3.4 B). At forces below 12 pN, Top7 is mostly folded (N) so release proceeds largely through the mechanical path, whose rate, \( k_R \), increases with increasing force; at forces higher than 12 pN the protein spends shorter and shorter times in the folded state, so release occurs more and more via the slower, spontaneous path, \( k_0 \), from the unfolded state.

### 3.4 Discussion

Our results provide evidence that the translocon must generate at least 10 pN of force to relieve SecM-induced ribosomal arrest. It appears that another translocase, ClpX, operates
Figure 3.3: Top7 refolds against an applied mechanical load. (A) Example force ramp cycles for a single Top7 molecule. Pulling is shown in red, relaxing in blue. Successive cycles are offset along the x-axis for display purposes. (B and C) Folding and unfolding force distributions, respectively, for Top7 at a pulling speed of 100 nm/s. Black line: Distributions reconstructed from the force-dependent rates in (D). The unfolding-force distribution in (C) is right-censored because the maximum force in pulling experiments was set at 45 pN to avoid tether rupture. (D) Force-dependent rates of folding and unfolding extracted from the distributions in (B and C). Dashed lines: Fit of Bells model to the force-dependent rates. For folding, Δx⁺: 6 nm (95% CI: 4 nm, 8 nm) and k₀: $1 \times 10^6$ s⁻¹ (95% CI: $0.04 \times 10^6$ s⁻¹, $30 \times 10^6$ s⁻¹). For unfolding, Δx⁺: 0.4 nm (95% CI: 0.3 nm, 0.6 nm) and $k₀$: $0.01$ s⁻¹ (95% CI: $0.003$ s⁻¹, $0.03$ s⁻¹).
Figure 3.4: **Kinetic model for folding-induced release of stalled ribosomes.** (A) Kinetic scheme illustrating the pathway to release of translation arrest. The RNC can transit reversibly between the native and unfolded states, with rate constants $k_N(F)$ and $k_U(F)$. Once folded, the nascent protein both generates and experiences a force, “F”, which can drive it either irreversibly to the “released” state, with rate constant $k_R(F)$, or back to the unfolded state. In addition, the stall can be released via the irreversible spontaneous process from the unfolded state, with rate constant $k_0$, which is independent of force. PDB: 1QYS. (B) $P_{force-dependent}$ is plotted as a function of the folding force.
in a similar force regime [113, 114]. Given that a number of polypeptide sequences are known to stall the ribosome [34, 100, 115, 116], our results suggest that force can play a generally important role as a regulator of elongation. In instances where ribosome-nascent chain interactions are less robust than the SecM system, folding could play an important role in modulating translation elongation and vice versa. For example, if such interactions occur near the C-terminus of a newly synthesized domain, elongation would slow down when the polypeptide segment just outside the ribosome begins to acquire stable structure, permitting the folding of this segment to be completed before more of the polypeptide is synthesized. Likewise, folding of the domain, and the force so generated, may provide the signal for speeding up elongation through and beyond the regulatory signal. A force could be generated not only by nascent chain folding, but also by the binding of partners to the nascent chain outside the ribosome. In eukaryotic cells, ribosome profiling experiments have shown that the chaperone Hsp70 relieves global stalling of ribosomes near the beginning of genes [117, 118], perhaps by generating a pulling force similar to the scenario of Hsp70 binding to protein aggregates [119]. Thus, force generated by either nascent chain folding or chaperone binding could constitute an important feedback mechanism to tune elongation to folding.

3.5 Methods

Preparation of SecM-stalled RNCs for optical tweezers experiments

Stalled RNCs with covalently attached dsDNA oligonucleotides and N-terminally biotinylated nascent chains were prepared using the PURExpress Δ Ribosome Kit (New England Biolabs) as described [106] with several modifications. Instead of truncated mRNAs, RNCs were stalled at the SecM sequence. After the initial translation reaction, RNCs were treated with 40 \( \mu \text{M} \) puromycin, to release nascent chains that were not SecM-stalled. RNCs were sedimented through a 1 M sucrose cushion in a polymix buffer containing 20 mM HEPES-KOH pH 7.5, 5 mM MgCl\(_2\), 0.5 mM CaCl\(_2\), 5 mM NH\(_4\)Cl, 95 mM KCl, 1 mM spermidine, 8 mM putrescine, 0.1 mM DTT, 0.01 mM PMSF and 0.1 mM benzamidine, and stored at -80\( ^\circ\) C in the same buffer until the day of experiments.

Optical tweezers experiments with SecM-stalled RNCs

Immediately before experiments, SecM-stalled RNCs modified with a covalently attached dsDNA oligonucleotide were ligated to a dsDNA oligonucleotide on a polystyrene microsphere, as described [106].

To make the DNA “handle” bead needed for forming tethers, an \( \sim 2\text{kb} \) dsDNA was synthesized by PCR with a forward primer labeled with a biotin molecule and a reverse primer labeled with two digoxigenin molecules. The second digoxigenin provides additional
mechanical stability when the tether is brought under tension. Anti-digoxigenin coated beads were incubated at room temperature with the biotin/digoxigenin labeled dsDNA, then with streptavidin, before injection into the optical tweezers chamber. Tethers were formed by bringing the two beads in close proximity to each other. Optical tweezers experiments were performed in a polymix buffer containing 20 mM HEPES-KOH pH 7.5, 5 mM MgCl₂, 0.5 mM CaCl₂, 5 mM NH₄Cl, 95 mM KCl, 1 mM spermidine, 8 mM putrescine, 0.1 mM DTT, 0.01 mM PMSF and 0.1 mM benzamidine, and either in the presence or absence of puromycin, EF-G and GTP.

Only tethers displaying the reversible “hopping” signature of the CaM nascent chain at ~7pN were considered, ensuring tethering of the desired RNCs. After a tether was identified, the force was raised to the force of interest. During data analysis, the time between raising the force and tether rupture was recorded. If tethers displayed any unexpected extension changes during the measurement, the molecule was not included in the analysis, as such extension changes might indicate mechanical disruption of some portion of the assembly.

The rupture lifetimes were well described by a single-exponential, as expected for a single rate-determining step for the reaction. At forces below 30 pN, a fraction of the molecules had not ruptured by our cutoff time of 10 min. We used a right-censoring, single-exponential maximum-likelihood estimator (MLE) to calculate the time constant for stalling. The reported rates of release are the reciprocals of the time constants. We fit the force-dependent rates to Bells model [110]  \[ k(F) = k_0 e^{\frac{F \Delta S}{k_B T}} \] with \( F = \) force, \( k_B = \) Boltzmann constant, \( T = \) absolute temperature.

Puromycin, EF-G and GTP concentration in optical tweezers experiments

For optical tweezers experiments, we used puromycin dihydrochloride from Sigma-Aldrich. According to the manufacturer, \( \lambda_{\text{max}} = 266 - 270 \) nm and \( \varepsilon = 18.5 - 20.5 \) cm\(^{-1} \) mM\(^{-1} \). Using the middle value for both parameters (\( \lambda_{\text{max}} = 268 \) nm and \( \varepsilon = 19.5 \)), we estimate our puromycin concentration at 411 \( \mu \)M. Using the Michaelis-Menten parameters for puromycin reactivity of the posttranslocation state [120], we estimate a rate of 1.9 s\(^{-1} \) for the puromycin reaction, several orders of magnitude faster than the restart rates measured in the optical tweezers experiments.

EF-G was added to a final concentration of 1 \( \mu \)M. Using the Michaelis-Menten parameters for EF-G catalyzed translocation [121], we estimate a rate of 1.1 s\(^{-1} \) for translocation, also several orders of magnitude faster than measured restart rates. GTP was added to a final concentration of 1 mM.

In vivo arrest release

Plasmids for expression of the GFP reporter constructs shown in Figure 3.2 A were generated in the following way: Oligonucleotides encoding linker sequences from 4 to 28 amino acids in
length connecting Top7 and SecM17 were assembled as shown in Figure 3.11. A plasmid was created that encodes Top7, SecM17 and GFPuv under the control of a T7 promoter. SapI restriction sites were introduced near the 3’ end of the Top7 coding sequence and near the 5’ end of the SecM17 coding sequence. Digestion of this plasmid with SapI leaves 3-nucleotide overhangs that are compatible with the terminal overhangs of the oligonucleotides encoding the linker sequences. Thus, the procedure yields a collection of plasmids that encode Top7 followed by a 4 - 28 amino acid linker sequence (in one amino acid increments; see Figure 3.11 for details) that connects to the SecM17 sequence followed by GFPuv.

The plasmid collection was transformed into Lemo21(DE3) cells (New England Biolabs) according to the manufacturer’s instructions. Transformed bacteria were spread on LB agar plates containing 100 µg/ml ampicillin (for maintenance of the Top7-SecM17-GFP expression construct), 30 µg/ml chloramphenicol (for maintenance of the Lemo system) and 500 M of each isopropyl β-D-1-thiogalactopyranoside and L-rhamnose to induce expression at appropriate levels. Colonies were grown for 30-40 hours at room temperature and plates were imaged under UV-illumination. A fraction of the clones exhibited green fluorescence. Blue autofluorescence was observed for colonies that did not exhibit green fluorescence.

We analyzed plasmid DNA sequences for a set of colonies that were picked based on their green fluorescence under UV illumination (Figure 3.2 D). To obtain the distribution of linker lengths irrespective of green fluorescence (Figure 3.13), we picked colonies after growing them under repressive conditions. For analysis of plasmid DNA, individual colonies were subjected to colony PCR using primers that bind in the T7 promoter region of the plasmid (forward primer) and in the GFPuv coding sequence (reverse primer), generating PCR products of ~900 bp in size. PCR products were subjected to standard DNA sequencing. Sequencing reads covered the sequences encoding Top7, the linker region, the SecM stalling sequence, and the N-terminal region of GFP. The linker lengths reported in Figure 3.2 D and 3.13 were determined from this sequence information.

**Optical tweezers force ramp experiments for Top7**

The Top7 open reading frame was subcloned into a vector with an N-terminal Avi-tag and a C-terminal ybbR tag. Top7 was expressed, purified, derivatized and attached to polystyrene beads as described [106]. To make the second bead needed for forming tethers, an ~2kb biotinylated dsDNA strand was synthesized on the bead in a PCR-like reaction as described [106]. These beads were incubated with streptavidin before injection into the optical tweezers chamber. Tethers were formed by bringing the beads in close proximity to each other. Experiments were performed in 25 mM HEPES-KOH pH 7.5, 150 mM KCl and 5 mM Mg(OAc)$_2$.

**Observed contour length of Top7 in optical tweezers experiments**

The identity of Top7 in single-molecule pulling experiments was confirmed by the contour length of the molecule calculated from observed folding and unfolding transitions. Applying
the worm-like chain model to our experimental data, we calculated a contour length of 35.0 ± 2.3 nm (SD) for folding, and 33.7 ± 1.4 nm (SD) for unfolding. The expected contour length is 0.365 nm/aa ×93 aa = 33.9 nm.

Transformation of folding force histogram to force-dependent rates

In order to extract the force-dependent folding rates from the folding force histogram for Top7, we modified equation [10](ref [72]) (originally prescribed for unfolding) for folding. Rates were calculated by changing the limits of integration and imposing a negative pulling velocity to account for the relaxation of the polypeptide from longer to shorter extensions.

Transformation of right-censored unfolding force histogram to force-dependent rates

To determine force-dependent unfolding rates from the unfolding force histogram for Top7, we modified equation [10](ref [72]) for a right-censored unfolding distribution. Modification of the theory was necessary because we only sampled a portion of the unfolding distribution in our force ramp experiments, in order to avoid tether rupture at high forces. Thus, for a fraction of force ramp pulling cycles, we did not observe an unfolding event. In such a case, the unfolding event was classified as occurring at a force higher than the maximum force sampled in the experiment. In calculating the force dependent rates according to equation [10], we included these events in the summation term. Rates were only calculated for forces less than the maximum force.

Estimating the effective stall release rate, $k_{Reff}(F)$

As opposed to the scenario on the ribosome in which the nascent protein is tethered only by its C-terminus and excluded from the tunnel, a protein tethered by both termini can be mechanically denatured from either end. Thus, the force needed to unfold the protein when pulling it from both ends will represent a lower bound for the force required when the molecule is only tethered by one end and excluded from the tunnel. Similarly, the force developed by the folding protein in the optical tweezers geometry represents a lower bound for that generated by the nascent chain on the ribosome. Additionally, a fraction of ribosomes are likely inactive and are not released from stalling by the application of force in the optical tweezers; thus, we expect that the determined release rates represent an underestimate of the true values. For both of the aforementioned reasons, the calculated rate of release in response to folding likely underestimates the rate in vivo.
Approximate solution for $k_{Reff}(F)$ assuming equilibrium between N and U

An approximate solution for $k_{Reff}(F)$, assuming equilibrium between N and U, is valid for $k_U \gg k_R$. This condition is satisfied for the relevant force range (see Figure 3.15). The set of time-dependent differential equations describing the kinetic scheme is:

$$\frac{\partial U(t, F)}{\partial t} = -k_N(F) \times U(t, F) - k_0 \times U(t, F) + k_U(F) \times N(t, F)$$

$$\frac{\partial N(t, F)}{\partial t} = -k_U(F) \times N(t, F) - k_R(F) \times N(t, F) + k_N(F) \times U(t, F)$$

$$\frac{\partial R(t, F)}{\partial t} = k_R(F) \times N(t, F) + k_0 \times U(t, F)$$

Because $k_U \gg k_R$ for the force range under consideration, we can define: $K_{eq} = \frac{k_N(F)}{k_U(F)}$ Using this $K_{eq}$, $\frac{\partial U(t, F)}{\partial t}$ can be expressed solely as a function of $U(t, F)$ and we can integrate with respect to time to find the solution for $U(t, F)$:

$$U(t, F) = U(0, F) e^{-k_0 t}$$

And similarly, for $N(t, F)$:

$$N(t, F) = N(0, F) e^{-k_R(F) t}$$

Substituting these expressions and $K_{eq}$ into the expression for $\frac{\partial R(t, F)}{\partial t}$:

$$\frac{\partial R(t, F)}{\partial t} = N(0, F)(k_R(F)e^{-k_R(F)t} + \frac{k_0}{K_{eq}}e^{-k_0 t})$$

Because we are interested in the initial rate of entry to “R”, we evaluate at $t = 0$:

$$\frac{\partial R(0, F)}{\partial t} = N(0, F)(k_R(F) + \frac{k_0}{K_{eq}})$$

The generic rate law is of the form:

$$\frac{\partial R(0, F)}{\partial t} = k_{Reff}(F)(N + U)$$

Thus, to get $k_{Reff}(F)$, we divide $\frac{\partial R(0,F)}{\partial t}$ by $(N + U)$ and simplify:

$$k_{Reff}(F) = f_0(k_R + \frac{k_Uk_0}{k_N})$$

where $f_0$ is the fraction of natively folded protein.
CHAPTER 3. MECHANICAL FORCE RELEASES NASCENT CHAIN-MEDIATED
RIBOSOME ARREST IN VITRO AND IN VIVO

Exact solution for $k_{Reff}(F)$

The effective release time, $\tau_{Reff}(F)$, can be expressed as an infinite sum over all possible paths to reach the released state, with the time of each path weighted by the probability of taking that path. Since the released state can be reached either from the unfolded state via the spontaneous process, or from the folded state via the force-dependent process, there are two sum terms. Here, we assume the protein starts in the unfolded state, although since $k_U(F) \gg k_R(F)$ over the force range considered, the results do not depend significantly on the initial state of the protein.

$$\tau_{Reff}(F) = \sum_{i=0}^{\infty} p_1((1-p_1)(1-p_2))^i(i(\tau_U + \tau_N) + \tau_U) + \sum_{i=0}^{\infty} p_2(1-p_1)^{i+1}(1-p_2)^i((i+1)(\tau_U + \tau_N))$$

Here, $p_1$ is the branching probability for transiting from the unfolded state to the released state, and $p_2$ is the branching probability for transiting from the folded state to the released state:

$$p_1 = \frac{k_0}{k_0 + k_N}$$
$$p_2 = \frac{k_R}{k_R + k_U}$$

$\tau_U$ and $\tau_N$ are the lifetimes of the unfolded and folded states, respectively:

$$\tau_U = \frac{1}{k_0 + k_N}$$
$$\tau_N = \frac{1}{k_U + k_R}$$

The sums converge to give:

$$\tau_{Reff}(F) = \frac{k_N + k_R + k_U}{k_0(k_U + k_R) + k_Rk_N}$$

Thus,

$$k_{Reff}(F) = \frac{1}{\tau_{Reff}(F)} = \frac{k_0(k_U + k_R) + k_Rk_N}{k_N + k_R + k_U}$$

This expression simplifies to the approximate solution (see section above) in the limit $k_U \gg k_R$. 

Figure 3.5: **Primary sequence diagram for optical tweezers experiments.** The C-terminus of CaM is positioned 58 aa residues downstream of the peptidyl-tRNA at the point of SecM stalling, allowing the domain to emerge from the ribosome exit tunnel and away from the surface of the ribosome. The arrow indicates the stalling position of ribosomes on the mRNA template. Following the terminal stalling proline, there are 34 sense codons and a stop codon.
Figure 3.6: **Preparation of SecM-stalled ribosomes.** Stalled ribosomes were prepared in an in vitro translation system (PURE Express, NEB). Ribosomes were pelleted for 40 minutes at 200,000 g. The pellet was re-dissolved in polymix buffer and allowed to react with puromycin, EF-G and GTP or RNaseA for 10 minutes. The upper band is RNase-sensitive, indicating it is a peptidyl-tRNA species and is ribosome-bound. However, the same band is resistant to puromycin, indicating that it represents a SecM-stalled nascent chain. Nascent chains are biotinylated via an Avi tag at the N-terminus of the protein and are detected by western blotting with a streptavidin-horseradish peroxidase probe after PAGE gel electrophoresis.
Figure 3.7: Arrest release of SecM-stalled ribosomes in bulk after prolonged time. (A) Bulk arrest release assay in the absence of force. In an initial translation reaction using the PURE system, nascent chains are 35S-labeled and ribosomes translate to the SecM stall site. The reaction is then diluted into a polymix buffer containing 1 mM GTP, 1 M EF-G and 500 M puromycin. Arrest release is apparent as conversion of peptidyl-tRNA to peptidyl-puromycin in a gel-shift assay. (B) Fit of the peptidyl-tRNA to a single-exponential decay function, and the peptidyl-puromycin to a single exponential cumulative distribution function. The reported rate of \((3.9 \pm 0.1) \times 10^{-4} \text{ s}^{-1} \) (SEM) is the mean rate of peptidyl-tRNA decay determined from three independent experiments.
Figure 3.8: Schematic for the optical tweezers stall-release experiment showing A, P and E sites of the ribosome. Initially, SecM-stalled RNCs are resistant to treatment with puromycin. In the optical tweezers, force is applied to the nascent chain, catalyzing peptidyl transfer of the nascent polypeptide to the prolyl-tRNA<sub>Pro</sub> in the A site of the ribosome. Translocation is then catalyzed by EF-G, freeing the A site for reactivity with puromycin. The puromycin reaction leads to tether rupture. The observed time for tether rupture in the optical tweezers is a convolution of three reactions in series; however, the last two steps are very fast compared to the measured times. The rates shown for the EF-G and puromycin reactions are calculated from Michaelis-Menten parameters [120, 121].
Figure 3.9: Control experiments in the absence of puromycin, EF-G and GTP (polymix buffer only) revealed a low rate of background rupture due to breakage of the molecular assembly for reasons other than stalling rescue.
CHAPTER 3. MECHANICAL FORCE RELEASES NASCENT CHAIN-MEDIATED RIBOSOME ARREST IN VITRO AND IN VIVO

A.

B.

C.

D.

Figure 3.10 (preceding page): Alternative method for calculating force-catalyzed arrest release rates takes into account non-specific tether rupture. Correcting for nonspecific rupture events did not significantly affect the measured rates. (A) Cumulative probability distributions for tether rupture at each force in puromycin/EF-G/GTP (solid lines) or buffer only (dashed lines). (B) Cumulative rupture probability curves are re-plotted, this time taking into account non-specific rupture probability. The buffer-only distributions are subtracted from the puromycin/EF-G/GTP distributions at each respective force, and the resulting curves are re-normalized by the probability of tether survival in the buffer-only case. The resulting distributions are fit to a single-exponential cumulative distribution function to determine the force-dependent lifetimes. Color coding as in (A) (C) Error is determined by bootstrapping. At each force, N lifetimes are drawn from the pool of experimentally measured lifetimes at that force, where N is the number of measurements in the data set. In this sampling process, each lifetime that is drawn is subsequently replaced in the pool of lifetimes, so it is possible to sample each point more than once (sampling with replacement). This new set of lifetimes represents a bootstrapped dataset, and is done for both the puromycin/EF-G/GTP condition and the buffer-only condition. Curves are re-normalized and fit to extract the lifetime, as described in (A). This procedure is repeated 500 times at each force, and the resulting lifetimes are plotted in a histogram. Color coding as in (A) (D) Comparison of rates determined by MLE fitting (red) and the method described here (black). Blue dot: Rate of release measured in bulk in the absence of force.
Figure 3.11: Assembly of oligonucleotides encoding linker sequences. (A) Oligonucleotide sequences encoding GS-linker polypeptides that connect Top7 and SecM17. Annealed double-stranded oligonucleotides are shown with the translated sequence below. The 3’- and C-terminal portion of Top7 encoded by the oligonucleotides are colored orange, the 5’- and N-terminal portion or SecM17 in red. (B) Scheme for generating fragments encoding linker sequences from 4 amino acids ($n = 0, I_1$) to 28 amino acids ($n = 4, I_5$) in length. Oligonucleotides $I_1 - I_5$ contain the 3’-terminal sequence of the Top7 ORF followed by a 4 nucleotide overhang. This overhang is complementary to one overhang of oligonucleotide $E$, allowing annealing of $I_1 - I_5$ to $E$. The other end of $E$ can anneal to either $E$ or $T$. Ligation of the annealed oligonucleotides yields fragments with 3-nucleotide overhangs that were ligated into the a plasmid backbone that allows expression of Top7-linker-SecM17-GFP fusion proteins.
Figure 3.12: The ribosome does not influence folding of nascent Top7. Blue: Force-dependent folding lifetimes for Top7 (as shown in main figure 3D). Black: Ribosome-bound nascent Top7 with 100 aa between the C-terminus of Top7 and the P site peptidyl-tRNA. Red: Ribosome-bound nascent tandem Top7. For the tandem case, two Top7 domains have been translated, with 40 aa between the C-terminus of the C-terminal domain and the P site peptidyl-tRNA. The C-terminal domain is thus barely outside the ribosome exit tunnel. We do not observe a difference in folding rates for this construct compared to free Top7, despite its close proximity to the ribosome surface. Folding rates were calculated by modifying equation [10] (ref [72]) (originally prescribed for unfolding) (see Materials and Methods).
Figure 3.13: Sequencing plasmid DNA from colonies irrespective of whether or not they exhibited fluorescence recovered the entire distribution of linker lengths in the plasmid collection.

Figure 3.14: Experimental geometry for pulling experiments with single Top7 molecules. The structure of Top7 was obtained from PDB ID 1QYS.
Figure 3.15: The force-dependent rates used in the kinetic scheme are plotted together. $k_N(F)$ and $k_U(F)$ were determined by Bells model fits to the single-molecule data with the Top7 protein (Figure 3.3 D). $k_R(F)$ was determined by a Bells model fit to the single-molecule SecM restart data (Figure 3.1 D). $k_0$ is the rate at zero-force (given by the intersection of the black line with the y-axis).
Figure 3.16: The effective restart rate as a function of force, \( k_{Reff}(F) \), is a peaked distribution, with a maximum at \( \sim 10 \) pN.
Chapter 4

Development of an Optimized In Vitro Translation System for Single-Molecule Experiments

4.1 Introduction

In vitro translation systems have been used extensively to study various aspects of protein synthesis under well-defined conditions. For example, the concentrations of translation factors and small molecules can be varied in order to determine the kinetics of translation sub-steps. One vexing issue surrounding the use of such systems is that the observed rates of protein synthesis can vary widely between in vitro and in vivo systems. Specifically, while in vivo rates have been estimated at $\sim 10$ aa/s [21], rates in in vitro single-molecule experiments have not exceeded 1 aa/s [20, 122]. Anecdotally, a number of laboratories have struggled, with limited success, to increase translation rates.

This issue has raised two concerns. First is the issue of physiological relevance. A ribosome that translates more than one order of magnitude slower than in the cell may lead to misinterpretation of results. For example, if one is investigating co-translational folding of the nascent chain, slow translation may lead to alternate folding pathways or misfolding that is not observed in vivo. Secondly, reduced translation rates lead to technical difficulties in single-molecule experiments. For example, in optical tweezers experiments, monitoring polypeptide elongation is challenging due to the small increase in length upon addition of an amino acid. In many instruments, this rate of increase in length is on the same order as the drift of the instrument. This makes detection problematic. Additionally, in fluorescence experiments, photo-bleaching of fluorophores limits the experimental observation time. For these reasons, it is important to achieve a reconstituted, in vitro translation system with in vivo-like rates for single-molecule experiments.

Recent studies from the laboratory of Dr. Måns Ehrenberg have described a speed-optimized in vitro translation system [123], and have systematically dissected the effects of
buffer [123], pH [17] and [Mg\(^{2+}\)] [18, 19] on translation rates. In collaboration with Dr. Ehrenberg, we sought to determine why translation is slow in single-molecule experiments, and what can be done to increase rates. While part of the “rate gap” is due to temperature (many single-molecule experiments are carried out at 25° C, instead of 37° C), this likely does not account for the full discrepancy. Because single-molecule experiments are time-consuming and many molecules must be obtained under each condition to make conclusions, we decided to use the ensemble techniques developed in Dr. Ehrenberg’s laboratory for rate-optimization experiments. Most of the work described in the remainder of this chapter was carried out in the Ehrenberg laboratory in the Department of Cell and Molecular Biology at Uppsala University in Sweden.

4.2 Results

Maximal translation rates measured in single-molecule optical tweezers experiments are \(~0.5\) aa/s (Figure 4.1) [20]. This is most likely not related to the single-molecule setup itself (photo-damage, surface effects, etc.), as an ensemble gel-shift assay yielded similar average rates (Figure 4.2). Since the Ehrenberg laboratory has achieved near-\textit{in vivo} rates at 37° C, we started our study by performing experiments in the standard translation system of their laboratory. The results of these experiments at 25° C would provide us with a benchmark of the maximum \textit{in vitro} rates that we could hope to achieve at this temperature.

We started by measuring the kinetics of dipeptide formation at 25° C. By “dipeptide formation”, we actually mean all of the steps starting with initial selection and ending with peptidyl transfer, excluding translocation (Figure 4.3). Experiments were performed in a quench flow instrument, which rapidly mixes two solutions and then quenches the reaction, in this case with an acid solution. With this instrument, the reaction can be quenched as quickly as 2 ms after mixing. After the samples are obtained, the peptidyl-tRNA is hydrolyzed with a base solution, and the peptide products are separated and quantified by High-Performance Liquid Chromatography (HPLC) (Figure 4.4). These experiments were carried out similarly to what has been previously described [123], using “XR7” mRNA, in which the first three codons are AUG UUU UAC, which codes for fMet-Phe-Tyr. The two solutions that are mixed in the quench flow instrument are the “initiation mix”, which contains initiated ribosomes with fMet-tRNA\(^{fMet}\) in the P site, and the “elongation mix”, which contains, among other things, EF-Tu·GTP·Phe-tRNA\(^{Phe}\). The buffer used is the standard phosphate-buffered polymix used in the Ehrenberg laboratory [123], and in dipeptide formation experiments, ternary complexes were always present in excess over initiated ribosomes.

The rate of dipeptide formation was monitored at four different ternary complex concentrations, and single exponential accumulation functions were fit to the data (Figure 4.5). Rates increased with increasing concentration in a Michaelis-Menten fashion, with \(K_M = 1.5\) \(\mu\text{M}\) and \(k_{cat} = 25\ \text{s}^{-1}\).

Since these results exclude the translocation step, we next sought to measure translocation rates under saturating concentrations of EF-G (20 \(\mu\text{M}\)). In order to measure the rate
Figure 4.1: Results of a EF-Tu·GTP·aa-tRNA (ternary complex) titration on translation rates [20] at what was measured to be saturating [EF-G] in these experiments. Translation is measured in a single-molecule optical tweezers assay by tethering the ends of an mRNA hairpin to polystyrene microspheres. When the ribosome translates a codon within the hairpin, it releases 6 bases of single-stranded RNA, which is observed as a change in distance. Michaelis-Menten curves are fit to the data. $v_{ss}$: velocity on single-stranded mRNA. $v_{ds}$: velocity on double-stranded mRNA. For $v_{ss}$: $V_{max} = 0.47 \pm 0.04$ codons/s, $K_m = 5 \pm 2$ nM. For $v_{ds}$: $V_{max} = 0.28 \pm 0.02$ codons/s, $K_m = 2.5 \pm 0.7$ nM.

of translocation (which in this context includes EF-G binding, GTP hydrolysis and the actual movement of the mRNA), we implemented a recently described tripeptide formation assay [18]. The method is outlined in Figure 4.6.

The results of the tripeptide formation assay are shown in figure 4.7. In order to extract kinetic rates for the steps outlined in figure 4.6, we solved the following system of differential equations describing the kinetic scheme:

$$\frac{d[fM]}{dt} = -k_{peptide1} \times [fM]$$

$$\frac{d[fMF]}{dt} = k_{peptide1} \times [fM] - k_{translocation} \times [fMF]$$

$$\frac{d[fMF*]}{dt} = k_{translocation} \times [fMF] - k_{peptide2} \times [fMF*]$$

$$\frac{dfMFY}{dt} = k_{peptide2} \times [fMF*]$$
Figure 4.2: Translation rates in the PURE system (a commercial, \textit{in vitro} translation system purchased from New England Biolabs (NEB)) are comparable to those obtained in single-molecule experiments. (A) In the PURE system depleted of cysteine, the ribosome translates to the first cysteine codon in the mRNA. The mRNA contains (from N to C-terminus): an Avi tag for biotinylation (used in other experiments), a flexible serine/glycine linker, two domains of the \textit{de novo} designed protein Top7 \cite{112}, two cysteines, followed by 107 sense codons (the mRNA lacks a stop codon, so that ribosomes that reach the end of the template stall, preventing re-initiation). The ribosome is given 40 min to initiate and translate to the cysteine stalling site (“mid-stalled”). Then, 35S-labeled cysteine and unlabeled cysteine are added to the reaction. The reaction is incubated at 25° C and timepoint samples are taken into an RNaseA/EDTA solution to remove the tRNA moiety, then added to SDS sample buffer. The samples are run on a PAGE gel, exposed to a phosphor screen and imaged using a typhoon instrument. (B) The relative signal of the “terminally-stalled” product is quantified and plotted as a function of time (red dots). In order to estimate rates, a Monte Carlo simulation was performed. Many single ribosome trajectories are simulated and the accumulation of full-length product is plotted as a function of time. The time constant for the reaction, \( \tau \), was varied until the results of the simulation overlaid with the experimental results. Using this method, \( \tau = 1.7 \text{ s} (k = 0.6 \text{ s}^{-1}) \).
CHAPTER 4. DEVELOPMENT OF AN OPTIMIZED IN VITRO TRANSLATION SYSTEM FOR SINGLE-MOLECULE EXPERIMENTS

Figure 4.3: The translation substeps starting with initial selection and ending with peptidyl transfer are probed in “dipeptide formation” experiments [123].

Figure 4.4: A schematic for quench flow experiments to monitor the progress of a translation reaction.
Figure 4.5: (A) At four different concentrations of ternary complex, the progress of dipeptide formation was monitored over time. Single-exponential accumulation functions were fit to the data to extract rates. (B) Rates extracted from the data shown in (A) were fit by a Michaelis-Menten function, with $K_M = 1.5 \, \mu M$ and $k_{cat} = 25 \, \text{s}^{-1}$.

We fitted the data to the solution of the system of differential equations to extract kinetic rates (Figure 4.8). This yielded $k_{peptide1} = 13.0 \, \text{s}^{-1}$, $k_{peptide2} = 18.9 \, \text{s}^{-1}$, $k_{translocation} = 2.2 \, \text{s}^{-1}$.

Taken together, dipeptide formation and translocation, as defined in the above experiments, constitute one full cycle of elongation. Combining the $k_{cat}$ for dipeptide formation (25 s$^{-1}$) with the measured rate of translocation under saturating conditions (2.2 s$^{-1}$), we obtain an elongation rate of $\sim 2$ s$^{-1}$ under saturating factor concentrations. This is $\sim 4$-fold faster than the maximal velocity measured in the single-molecule experiments 4.1. This raises the question of what is causing the discrepancy between the single-molecule and the bulk experiments. Recent evidence has established an inverse correlation between free [Mg$^{2+}$] and translocation rates [18]. The single-molecule experiments (Figure 4.1) and the gel-shift assay (Figure 4.2), which both measured rates of $\sim 0.5 \, \text{s}^{-1}$, contained roughly 8 mM and 3 mM free [Mg$^{2+}$], respectively (after accounting for chelation by nucleotide triphosphates). In contrast, the ensemble measurements described in the dipeptide and tripeptide formation assays contained only 1.3 mM free [Mg$^{2+}$] (after accounting for chelation by both nucleotide triphosphates and phosphoenolpyruvate). We next measured the [Mg$^{2+}$]-dependence of translocation and dipeptide formation rates at 25$^\circ$ C.

We first probed the [Mg$^{2+}$]-dependence of $\tau_{total}$, which describes the total time of tripeptide formation (Figure 4.9). We observed an increase of $\sim 8$-fold of $\tau_{total}$ when increasing the [Mg$^{2+}$] from 1.3 mM to 4.6 mM. Because $k_{translocation} \ll k_{peptide}$, this difference is likely explained by a change in $k_{translocation}$. To verify this hypothesis, we measured the [Mg$^{2+}$]-dependence of $k_{peptide1}$, and found that it was invariant to [Mg$^{2+}$] (Figure 4.10). These results
**CHAPTER 4. DEVELOPMENT OF AN OPTIMIZED IN VITRO TRANSLATION SYSTEM FOR SINGLE-MOLECULE EXPERIMENTS**

![Diagram](image)

**Figure 4.6: A scheme describing a tripeptide formation assay to measure translocation rates.** In this scheme, fM represents the initiated ribosome, with fMet-tRNA$^{fMet}$ in the P site. fMF represents the post-peptidyl transfer, pre-translocation ribosome, with peptidyl-tRNA in the A site and deacylated Phe-tRNA in the P site. fMF* represents the post-translocated ribosome, with peptidyl-tRNA in the P site, and fMFY represents the post-peptidyl transfer ribosome with peptidyl-tRNA in the A site and deacylated Tyr-tRNA in the P site. In the experiment, the goal is to measure $k_{\text{translocation}}$. In order to do this, the progress of three reactions are measured: 1) The time for dipeptide formation (from fM to fMF), 2) The time for tripeptide formation (from fM to fMFY), and 3) The time for the second dipeptide formation (from fMF* to fMFY). For step 1, experiments were performed as described earlier in this chapter. For steps 1 and 2 (measured in parallel in the same reaction), experiments were performed as described earlier, except that tyrosyl-tRNA, tyrosyl-tRNA synthetase and tyrosine were included in the reaction so that the ribosome can synthesize both dipeptide and tripeptide, and the concentration of both dipeptide and tripeptide was monitored at each step. For step 3, ribosomes were stalled after translocation with fMF dipeptidyl-tRNA in the P site of the ribosome, and the concentration of tripeptide was monitored over time.

indicate that translocation rates are highly sensitive to [Mg$^{2+}$], and that high [Mg$^{2+}$] may limit overall elongation rates in single-molecule experiments.

To test whether we could increase translation elongation rates at the single-molecule level by decreasing [Mg$^{2+}$], we performed optical tweezers experiments (Figure 4.11) [20] in which ribosomes translate through a hairpin structure coding for polyvaline (50 sense codons). The translation system contains 1.3 mM free [Mg$^{2+}$], 4 $\mu$M tRNA$^{val}$, 8 $\mu$M EF-Tu, 10 $\mu$M EF-G, 1 $\mu$M EF-Ts, 1 $\mu$M valine-tRNA synthetase, 500 $\mu$M valine, 0.05 mg/ml pyruvate kinase, 10 mM phosphoenolpyruvate, 1 mM DTE, 1 mM ATP, 1 mM GTP and RNase Out (Life Technologies) in a phosphate-buffer polymix [123]. Under these conditions, we observe a mean translation rate of $\sim$1.4 codons/s, an increase in $v_{\text{max}}$ of $\sim$3-fold over previous conditions (Figure 4.12). These results demonstrate that translation rates in single-molecule experiments can be increased by reducing [Mg$^{2+}$].

Our ensemble, single-turnover measurements indicate that under optimal *in vitro* conditions, the ribosome can translate at $\sim$2 codons/s. However, we wanted to verify this...
Figure 4.7: (A) The results for steps 1 and 2 of the tripeptide assay. Red dots: dipeptide; Blue dots: tripeptide; Black dots: the sum of dipeptide and tripeptide. (B) The results of step 3 of the tripeptide assay. Tripeptide formation is measured starting with the ribosome in the post-translocation state and dipeptidyl-tRNA in the P site of the ribosome.

Figure 4.8: (A) Data and fits for steps 1 and 2 of the tripeptide assay. Blue dots: tripeptide formation; Black dots: the sum of dipeptide and tripeptide. (B) Data and fit for step 3 of the tripeptide assay (formation of tripeptide starting from the ribosome in the post-translocation state and dipeptidyl-tRNA in the P site). Fits yielded $k_{\text{peptide}1} = 13.0 \text{ s}^{-1}$, $k_{\text{peptide}2} = 18.9 \text{ s}^{-1}$, $k_{\text{translocation}} = 2.2 \text{ s}^{-1}$.
Figure 4.9:  (A) Data and fits for tripeptide formation at varying $[\text{Mg}^{2+}]$. The data are fit using an expression for the total time of tripeptide formation, $\tau_{\text{total}}$ (a linear combination of $k_{\text{peptide}1}$, $k_{\text{peptide}2}$ and $k_{\text{translocation}}$), which follows from the solution to the system of differential equations describing the kinetic scheme. (B) $\tau_{\text{total}}$, extracted from the fits shown in (A). Because $k_{\text{translocation}} \ll k_{\text{peptide}1}$, the difference in $\tau_{\text{total}}$ is likely due to a difference in $k_{\text{translocation}}$ (this is further reinforced by data in figure 4.10, indicating that $k_{\text{peptide}}$ is invariant to $[\text{Mg}^{2+}]$ in over the concentration range shown here).

Figure 4.10:  (A) Data and fits for dipeptide formation at varying $[\text{Mg}^{2+}]$. The data are fit as in figure 4.5 A. (B) $k_{\text{peptide}1}$, extracted from the fits shown in (A).
observation would hold in multiple-turnover experiments, in which the ribosome translates both a wider variety of codons and proceeds deeper into an open reading frame. For this purpose, we used an mRNA encoding a heptapeptide of seven distinct amino acids (Figure 4.13). We used saturating [EF-G] (20 µM) so that the rate-limiting step of translocation would be maximal. Additionally, we used *E. coli* total tRNA with 20 µM EF-Tu, and added only the aminoacyl-tRNA synthetases needed to decode the message.

Using HPLC, we separated the products of the reaction into seven distinct peaks (Figure 4.14). By solving the kinetic scheme for a six-step sequential reaction and fitting the concentrations of each species to its time-dependent equation, we extracted the kinetic rates for each of the six steps (Figure 4.15). The mean elongation rate for translation of this peptide sequence is \( \sim 1.9 \) codons/s, in line with the single-turnover experiments and the single-molecule experiments performed at 1.3 mM [Mg\(^{2+}\)].

Upon examination of the rates determined in the heptapeptide experiment, we noticed a \( \sim 8 \)-fold variation in elongation rate, depending on the codon translated. Recent studies have suggested that global variations in elongation rate can be explained by a combination of
CHAPTER 4. DEVELOPMENT OF AN OPTIMIZED IN VITRO TRANSLATION SYSTEM FOR SINGLE-MOLECULE EXPERIMENTS

Figure 4.12: Example traces of the single ribosomes translating through an mRNA hairpin at 1.3 mM free [Mg\(^{2+}\)]. The four example traces are offset along the time axis for display purposes. Some “backwards” drift is evident in the traces due to equilibration after flowing the translation mix into the micro-chamber. However, this does not affect the determination of initial and final times, which is used to calculate the average translation rate through the hairpin. Each one-codon step releases 6 bases of ssRNA, which at this force (17 pN) is 2.7 nm. The expected extension change if the ribosome translates all 50 valines is ~135 nm. Mean translation rates, obtained by dividing the number of codons translated by the total time needed to translate, are (from left to right) 0.9, 1.5, 1.3 and 1.9 codons/s.

differential tRNA concentrations and the nature of base-pairing between the codon and anticodon (Watson-Crick vs. wobble) [23]. To see if our observed variations can be explained by these factors, we plotted the elongation rate for each codon against the calculated concentration of cognate ternary complex (based on the fraction of each tRNA in E. coli total tRNA [124] and the concentration of total tRNA in our experiments). Additionally, we identified whether each codon interacted with its cognate tRNA via Watson-Crick or wobble base-pairing. Intriguingly, rates cannot be explained by differential tRNA concentration or the nature of codon-anti-codon base pairing (Figure 4.16). One hypothesis to explain the observed trend is that codon-specific translocation rates, not decoding rates, determine variations in elongation kinetics. The observation that translocation is rate-limiting under saturating concentrations of ternary complex and EF-G fits nicely with this hypothesis. Alternatively, it is possible that in the cell, the concentrations of some tRNAs are sub-saturating and tRNA selection becomes rate-limiting. Further experiments will be needed to determine whether translocation rates vary significantly based on context.
A.

\[
\text{MKITQLF}
\]

...AUG AAA AUC ACU CAG CUG UUU UAA...

B.

\[
fM \xrightarrow{k_1} fMK \xrightarrow{k_2} fMKI \xrightarrow{k_3} fMKIT \xrightarrow{k_4} fMKITQ \xrightarrow{k_5} fMKITQL \xrightarrow{k_6} fMKITQLF
\]

Figure 4.13: (A) Amino acid and mRNA sequence for the heptapeptide experiment. (B) Kinetic scheme for the heptapeptide experiment. The addition of each amino acid is described by a single rate constant, representing a full elongation cycle.

Figure 4.14: **Products from heptapeptide experiment are separated into distinct peaks by HPLC.** Peptides are separated by hydrophobicity; the more hydrophobic the peptide, the longer the elution time.
Figure 4.15: Data and fits for the time-dependent concentrations of products in the heptapeptide experiment. Color-coded data points indicate the normalized concentration of each product (color coding given in upper right of the figure). The data are fit to the time-dependent equations for each species. Rates for each step are given in the figure. Note that $k_1$ is considerably faster than all of the other rate constants because it lacks a translocation step.

4.3 Discussion

In an effort to develop a speed-optimized translation system for single-molecule experiments, we have used ensemble kinetic measurements to determine some of the factors influencing \textit{in vitro} translation rates. We find that at high factor concentration, translocation is the rate-limiting step, and that this step is highly sensitive to $[\text{Mg}^{2+}]$. By decreasing $[\text{Mg}^{2+}]$ to 1.3 mM, maximal translation rates in single-molecule experiments can be increased $\sim$3-fold, to $\sim$1.5 codons/s. Most of the remaining “rate gap” between single-molecule and \textit{in vivo} experiments is likely explained by the difference in temperature ($25^\circ$ C instead of $37^\circ$ C). Experiments performed at low-$[\text{Mg}^{2+}]$ concentrations are thus likely to be more physiologically relevant. Additionally, the faster rates achieved under these conditions will make detection easier in both single-molecule fluorescence and optical tweezers experiments. Interestingly, we find that variations in elongation rate along our short open reading frame cannot be explained by differential tRNA concentration or the nature of codon-anti-codon base pairing. This observation, combined with the finding that translocation is the rate-limiting step at
Figure 4.16: Rate variation in heptapeptide experiment is not explained by differential tRNA concentration or wobble base-pairing. The concentration of each cognate ternary complex is calculated by multiplying the total tRNA concentration by the fraction of each tRNA present in total tRNA [124]. Red points indicate wobble base-pairing.

High factor concentrations suggests that it is not necessarily helpful to attempt to increase rates by "codon optimization"—that is, the choice of synonymous codons that have supposedly faster decoding rates. Further ensemble and single-molecule experiments will be needed to determine what factors influence translocation rates.
Chapter 5

Conclusion

In this work, we have investigated several aspects of protein synthesis by the ribosome. On the one hand, the ribosome can modulate folding of nascent polypeptide chains. This finding, as well as the development of a technique to probe nascent protein folding, pave the way for a deeper understanding of how proteins fold in the cell. We have also demonstrated that mechanical force applied to the nascent chain can release peptide-mediated ribosome stalling. These results establish a regulatory role for force during translation. Such a force might be generated by a protein folding near the exit tunnel of the ribosome, as observed in our studies with the Top7 protein. Combined with nascent peptide sequences that reduce translation rates by interacting with the exit tunnel, such a force might help tune elongation rates to folding transitions of the nascent chain.

We foresee a number of exciting ramifications of our findings. Many proteins likely begin to fold *de novo* during synthesis by the ribosome. Thus, the folding landscape is dynamic with time as the polypeptide emerges from the exit tunnel, causing an evolution of accessible states and interactions with the ribosome. The folding-on-the-ribosome work described here was performed using T4 Lysozyme, which is a single-domain protein that requires that its entire sequence be outside of the ribosome to observe any folding whatsoever. We are investigating how the folding landscape changes as a function of chain length for more complex, multi-domain proteins. Such proteins will likely allow observation of on- and off-pathway folding intermediates. We want to understand how folding is mediated by intra-chain and ribosome-nascent chain interactions. Does one domain begin to fold during synthesis of the other domain? Does the ribosome change the folding pathway of either domain relative to the free domain? Do domains interact with each other during synthesis before folding to the native state? Can we observe folding transitions involving sub-domain structures before a full domain has emerged from the ribosome?

In another layer of complexity, the rate of translation, and thus the emergence of the nascent chain from the ribosome, does not occur at a constant rate. Rather, protein synthesis rates vary widely depending on codon usage, mRNA secondary structure and ribosome-nascent chain interactions within the exit tunnel. These variations have consequences for regulation of gene expression and folding of the nascent polypeptide. We plan on investigat-
ing how translation rates are coupled with folding transitions of the nascent chain. Does the ribosome slow down in order to allow folding of newly synthesized domains? If translation kinetics are modified, for example by synonymous mutations in the gene, does the folding pathway change? Does the protein misfold? To answer these questions we are adapting our experimental setup to follow nascent chain elongation in real time. We will be able to detect folding and elongation simultaneously, and to probe the coupling between these processes. For the proposed experiments, the speed-optimized \textit{in vitro} translation system described in chapter 4 will provide more physiologically relevant elongation rates.

Finally, folding of the nascent chain depends on co-translational events such as binding of chaperones, translocation across or insertion into membranes, and docking of cis- or trans-protein domains. Ultimately, we want to understand the process of folding as it occurs in a cellular environment, taking into account all of the features of a co-translational energy landscape. Additionally, co-translational events can generate mechanical forces on the nascent chain. We want to understand how the ribosome responds to these mechanical forces in different contexts. How does the ribosome respond to mechanical force in the absence of peptide stalling sequences? Does force reduce or increase translation rates? Now that many peptide stalling sequence are known, what is the regulatory function of these sequences? Which peptide-stalling sequences are sensitive to force? Do some of these sequences respond to folding of the nascent chain? While some of these questions can be addressed by single-molecule methods, other complementary techniques, namely system-wide experiments such as ribosome profiling, will provide insight into these processes. We are looking forward to exciting discoveries in the nascent protein folding and translational regulation fields.
Bibliography


