Cotranslational pulling forces alter outcomes of protein synthesis

Thesis by

Matthew Holden Zimmer

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended July 21st, 2021

ACKNOWLEDGEMENTS

Although there is only one name on the title page, this thesis is a reflection of the support and advice of many individuals to whom I am deeply grateful. Thanks first of all to Tom, my advisor, for guiding both my projects and my development as a scientist far beyond where they could have gone without him. Michiel's mentorship throughout the first half of my PhD was truly invaluable and even after his graduation he remained an essential resource for all things biophysics and gaming. Bil deserves special mention for the number of roles he played supporting and guiding me throughout my graduate career, including as a collaborator, committee chair, and BMB option representative. Thanks also to my collaborators, particularly Jonathan, Rebecca, and Gunnar, for broadening the perspectives of my science and this thesis. My PhD experience would not have been nearly as fun or productive without the amazing Miller group members, both past and current.

No matter the ups and downs of research, my wife Lulu was always present and always a comfort. It is hard to imagine grad school without her love, support, and encouragement. My parents also played an essential role, guiding me on the path toward exploration and creativity from a young age and showing me that discovery is not only possible, but fun.

For all my family, friends, and colleagues, mentioned and unmentioned, again, thank you.

ABSTRACT

As nascent proteins are synthesized by the ribosome, interactions between the nascent protein and its environment can create pulling forces that are transmitted to the ribosome's catalytic center. These forces can affect the rate and outcomes of translation. We use atomistic and coarse-grained simulation to characterize the origins of pulling forces, the propagation of force to catalytic center of the ribosome, and the effects of force on synthetic outcomes. We uncover a novel form of pulling force-mediated regulation in which the forces generated by the integration of a transmembrane helix induce frameshifting in a viral polyprotein. Computational force measurements of hundreds of mutant viral sequences in combination with deep mutational scanning experiments reveal the structural and sequence-level features that enable this powerful regulatory mechanism. Force measurements are also used to provide a molecular picture for complex pulling force experiments on multispanning membrane proteins. In particular, we identify signatures of cotranslational helix packing interactions and the translocation of surface helices. To understand how forces are propagated through the nascent protein in the ribosomal exit tunnel, we ran and analyzed hundreds of microseconds of atomistic molecular dynamics with an applied pulling force on the nascent protein. The simulations reveal how the secondary structure of nascent proteins and their interactions with the ribosome control force propagation. The inhibition of force transduction by nascent protein-ribosome interactions explains how amino acids tens of angstroms away from the catalytic center of the ribosome can still influence the force-induced restart of stalled ribosomes.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Patrick J Carmody, Matthew H. Zimmer, Charles P Kuntz, Haley R Harrington, Kate E Duckworth, Wesley D Penn, Suchetana Mukhopadhyay, Thomas F Miller III, and Jonathan P Schlebach. Coordination of -1 Programmed Ribosomal Frameshifting by Transcript and Nascent Chain Features Revealed by Deep Mutational Scanning. *bioRxiv*, 2021. doi: 10.1101/2021.03.11.435011.
MHZ and PJC contributed equally to this work. MHZ conceptualized and performed all simulations, prepared and analyzed simulated and experimental data, and participated in the writing of the manuscript.

[2] Haley R. Harrington, Matthew H. Zimmer, Laura M. Chamness, Veronica Nash, Wesley D. Penn, Thomas F. Miller III, Suchetana Mukhopadhyay, Jonathan P. Schlebach, Thomas F. Miller, Suchetana Mukhopadhyay, and Jonathan P. Schlebach. Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein. *Journal of Biological Chemistry*, 295 (20):6798–6808, 2020. doi: 10.1074/jbc.RA120.012706.
MHZ conceptualized and performed all simulations, prepared and analyzed computational data, and participated in the writing of the manuscript.

[3] Felix Nicolaus, Ane Metola, Daphne Mermans, Amanda Liljenström, Ajda Krč, Salmo Mohammed Abdullahi, Matthew H. Zimmer, Thomas F. Miller III, and Gunnar von Heijne. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *eLife*, 10:1–16, 2021. doi: 10.7554/eLife.64302.
MHZ conceptualized and performed all simulations, prepared and analyzed computational data, and participated in the analysis of experimental data and preparation of the manuscript.

[4] Matthew H. Zimmer, Michiel J.M. Niesen, and Thomas F. Miller. Force transduction creates long-ranged coupling in ribosomes stalled by arrest peptides. *Biophysical Journal*, 2021. doi: 10.1016/j.bpj.2021.03.041.
MHZ conceptualized and performed the research, prepared and analyzed data, and wrote the manuscript.

TABLE OF CONTENTS

Acknowledgements . . . iii
Abstract . . . iv
Published Content and Contributions . . . v
Table of Contents . . . v
List of Illustrations . . . vii
List of Tables . . . ix
Nomenclature . . . x
Chapter I: Introduction . . . 1
Chapter II: Force transduction creates long-ranged coupling in ribosomes stalled by arrest peptides . . . 4
  2.1 Introduction . . . 5
  2.2 Results and discussion . . . 7
  2.3 Conclusions . . . 23
  2.4 Methodology . . . 24
Chapter III: Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein . . . 29
  3.1 Introduction . . . 30
  3.2 Results and discussion . . . 32
  3.3 Conclusions . . . 39
Chapter IV: Coordination of -1 Programmed Ribosomal Frameshifting by Transcript and Nascent Chain Features Revealed by Deep Mutational Scanning . . . 48
  4.1 Introduction . . . 48
  4.2 Results . . . 50
  4.3 Discussion . . . 61
  4.4 Methodology . . . 64
Chapter V: Residue-by-residue analysis of cotranslational membrane protein integration in vivo . . . 66
  5.1 Introduction . . . 67
  5.2 Results . . . 67
  5.3 Discussion . . . 74
  5.4 Methodology . . . 75

LIST OF ILLUSTRATIONS

Number *Page*
2.1 SecM simulation system . . . 8
2.2 Identifying states in restarting pathway . . . 10
2.3 State assignment with differing pulling velocity . . . 10
2.4 Dihedral angle changes in restart pathway . . . 11
2.5 Identifying state corresponding to restart of synthesis . . . 12
2.6 Inferring energy landscape of stall breaking . . . 15
2.7 Graphical depiction of SecM energy landscape . . . 16
2.8 Mutation of key SecM amino acids . . . 19
2.9 VemP restarting pathway . . . 20
2.10 Tension propagation in VemP . . . 21
2.11 Dihedral angle straightening in VemP restart pathway . . . 22
3.1 Structure and topological properties of the alphavirus polyprotein . . 31
3.2 Topological properties of the major form of the SINV structural polyprotein . . . 43
3.3 Putative model for the interplay between translocon-mediated membrane integration and ribosomal frameshifting . . . 44
3.4 Influence of sequence modifications on -1 programmed ribosomal frameshifting . . . 45
3.5 Forces and topologies from simulated polyprotein biosynthesis . . . 46
3.6 Model for the interplay between topology, pulling force, and programmed ribosomal frameshifting . . . 47
4.1 Generating and probing a library of polyprotein mutants using deep mutations scanning . . . 51
4.2 Impact of slip-site and stem-loop mutations on ribosomal frameshifting 52
4.3 Impact of nascent chain mutations on ribosomal frameshifting . . . 54
4.4 Impact of TM2 mutations on -1PRF, membrane integration energetics, and pulling forces . . . 55
4.5 Structural context and conformational dynamics of nascent TM2 . . . 57
4.6 Impact of codon modifications on ribosomal frameshifting . . . 58
4.7 Impact of translation kinetics on nascent chain pulling forces . . . 59

4.8 Correlation between tRNA abundance and PRF after controlling for hydrophobicity 60
5.1 Force profiling technique 68
5.2 EmrE force profile 69
5.3 GlpG force profile 77
5.4 BtuC force profile 78

LIST OF TABLES

Number *Page*
2.1 Optimized parameters describing SecM energy landscape 16
2.2 Number of SecM simulations per pulling velocity and mutant 27
3.1 Simulated integration forces and topologies of related alphavirus polyproteins 39
4.1 Regression statistics when predicting frameshifting rates of TM3 mutants 61

NOMENCLATURE

f_{FL}. Fraction full-length, referring the fraction of restarted ribosomes during a arrest peptide experiment.

AA. Amino acid.

AP. Arrest peptide.

CGMD. Coarse-grain molecular dynamics.

CI. Confidence interval.

Cotranslational. Occurring during the process of protein synthesis.

Cryo-EM. Cyrogenic electron microscopy.

DMS. Deep mutational scanning.

eGFP. A green fluorescent protein.

ER. Endoplasmic reticulum.

FACS. Fluorescence activated cell sorting.

FP. Force profile.

FPA. Force profile analysis.

FSAP. Force-sensitive arrest peptide.

gGFP. Glycosylatable green fluorescent protein, which will not fluoresce if glycosylated.

IRES. Internal ribosome entry site.

Lep. Leader peptidase, often cloned into membrane proteins to improve expression or control topology.

Loading rate. The rate of increase in the pulling force applied to the end of the nascent chain. Units of force per time.

MD. Molecular dynamics.

mKate. A red fluorescent protein.

Multispanning membrane protein. A membrane protein with multiple transmembrane helices.

$\mathbf{N_{in}}$. Topology of a membrane protein in which the N-terminus is on the cytosolic side of the membrane.

N_{out}. Topology of a membrane protein in which the N-terminus is on the lumenal or extracellular side of the membrane.

NTD. N-terminal domain.

PDB. Protein data bank, typically used to indicate the ID of published protein structure.

PRF. Programmed ribosomal frameshifting. -1PRF refers specifically to a frameshift into the -1 reading frame.

PTC. The peptidyl transferase center of the ribosome, where new peptide bonds are synthesized.

SD. Standard deviation.

SecM. Arrest peptide sequence that stalls the ribosome as it is synthesized. The stall is relieved in the presence of a sufficiently strong pulling force.

SINV. Sindbis virus.

Slip site. A heptanucleotide sequence upon the translation of which the ribosome may undergo programmed ribosomal frameshifting.

TM. Transmembrane helix. Alternatively represented as TMH.

Translocation. Passage across the cell membrane.

WT. Wild type.

Chapter 1

INTRODUCTION

During the translation of nascent proteins, the interactions between the protein and its environment can generate pulling forces that propagate through the nascent protein to the peptidyl transferase center (PTC) of the ribosome. These pulling forces are capable of altering the rate and even outcomes of protein synthesis. In this manner, pulling forces allow the ribosome-nascent chain complex to sense and respond to its environment without involving any extra signalling intermediates or transcriptional processes. These unique capabilities and their applications in both naturally evolved and engineered contexts motivate efforts to understand the generation and propagation of cotranslational pulling forces, as well as their potential influences on synthetic outcomes.

The first identified manifestation of pulling force-mediated regulation was in an arrest peptide, SecM [1]. Typically, arrest peptides stall translation indefinitely upon their synthesis. In the case of SecM, however, a pulling force on the nascent chain is able to relieve the stall and allow translation to recommence[2]. Physiologically, this allows SecM to regulate the expression of the translocase SecA, because restart of SecM-stalled translation is dependent on the presence of SecA and the pulling force it exerts on the nascent protein [1, 3]. In an example of how the utility of force-mediated regulatory mechanisms can be extended far beyond their evolved roles, force-sensitive arrest peptides form the basis of our understanding of how nascent chains interact with their environment to generate pulling forces. By appending the SecM stall sequence to protein sequences of interest, one can experimentally measure the fraction of stalled ribosomes that restart and thus estimate the forces generate by the protein of interest[4]. Through this technique, researchers have identified that pulling forces are generated by myriad cotranslational processes, including cotranslational folding, membrane integration, translocation of charged particles across a potential, and interactions with other proteins[2, 4–9].

Despite the seeming ubiquity of cotranslational pulling forces, their known biological consequences remain few. Beyond enabling the restart of arrest peptides such as SecM, pulling forces have also been suggested to be able to alter the rate of protein synthesis [10, 11]. Altering translation rate can alter the structure and function of newly synthesized proteins[12]. However, a naturally evolved example of a protein where pulling forces impact translation rate and thus protein function has not yet been discovered.

The relative sparsity of identified applications of pulling force-mediated regulation likely arises from the many difficulties in experimentally measuring cotranslational forces and their consequences. As discussed above, pulling forces originate from a wide variety of physical phenomena. Further complicating characterization, cotranslational pulling forces are inherently transient and out of equilibrium, as the ribosome is adding new amino acids to the nascent protein. The magnitude of the forces and the response of the ribosome are also stochastic, requiring a large number of independent trials to characterize. The minority products can be in low concentration, making identification challenging.

Molecular dynamics (MD) simulations naturally address many of these issues. Forces are calculated at every timestep in order to evolve the atom positions, and they can be recorded as the simulation progresses. Rerunning simulations to calculate the probability distribution of forces is trivially parallelizable, given sufficient resources. Simulations can be pushed out of equilibrium in a well controlled manner by adding additional amino acids or by applying pulling forces. All these advantages are exploited in chapter II of this thesis to characterize the molecular pathway for force-induced restart of ribosomes stalled by arrest peptides. By tracking the propagation of an applied pulling force along the nascent chain, we discover how secondary structure and interactions between the nascent protein and the ribosome impede force propagation, allowing amino acids tens of angstroms away from the catalytic center of the ribosome to influence the restart pathway.

Moving beyond stalled ribosomes presents a major challenge to molecular dynamics due to the many order of magnitude difference between the femtosecond timescale of molecular motions and the second timescale of translation. To directly simulate translation and capture the interplay of slow timescale events such as membrane integration and protein synthesis requires a heavily coarse-grained model. In this thesis, we extend and employ a model previously developed in the group which condenses three amino acids into a single particle and explicitly represents the nascent protein, ribosome, and translocon while treating the cytosol and lipid membrane implicitly [13, 14]. These approximations allow simulations to reach the minute timescale on a single core while still retaining sufficient fidelity to accurately calculate forces arising from processes such as membrane integration or translocation of charges across a potential [15]. This model provides the unique capability to rapidly predict both the pulling forces and molecular underpinnings for long timescale cotranslational processes.

These capabilities led to the discovery of a novel form of pulling-force mediated regulation in alphaviruses, as described in chapters 3 and 4. We show how the pulling forces exerted from the integration of a viral polyprotein correlate with experimentally measured frameshifting ratios. This leads us to propose and validate a model in which the successful integration of a transmembrane helix causes the production of a frameshifted version of a protein, while failure to integrate leads to the production of the main-frame variant. The sensitivity to pulling forces allows the production of two different protein sequences with two different topologies, from a single mRNA. Further studies on hundreds of mutants of polyprotein sequences uncovered the physicochemical features that enable this novel regulatory mode. The conformations and pulling forces observed in the mutants emphasize the importance of the N-terminus of the integrating viral helix and the impact of slowing the translation rate with codons with low abundance cognate tRNAs.

Coarse-grained simulations were additionally used in Chapter V to interpret the highly complex experimental force profiles obtained from the integration and co-translational folding of multispanning membrane proteins. By matching force profiles between experiment and simulation, we provide a molecular picture for the experimentally observed force peaks. Altering the physics in the simulations, for example by turning the membrane potential on or off, allows us to identify the specific processes that are responsible for some of the force peaks. Features observed in the experimental force profiles that have no simulated counterpart can then restricted to originate from physical process not present in the models, such as helical packing interactions or amphiphilic surface helices.

Chapter 2

FORCE TRANSDUCTION CREATES LONG-RANGED COUPLING IN RIBOSOMES STALLED BY ARREST PEPTIDES

Adapted from:

Matthew H. Zimmer, Michiel J.M. Niesen, and Thomas F. Miller. Force transduction creates long-ranged coupling in ribosomes stalled by arrest peptides. *Biophysical Journal*, 2021. doi: 10.1016/j.bpj.2021.03.041.

Force-sensitive arrest peptides regulate protein biosynthesis by stalling the ribosome as they are translated. Synthesis can be resumed when the nascent arrest peptide experiences a pulling force of sufficient magnitude to break the stall. Efficient stalling is dependent on the specific identity of a large number of amino acids, including amino acids which are tens of angstroms away from the peptidyl transferase center (PTC). The mechanism of force-induced restart and the role of these essential amino acids far from the PTC is currently unknown. We use hundreds of independent molecular dynamics trajectories spanning over 120 μs in combination with kinetic analysis to characterize multiple barriers along the force-induced restart pathway for the arrest peptide SecM. We find that the essential amino acids far from the PTC play a major role in controlling the transduction of applied force. In successive states along the stall-breaking pathway, the applied force propagates up the nascent chain until it reaches the C-terminus of SecM and the PTC, inducing conformational changes that allow for restart of translation. A similar mechanism of force propagation through multiple states is observed in the VemP stall-breaking pathway, but secondary structure in VemP allows for heterogeneity in the order of transitions through intermediate states. Results from both arrest peptides explain how residues that are tens of angstroms away from the catalytic center of the ribosome impact stalling efficiency by mediating the response to an applied force and shielding the amino acids responsible for maintaining the stalled state of the PTC.

2.1 Introduction

Regulation of protein expression levels is an essential part of maintaining cellular homeostasis. While this is most commonly performed at the nucleotide level by controlling rates of transcription initiation or ribosome binding, regulation at the translational level allows for immediate control of protein levels [16]. One method of translation-level regulation utilizes arrest peptides, which are short amino acids sequences that induce a stall in ribosomal translation when synthesized, until an external signal breaks the stall [17]. Force-sensitive arrest peptides (FSAPs) are a class of arrest peptides that allow for the restart of translation with a force-dependent rate [18]. Physiologically, these peptides act as a force sensor associated with the integration of nascent proteins into the membrane or translocation across the membrane. For example, the *E. coli* FSAP SecM is able to regulate expression of the translocase SecA because restart of SecM-stalled translation is dependent on the pulling force exerted by SecA on the nascent peptide [19]. SecM has also been used as a powerful biophysical tool to measure in vivo forces arising from membrane integration [4], cotranslational folding [5, 20], and charge translocation [8]. The relationship between force and restart rate was quantitatively characterized through optical tweezer experiments on SecM-stalled ribosomes [18] Understanding the mechanism through which forces propagate up the nascent chain and break the stall of FSAPs will not only shed light on an essential mode of regulation in biology, but also provide insights for the design of mutants that respond to forces with greater dynamic range and sensitivity, thereby enabling wider application of FSAPs as *in vivo* force sensors [21, 22].

Considerable progress has been made in understanding the molecular mechanism behind FSAP stalling. SecM was the first FSAP identified and remains one of the best characterized. The requisite sequence in SecM for stalling has been narrowed down to a stretch of seventeen amino acids, of which nine amino acids are essential for efficient stalling[1, 23]. Cryo-EM structures have identified several SecM stalled states with different ribosome conformations and tRNA occupancies [24, 25]. A common feature of the different stalled ribosomes is that their conformations around the peptidyl transferase center (PTC) resemble the structure of the uninduced state of the ribosome characterized by Schmeing et al in which the P-site tRNA is protected from nucleophilic attack [26]. The importance of particular residues is made clear by the SecM-stalled structures; for example, the essential C-terminal R163 interacts with several neighboring nucleotides to distort the PTC geometry [27]. However, other amino acids essential for efficient stalling, such as F150 and W155, are more than ten angstroms away from the PTC and have no obvious role in altering the conformation of the ribosome. Although structural data show the essential residues far from the PTC are interacting with ribosomal proteins and RNA in the exit tunnel, there are no large scale rearrangements in the ribosome that could lead to propagation of this signal through the ribosome [24, 28–30]. Instead, these residues are suggested to increase stalling by precisely arranging the conformation of C-terminal portion of the nascent chain or by controlling the degree of compaction in exit tunnel [22, 31]. The molecular details of how these essential C-terminal amino acids induce this arrangement have not yet been uncovered.

More recently, the structures of ribosomes stalled by VemP and MifM, two other FSAPs, have been solved [30, 32]. These peptides share many of the same features as SecM, including the stabilization of the uninduced state of the ribosome and the presence of many essential amino acids positioned far down the exit tunnel [26]. Despite these similarities, there are also significant differences between these more recently characterized peptides and SecM, most notably the secondary structure and compaction of the nascent chain in VemP and the species specificity of MifM.

While the available structures provide insight into how arrest peptides stall translation, the mechanism of force-dependent restart is less well understood. Experiments using optical tweezers were able to quantify the relationship between the applied force and restart rates [18], while a series of mutant arrest sequences identified amino acids critical to maintaining the stall in both the presence and absence of an applied force [22, 31, 33]. However, these data are insufficient to reconstruct the molecular mechanism of force-induced restart. Since most experiments only report on when translation is able to restart, little is known about the conformational changes that FSAPs undergo prior to restart, including the number of transitions, their relative barrier heights, and whether they progress in a consecutive, random, or concerted manner. In the current work, a molecular picture of the stall-breaking processes is revealed by microsecond-timescale molecular dynamics (MD) simulations of the SecM and VemP FSAPs with an applied force on the N-terminus. Analysis of over 150 independent stall-breaking events uncovered a conserved multi-step pathway in which pulling forces are unable to disrupt the conformation around the PTC until key interactions formed by the N-terminal regions of the arrest peptides are broken. We characterize the locations and height of these barriers, and propose a role for the essential amino acids far from the PTC in which they control force propagation. This suggests a mutation strategy to produce FSAPs with altered force sensitivity.

Comparison between the stall-breaking pathways of SecM and VemP highlights how interactions between the nascent peptide and the ribosome alter force transduction differently than intra-chain interactions. Beyond FSAPs, cotranslational forces have been shown to be able to influence the rates and even outcomes of translation [10, 34]; thus, a better understanding of the factors that control force propagation in the exit tunnel also provides valuable insight into how ribosomes are able to sense and respond to their environment via the nascent protein.

2.2 Results and discussion

Pulling-force trajectories

The force-induced restart of stalled ribosomes is a non-equilibrium biomolecular process that occurs on the timescale of seconds to minutes. Microsecond-timescale molecular dynamics with an applied force on the nascent protein provides insight into these non-equilibrium effects and allows for the direct visualization of the motions and forces of the nascent chain throughout the stall-breaking pathway. Simulations are initialized using cryo-EM structures of SecM and VemP stalled ribosomes. SecM stalls the ribosome in many states during synthesis and several of these states have been characterized structurally [24, 25, 35]. Simulations of SecM were started from the structure corresponding to the earliest point in translation (PDB: 3JBU), in which Pro-tRNA has not yet bound to the ribosome (Fig. 2.1A). To reduce computational cost, the ribosomes in both SecM and VemP structures are truncated beyond a 23 Å radius around the nascent chain, with an outer shell of 3 Å restrained to preserve the structure of the ribosome (Fig. 2.1B, Methods). With the exception of the areas near the nascent peptide and the PTC, the conformation of the 50S subunit of SecM-stalled and unstalled ribosomes are similar [24, 25, 36], so the dynamics of the ribosome 20 Å away from the PTC is not expected to influence the restart pathway. Following truncation, the starting structures were relaxed using 1 μs of equilibrium MD. During equilibration, the initial conformation of the nascent protein and ribosome around the PTC was well conserved

To simulate force-induced stall breaking, we apply a force on the nascent chain by placing the N-terminal alpha carbon in a harmonic potential, with the minimum energy position of the harmonic potential moving out of the exit tunnel at constant velocity. This protocol mimics the commonly used force-ramping protocol in optical tweezer experiments [18, 37, 38]. To observe stall-breaking on a timescale accessible to simulation, it is necessary to pull with a faster force-ramping protocol than is done experimentally. We run simulations over two orders of magnitude in force loading rates (0.5, 5, and 50 pN/ns). These loading rates lead to rupture forces between 100 to 1000 pN and timescales from 50 ns to 5 μs. Although this is faster than the presumed experimental timescale of seconds [18, 35], the wide range of pulling forces applied allows us to evaluate the consistency of our trajectories with varying simulation timescale. Additionally, our simulations match experimental estimates of the free energy barrier to stalling and the effects of mutations, supporting the assumption that the pathway of stall-breaking is unchanged by the accelerated timescale of our simulations. Detailed analysis of over one hundred independent trajectories allows for more detailed and statistically robust analysis than previous simulations of stalling peptide restart which consisted of two force pulling trajectories of 10 ns each [27]. The large number of long trajectories (over 120 μs total) was enabled by the use of the Anton2 computer [39]. Descriptions of each trajectory are provided in Table 2.

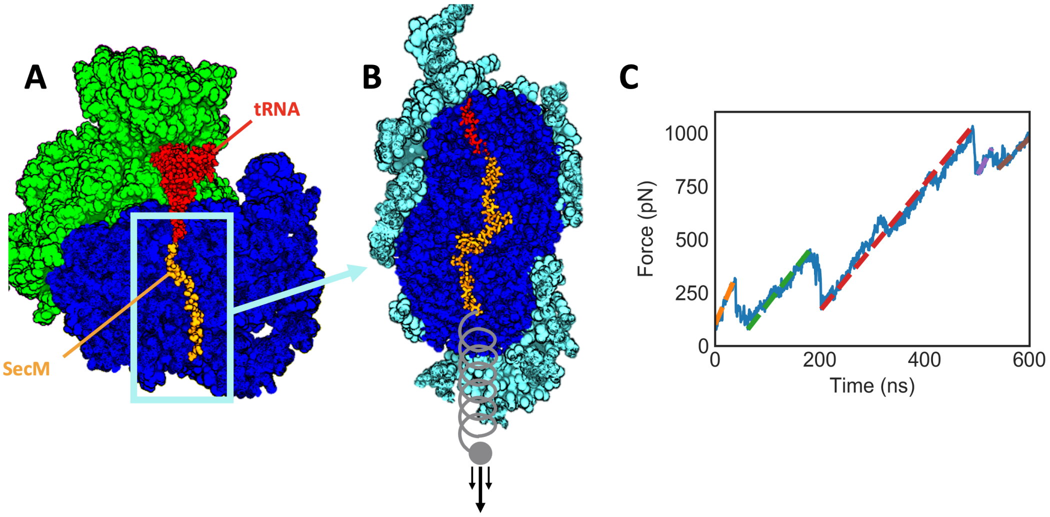


Figure 2.1: Description of the simulation system and typical output. A) Cryo-EM structure of a SecM stalled ribosome (PDB: 3JBU). Only the region of the ribosome surrounding the nascent chain is included in our simulations. B) The simulated part of the ribosome, with atoms shown in cyan harmonically restrained. A ramping force is applied to the N-terminus of the nascent chain. C) The result of a SecM pulling trajectories. Several regions of linear extension can be identified (colored dashed lines), corresponding to different stable conformations of the nascent chain. Note that, for visual clarity, the structures in A and B do not show atoms that are nearer to the viewer than the nascent chain.

Force propagation through stall-breaking intermediates

We first consider force-pulling simulations involving the SecM sequence, as SecM is the best characterized FSAP and has seen the most use in force-measurement experiments, enabling us to validate our methodology and make readily testable predictions. The force traces indicate that several distinct conformational changes occur within a typical trajectory (Fig. 2.1C). The changes are evidenced by the sharp drops in the force trace, corresponding to moments in which the nascent chain extends along the direction of the applied force. Regions of the force trace in between these drops with linearly ramping force correspond to metastable intermediate states along the stall-breaking pathway. To identify states that are consistently observed across the 5 pN/ns trajectories, the linear rate of force increase (70 pN/Å) is subtracted from each force trace (Fig. 2.2A, Methods). This allows for direct comparison between force traces even though the times at which conformational changes occur are stochastic [40].

Combining the results of the ensemble of pulling-force trajectories reveals a distribution of applied forces with five peaks, with each peak corresponding to a conformational state that is robustly observed across the trajectories (Fig. 2.2B). These data are then fit by a mixture distribution consisting of the sum five Gaussians. Each frame of the trajectories is assigned to one of the five states based on the force at that frame. Each state may contain sub-states that give rise to small drops in the force trace within a given state (e.g. in Fig. 2.1C); however, these sub-states are not consistent across different trajectories. Since these sub-state transitions are not conserved in the stall-breaking pathway and their infrequent occurrence makes robust analysis difficult, they were not further characterized. Repeating the above analysis on simulations conducted with a faster, 50 pN/ns loading rate reveals a distributions of external forces that are also well-fit assuming a five state model (Fig. 2.3).

With thousands of configurational samples per state, the molecular motions that govern the transitions between the observed states can be characterized. Measuring the changes in dihedral angles between consecutive states on an per-amino acid basis (Fig. 2.4) reveals that the conformational changes between states are driven by rotations of only a few adjacent amino acids per transition. For all of the amino acids that undergo changes in dihedral angle, the rotation observed involves the straightening of the ψ dihedral angle from a bent conformation to a 0 degree planar conformation. This extends the nascent chain in the direction of the applied force. The set of residues that rotate in each transition progresses toward the C-terminus

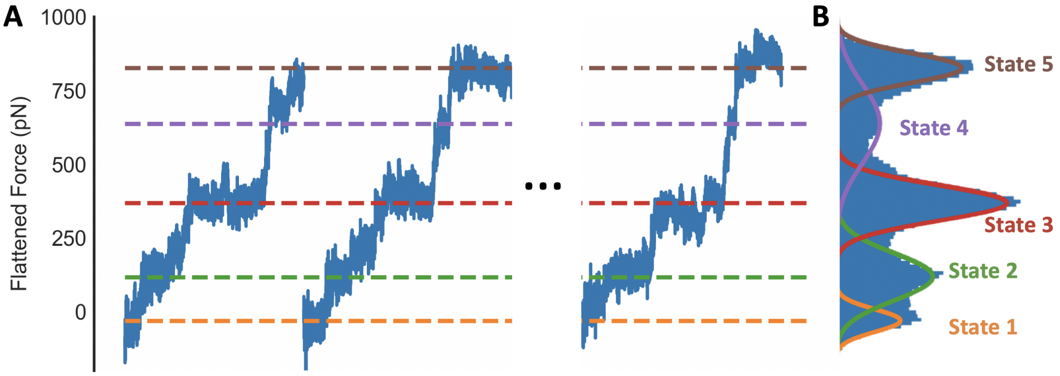


Figure 2.2: Identifying states based on force traces A) Force traces from three representative simulations of SecM after a force ramp of 5 pN/ns has been subtracted. Dashed lines indicate the peaks of the Gaussians which have been fit to the data in panel B. B) Histogram of the flattened force traces of all 30 trajectories run with a force loading rate of 5 pN/ns.

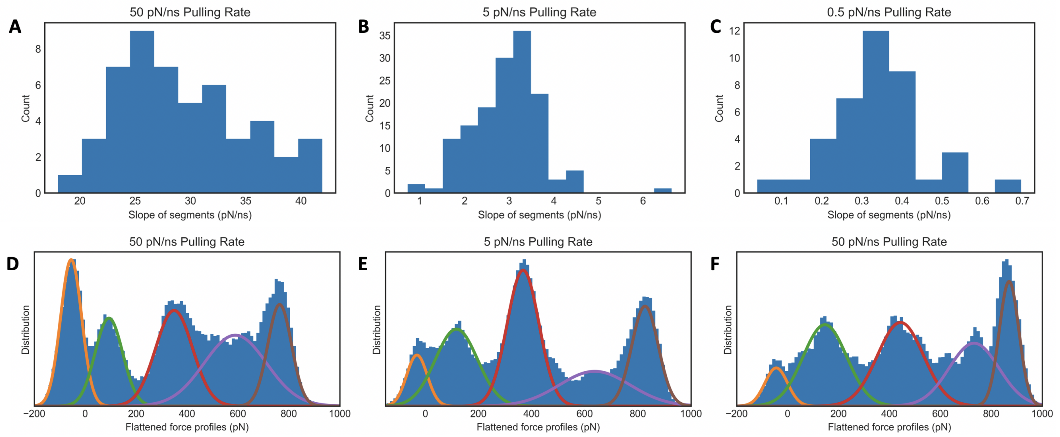


Figure 2.3: Identifying states based on force profiles. A) For all 30 50 pN/ns trajectory, linear segments were identified as described in the Methods. The distribution of the slopes of these segments is plotted. B) Same as A, but for all 30 5 pN/ns trajectories. C) Same as A, but for all 5 0.5 pN/ns trajectories. D) The median of the slopes of the linear segments plotted in A were subtracted from each force profile, generating flattened profiles as in Fig. 2.2A. The distribution of flattened force profiles is plotted. Superimposed is the results of fitting the distribution to the sum of five Gaussians. Each Gaussian is plotted separately. E) Same as D, but for 5 pN/ns trajectories. F) Same as D, but for 0.5 pN/ns trajectories.

of nascent chain with each successive transition, from Ser151 rotating in state 1, Ser157 and Gln158 in state 2, Gly161 in state 3, and finally to Arg163 and Ala164 rotating at the C-terminus in state 4. After a state transition, there is no evidence for any of the fully extended amino acids reforming interactions with the ribosome.

The molecular descriptions of the states and the transitions between them are almost identical when simulations are performed with a 10x greater force ramping rate (Fig. 2.4). The same analysis performed on trajectories run with the slower ramping rate of 0.5 pN/ns identified nearly identical conformational changes for each state transition, albeit with some overlap between the peaks that is likely due to the difficulty in assigning states with only five trajectories. The rotation of the specific amino acids listed above defines a robust molecular pathway that describes that conformational changes SecM undergoes when subjected to an applied force on the N-terminus. It is important to note, however, that the amino acids undergoing dihedral rotation between states are not necessarily responsible for the free energy barrier between states. The key interactions that maintain the states will be investigated in a later section using computational mutagenesis.

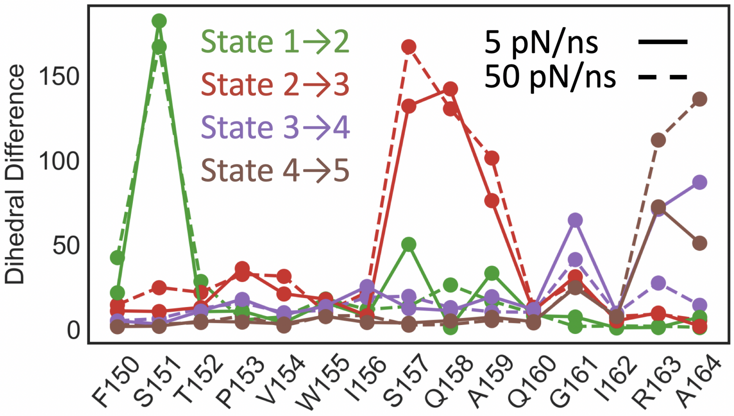


Figure 2.4: Conformational changes are quantified for each amino acid in terms of the distance of the ϕ and ψ dihedral angles between two states. The key differences between each state can be described by rotations in only a few amino acids per state.

Connecting peptide conformation to restart of translation

We now identify the state during force propagation that first allows the restart of translation. The five amino acids at the C-terminus of SecM are all essential to its stalling behavior, as well as being the closest to the site of new bond formation [22]. It is therefore expected that only conformations in which the essential C-terminal residues have been affected by the force could allow restart of synthesis. The most dramatic movement of these amino acids occurs during the transition from state 3 to 4 when Arg163 and Ala164 straighten (Fig. 2.4). However, this transition is unlikely to occur under physiological conditions as it coincides with an unphysical extension of the tRNA. Characterizing the movement of the tRNA nucleotide which is covalently bonded to the nascent protein reveals that the tRNA maintains a stable position until the transition to state 4, at which point it is pulled several angstroms into the exit tunnel (Fig. 2.5A). This extension is enabled by the breaking of the stacking interactions between the nucleobases in the tRNA in a manner which, to our knowledge, has not been observed *in vivo*.

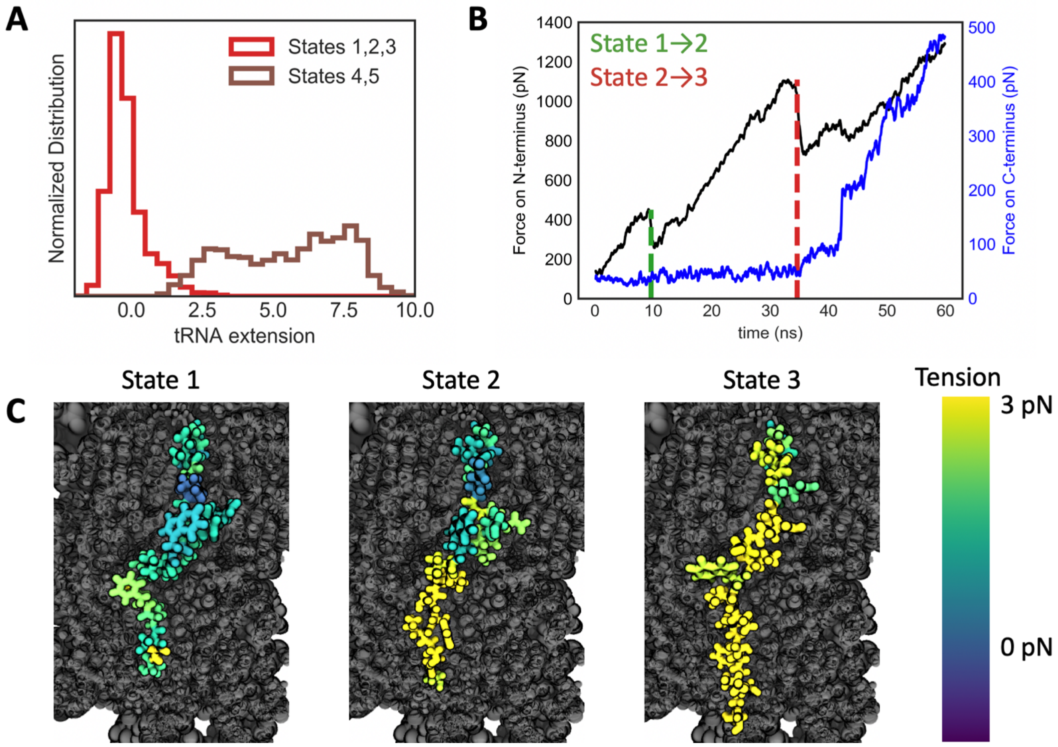


Figure 2.5: Identifying which states could correspond to restart of synthesis A) Distributions of the extension of the 3' hydroxyl group of the tRNA nucleotide which is bonded to the nascent chain beyond its position during equilibration. Extension is measured in the direction of pulling. States 4 and 5 are unphysically extended. The data are taken from the 30 independent 5 pN/ns trajectories. B) Simulations of SecM were run with the force on the N-terminus ramped at 50 pN/ns as previously, but now with the C-terminal Arg163 fixed by a harmonic restraint. The pulling force is plotted in black and the amount of force that is acting on Arg163 is plotted in blue. The force only increases above its basal level until after the transition to state 3. The transitions between states 1 to 2 and 2 to 3 were determined by the frame in which residues Ser151 and Ser157 respectively straighten. C) Characteristic conformations of each state are shown, colored by the magnitude of the tension along the nascent chain relative to the tension in the equilibration trajectory. Tension only reaches the C-terminal amino acids once state 3 is reached.

To further evaluate the possible relevance of the 3-to-4 transition, we estimated the height of the free energy barrier by fitting the observed distribution of state transition times to a probabilistic model of force-induced barrier crossing proposed by Bullerjahn et al (Fig. 2.6, 2.7, Table 1) [41]. This leads to an estimate for the barrier height of 120 kT (90% CI: 110-160 kT), which could not be crossed on biologically relevant timescale without forces far stronger than what can be exerted by SecA, estimated to be around 10 pN [42]. The analyses of the conformational changes and the energetics of the state 3 to 4 transition both indicate that state 4 is physiologically inaccessible and therefore that restart must occur in state 3 or before.The following section details the methodology and validation of these estimates.

Estimation of energy landscape

Given a number of repeated and stochastic trajectories of force-induced disruption of a bond, it is possible to estimate the free energy landscape of that bond. This estimation is typically performed by assuming one-dimensional double-well energy landscape that the system explores through thermal fluctuations, just as in Kramer's theory of chemical reaction rates[43]. The applied force distorts the underlying free energy landscape of the bond, acting to reduce the height of the energy barrier and increase the rate of crossing. Bond rupture occurs when the free energy barrier is crossed.

Numerous theories have been developed that provide strategies to estimate an energy landscape given different pulling protocols and force regimes [44–46]. In this work, we use the framework developed by Bullerjahn, Strum, and Kroy [41] which has the advantage of being directly applicable to the force loading protocol used in our molecular dynamics simulations, namely the application of external force via a stiff and rapidly moving harmonic potential. Their theory predicts the distribution of observed rupture forces given the parameters describing the energy landscape and the pulling protocol, i.e. $p(F|x^{\ddagger}, \Delta G^{\ddagger}, D; k_s v)$. $x^{\ddagger}$ is the distance from the minimum of the bound state to the location of the barrier maximum, ΔG is the height of the free energy barrier, D is the diffusion constant, and $k_s v$ is the force loading rate used in simulation. $k_s v$ is chosen when setting up the simulation, while the remaining three parameters need to be estimated. This is done by finding the value of the parameters that maximizes $p(F|x^{\ddagger}, \Delta G^{\ddagger}, D; k_s v)$. The functional form for this probability is provided in [41], and assumes the underlying free energy landscape to be harmonic with a cusp at the barrier maximum.

Optimizing this probability to find $x^{\ddagger}$, $\Delta G^{\ddagger}$, and D simultaneously produced unphysical results with estimates of $\Delta G^\ddagger$ and D both orders of magnitude higher than expected for all transitions. The poor estimates are likely due to insufficient samples of stall-breaking events. Instead of performing a fully flexible fit, we first estimate the spring constant of the underlying harmonic energy landscape k_l, which fixes the ratio of $x^{\ddagger 2}$ to $\Delta G^\ddagger$ as $k_l = \frac{2\Delta G^\ddagger}{x^{\ddagger 2}}$. The estimate for the spring constant of the underlying landscape is obtained by first expressing the effective potential as the sum of the underlying landscape plus the moving harmonic trap, i.e.

$$U(x,t) = k_l x^2 + k_s(x - v_s t)^2$$

, where k_s is the spring constant of the applied harmonic potential and v_s is the speed at which the applied potential is moved. Because the harmonic trap is moving with time, the energy minimum of the effective potential moves as well, specifically

$$x_{min}(t) = \frac{k_s v_s t}{k_s + k_l}$$

. As the system tracks the moving minimum of the effective potential, the expected value of the applied force at a given time varies as:

$$\mathbb{E}[F(t)] = k_s\left(vt - x_{min}(t)\right) = k_s vt\left(1 - \frac{k_s}{k_s + k_l}\right)$$

.

The applied force is an experimental observable, and all the other parameters besides k_l are known. This means that the spring constant of the underlying constant can be simply estimated from the slope of the applied force as a function of time:

$$k_l = \frac{k_s}{1 - \dot{F}/k_s v}$$

,

where $\dot{F}$ is the slope of the applied force. The estimation of $\dot{F}$ is described in Methods and results are shown in Fig. 2.3. The estimate of k_l is informative because the spring constants of the underlying landscape and the applied potential are of a similar scale. This is not typically the case for optical tweezer experiments, for which $k_l \gg k_s$ and therefore $\mathbb{E}[F(t)] \approx k_s vt$, which has no k_l dependence.

Estimating k_l removes one free parameter from the optimization problem, as $x^\ddagger$ and $\Delta G^\ddagger$ are linked through $k_l = \frac{2\Delta G^\ddagger}{x^{\ddagger 2}}$. Still, fitting the observed forces at rupture to $p(F|x^\ddagger, \Delta G^\ddagger, D; k_s v)$ produced unphysical parameters. To further reduce the complexity of the fit, $x^{\ddagger}$ is fit to the median of the rupture forces while D is held constant. By repeating this fit over each of the three pulling forces and finding the value of D which leads to the best agreement between predictions of $x^{\ddagger}$, both $x^{\ddagger}$ and D can be estimated (Fig. 2.6).

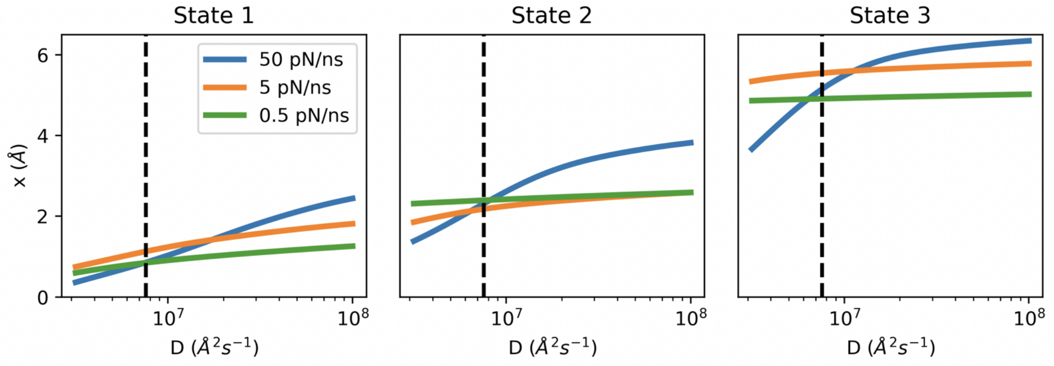


Figure 2.6: Each plot shows the optimum value for $x^{\ddagger}$ given a fixed value for the diffusion constant. Each plot corresponds to the barrier for exiting a given state, and the fits are performed separately for trajectories of different force ramping rates. The dashed lines indicate the value of the diffusion constant that produces the best agreement between estimates of $x^{\ddagger}$ across all three pulling speeds and states.

The above methodology allows fitting of a single-barrier landscape, however the the pathway of force-induced SecM restart clearly exhibits multiple, sequential barriers. To fit barriers beyond the first, it is necessary to find the location of the minimum of the state before the barrier of interest. This is done by assuming that the distance from the previous minimum to the barrier is the same as the distance from the barrier to the subsequent minimum (Fig. 2.7). This significant and simplistic assumption is necessary because our trajectories do not exhibit the rebinding events that would be necessary to gain more detailed information on the shape of the landscape between a barrier and the subsequent minimum. Despite the simplicity of the assumption, we do not observe any states rupturing before their predicted minima, which would be clear indicator of a poor choice of landscape. Having established a method to identify the minima of a state given the parameters estimated for the previous state, we can iteratively estimate the landscape of each state (Fig. 2.6, Table 2.1).

The estimated parameters clearly indicate that the barrier between state 1 and 2 cannot be the sole barrier before restart of translation, as it can be overcome by thermal motion with a rate of over $10^4\ s^{-1}$. The crossing of the barrier between states 3 and 4 cannot be required for restart of translation because the crossing time is far too slow at physiological forces, e.g. Bell's law predicts that with a force

Table 2.1: Optimal parameters that describe that energy barrier for leaving states 1-3. Confidence intervals were determined by bootstrapping over trajectories.

	$x^\ddagger$ (Å)	$x^\ddagger$ 90% CI	$\Delta G^\ddagger (k_B T)$	$\Delta G^\ddagger$ 90% CI	$k_0 (s^{-1})$	k_0 90% CI
State 1	1.0	0.83	4.2	3.3	1.0×10^6	6.3×10^4
		1.3		8.1		2.3×10^6
State 2	2.5	2.2	27	21	4.3×10^{-4}	6.7×10^{-25}
		3.9		76		8.9×10^{-2}
State 3	5.3	5.0	120	110	2.8×10^{-45}	2.7×10^{-61}
		6.0		160		1.6×10^{-40}

of 10 pN the barrier will be crossed with a rate $k_0 e^{Fx^\ddagger/k_B T} = 10^{-44} s^{-1}$. Note that both barriers are still implausibly relevant even at the ends of the 90% confidence intervals.

By contrast, the barrier between state 2 to 3 has a more physiologically sensible rate constant of 4.3×10^{-4} which drops to 5.6×10^{-3} with a 10 pN force, assuming Bell's law. Although the confidence intervals on these estimates are wide, this barrier appears to be substantial yet still physiologically surmountable. In combination with the fact that crossing this barrier allows the force to affect the conformation at the C-terminus of SecM, the barrier between state 2 and 3 is likely to be the key barrier to force induced restart of translation.

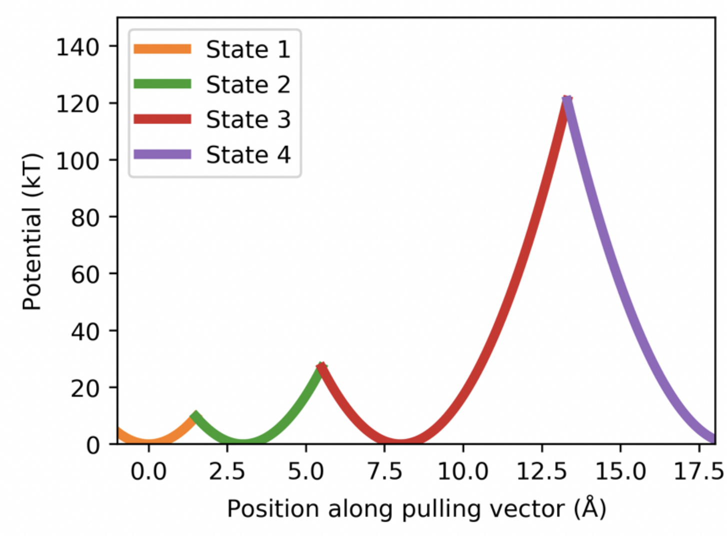


Figure 2.7: Estimated potential energy landscape for force-induced SecM restart. Each potential energy well is colored by the state it corresponds to.

Force and tension propagation

For restart to occur, the conformation of the amino acids at the PTC must be disrupted by the applied force. This occurs only in state 3 and later (Fig. 2.4), suggesting that the transition from state 2 to 3 is necessary for restart of synthesis to occur. To precisely determine when the amino acids at the PTC experience pulling forces, simulations are run with a harmonic restraint on the alpha carbon of the essential Arg163 (Fig. 2.5B). The displacement of the restrained alpha carbon reports on the force that atom experiences. Until SecM enters state 3, no significant forces are felt by Arg163 and restart of translation cannot occur. This held true across several different simulation protocols, including with two different loading rates and when the restraint was placed on Gly165 rather than Arg163. The inhibited force transmission in SecM contrasts with previously published results that measured force transmission in a non-stalling peptide and found no significant difference between the forces at the N- and C-termini [10].

To further confirm the mechanistic significance of the transition from state 2 to state 3, the tension along the backbone of the nascent chain is measured (Methods). Tension in the nascent chain around the PTC only increases after state 3 is reached (Fig. 2.5C). Analysis of the computed tensions as a function of position along the nascent chain leads to the same conclusions as those reached above, while having the advantage of not requiring an artificial restraint and providing measurements from all SecM amino acids, instead of just the restrained atom. However, the tension fluctuates dramatically with time and meaningful comparisons can only be made after averaging over all the frames in a state.

Both metrics of tracking the intra-chain forces during a pulling trajectory indicate that overcoming the first barrier and entering the second state could not lead to a restart of translation. Only once the third state is entered does the force reach the C-terminus and enable disruption of the interactions around the PTC that hold the stall in place. Additionally, our estimate of the free energy barrier between states 2 and 3 is 27 kT (90% CI: 21-76 kT), which makes this transition the last energetically accessible step. The 4 kT (90% CI: 3-8 kT) barrier between states 1 and 2 can be overcome by thermal motion, whereas the 120 kT (90% CI: 110-160 kT) barrier between steps 3 and 4 is physiologically insurmountable, requiring the disruption of tRNA stacking. Taken together, this indicates that arrival at state 3 is both necessary and sufficient for stall-breaking.

Force transduction in mutant arrest peptides

The extent to which mutations in SecM affect the response to an applied force is poorly understood. The majority of previous studies of the stalling of SecM mutants were performed in the absence of the translocation machinery that physiologically exerts the forces necessary to relieve the stall [1, 22, 31]. These experiments identify which amino acids are essential for stalling, but do not indicate which mutants alter force sensitivity. An extensive study on the force-responsiveness of a series of mutants of a SecM sequence derived from *Mannheimia succiniciproducens* (Ms-SecM) identified a number of sequences with enhanced resistance to pulling force, and these variants have found subsequent use in characterizing biophysical processes that induce a wide range of forces [9]. However, the substantial differences in sequence between Ms-SecM and the more commonly studied *E. coli* SecM, combined with the lack of a structure of Ms-SecM, make it difficult to rationalize the effects of the mutants in molecular terms. We thus employ MD simulations to explore how mutants alter the barriers to force propagation along the nascent chain. Several amino acids in SecM were individually mutated to alanine, equilibrated for 250 ns, then re-simulated with a force loading rate of 50 pN/ns. From these trajectories, the work required to pull the mutant into the restart-competent state 3 is calculated. Work is a convenient metric because it includes the effect of mutations on both the state 1 to 2 and 2 to 3 transitions. To focus on the effects due to force propagation rather than the conformational changes at the PTC that may occur upon mutation, mutations were made only to amino acids more than 10 Å away from the PTC. Five mutants were simulated, three that were found experimentally to reduce stalling efficiency in the absence of an applied force (F150A, W155A, I156) and two that were not observed to have an effect on stalling (L149A, T152A). The three mutants that required the least amount of work to restart were the same three that were found to experimentally reduce stalling efficiency (Fig. 2.8B) [1].

The findings presented so far have broad implications for the process of force-induced restarting of stalled peptides. Firstly, it provides an explanation for why the specific identity of amino acids is essential even when they are tens of angstroms away from the PTC, as the force is unable to propagate beyond these amino acids until specific interactions with the ribosomal exit tunnel are broken (Fig. 4A). The interactions formed by these residues should also be capable of blocking or damping fluctuations on the chain, which provides a connection to mutagenesis studies that investigate stalling behavior at zero applied force [1, 23, 31]. This mechanism also suggests that mutating these essential amino acids lower in the chain would be the

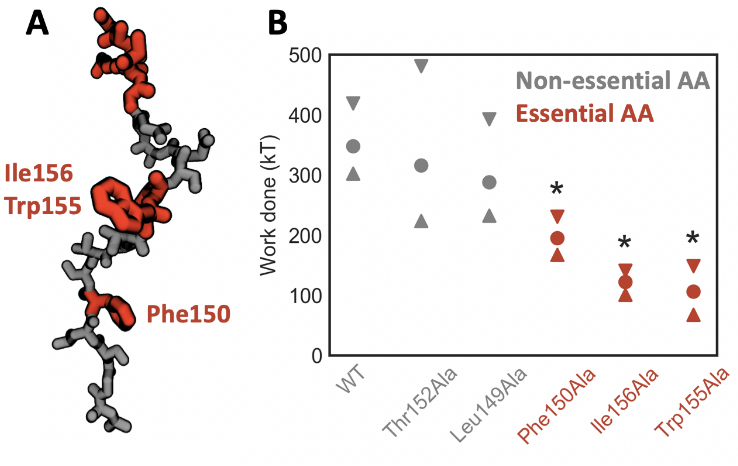


Figure 2.8: Exploring the role of the essential amino acids in the restart process A) Snapshot of SecM with the amino acids whose identity is essential to effective stalling colored in red. B) Mean work done to reach the restart-competent state 3. The mutants colored grey are those which were experimentally found to have little effect on stalling, while the red mutants led to a significant reduction in stalling. Error bars indicate 95% confidence intervals on the mean. Stars indicate mutants for which the mean of the work done was significantly different from the WT work according to a t-test with unequal variances and a threshold of $p < 0.01$

most effective strategy for tuning the force response properties of SecM. Mutations of the essential residues at the C-terminus, such as Arg163, would be more likely to abrogate the stall entirely rather than change its sensitivity to an applied force. Additional variants of SecM with altered response to force would aid the use of FSAPs as in vivo force sensors, since the diversity of biological processes that arrest peptides can report on is limited by the dynamic range of the reporters [33].

Alternative stall sequences

Simulations of SecM suggest a mechanism in which interactions between the nascent chain and the ribosome inhibit force propagation to the PTC and prevent restart of translation until those interactions are disrupted. To evaluate the generality of this mechanism, a second force-sensitive arrest peptide was investigated, VemP. As with SecM, force-sensitive stalling of VemP regulates the expression of protein translocation components [47]. However, a recent structure of VemP stalled in the ribosome exit tunnel showed that VemP adopts a conformation in the exit tunnel highly dissimilar to SecM [30] (Fig. 2.9A). Unlike SecM, the VemP nascent chain is highly compacted with clear secondary structure; two helical regions in the upper and lower sections of the peptide. The upper helix is principally responsible for inducing the stall, with several amino acids interacting with ribosomal nucleotides in such a way as to block peptide bond formation. It has been proposed that an force applied to the nascent chain could unravel the upper helix, thereby disrupting these interactions with the ribosome and enabling the restart of translation [30]. Due to the very different stalled conformation of VemP, it provides a stringent test of the generality of the mechanism uncovered for the force induced-restarting of SecM.

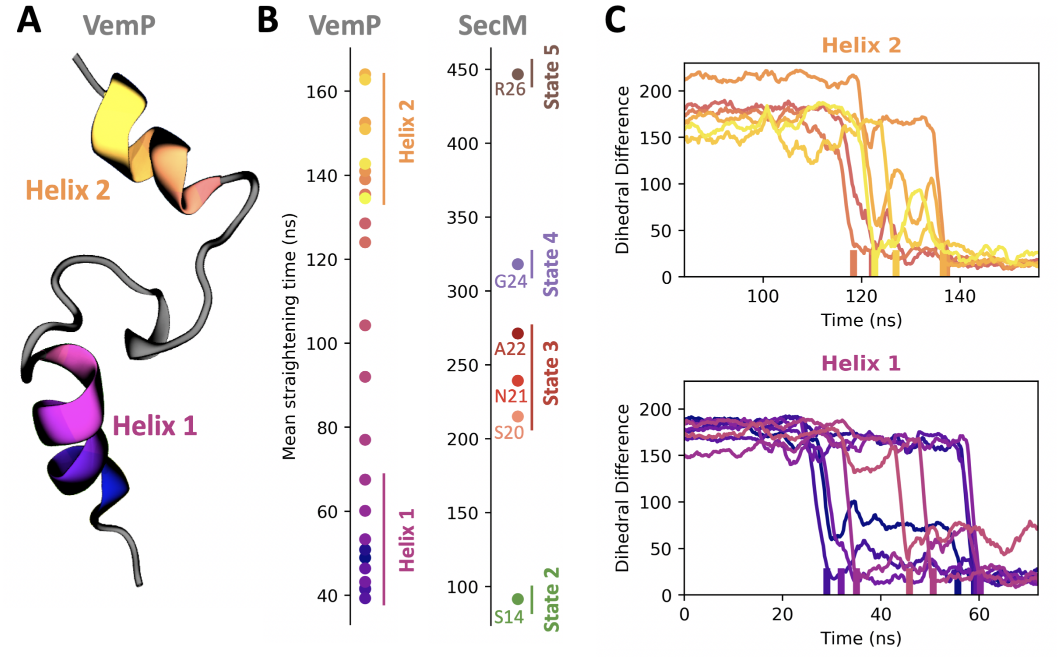


Figure 2.9: Force-induced conformational changes in VemP. A) Structure of the stalled VemP nascent chain, with the two helical regions colored. B) Straightening time for different amino acids in VemP and SecM averaged over 30 trajectories, using a force loading rate of 35 pN/ns for VemP and 50 pN/ns for SecM. C) Time series of the deviation of the dihedral angles in helices 1 and 2 from the fully extended conformation, taken from a representative trajectory. The amino acids within a helix do not straighten in consecutive order.

Following the same protocol used to study SecM, simulations of VemP were prepared from its cryo-EM structure (PDB: 5NWY), with the ribosome cut-off and restrained beyond a 23 Å radius. Force was applied on the N-terminus along the direction of the exit tunnel with a force-loading rate of 35 pN/s. Due to the high degree of compaction in VemP, if the nascent chain were fully extended from its starting configuration the N-terminus would extend well outside the ribosome exit tunnel and require a large simulated periodic box. To reduce the total simulation volume, a pulling trajectory is divided into three parts and the fully extended portion of the nascent chain is removed at the end of each third (Methods). This relies on the assumption that the fully extended nascent chain will not inhibit the propagation of the applied force.

The N-terminal helix unravels strictly before the C-terminal helix, as measured by the average time at which the dihedral angles of each residue straighten (Fig. 2.9B). This suggests that, as with SecM, interactions formed further from the PTC must be broken before force can propagate to residues nearer the PTC. Tracking tension throughout the simulations also indicates that the tension in the C-terminal helix near the PTC does not increase until the N-terminal helix is unraveled (Fig. 2.10).

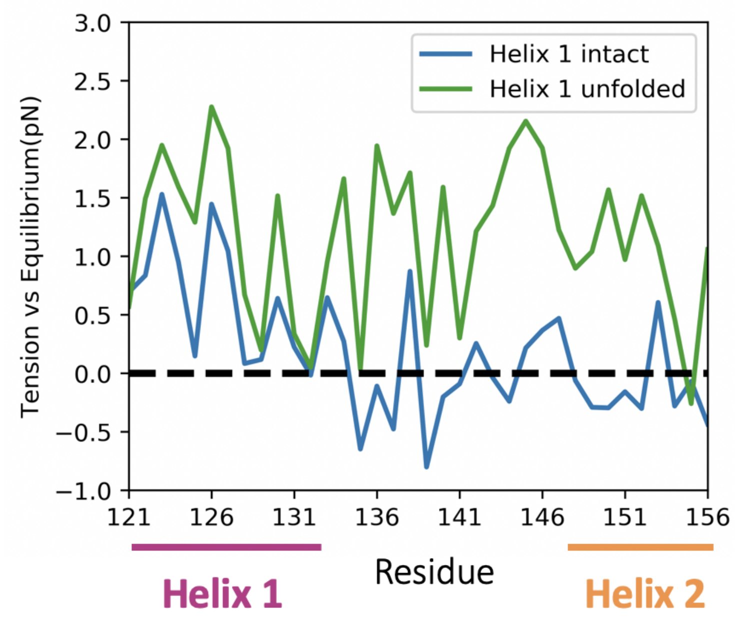


Figure 2.10: The tension along the VemP backbone relative to the tension during equilibration is averaged over all frames with helix 1 either intact or unfolded, then plotted per amino acid. The unfolding of helix 1 is measured by the straightness of Thr135 at the C-terminus of helix 1. When the dihedral angle is within 30°of 0°for more than 3 ns, the helix is defined as unfolded. While helix 1 is intact, the pulling force does not induce additional tension in helix 2, implying the propagation of force up nascent chain is inhibited. Only once helix 1 is unfolded does helix 2 experience increased tension.

However, the presence of secondary structure in VemP also leads to notable differences in the mechanism of restart. One such difference is that the straightening of amino acids *within* a helix does not necessarily occur consecutively. This can be seen both in the straightening times of the amino acids averaged over all trajectories (Fig. 2.9B) and individual trajectories (Fig. 2.9C). In Fig. 2.9, the amino acids and the markers for their transitions are colored from darker to lighter from the N-terminus to the C-terminus. If the order of straightening was strictly consecutive, the ordering of colors in Fig. 2.9 B and C would become monotonically lighter with time. This is evidently not the case within both helices 1 and 2. In fact, not only is the order of unraveling of different turns of a helix not consecutive, it is also not conserved across independent trajectories simulated with the same protocol (Fig. 2.11).

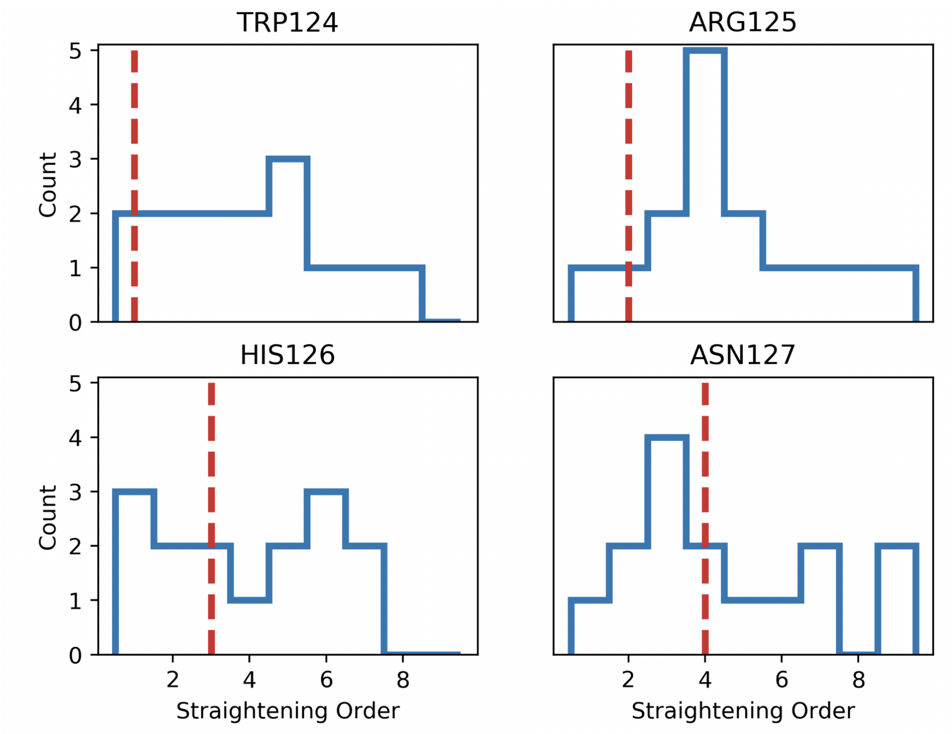


Figure 2.11: The order in which the dihedral angles of the first four amino acids in VemP helix 1 straightens is determined for each trajectory and the distribution of these orders is plotted. The expected order if the amino acids straightened consecutively is indicated with the red dashed line. For example, Arg125 is most often the 4th amino acid in helix 1 to straighten, despite being the 2nd amino acid from the helix N-terminus. All four dihedral angles show significant heterogeneity in the order of straightening across different trajectories.

The underlying reason for the difference in ordering between SecM and VemP arises from the fact that the interactions found in SecM are between the nascent chain and the ribosome, whereas the interactions within a VemP helix are between different VemP amino acids. Interactions between the nascent chain and ribosome hold the nascent chain in place relative to the exit tunnel. Intra-chain interactions in a helix still allow the helix to move as whole in response to a force, which allows force to propagate through the helix. This means that interactions within and directly C-terminal of a helix break in order of their free energy barriers. For example, if there is a weak interaction between the nascent chain and the exit tunnel C-terminal of a helix, then that weak interaction can be broken by an applied force before the alpha helix is disrupted. Conversely, if the exit tunnel-nascent chain interaction is stronger than the intra-helical interactions, then the alpha helix will be disrupted first. The similar free energy barriers for unfolding different turns of the same helix is likely why heterogeneity is observed in the breaking order within a helix between trajectories. This difference between inter- and intra-chain interactions can also explain why the middle portion of VemP, which lacks secondary structure, straightens in consecutive order.

Due to the greater cost of simulations of the longer VemP peptide, there are still several open questions regarding how its secondary structure responds to a pulling force. For instance, we were not able to investigate how the order of unraveling within a helix depends on an applied force, although we hypothesize that the heterogeneity would only increase at lower forces as thermal motion becomes more important [48]. The lack of trajectories at longer timescales also prevents estimation of free energy barriers, as was done with SecM. While challenging, this could be a particularly fruitful direction for future simulations as the barrier height determines whether nascent chain-ribosome or intra-chain interactions will break first. Additionally, although it was not observed in these trajectories, steric interactions between an alpha helix and a constricted region in the exit tunnel may force consecutive unraveling of the alpha helix as it needs to pass through the constriction zone. Despite these open questions, simulations of VemP revealed clear similarities and differences as compared with SecM. VemP exhibits non-consecutive breaking of interactions within a helix, while the interactions in SecM are disrupted consecutively from the N-terminus to the C-terminus. Importantly, both stalling peptides still require N-terminal interactions between the nascent chain and the exit tunnel to be broken before the critical conformation stalling the PTC can be disrupted.

2.3 Conclusions

The results presented here reveal a mechanism for force-induced restart of ribosomal translation in which cotranslational force propagation is governed by both interactions between the nascent chain and the ribosome and intra-nascent chain interactions. Simulations identified a pathway of sequential conformational changes in SecM that are required in order for a force applied at the N-terminal end of SecM to propagate to the PTC. This pathway was found to be consistent over two orders of magnitude of force ramping rate, and mutations that were experimentally found to reduce the stalling efficacy of SecM in the absence of force also reduced the work required to progress through the various conformational changes. Analysis of the VemP restarting pathway revealed a distinct force propagation mechanism in which intra-helical interactions are broken stochastically, while helix 1 was necessarily unraveled before helix 2 could be unfolded by the applied force. The results from both SecM and VemP highlight a previously underappreciated distinction between amino acids in FSAPs which induce stalling through specific interactions with the ribosome around the PTC and those which stabilize the conformation of the nascent chain and shield the residues at the PTC from external forces. We believe the latter interactions should be the principal target of mutagenesis aimed to alter the force sensitivity of FSAPs, since mutations to amino acids around the PTC are likely to reduce stalling efficiency even in the absence of an applied force.

Understanding cotranslational force propagation is not only essential for understanding the restarting of FSAPs but also other cotranslational processes that are affected by pulling forces, such as translation rate or frameshifting [10, 34] Comparison between the pulling force simulations of SecM and VemP emphasizes the role of interactions between the nascent chain and the exit tunnel which block further propagation of force up the nascent chain. This is contrasted with intra-chain interactions such as in alpha helices which can move as a unit and therefore do not inhibit the propagation of force and allow breaking of interactions in non-consecutive order. The force-induced restarting pathways of SecM and VemP both highlight the need to consider potential interactions between the nascent chain and the ribosome when estimating the cotranslational forces experienced by the ribosome [49, 50]. Fortunately, characterization and understanding of interactions between a nascent peptide and the exit tunnel are becoming increasingly available thanks to the rapid expansion of cryo-EM structural data. Future simulations will provide an essential connection between the study of nascent chain behavior in the exit tunnel [6, 51, 52] and the role of forces in influencing cotranslational behavior[34, 50].

2.4 Methodology

Simulation set-up

Simulations of SecM were initialized from the 3.6 Å resolution PDB structure 3JBU [24]. System preparation was performed using Schrödinger Maestro. First, the covalent peptidyl-tRNA bond was manually restored. The structure was then truncated by removing amino acids with any heavy atoms 24 Å beyond the region of interested, defined as the SecM nascent chain and first three nucleotides of the P-site tRNA. This region was chosen to be sufficiently large to minimize the influence of the truncated region, as they lie beyond the cut-off distance for the non-electrostatic interactions of the MD forcefield. Additionally, high resolution structures of stalled ribosomes do not suggest large-scale conformational changes beyond the PTC [24, 28–30], in contrast to earlier studies analyzing lower resolution stalling peptide structures [53]. Any amino acid or nucleotide monomers created by the truncation were removed. The SecM nascent chain was cut off at Leu149, and a SerGlySer linker was appended to the N-terminus. Hydrogens were modeled in, and N- and C-termini newly created by the truncation were capped with acetyl and N-methyl amide groups, respectively. Minimization was performed to alleviate clashes formed by the newly added capping groups. The structure was then solvated with TIP3P water and neutralized with Mg^{2+}. Solvation filled the orthorhombic periodic simulation box, which was set to extend 10 Å beyond the truncated ribosome. The ribosome was oriented to minimize the volume of the box. To retain the structure of the ribosome despite the truncation, all non-solvent heavy atoms 20 Å away from the region of interest were restrained with 1 kcal/mol/Å harmonic restraints. RATTLE restraints were applied to enable production timesteps of 2.5 fs. The full system consists of 117k atoms. Initial simulation was run for 30 ns using the AMBER99SB-ILDN forcefield [54], the Desmond simulation package [55], and a timestep of 1.5 fs. The same steps were used to prepare the VemP structure from the 2.9 Å resolution PDB structure 5NWY, with no truncation of the nascent chain [30]. Extensive equilibration was then performed on Anton2. All Anton simulations were conducted with a timestep of 2.5 fs. SecM was equilibrated for 1000 ns and VemP for 500 ns. Simulations of mutant SecM were prepared starting from the SecM structure after equilibration. After mutations were introduced, each structure was re-equilibrated for 250 ns.

Force-pulling protocol

The pulling direction was determined by identifying the vector that extends from the N-terminal alpha carbon of SecM or VemP and maximizes the distance between any point along the vector and the atoms of the ribosome. The optimal vector passes through the exit tunnel and maintains a distance of greater than 6 Å from any ribosomal atoms at all times.

Pulling is performed by applying a 1 kcal/mol/Å harmonic potential to the N-terminal alpha carbon that constrains the the distance between the alpha carbon and a point 30 Å away from the alpha carbon in the direction of pulling vector. Initially, this distance is set to the equilibrium distance of 30 Å, such that there is no force applied to the alpha carbon. This distance is decreased linearly with time, applying increasing force to the N-terminal alpha carbon. Initial structures are drawn randomly from the last quarter of the equilibration simulations. We explore different force loading rates of 0.5, 5, and 50 pN/ns. Simulations for were run 6.5 μs, 650 ns, and 78 ns respectively. These times correspond to halting the simulation when the minimum of the harmonic potential used to apply the pulling force was 45 Å away from initial position of the SecM N-terminus. For the fastest force loading rate of 50 pN/ns, this was extended to 55 Å to ensure all conformational transitions were observed. 30 independent simulations were performed at the force loading rates of 5 and 50 pN/s. Due to the longer simulations required, only 5 trajectories were run with a force loading rate of 0.5 pN/ns. For all simulations, frames are recorded every 1.5 ns. See Table 2 for a list of all simulations performed.

Simulations to measure the force at the C-terminus of SecM were performed as described above, but with the addition of a 1 kcal/mol/Å harmonic restraint on the alpha carbon of Arg163 or Gly165. The minimum of the harmonic potential was centered on the alpha carbon's position in the cryo-EM structure. 25 independent trajectories were run with a loading rate of 50 pN/s and C-terminal restraint on Arg163, 10 trajectories with a loading rate of 5 pN/s and C-terminal restraint on Arg163, and 10 trajectories with a loading rate of 50 pN/s and C-terminal restraint on Gly165. Simulations of mutant SecM were performed with a force loading rate of 50 pN/ns. 20 independent trajectories were generated for each mutant.

Simulations of VemP were performed analogously to SecM, albeit a force loading rate of 35 pN/ns. Initially, 30 simulations were run until the minimum of the pulling harmonic potential was 40 Å away from the initial position of the VemP N-terminal alpha carbon, corresponding to a simulation time of 80 ns. However, this was insufficient to disrupt the upper helix of VemP and further extension would bring the N-terminus of VemP near the periodic boundary of the simulation. To overcome this, after each trajectory was pulled for 80 ns, the simulation was halted and all amino acids of the nascent chain which were fully extended (i.e ϕ and ψ dihedral angles are within 30 degrees of straight (180°) and remains so for 10 frames) were removed. Simulation was then resumed, with the distance restraint adjusted such that the applied force at the end of the previous simulation and start of the new simulation were the same. This procedure was repeated until all VemP amino acids were fully extended. No more than two rounds of truncation and restart were required to fully extend the nascent chain.

Table 2.2: Enumeration of each set of simulation discussed in this work, providing the number of independent replicas and the simulation time per replica

Description	# Replicas	Time per replica (ns)
SecM Equilibration	1	1000
SecM 50 pN/ns pull	30	78
SecM 5 pN/ns pull	30	600
SecM 0.5 pN/ns pull	5	6000
SecM Fixed Arg26 50 pN/ns pull	25	78
SecM Fixed Arg26 5 pN/ns pull	10	600
SecM Fixed Gly28 50 pN/ns pull	10	78
Mutant equilibration (per mutant)	1	250
Mutant 50 pN/ns pull (per mutant)	20	78
VemP Equilibration	1	500
VemP 35 pN/ns pull	30	200

Trajectory analysis

Assigning states based on force traces

To enable identification of states over the ensemble of trajectories, force traces are compared after subtracting the linear rate of increase of force with time. Note that this rate of increase is not equivalent to the force loading rate when the spring constant of potential used to apply the pulling force is greater than the curvature of the energy landscape around a state [41], as was found to be the case in our simulations. To determine the observed rate of force increase, the beginning and end points of a linear segment are identified by finding all the local minima and maxima in a given trace. To avoid identifying spurious extrema caused by noise in the force trace, the trace is smoothed with a moving average filter and any extrema within 50 frames of another extremum are neglected, thus enforcing a minimum length for the identified linear segments (see Fig. 2.1C for example results). The slope of these segments can then be calculated, revealing a singly peaked distribution with fat tails (Fig. 2.3 A-C). The distribution of slopes from trajectories with different pulling rates are comparable after dividing by the pulling velocities, i.e. calculating the force increase per angstrom of trap movement. The median of the combined distributions is then used as the slope to flatten each of the force traces, which are subsequently binned and fit to a mixture distribution composed of the sum of Gaussians in order to define the five states (Fig. 2.3 D-F). The assumption that a single rate of force increase is appropriate for all of the states is equivalent to assuming that the curvature of the energy landscape in each state is similar. Each frame is assigned to one of the five states, based on which of the five Gaussians has the highest probability density at the force associated with that frame.

Conformational analysis

After each frame has been assigned to a state based on the force trace, the conformational changes between states are characterized in terms of the dihedral angles. Specifically, for a given state i and amino acid a, the dihedral distance is defined by $r_{i,a} \equiv \sqrt{(\overline{\psi_{i,a}} - \overline{\psi_{i-1,a}})^2 + (\overline{\phi_{i,a}} - \overline{\phi_{i-1,a}})^2}$. Averages of a dihedral angle are taken over all frames belonging to a given state and are calculated as circular quantities, i.e. angles are converted to Cartesian coordinates, averaged, and then converted back to an angle.

Calculation of tension and work

To calculate the tension on the nascent chain, the forces induced by extension or contraction of the backbone covalent bonds are recorded. In particular, for each amino acid in the protein backbone, the bond distances between the amide nitrogen to alpha carbon, between the alpha carbon to carbonyl carbon, and between the carbonyl carbon and the subsequent amino acid's amide nitrogen are recorded. These distances were then converted into forces based on the AMBER99SB-ILDN forcefield parameters [54]. The forces from these three bonds are then averaged into a single quantity for each amino acid. This calculation is performed over all all the frames in each state, as well as the equilibration trajectory in which no pulling force was applied. Tension on each amino acid is compared relative to the tension during equilibration and reported in Fig. 2.5C and Fig. 2.10.

The work to reach a state is calculated as the integral of the applied force from the start of a trajectory until all the dihedral conformational changes characteristic of the given state are observed. Specifically, state 2 is reached when Ser151 is straightened, state 3 is reached when Ser157, Gln158, and Ala159 are all straightened, and state 4 when Arg163 is straightened. A dihedral angle is straightened when the ψ dihedral angle is within 30 degrees of straight (180°) and remains so for 10 frames.

Chapter 3

COTRANSLATIONAL FOLDING STIMULATES PROGRAMMED RIBOSOMAL FRAMESHIFTING IN THE ALPHAVIRUS STRUCTURAL POLYPROTEIN

Adapted from:

Haley R. Harrington, Matthew H. Zimmer, Laura M. Chamness, Veronica Nash, Wesley D. Penn, Thomas F. Miller III, Suchetana Mukhopadhyay, Jonathan P. Schlebach, Thomas F. Miller, Suchetana Mukhopadhyay, and Jonathan P. Schlebach. Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein. *Journal of Biological Chemistry*, 295(20): 6798–6808, 2020. doi: 10.1074/jbc.RA120.012706.

Viruses maximize their genetic coding capacity through a variety of biochemical mechanisms, including programmed ribosomal frameshifting (PRF), which facilitates the production of multiple proteins from a single mRNA transcript. PRF is typically stimulated by structural elements within the mRNA that generate mechanical tension between the transcript and ribosome. However, in this work, we show that the forces generated by the cotranslational folding of the nascent polypeptide chain can also enhance PRF. Using an array of biochemical, cellular, and computational techniques, we first demonstrate that the Sindbis virus structural polyprotein forms two competing topological isomers during its biosynthesis at the ribosome-translocon complex. We then show that the formation of one of these topological isomers is linked to PRF. Coarse-grained molecular dynamics simulations reveal that the translocon-mediated membrane integration of a transmembrane domain upstream from the ribosomal slip site generates a force on the nascent polypeptide chain that scales with observed frameshifting. Together, our results indicate that cotranslational folding of this viral protein generates a tension that stimulates PRF. To our knowledge, this constitutes the first example in which the conformational state of the nascent polypeptide chain has been linked to PRF. These findings raise the possibility that, in addition to RNA-mediated translational recoding, a variety of cotranslational folding or binding events may also stimulate PRF.

3.1 Introduction

Viruses have evolved numerous mechanisms to exploit the host machinery to increase the coding capacity of their highly constrained genomes. There are at least 27 viral genera that utilize programmed ribosomal frameshifting (PRF) 4 to produce multiple proteins from a single transcript. PRF is genetically encoded and minimally requires a portion of the transcript that contains a repetitive "slippery" heptanucleotide sequence (slip site) followed by a region that forms stimulatory RNA secondary structures (an ensemble of stem loops and/or pseudoknots) [56, 57]. A collision between the translating ribosome and the stimulatory secondary structure increases the kinetic barrier to translocation, which causes the ribosome to dwell on the slip site [58–62]. During this pause, the t-RNA that is annealed within the ribosomal P-site (and most often also the t-RNA in the A-site) [63] begins to sample alternative base-pairing interactions that shift the reading frame of the ribosome [64]. Based on these mechanistic considerations, PRF is typically believed to be mediated at the level of RNA structure. Nevertheless, recent reports have also found that the efficiency of PRF can be tuned by a variety of regulatory proteins and/or miRNA [65–67]).

PRF is utilized to temporally and stoichiometrically regulate protein production during viral replication and assembly. For instance, the alphavirus structural proteins are most often produced from a single polyprotein that is cleaved into the capsid (CP), E3, E2, 6K, and E1 proteins (Fig. 3.1A) [68]. The E2 and E1 proteins are membrane glycoproteins that heterodimerize early in the assembly pathway. These dimeric units then form trimeric spike complexes, traffic to the plasma membrane, and initiate viral budding [69–71]. A programmed ribosomal frameshift into the -1 reading frame (-1PRF) occurs with 10–15% efficiency during the translation of the 6K protein and gives rise to a secondary form of the polyprotein. This frameshifted polyprotein contains the TransFrame (TF) protein [68, 72], a known virulence factor [73–76], in place of the 6K and E1 proteins (Fig. 3.1B). Because -1PRF precludes the translation of E1, the efficiency of ribosomal frameshifting (1–48% in alphaviruses) [77] influences the stoichiometric ratio of the E1 and E2 glycoproteins and the net accumulation of spike complexes. Current evidence suggests that -1PRF is stimulated by a canonical poly-U slip site and a downstream RNA hairpin [78]. However, an effort to map the stimulatory RNA structures within alphavirus polyproteins revealed that deletions within the predicted hairpin region are capable of reducing the efficiency of -1PRF but appear to be insufficient to knock out frameshifting completely [77]. This observation suggests that there may be multiple regulatory mechanisms that mediate -1PRF within the alphavirus structural polyprotein.

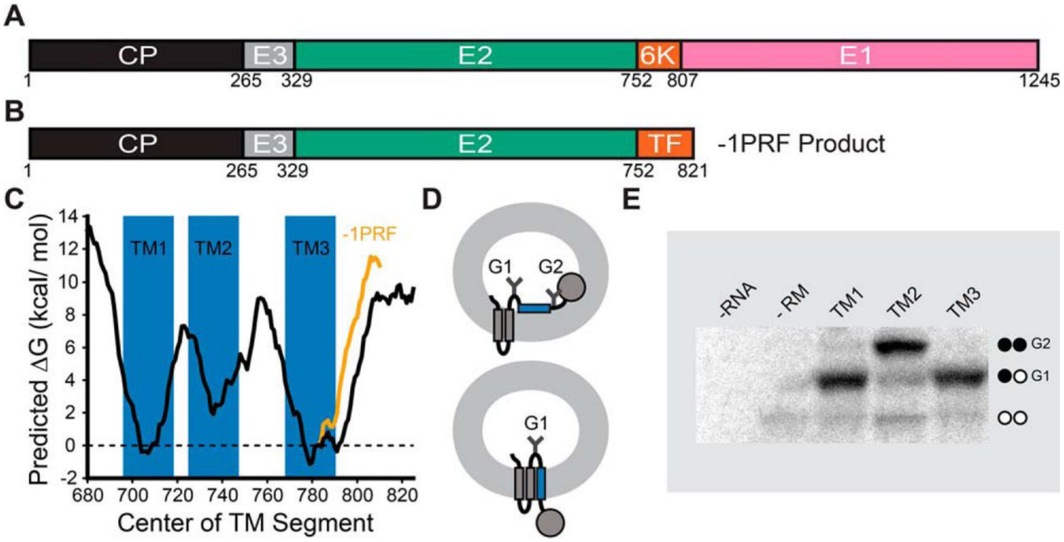


Figure 3.1: Structure and topological properties of the alphavirus polyprotein. A) a cartoon depicts the relative size and orientation of proteins within the major form of the alphavirus structural polyprotein. B) a cartoon depicts the relative size and orientation of proteins within the frameshifted form of the alphavirus structural polyprotein. C) a portion of the SINV structural polyprotein spanning the E2, 6K, and E1 proteins was scanned with the ΔG predictor using a 23-residue window (33). The predicted free energy difference associated with the cotranslational membrane integration of every possible 23-residue segment within the major form (black) and frameshifted form (orange) of the structural polyprotein is plotted as a function of the position of its central residue. The position of each predicted TM domain is indicated in blue. D) a cartoon depicts the manner in which the topological preferences of the guest TM domain (blue) influence the glycosylation state of the chimeric Lep protein. The glycosylation machinery is on the interior of the vesicle (the microsomal lumen). Membrane integration of the guest domain results in the production of a singly glycosylated product (bottom). The misintegration of the guest domain results in the production of a doubly glycosylated product (top). E) chimeric Lep constructs bearing putative TM domains from the SINV structural polyprotein were produced by in vitro translation in the presence of canine rough microsomes and analyzed by SDS-PAGE. A representative gel reveals the relative abundance of singly (G1) and doubly (G2) glycosylated translation products for each construct. Control reactions containing no RNA (no protein) and no rough microsomes (untargeted protein) are shown for the sake of comparison. These trends were consistently observed across five independent replicates.

-1PRF occurs during synthesis and processing of the nascent alphavirus structural polyprotein at the endoplasmic reticulum (ER) membrane. Following autoproteolytic cleavage of CP in the cytosol, a signal peptide at the N terminus of the E3 protein directs the nascent polyprotein to the ER lumen, where processing of the downstream proteins occurs. Localization of these segments within the lumen is essential to ensure that the E3, E2, and E1 ectodomains form their native disulfides and undergo glycosylation [70, 71, 79]. Post-translational modifications are also critical for TF, which must be palmitoylated to reach the plasma membrane and incorporate into the viral envelope [80]. The palmitoylated cysteines in TF are positioned near the edge of a putative transmembrane (TM) domain that is found in both TF and 6K [80, 81]. Although these residues are present in both proteins, they are only palmitoylated in the context of the frameshifted polyprotein [80]. Considering that palmitoylation only occurs on the cytosolic face of cellular membranes [82], the distinct modification state of the two forms of the polyprotein is therefore suggestive of an underlying difference in their topologies. In this study, we set out to gain insight into the interplay between -1PRF and the topology of the structural polyprotein. We first mapped the topology of the Sindbis virus (SINV) structural polyprotein. Our results demonstrate that the structural polyprotein forms two topological isomers. The predominant topology features two TM domains upstream of the -1PRF site, and its formation coincides with production of the 6K protein. Alternatively, the minor topology contains an additional TM domain upstream from the -1PRF site that is linked to the production of TF. Using protein engineering in conjunction with coarse-grained molecular dynamics (CGMD) simulations, we demonstrate that the efficiency of -1PRF depends upon the force generated by the translocon-mediated membrane integration of the extra TM domain within the minor topomer. Together, our observations highlight novel connections between the cotranslational folding, biosynthesis, and processing of the alphavirus structural polyprotein. Moreover, our findings reveal a novel mechanism that regulates the overall efficiency of -1PRF.

3.2 Results and discussion

Topological properties of the alphavirus structural polyprotein

The current model of the alphavirus structural polyprotein suggests that the E2 and 6K proteins each contain two TM domains [68, 83]. However, there are two caveats to this model. First, cryo-EM structures reveal that the E2 protein only contains a single TM domain in the context of the viral envelope [84, 85]. Although it has been speculated that a second TM domain within E2 is somehow extruded from the membrane during processing, the marginal hydrophobicity of this segment also raises the possibility that it may fail to undergo translocon-mediated membrane integration in the first place. Second, the hydrophobic portion of the SINV 6K protein is only 35 residues in length, which is quite short for a segment containing two putative TM domains and a loop. These ambiguous topological signals suggest that this portion of the polyprotein is frustrated and could potentially form multiple topological isomers [86], as has been suggested for the coronavirus E protein [87].

To survey the topological preferences of the E2-6K region, we scanned its sequence using a knowledge-based algorithm that predicts the energetics associated with the transfer of polypeptide segments from the translocon to the ER membrane (ΔG predictor) [88]. Energetic predictions suggest that only the regions corresponding to the first hydrophobic segments within the E2 (TM1) and 6K (TM3) proteins are sufficiently hydrophobic to undergo robust membrane integration ($\Delta G < 0$; Fig. 3.1C). In contrast, the translocon-mediated membrane integration of the second hydrophobic segment within E2 (TM2) is predicted to be inefficient (Fig. 3.1C). To test these predictions, we measured the translocon-mediated membrane integration of each putative TM domain using a glycosylation-based translocation assay [89]. Briefly, the sequences of each individual putative TM domain were cloned into a chimeric leader peptidase (Lep) reporter protein [89]. Membrane integration of the putative TM segment (blue helix in Fig. 3.1D) results in the modification of only a single glycosylation site in Lep, whereas the passage of the segment into the lumen results in the modification of two glycosylation sites (Fig. 3.1D). Chimeric Lep proteins were produced by in vitro translation in the presence of canine rough microsomes, which contain native ER membranes and translocons. Consistent with predictions, Lep proteins containing TM1 and 3 acquire a single glycosyl modification, which suggests that these segments undergo robust translocon-mediated membrane integration (Fig. 3.1E). In contrast, the translocon-mediated membrane integration of the second putative TM domain of E2 (TM2) is significantly less efficient (Fig. 3.1E). These observations suggest that the E2 and 6K proteins are each likely to contain a single TM domain (TM1 and 3; Fig. 3.2A). It should also be noted that -1PRF only modifies the sequence of the loops downstream from these TM domains (Fig. 3.1C, orange line) and is therefore unlikely to impact their topological preferences.

Based on the computational and biochemical results in Fig. 3.1, we generated a topological model of the SINV structural polyprotein (Fig. 3.2A). This model correctly places the E2 and E1 ectodomains within the ER lumen and places the two palmitoylated cysteine residues in E2 (Cys716 and Cys718) within the cytosol

[90, 91]. To probe the topological preferences of the SINV polyprotein in the cell, we produced and characterized a series of reporter constructs that begin with the E3 protein and end at the C-terminal edge of each of the three putative TM domains within E2 and 6K (Fig. 3.2A). Each of these fragments was genetically fused to a C-terminal cassette containing a short GS linker and glycosylatable GFP (gGFP) gene, which contains two glycosylation sites within the core of the enhanced GFP (eGFP) protein [92]. Topological signals that direct the gGFP protein into the cytosol will produce a fluorescent gGFP, whereas the glycosylation of gGFP within the ER lumen renders the protein nonfluorescent (Fig. 3.2B). Each construct was then expressed in HEK293T cells, and flow cytometry was used to quantify the fluorescence intensity of the gGFP reporter at a consistent expression level, as judged by the intensity of a bicistronic reporter protein. Expression of the reporter constructs containing gGFP downstream from TM1 and 2 (see Fig. 3.2A) generates fluorescent gGFP (Fig. 3.2C), which suggests that the C termini of these TM domains reside within the cytosol. In contrast, the reporter construct with gGFP downstream from TM3 (after Arg785) exhibits an attenuated GFP signal at an equivalent expression level (Fig. 3.2C), which suggests that the gGFP fused to the C terminus of TM3 is projected into the ER lumen. Placement of the gGFP after the full stretch of hydrophobic amino acids in the 6K protein (after Tyr807; see Fig. 3.2A) also results in an attenuated gGFP signal (TM3+ in Fig. 3.2C), which confirms that the full stretch of hydrophobic residues near TM3 only spans the membrane once. Thus, results from this cellular reporter assay (Fig. 3.2C) are consistent with predictions (Fig. 3.1C), in vitro translation data (Fig. 3.1E), and the model shown in Fig. 3.2A. These observations together confirm that the E1 and 6K proteins each contain a single TM domain in the most abundant form of the polyprotein.

Link between topology and -1 PRF

The topological properties of the structural polyprotein described above have implications for the manner in which it is processed at the ER membrane. Our model suggests the cluster of unmodified cysteines in the 6K protein (Cys786, Cys787, Cys789, and Cys790) reside at a C-terminal portion of TM3 that is projected into the ER lumen (Fig. 3.3A) and is therefore inaccessible to palmitoylating enzymes. However, these same residues are palmitoylated in the TF protein [80], which suggests that the orientation of TM3 must become inverted upon frameshifting to expose them to the cytosolic leaflet. Such an inversion could potentially occur as a consequence of the membrane integration of TM2 (Fig. 3.3B), which exhibits a weak propensity to undergo translocon-mediated membrane integration (Fig. 3.1E). Furthermore, the efficiency associated with the translocon-mediated membrane integration of TM2 (≈20%; Fig. 3.1E) is comparable with the frequency of -1PRF in the SINV polyprotein (≈16%) [77]. Taken together, these observations potentially suggest a connection between the formation of a secondary topomer and -1PRF.

Based on these observations, we hypothesize that the translocon-mediated membrane integration of TM2 is mechanistically linked to -1PRF and the translation of the TF protein. Our model suggests mutations that alter the translocon-mediated membrane integration of TM2 should have a direct impact on -1PRF (Fig. 3.3). To test this hypothesis, we assessed whether mutations that alter the hydrophobicity of TM2 also influence -1PRF. We designed two double mutants that alter the hydrophobicity of TM2 by introducing hydrophobic residues at native polar residues, and vice versa. Both energetic predictions and in vitro translation measurements suggest the introduction of two nonnative leucine residues into TM2 (T738L/S739L, LL mutant) enhances the translocon-mediated membrane integration of TM2 (predicted $\Delta\Delta G = -1.7$ kcal/mol). In contrast, the introduction of two glutamate residues into TM2 (V735E/I736E, EE mutant) is predicted to increase its transfer free energy by 3.3 kcal/mol, which should significantly reduce its membrane integration efficiency.

To determine whether the cotranslational membrane integration of TM2 impacts translational recoding, we measured the effects of these substitutions on ribosomal frameshifting. PRF is most commonly measured using Dual-Luciferase reporters, which fuse luciferase domains to the 5’ (Renilla luciferase, 0-frame) and 3’ (firefly luciferase, -1 frame) of the gene of interest. We generated a series of reporter constructs in which translation begins at the native E3 signal peptide and continues until the ribosome reaches a fluorescent mKate fusion domain that is encoded in the -1 reading frame downstream from the PRF site. To control for variations in transfection efficiency at the single-cell level, we included a downstream IRES cassette that drives the bicistronic expression of GFP from the reporter transcript. Reporter constructs encoding TM2 variants of the polyprotein were expressed in HEK293T cells, and cellular mKate intensities were quantitatively compared across cells within a discrete range of IRES-GFP intensities by flow cytometry. The average mKate intensity among cells expressing a reporter construct bearing mutations that knock out the native ribosomal slip site (UUUUUUA → GUUCCUA, SSKO) is 79 ± 5% (n = 3, mean intensity change ± S.D.) lower than that among cells expressing the WT form of the reporter (Fig. 4A), which confirms that mKate intensities reflect the efficiency of -1PRF. The EE substitutions in TM2 decrease mKate intensity by 61 ± 16% relative to WT (n = 3; Fig. 3.4A). In contrast, the LL substitutions increase the mKate intensity by 30 ± 11% (n = 3; Fig. 3.4A). Differences in the distributions of cellular intensities were found to be both reproducible and statistically significant (n = 3, Mann–Whitney U test, $\alpha = 0.001$). It should be emphasized that each of these reporters contains both the native slip site and stem loop regions and that these mutations do not alter their sequences. Thus, these findings demonstrate that -1PRF is sensitive to mutations that impact the membrane integration efficiency of TM2. Together, biochemical evidence suggests that TM2 is inefficiently recognized by the translocon (Fig. 3.1E), and cellular topology reporters suggest that this segment is most often localized within the cytosol (Fig. 3.2). Nevertheless, mutagenesis reveals that the propensity of the nascent chain to form a secondary topomer is positively correlated with -1PRF. These results are therefore suggestive of a mechanistic link between topogenesis and -1PRF.

Impact of nascent chain forces on ribosomal frameshifting

The apparent link between cotranslational folding and ribosomal frameshifting has implications for the mechanism of -1PRF in the SINV structural polyprotein. The portion of the transcript containing the EE and LL mutations is over 100 nucleotides upstream from the ribosomal slip site and should therefore not perturb the stimulatory RNA structures that are currently believed to modulate -1PRF [57, 77, 78]. These mutations instead alter the portion of the nascent chain that falls just outside of the ribosomal exit tunnel during -1PRF, which suggests that the nascent chain itself may stimulate frameshifting. Although it has yet to be implicated in ribosomal frameshifting, the cotranslational membrane integration and/or folding of the nascent chain is known to generate tension on the ribosome [2, 4, 93, 94]. Furthermore, the C-terminal residue of TM2 is positioned 45 residues upstream of the slip site, which corresponds to a distance that should maximize the tension on the nascent chain at the moment the slip site occupies the ribosomal active site [4, 93]. Previous investigations have demonstrated that the force generated by the membrane integration of the nascent chain is sharply dependent upon this spacing [4, 93]. Therefore, to assess the potential role of this force in ribosomal frameshifting, we generated a set of SINV -1PRF reporter constructs (used in Fig. 3.4A) containing a series of insertions and deletions that alter the distance between TM2 and the ribosomal slip site. Reporter constructs were then expressed in HEK293T cells, and -1PRF reporter intensities were quantitatively compared at a uniform expression level by flow cytometry. A comparison of reporter intensities reveals that -1PRF is maximized at the WT distance of 45 residues (Fig. 3.4B). In all cases, deletions and insertions that change the distance between TM2 and the slip site result in large, statistically significant reductions in the relative intensity of the -1PRF reporter (Fig. 3.4B, n = 3). Moreover, the insertion of a 10-residue G/S linker decreases the intensity of the frameshift reporter by 76 ± 8% (n = 3), which suggests that the membrane integration of TM2 is likely to be the primary driver of -1PRF within the SINV structural polyprotein. Together, these findings suggest that topological signals within the SINV structural polyprotein generate a mechanical force that stimulates -1PRF.

To further explore the interplay between sequence, topology, and force, we carried out CGMD simulations of the translation and translocon-mediated membrane integration of the nascent structural polyprotein [13, 14]. In these simulations, three-residue segments of the nascent chain are modeled as individual beads with physicochemical properties based on their constituent amino acids. New beads are translated at a rate of 5 amino acids/s and emerge from the ribosome-translocon complex into an environment with an implicit representation of the bilayer and cytosol [14]. These simulations were previously found to sufficiently recapitulate several aspects of cotranslational membrane protein folding, including the formation of topological isomers and the generation of tension on the nascent chain [14, 15, 95]. CGMD simulations of SINV polyprotein biosynthesis suggest that the nascent chain samples several different topological isomers (Fig. 3.5A), and that its topological heterogeneity persists after the polyprotein has cleared the translocon. TM2 undergoes translocon-mediated membrane integration (Fig. 3.5A, right) in only 44 ± 4% of the CGMD trajectories in which TM1 is correctly integrated into the membrane. Consistent with expectations, CGMD simulations suggest that the membrane integration efficiency of TM2 is enhanced by the LL mutations (51 ± 4%) and reduced by the EE mutations (11 ± 3%). This finding provides additional evidence that the topological frustration within this domain (see Fig. 3.3) arises primarily from its marginal hydrophobicity.

To evaluate the connection between pulling forces on the nascent chain and frameshifting, we measured the tension on the nascent chain at the point of elongation when the slip site occupies the ribosomal active site [15]. Pulling forces were highest for the LL variant, which averaged 4.2 pN higher than the WT. In contrast, the EE mutations reduce the pulling force on the nascent chain by an average of 2.1 pN relative to WT.

These results are consistent with the hypothesis that differences in frameshifting arise from the effects of these mutations on the pulling force on the nascent chain. Simulations of polyprotein variants bearing insertions or deletions that alter the distance between TM2 and the ribosomal slip site indicate that the native distance (45 residues) is nearly optimal for the transmission of pulling force through the nascent chain, which is consistent with the observed patterns in frameshifting (Fig. 3.4B). Overall, we find that the -1PRF efficiency associated with each polyprotein variant roughly scales with corresponding mean pulling force measurements from CGMD simulations (Pearson's R = 0.74, p = 0.036; Fig. 3.5B), which strongly suggests that the pulling forces generated by the translocon-mediated membrane integration of the nascent chain stimulate -1PRF.

An analysis of the spectrum of topological states sampled during translation reveals that the magnitude of the pulling force transmitted to the ribosome scales with the number of beads that occupy the translocon (Fig. 3.5C). This finding suggests that pulling forces are generated by the movement of hydrophobic transmembrane segments from the protein-conducting channel of the translocon to the hydrophobic membrane core, as has been established previously [4, 15, 93]. The apparent variation in the conformation of TM2 at the translocon provides an explanation of the observed differences in pulling forces. Simulations suggest that differences in pulling forces arise from variations in the distribution of topological isomers that form during translation of these variants (Fig. 3.5D). The LL mutant predominately samples conformations where the majority of TM2 beads are in the translocon (Fig. 3.5A, right), whereas the EE mutant almost exclusively adopts conformations in which the majority of TM2 beads fall outside of the translocon and within the cytosol (Fig. 3.5A, left). As passage through the translocon is a prerequisite for membrane integration, the relationship between pulling forces and residence of the nascent chain within the translocon is consistent with our model for structural polyprotein biogenesis (Fig. 3.3) and confirms that the translocon-mediated membrane integration of TM2 stimulates -1PRF.

Survey of frameshifting elements among alphavirus structural polyproteins

Our model suggests that the hydrophobicity of TM2 and its distance from the slip site are the key determinants of the -1PRF efficiency within the SINV structural polyprotein. To assess whether this mechanism is likely to be operative within other alphaviruses, we surveyed six related structural polyproteins for similar sequence elements. Sequence scans carried out with the ΔG predictor reveal that each form of the alphavirus structural polyprotein contains a marginally hydrophobic TM domain upstream from the ribosomal slip site. Predicted transfer free energies associated with the translocon-mediated membrane integration of these putative TM domains range from +1.4 to +2.7 kcal/mol (Table 1), which suggests that the translocon-mediated membrane integration of these segments is likely to be inefficient. Consistent with this notion, CGMD simulations of the translation of these polyproteins indicate that the membrane integration efficiency of these segments ranges from 33 to 64% (Table 1). Furthermore, these marginally hydrophobic TM domains reside 44–52 residues upstream of their corresponding -1PRF sites (Table 1), which suggests that the tension generated by their translocon-mediated membrane integration is likely to be propagated back to the slip site [4, 93]. Force measurements derived from CGMD simulations of polyprotein synthesis suggest that the tension in the nascent chain when the slip site occupies the ribosome is comparable with or greater than the tension generated during translation of the SINV variants characterized herein (Table 1). Taken together, these findings suggest that this -1PRF mechanism is likely to be conserved across the alphavirus genus. Additional investigations are needed to determine how nascent chain- and RNA-mediated -1PRF mechanisms are balanced against one another and how this mechanistic diversity ultimately influences viral evolution.

Table 3.1: Simulated integration forces and topologies of related alphavirus polyproteins. Integration refers the probability of TM2 integrating into the membrane according to simulation. A.A. to slip refers to the number of amino acids between TM2 and the slip site.

Strain	ΔG of TM2	Integration	A.A. to slip	Mean force
	(kcal/mol)		(# residues)	(pN, 95% CIs)
Sindbis	1.9	0.44	45	29.0 (28.8-29.1)
Eastern equine	2.4	0.51	45	31.1 (31.0-31.2)
Middleburg	1.6	0.33	52	34 (33.9-34.1)
Sleeping disease	2.7	0.33	44	25.6 (25.5-25.8)
Southern elephant seal	2.1	0.34	47	25.2 (25.0-25.3)
Semliki forest	2.1	0.64	49	33.2 (33.1-33.4)
Venezuelan equine	1.4	0.33	44	28.6 (28.5-28.7)

3.3 Conclusions

Using an array of biochemical, cellular, and computational methods, we show that the nascent SINV structural polyprotein forms a spectrum of topological intermediates during biosynthesis and that -1PRF is primarily driven by the translocon-mediated membrane integration of a marginally hydrophobic TM domain within the E2 protein. We also provide evidence to suggest that this mechanism is likely to be conserved across the alphavirus genus. To date, the mechanistic basis of -1PRF has been generally attributed to the kinetic effects of mechanochemical forces that are generated by structural elements within the mRNA. Indeed, we do find that PRF in the SINV structural polyprotein depends upon the RNA stem loop downstream of the slip site. Nevertheless, it is clear that the forces generated by the translocon-mediated membrane integration of TM2 dramatically enhance the frameshifting efficiency. To our knowledge, the findings reported herein constitute the first instance in which forces generated by conformational transitions in the nascent polypeptide chain have been implicated in the efficiency of PRF. Although additional investigations are needed to elucidate how pulling forces in the nascent chain physically stimulate -1PRF, a causative role for tension in both the transcript and nascent chain seems plausible, given that ribosomal frameshifting fundamentally arises from the movement of the tRNA with respect to the mRNA. It seems likely that cotranslational folding is one of many regulators, which include both host and viral effectors, that tune the net efficiency of PRF. This creates the potential for mechanistic diversity that could provide an evolutionary benefit for alphaviruses, as -1PRF is rendered tunable by either downstream or upstream mutations that impact the stability of the mRNA hairpin or the conformational ensemble of the nascent chain, respectively. This flexibility could also potentially provide the virus with a means of maintaining desired -1PRF levels in the presence of host factors that globally regulate -1PRF through mRNA interactions[67].

It should be noted that the implications of these findings potentially extend beyond the realm of viral proteins. A wide variety of molecular transitions have been found to generate tension within the nascent chain, including the folding of soluble domains near the ribosomal exit tunnel [2, 49] and the translocon-mediated membrane integration of nascent TM domains [4, 93]. These observations suggest that the tension in the nascent chain should fluctuate as the structural features emerge from the ribosome (Fig. 3.6A), which may therefore provide the ribosome with a readout for the progress of cotranslational folding. In the case of the SINV structural polyprotein, the topological frustration within the nascent chain leads to the production of two competing topomers that generate distinct pulling forces on the ribosome in a manner that ultimately impacts the fidelity and processivity of translation (Fig. 3.6, B and C). This translational feedback constitutes a new form of mechanochemical allostery on the ribosome [49]. Additional investigations are needed to explore the potential relevance of this type of cotranslational feedback to protein homeostasis. Indeed, interactions between the nascent chain and molecular chaperones are known to ratchet polypeptides across the membrane [96, 97] and may therefore contribute to pulling forces. This could potentially account for the fact that the deletion of components of the ribosome-associated chaperone complex has been found to attenuate -1PRF in yeast [98]. Future investigations are needed to evaluate the full range of -1PRF effectors and how these are potentially exploited for regulatory purposes.

Methodology

Computational predictions of topological energetics

The energetics associated with the translocon-mediated membrane integration of the nascent structural polyprotein were carried out using the ΔG predictor[89]. These predictions are generated using a window scan function that sums depth-dependent free energies associated with the transfer of amino acids from the translocon to the ER membrane[88]. Full sequence scans of varying window size were used to compare the predicted transfer free energies associated with each segment within the polyprotein to identify the segments that are most likely to undergo translocon-mediated membrane integration.

Coarse-grained simulations of polyprotein translation

CGMD simulations are based on a previously developed and tested approach [13, 14]. Briefly, simulations are carried out in the context of a coarsened representation of the ribosome exit tunnel, Sec translocon, and nascent chain. The nascent chain is composed of beads that each represent three amino acids, and new beads are sequentially added to the nascent chain to explicitly simulate translation. Translation occurs at a rate of 5 aa/s, which mimics the rate of translation by eukaryotic ribosomes. Each bead interacts with the translocon, ribosome, and other beads in a manner that depends on the hydrophobicity and charge of its composite amino acids. Interactions with the solvent and lipid bilayer are modeled implicitly. The ribosome and Sec translocon are fixed in place, with the exception of the lateral gate of the translocon, which stochastically switches between the open and closed conformations in a manner dependent on the free energy difference between the two configurations.

The geometries of the ribosome and translocon are based on cryo-EM structures[99]. Residue-specific interactions have been parameterized based on over 200 μs of simulations with the MARTINI forcefield. Fitting is performed using a Bayesian uncertainty quantification framework [100]. Integration is performed using overdamped Langevin dynamics with a diffusion coefficient of 253 nm^2/s and a time step of 300 ns. Despite the significant simplifications involved in this model, the CGMD model has proven accurate in capturing the integration probabilities, topology distributions, and forces experienced in previous studies [13–15].

To obtain the distribution of topologies for various polyprotein mutants, the translation and integration of each sequence was simulated 560 times. To reduce computational cost, simulations only included the first three TMs of the alphavirus polyprotein. To focus on the topological preferences of TM2, restraints were applied to enforce that TM1 adopts its native topological orientation. Pulling force measurements were performed by pausing translation when the -1PRF site resides within the ribosomal peptidyl transfer center. During this pause, the final bead was fixed in place, and the force on the bead exerted by the nascent chain was measured along the translocon channel axis. Due to the truncation of the exit tunnel in our model, the final bead corresponds to the amino acids 27 residues N-terminal of the -1PRF site. Translation is paused for 3 s, which is equivalent to the time it would take to translate five beads. This relatively short time window ensures that the distribution of polyprotein topologies is not affected by the pause. During this window, pulling forces were measured at a rate of 333 frames/s. To sample a wide range of topologies and conformations, each mutant was independently simulated 560 times. This protocol is analogous to force measurements performed in previous work, with the exception of a shortened pause duration [15].

Experimental methodology is available in the original publication on which this chapter is based[34].

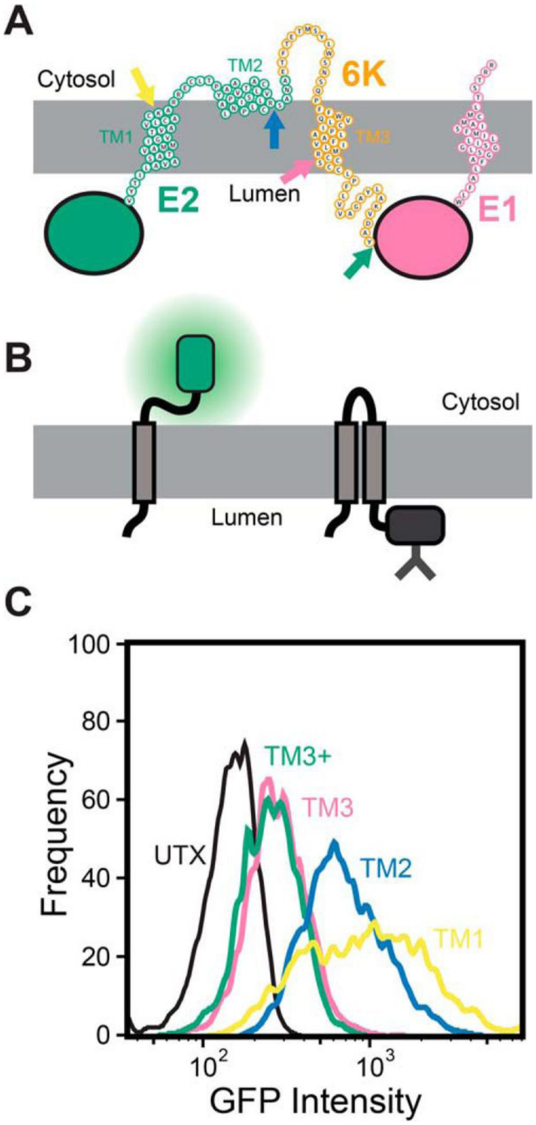


Figure 3.2: Topological properties of the major form of the SINV structural polyprotein. A) a cartoon depicts a putative topological model of the most abundant topology of the structural polyprotein that is consistent with computational and biochemical data. The positions at which gGFP reporter domains were fused to determine the cellular compartmentalization of the C-terminal portion of the segments corresponding to TM1 (yellow), TM2 (blue), TM3 (pink), and the C-terminal edge of the hydrophobic portion of 6K (TM3+, green) are indicated with arrows. B) a cartoon depicts the manner in which the cellular compartmentalization of the gGFP reporter domain alters its fluorescence. Topological signals that direct the gGFP domain into the cytosol will generate a fluorescent gGFP (left), whereas topological signals that direct the gGFP domain into the lumen (right) generate a glycosylated, nonfluorescent fusion domain. C) reporter constructs were analyzed by flow cytometry. A histogram from a representative trial depicts the gGFP intensities associated with 3,000 transfected cells.

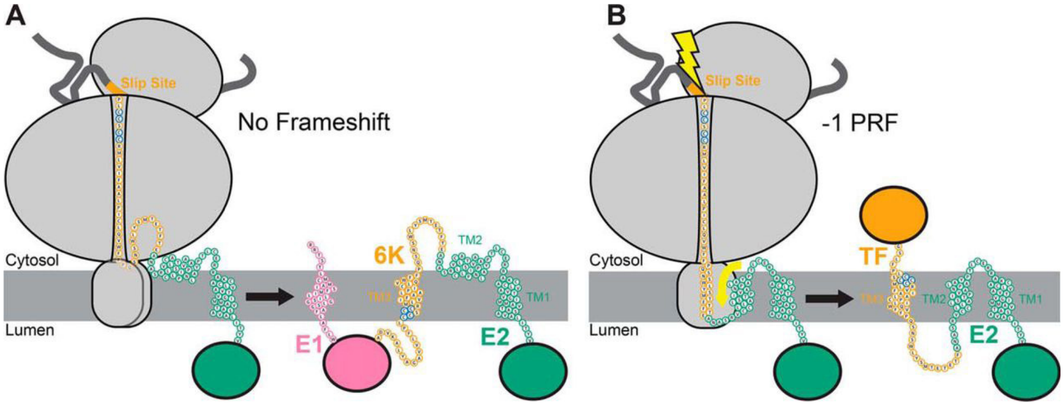


Figure 3.3: Putative model for the interplay between translocon-mediated membrane integration and ribosomal frameshifting. A cartoon depicts the putative manner in which cotranslational folding of the nascent structural polyprotein is linked to -1PRF. A) a cartoon depicts the topological properties of the major form of the nascent structural polyprotein. TM2 is too polar to robustly partition into the membrane during translation (left). As a result, the E2 protein only contains a single TM domain in the context of the major form of the structural polyprotein (right). Cysteine residues that are conditionally palmitoylated in the frameshifted form of the polyprotein are shown in blue. B) a cartoon depicts the topological properties of the frameshifted form of the nascent structural polyprotein. TM2 is hydrophobic enough to occasionally partition into the membrane during translation (left), which imposes a tension on the ribosome that stimulates -1PRF. As a result, the E2 protein contains two TM domains in the context of the frameshifted form of the structural polyprotein (right). Cysteine residues that are conditionally palmitoylated in the frameshifted form of the polyprotein are shown in blue.

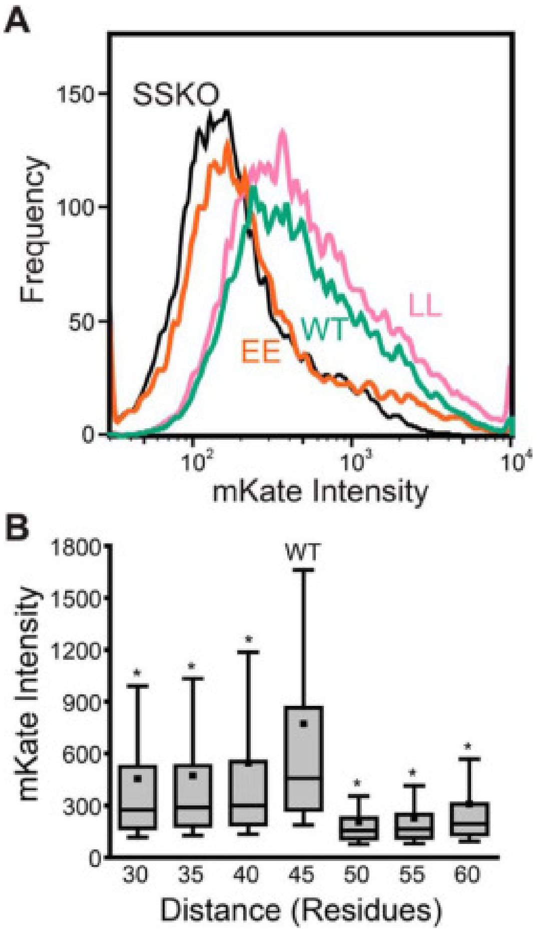


Figure 3.4: Influence of sequence modifications on -1 programmed ribosomal frameshifting. A fragment of the SINV polyprotein containing mKate fused in the -1 reading frame was used to compare the effects of sequence modifications on -1PRF levels. A) -1PRF reporter constructs containing the WT (green), the EE double mutant (orange), and the LL double mutant (pink) TM2 sequence were were analyzed by flow cytometry. A histogram depicts the mKate intensities associated with 3000 cells. The intensity distribution of cells expressing a reporter construct containing the WT version of TM2 and a mutated slip site (UUUUUUA → GUUCCUA, SSKO, black) is also shown for reference. These trends were found to be statistically distinct from those of WT at a significance level (α) of 0.001 according to a Mann–Whitney U test. B) -1PRF reporter constructs containing a series of deletions and G/S linker insertions within the loop between TM2 and -3 that alter the number of residues between the slip site and TM2 were compared by flow cytometry. The distribution of fluorescent mKate reporter intensities from a representative experimental replicate is shown for cells expressing each reporter construct. Analysis of intensity distributions using a Mann–Whitney U test suggests that the distributions of each test construct are statistically distinct from that of the WT reporter (45 residues) at a significance level (α) of 0.001 (*).

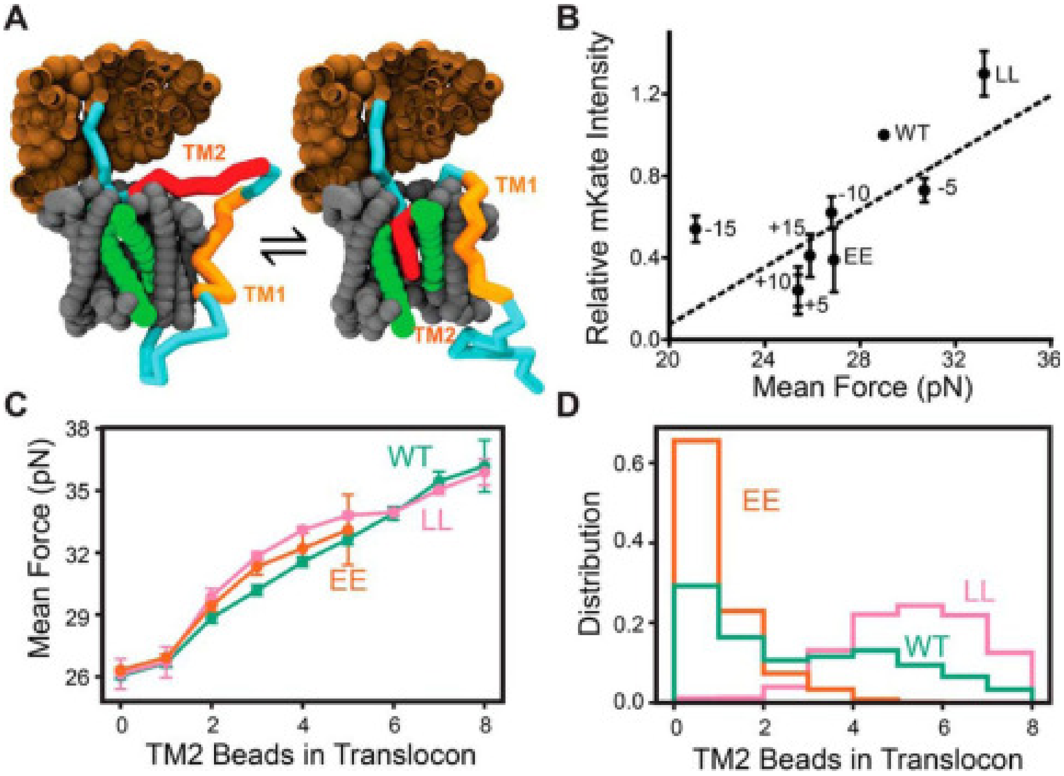


Figure 3.5: Coarse-grained molecular dynamics simulations of polyprotein biosynthesis. CGMD simulations were carried out to simulate biosynthesis of a series of SINV structural polyprotein variants, and the pulling force on the nascent chain was calculated at the point in which the ribosome occupies the slip site. A) representative snapshots from CGMD simulations are shown during translation at the slip site, which is the point during elongation at which pulling forces on the nascent chain are measured. The ribosomal exit tunnel is shown in brown. The translocon is shown in gray, and its lateral gate is highlighted in green. The nascent chain is shown in blue, except for the portions that correspond to TM1 and 2, which are highlighted in orange and red, respectively. B) -1PRF fluorescence reporter (mKate) intensity values for cells expressing a series of modified polyprotein variants were normalized relative to WT and plotted against the corresponding mean force measurements calculated from 560 CGMD trajectories. The identity of each variant along with a linear fit of the data (dashes) are shown for reference (Pearson's R = 0.74, p = 0.036). C) pulling force measurements are compared among topological isomers for the WT (green), LL (pink), and EE (orange) polyprotein variants in which the number of TM2 residues (or beads) located within the translocon was found to vary. Values reflect the mean force, and error bars represent a 95% confidence interval. D) a histogram depicts the number of TM2 residues (or beads) within the translocon among the conformational trajectories sampled during biosynthesis of the WT (green), LL (pink), and EE (orange) variants of the SINV polyprotein.

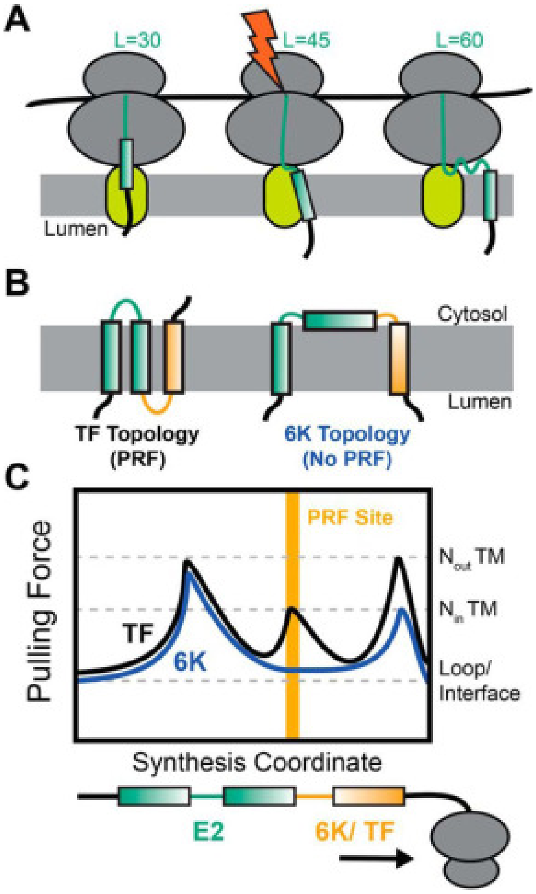


Figure 3.6: Interplay between topology, pulling force, and programmed ribosomal frameshifting. Cartoons depict the manner in which the translocon-mediated membrane integration of the nascent chain generates a fluctuation in pulling force that triggers PRF during synthesis of the SINV structural polyprotein. A) the pulling force generated by the translocon-mediated membrane integration of each TM domain generates a pulling force on the nascent chain that is maximized during the conjugation of the amino acid that lies ≈45 residues upstream of the C-terminal residue of the TM domain. B) the translocon-mediated membrane integration of TM2 is marginally efficient, which results in the formation of two topologies during translation of the SINV polyprotein. TM2 most often fails to undergo translocon-mediated membrane integration, which results in the formation of a topology featuring only two TM domains (TM1 and 3) in the form of the polyprotein containing the 6K protein. However, the translocon-mediated membrane integration of TM2 generates an alternative topology in the frameshifted form of the polyprotein containing the TF protein. C) a hypothetical plot of the fluctuations in the nascent chain pulling force during the translation of the two forms of the SINV structural polyprotein is shown. The translocon-mediated membrane integration of TM2 generates an extra pulling force on the nascent chain while the slip site occupies the ribosomal P-site, which stimulates -1PRF.

Chapter 4

COORDINATION OF -1 PROGRAMMED RIBOSOMAL FRAMESHIFTING BY TRANSCRIPT AND NASCENT CHAIN FEATURES REVEALED BY DEEP MUTATIONAL SCANNING

Adapted from:

Patrick J Carmody, Matthew H. Zimmer, Charles P Kuntz, Haley R Harrington, Kate E Duckworth, Wesley D Penn, Suchetana Mukhopadhyay, Thomas F Miller III, and Jonathan P Schlebach. Coordination of -1 Programmed Ribosomal Frameshifting by Transcript and Nascent Chain Features Revealed by Deep Mutational Scanning. *bioRxiv*, 2021. doi: 10.1101/2021.03.11.435011.

Programmed ribosomal frameshifting (PRF) is a translational recoding mechanism that enables the synthesis of multiple polypeptides from a single transcript. In the alphavirus structural polyprotein, -1PRF is coordinated by a “slippery” sequence in the transcript, an RNA stem-loop, and a conformational transition in the nascent polypeptide chain. To characterize each of these effectors, we measured the effects of 4,530 mutations on -1PRF by deep mutational scanning. While most mutations within the slip-site and stem-loop disrupt -1PRF, mutagenic effects upstream of the slip-site are far more variable. Molecular dynamics simulations of polyprotein biogenesis suggest many of these mutations alter stimulatory forces on the nascent chain through their effects on translocon-mediated cotranslational folding. Finally, we provide evidence that the coupling between cotranslational folding and -1PRF depends on the translation kinetics upstream of the slip-site. These findings demonstrate how -1PRF is coordinated by features within both the transcript and nascent chain.

4.1 Introduction

Programmed ribosomal frameshifting (PRF) is a translational recoding mechanism that occurs in all kingdoms of life. Though a handful of prokaryotic and eukaryotic PRF motifs have been identified, most of the well-characterized motifs are found within viral genomes [101]. Viruses utilize ribosomal frameshifting to increase their genomic coding capacity and to regulate the stoichiometric ratios of viral protein synthesis. Some rely on these motifs to coordinate genomic replication, while others utilize PRF to regulate the production of the structural proteins that mediate assembly [102]). In some cases, the frameshift products are themselves virulence factors that antagonize the host interferon response[73–76]. For these reasons, the efficiency of PRF, which is globally regulated by both host and viral proteins [66, 67], is often critical for infection and immunity.

The most common type of PRF involves a -1 shift in reading frame (-1PRF) and minimally requires a “slippery” heptanucleotide site in the transcript that typically takes the form X_1 XXY_4 YYZ_7, where X represents three identical nucleotides, Y is an A or U, and Z can be an A, C, or U [103]. This structure reduces the thermodynamic penalty associated with non-native codon-anticodon interactions in the -1 reading frame [64], which increases the intrinsic efficiency of ribosomal frameshifting by several orders of magnitude [104, 105]. Most -1PRF motifs also feature an mRNA stem-loop or pseudoknot that forms 6-8 bases downstream of the slip-site [106, 107]. This structure imposes a mechanical resistance to translocation that stalls the ribosome on the slip-site and provides time for the anticodon loops of the A- and/ or P-site tRNA to sample alternative base pairing interactions [59–62]). Frameshifting is also enhanced by the depletion of the t-RNA pool that decodes the YYZ_7 codon (“hungry” frameshifting), which provides the opportunity for the P-site tRNA to move out of frame [63]. In addition to mRNA and tRNA effectors, our group also recently found that -1PRF is enhanced by mechanical forces generated by the cotranslational folding of the nascent polypeptide chain [34]. This finding has since been echoed by recent evidence suggesting that the cotranslational folding of nsp10 alters -1PRF during translation of SARS-CoV-2 genome [108]. Thus, emerging observations suggest the efficiency of ribosomal frameshifting can be modified by a wide variety of factors ranging from the rate of translation to the structural properties of the transcript and nascent chain. Nevertheless, a unified understanding of how the features of the mRNA, tRNAs, and nascent polypeptide chain cooperate to tune the efficiency of PRF is needed [109].

In this work, we utilize deep mutational scanning (DMS) to identify additional sequence constraints associated with -1PRF in the Sindbis virus (SINV) structural polyprotein. We report the effects of 4,530 mutations on the relative efficiency of -1PRF in the context an extended genetic reporter containing the native features that coordinate the biosynthesis and cotranslational folding of the nascent polyprotein at the endoplasmic reticulum (ER) membrane (Fig. 4.1A). Incorporating these elements into our reporter facilitated the identification of -1PRF modulators within both the transcript and nascent chain. As expected, mutational patterns within the slip-site and the region immediately downstream of the slip-site are consistent with the formation of a previously characterized RNA stem-loop [77, 110]. However, we also find that -1PRF is sensitive to numerous mutations within the ≈160 base region preceding the slip-site, which encode a pair of transmembrane (TM) domains that occupy the ribosomal exit tunnel and translocon during frameshifting. In conjunction with a suite of atomistic and coarse-grained molecular dynamics (CGMD) simulations [111], we provide evidence that many mutations within this region attenuate -1PRF through their effects on a folding intermediate that forms during translocon-mediated membrane integration of the nascent polyprotein. We also find that -1PRF appears to be sensitive to changes in the translation kinetics within the region immediately preceding the slip-site, which influences the pulling forces generated by the cotranslational folding of the nascent chain. Together, these findings provide a detailed description of the sequence constraints of -1PRF within the alphavirus structural polyprotein.

4.2 Results

High-throughput evaluation of programmed array of -1 PRF reporters

To probe the sequence constraints of -1PRF, we employed DMS to map the effects of mutations on ribosomal frameshifting within the SINV structural polyprotein. Mutagenic effects were assessed in the context of a reporter that selectively generates a fluorescent protein (mKate) as a result of -1PRF during translation and processing of the SINV structural polyprotein (Fig. 1A). The design of this reporter is similar, in principle, to those described in a recent high-throughput investigation of PRF [109], except that we included a much larger fragment of the transcript (1.7 kb, 557 amino acids) and opted to track expression using an IRES-eGFP cassette rather than an N-terminal fusion domain. This design preserves the integrity of the native signal peptide in the E3 protein (Fig. 4.1A), which is critical for the targeting and maturation of the nascent polyprotein at the ER membrane. To survey the effects of mutations on the reporter signal, we first generated a one-pot genetic library by randomizing each 0-frame codon within a region beginning at the N-terminal codon of the first TM domain (TM1) of the E2 protein (I696) and ending 23 codons downstream of the slip-site (P817, Fig. 4.1A). We then used this genetic library (7,614 total variants) to generate recombinant stable HEK293T cell lines in which each cell expresses a single variant from a unique Tet-inducible promoter within the genome, as previously described [112]. Importantly, the specificity of this

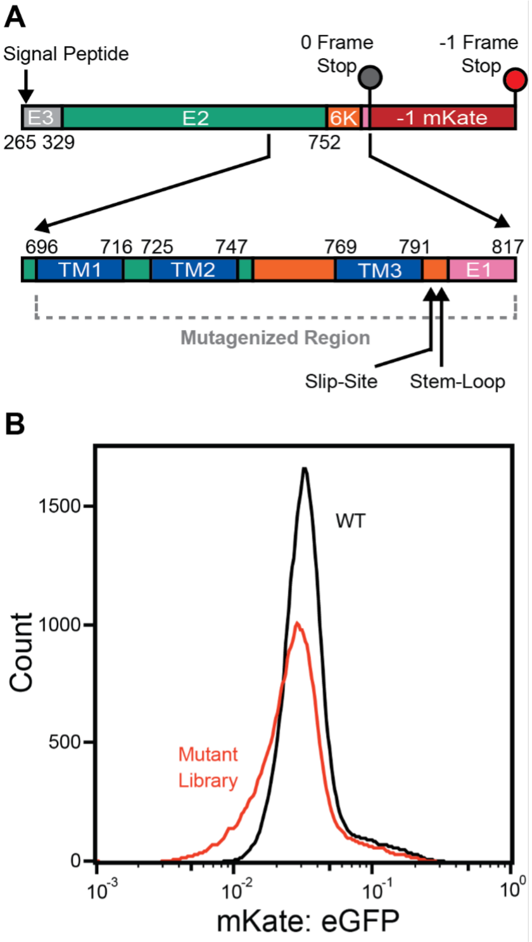


Figure 4.1: Library of ribosomal frameshifting reporter variants. A) A cartoon depicts the structure of a reporter that generates a fluorescent protein (mKate) as a result of -1PRF in the SINV structural polyprotein (top). A library of reporter variants was generated by randomizing every 0-frame codon within a region ranging from residues 696 to 817 (bottom). B) A histogram depicts the range of single-cell mKate:GFP intensity ratios among recombinant stable cells expressing the WT reporter (black) or the library of reporter variants (red).

recombination reaction ensures individual variants are expressed at similar levels within each cell. Nevertheless, we normalized mKate levels according to the IRES-eGFP intensity to account for small variations in expression, as has been previously described [113].

On average, stable cells expressing individual single-codon variants have an mKate: eGFP intensity ratio that is comparable to that of cells expressing the WT reporter (Fig. 4.1B). However, a sub-set of these cells express variants that generate intensity ratios outside the WT intensity range (Fig. 4.1B). This suggests that, while most mutations have minimal impact, the library contains many that modify the efficiency of -1PRF. To analyze the effects of individual mutations, we fractionated these cells according to their mKate:eGFP intensity ratios using fluorescence activated cell sorting (FACS), quantified the variants within each fraction by deep sequencing, then used these data to estimate the intensity ratio for each individual variant as described previously [114]. Intensity ratios were generally consistent across biological replicates when normalized according to the WT value (Pearson's R = 0.79-0.83).

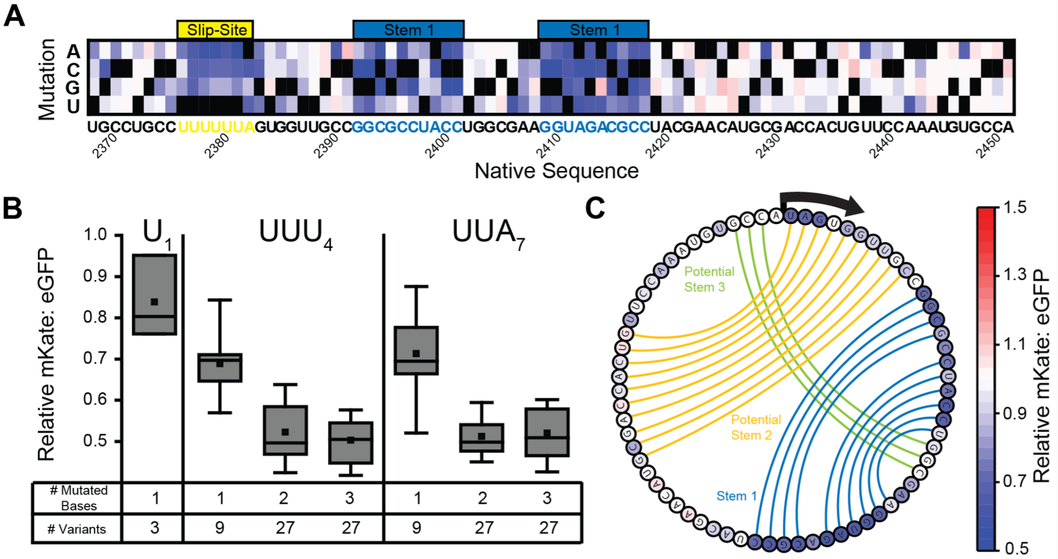


Figure 4.2: Impact of slip-site and stem-loop mutations on ribosomal frameshifting. A) A heatmap depicts the effects of every single nucleotide substitution (Y-coordinate) at each position (X-coordinate) within or near the slip-site on the mKate:eGFP ratio as determined by DMS. Ratios were normalized relative to that of WT, and black squares are shown in place of the native nucleobase at each position. The relative position of the slip-site (yellow) and stem 1 (blue) are shown for reference. B) A box plot depicts the distribution of relative mKate:eGFP values associated with variants bearing one, two, or three mutations at certain positions within the slip-site. C) Connecting lines depicting predicted base pairing interactions within stem 1 (blue), potential stem 2 (orange), and potential stem 3 (green) are shown in the context of the nucleotide sequence downstream from the slip-site [77]. Each nucleotide is colored according to the average relative mKate:eGFP value across the three mutations at each position.

Sequence constraints within the heptanucleotide slip-site

Slippery nucleotide sequences enable favorable codon-anticodon base pairing in the -1 reading frame, and the efficiency of -1PRF is therefore highly sensitive to mutations that disrupt the slip-site [115]. As expected, we find that every single-nucleotide variant (SNV) within the heptanucleotide slip-site (U_1 UUU_4 UUA_7) measurably decreases -1PRF (Fig. 4.2A, yellow). Furthermore, an analysis of 128 single, double, and triple mutants within the slip-site reveals that frameshifting decreases with an increasing mutational load within the UUU_4 (P-site) or UUA_7 (A-site) codons (Fig. 4.2B). These mutations presumably decrease the efficiency of -1PRF by increasing the free energy difference between the codon-anticodon base pairing in the 0 and -1 reading frames. However, we also find frameshifting to be sensitive to mutations within the adjacent codon upstream of the slip-site (Fig. 4.2A). This observation potentially suggests the observed PRF signal may arise from some combination of -1 and -4 frameshifting transitions, as has been observed for other PRF motifs [116]. It is also possible that the observed -1PRF signal arises from a mix of canonical frameshifting (two-tRNA) and "hungry" frameshifting (one-tRNA), where the P-site tRNA frameshifts prior to the decoding of the UUA_7 codon. The overall efficiency of both the HIV and Semliki forest virus -1PRF motifs is maximized when the relative abundance of the tRNA that decodes the UUA_7 codon is depleted [63]. Similarly, we find that -1PRF is decreased by all three A_7 mutations in the SINV slip-site (Fig. 4.2A), which should attenuate hungry frameshifting given that these codons are decoded by tRNA that are more abundant in these cells . Together, these observations highlight the essential role of the slip-site, yet suggest the observed -1PRF signal may arise from a spectrum of frameshifting transitions.

Sequence constraints within the stimulatory RNA stem-loop

Efficient ribosomal frameshifting requires a stimulatory RNA secondary structure that impedes translocation and increases the dwell time of the ribosome on the slip-site. Stimulatory structures within the sub-genomic RNA of a number of alphaviruses have been previously characterized [77, 78, 110]. Secondary structure predictions for SINV suggest the region adjacent to the slip-site is capable of forming a pseudoknot consisting of one primary stem-loop (stem 1), a secondary stem-loop within the spacer region (potential stem 2), and a short stem that basepairs within the loop of stem 1 (potential stem 3). However, experimental evidence suggests frameshifting is primarily driven by stem 1 [77]. Consistent with these findings, our map of mutational effects shows that PRF is highly sensitive to mutations within a region beginning 8 bases downstream of the slip-site (Fig. 4.2A). Moreover, the bases that pair within stem 1 are more sensitive to the effects of single nucleotide variations relative to those within potential stems 2 and 3 (Fig. 4.2C). Interestingly,

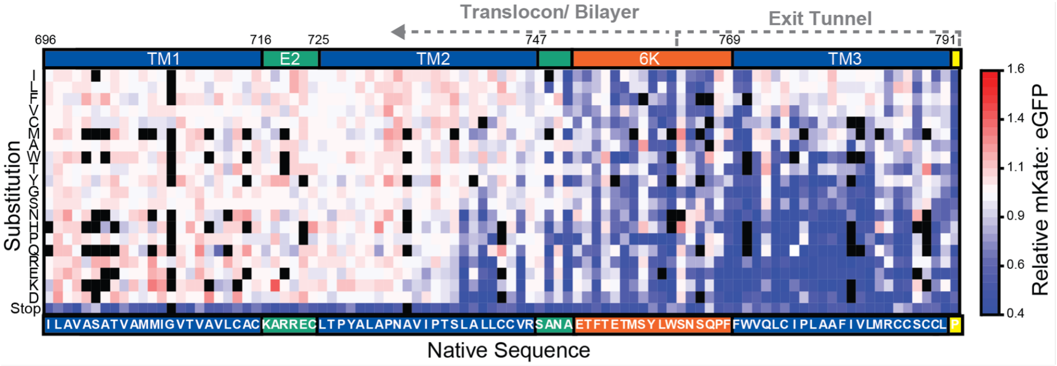


Figure 4.3: Impact of nascent chain mutations on ribosomal frameshifting. A heatmap depicts the effects of every amino acid substitution (Y-coordinate) at each position (X-coordinate) within the nascent polypeptide chain on the mKate:eGFP ratio as determined by DMS. Substitutions are arranged from most hydrophobic (top) to most polar (bottom). Ratios were normalized relative to that of WT, and black squares indicate mutations that lack coverage. The positions of each TM domain (blue) and the slip-site (yellow) as well as the domain boundaries for the E2 (green) and 6K (orange) proteins are shown for reference. The amino acid that corresponds to the first codon of the slip site (containing U1) is indicated in yellow. The residues that are likely to reside within the exit tunnel or outside of the ribosome during ribosomal frameshifting are indicated with dashed lines.

stem 1 contains a single mismatch (Fig. 4.2C), and PRF is enhanced by the only SNV that creates a G-C base pair at this position (A2414G, Fig. 4.2A). We also find that frameshifting is enhanced by two SNVs that extend the base of the stem-loop by an additional base pair (C2391A, U2419G, Fig. 4.2A). Together, these observations confirm that stem 1 is critical for PRF efficiency. While more stable equilibrium structure(s) could potentially form, our functional analysis suggests stem 1 is the dominant structure within the ensemble that stimulates -1PRF between rounds of ribosome-mediated unwinding [57].

Sequence constraints within the nascent polypeptide chain

We recently found that the mechanical force generated by the translocon-mediated membrane integration of the nascent chain enhances -1PRF in the SINV structural polyprotein [34]. Our DMS results reveal that ribosomal frameshifting is sensitive to a variety of mutations that alter the amino acid sequence within the portion of the nascent chain that has been translated at the point of frameshifting, including mutations to residues within the exit tunnel, between the exit tunnel and translocon, and within the second transmembrane domain of E2 (TM2, Fig. 4.3). Consistent with previous findings [34], our results show that the introduction of polar or charged

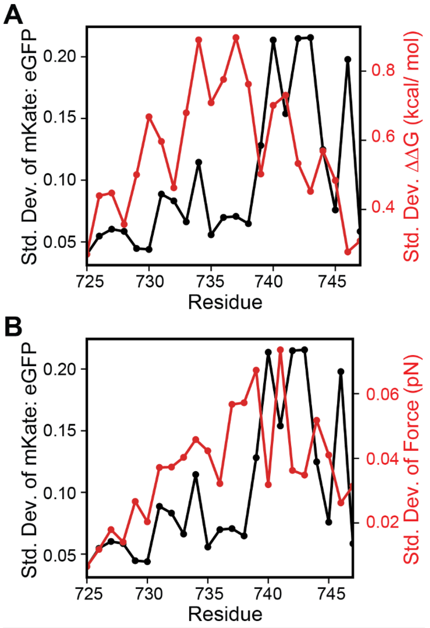
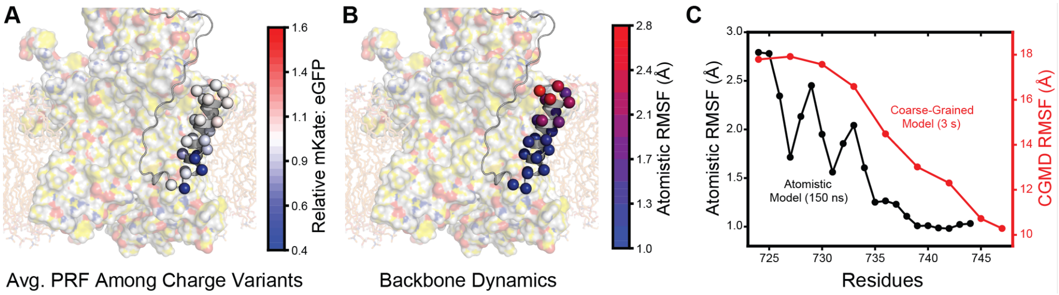


Figure 4.4: Impact of TM2 mutations on -1PRF, membrane integration energetics, and pulling forces. The standard deviation of the relative mKate:eGFP intensity ratio measurements for all 19 amino acid substitutions at each residue are plotted in relation to the corresponding standard deviation of (A) the predicted change in the transfer free energy associated with the translocon-mediated membrane integration of TM2 as determined by the ΔG predictor [88] and B) the relative pulling force generated by its translocon-mediated membrane integration as was determined by CGMD simulations.

side chains at certain positions within TM2 reduces -1PRF while mutations that enhance the hydrophobicity of TM2 generally increase -1PRF (Fig. 4.3). However, the effects of mutations on -1PRF appear to deviate from their predicted effects on the energetics of translocon-mediated membrane integration. Mutations to residues near the center of TM2 are predicted elicit the largest changes in the energetics of translocon-mediated membrane integration (Fig. 4.4A, red), which reflects the depth-dependence of amino acid transfer free energies [88, 117]. In contrast, variations in mKate:eGFP intensity ratios are most pronounced among mutations within the C-terminal residues of TM2 (Fig. 4.4A, black). This discrepancy suggests that, in addition to the hydrophobicity of TM2 and its corresponding transfer free energy, stimulatory pulling forces may hinge upon structural and/or dynamic constraints associated with its translocon-mediated membrane integration.

To explore the effects of these mutations on the nascent chain, we performed CGMD simulations of polyprotein translation that include coarse-grained representations of the ribosomal exit tunnel and the Sec translocon in an implicit lipid bilayer [118]. The nascent polypeptide is treated as a polymer of beads (3 AA/bead) that emerge from the ribosome exit tunnel at the physiological rate of translation (5 AA/s). The hydrophobicity and charge of each bead is derived from its three constituent amino acids. To quantify stimulatory pulling forces, we paused translation for three seconds once the ribosome reaches the slip-site and measured the force on the nascent chain as it explores the environment in, on, and around the translocon as has been previously described [15]. Consistent with previous findings [34], simulations of the 437 missense variants bearing mutations within TM2 reveal that mutations that attenuate pulling forces on the nascent chain generally reduce frameshifting; pulling forces are correlated with experimentally derived relative mKate:eGFP intensity ratios (Pearson's R = 0.48). Hydrophobic substitutions in TM2 increase the propensity of TM2 to enter the translocon and partition into the membrane in these simulations, which generates higher pulling forces and enhanced frameshifting. Conversely, polar mutations decrease translocon occupancy, membrane integration of TM2, and pulling forces. Importantly, both pulling forces from CGMD and the observed mKate:eGFP intensity ratios are most sensitive to mutations within the C-terminal residues of TM2 (Fig. 4.4B). This agreement suggests stimulatory forces originate from conformational transitions that are mediated by dynamic interactions between TM2, the translocon, and the lipid bilayer.

To evaluate the structural context of mutations that influence -1PRF, we constructed an atomistic model of TM2 nested within the translocon. An initial model was generated by mapping the polyprotein sequence onto a cryo-EM structure of a nascent chain intermediate that adopts an N_{in} topology comparable to that of TM2 in the context of the translocon [119]. After a 150 ns equilibration of the translocon-nascent chain complex within an explicit lipid bilayer, TM2 adopts a tilted orientation in which the N-terminal portion of TM2 projects into the lipids while its C-terminal residues remain wedged within the lateral gate of the translocon (Fig. 4.5A). The N-terminal residues that interact with the bilayer are generally more dynamic than the C-terminal residues that interact with the translocon (Fig. 4.5B). TM2 achieves a similar topological orientation and exhibits similar dynamic fluctuations regardless of how the amino acid sequence is initially mapped onto the structural template of the nascent chain. While the magnitude of the dynamic fluctuations observed on the time scales of the atomistic simulations is modest relative to coarse-grained simulations, the conformational dynamics observed within both models indicate that fluctuations within the C-terminal residues of TM2 are suppressed by interactions



Avg. PRF Among Charge Variants

Backbone Dynamics

Figure 4.5: Structural context and conformational dynamics of nascent TM2. A) The average relative mKate:eGFP values associated with mutations that introduce charged side chains at each position within TM2 are mapped on to an atomistic model of nascent TM2 at the approximate point of frameshifting. Residues are rendered as spheres and colored based on the relative mKate:eGFP ratio for charged substitutions as was determined by DMS. The image of the nascent chain structure is superimposed on top of the structure of the translocon and lipid bilayer for clarity. B) The root mean square fluctuation (RMSF) of each Cα during a 150 ns all-atom molecular dynamics simulation are mapped on to an atomistic model of nascent TM2 at the approximate point of frameshifting. Residues are rendered as spheres and colored according to RMSF. The image of the nascent chain structure is superimposed on top of the structure of the translocon and lipid bilayer for clarity. C) RF values from atomistic (Cα, black) and coarse-grained (bead indexed to sequence, red) molecular dynamics simulations are plotted for each position within TM2.

with the translocon (Fig. 4.5C), which is the final portion of the helix to partition into the bilayer. Mutations that enhance the polarity of these residues generally decrease -1PRF (Fig. 4.5A), presumably by strengthening interactions with the translocon and/ or disfavoring the transfer of these residues into the bilayer. Taken together, these observations suggest mutations in TM2 influence pulling forces, and by extension -1PRF, by modifying interactions that form during the translocon-mediated membrane integration of TM2.

Impact of translation kinetics on pulling forces and -1PRF

Ribosomal frameshifting is also highly sensitive to mutations within the region of the transcript encoding the residues between TM2 and the peptidyl transfer center (Fig. 4.3). Most of these mutations alter the portion of the nascent chain that resides within the ribosomal exit tunnel during -1PRF. With respect to the amino acid sequence, these mutagenic trends suggest native -1PRF levels are only maintained by mutations that preserve the hydrophobicity of this segment (Fig. 4.3). This may suggest this segment is capable of forming structural contacts within the exit tunnel (see Discussion). Nevertheless, it should also be noted that the translation of this

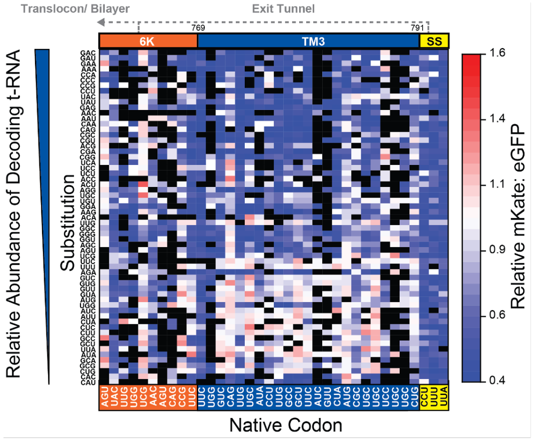


Figure 4.6: Impact of codon modifications on ribosomal frameshifting. A heatmap depicts the effects of every codon substitution (Y-coordinate) at each position (X-coordinate) within the nascent polypeptide chain on the mKate: eGFP ratio as determined by DMS. Substitutions are arranged with respect to the relative abundance of tRNA that can decode each codon in HEK293 cells. Ratios were normalized relative to that of WT, and black squares indicate mutations that lack sufficient coverage. The positions of TM3 (blue), the other 6K residues (orange), and the slip-site (yellow) are shown for reference. The residues that are likely to reside within the exit tunnel or outside of the ribosome during ribosomal frameshifting are indicated with dashed lines.

segment occurs as TM2 begins to emerge from the exit tunnel and interact with the translocon and/ or membrane. Modifications that alter the translation kinetics within this region could therefore also impact the membrane integration of TM2 and pulling forces on the nascent chain [111]. To evaluate the potential role of translational kinetics, we calculated mKate:eGFP intensity ratios associated with each individual codon substitution, then sorted the values for each codon according to the relative abundance of their decoding tRNAs in HEK293 cells. A heat map of intensity ratios reveals that -1PRF is generally attenuated by mutations within the 21 codons upstream of the slip-site that increase the relative abundance of decoding t-RNA (Fig. 4.6). Codons that are decoded by low-abundance tRNAs are generally tolerated within this region (Fig. 4.6), which suggests that efficient -1PRF hinges upon the slow translation of these codons.

The attenuation of frameshifting by mutations that increase the relative abundance of decoding tRNAs suggests the membrane integration of TM2 may be less efficient when translation is accelerated, as has been observed in other systems [111]. To assess how translational kinetics impact the membrane integration of TM2, we carried out a series of CGMD simulations of polyprotein biosynthesis in which the rate of translation was varied from 1 to 15 amino acids per second. An analysis of these trajectories reveals that the rate of translation has minimal impact on the final topology of TM2. Nevertheless, -1PRF is likely to be more closely associated with the magnitude of the pulling forces on the nascent chain rather than the equilibrium structure of TM2. To evaluate whether pulling forces are sensitive to translation kinetics, we varied the rate of translation within CGMD simulations and measured the pulling force on the nascent chain while the ribosome occupies the slip-site. One hundred independent trajectories were carried out at distinct translation rates ranging from 1 to 15 amino acids per second, and pulling forces were averaged over 0.8 s once the ribosome reaches the slip-site. A plot of the mean force against translation rate reveals that the pulling force trends upward as translation slows. These results provide secondary evidence suggesting the mechanical forces that stimulate -1PRF are likely attenuated at higher translation rates within this region. In conjunction the observed mutagenic trends (Fig. 4.6), our findings suggest that efficient frameshifting may be more efficient when translation of the region preceding the slip-site is slow.

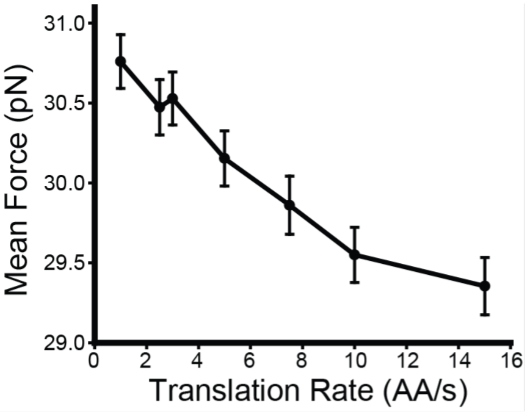


Figure 4.7: Impact of translation kinetics on nascent chain pulling forces. Coarse-grained molecular dynamics simulations of the translocon-mediated cotranslational folding of the SINV structural polyprotein were carried out varying the translation rate. Pulling forces on the nascent chain were recorded over 0.8 s while the ribosome occupies the slip-site. The mean force on the nascent chain across 100 independent trajectories is plotted against the corresponding translation rate. Error bars reflect the standard error of the mean.

Another possible explanation for the observed trends in Fig. 4.6 is that the underlying cause of the decrease in frameshifting is the hydrophobicity of the TM3 amino acids and not their translation rate. Since hydrophobicity and cognate t-RNA abundance are inversely correlated with a Spearman correlation coefficient of -0.66, we would expect to see the observed trend in cognate t-RNA abundance even if TM3 hydrophobicity was solely responsible for the high frameshifting rates. Disentangling the cause of the altered frameshifting rates requires statistical techniques beyond measuring correlations.

One approach is to look at the correlations in subgroups in which hydrophobicity is held constant. The redundancy of the genetic code makes this possible, since we can measure the correlation between tRNA abundance and PRF data within a group of synonymous codons. With this approach, the overall correlation between tRNA abundance and hydrophobicity is irrelevant because all the codons we are comparing encode the same amino acid with the same hydrophobicity. The disadvantage is that we are forced to measure correlations within small pools of data, since there are only a couple synonymous codons per amino acid. To mitigate this, we pool data either by measuring correlations within a given amino acids synonymous codons over all of TM3 (Fig. 4.8A), or by calculating the correlations per amino acid using only a single position then averaging the correlations over all amino acids (Fig. 4.8B). Both techniques suggest minimal correlation between tRNA abundance and PRF data once hydrophobicity is controlled for.

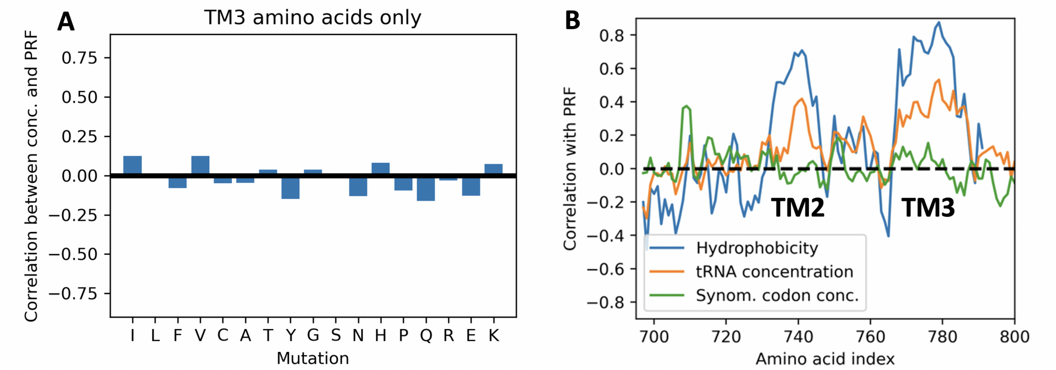


Figure 4.8: A) Correlation between tRNA abundance and DMS data within a group of synonymous codons coding for a given amino acid (identified on the x-axis). The inverse of the tRNA abundance is used, such that a positive correlation indicates more frameshifting with less abundance. B) Calculating the correlation between PRF and various predictors per position in the polyprotein. For the synonymous codons, the correlations are calculated per amino acid, then averaged. Again, the inverse of the tRNA abundance is used, such that a positive correlation indicates more frameshifting with less abundance.

An alternative method to disentangle two correlated variables is to run a multivariate regression which includes both variables as predictors. The multivariate model will take advantage of the unique information from each predictor. The resulting partial R^2 quantifies how much of the variance in the PRF data is explained as a result of adding an additional predictor. The statistics for hydrophobicity and tRNA abundance are summarized in Table 4.1. The low partial R^2 of tRNA abundance implies that the abundance does not usefully predict frameshifting after the contribution from hydrophobicity has been accounted for. Contrariwise, the partial R^2 of 0.2 for hydrophobicity indicates that using hydrophobicity to predict frameshifting is still useful even after the contributions from tRNA abundance have been taken out. Other measures of significance, such as the t-stat per predictor in the multivariate regression, agree that hydrophobicity is more influential.

Table 4.1: The R^2 column displays the fraction of explained variance for separate linear regressions predicting PRF data using hydrophobicity and abundance either jointly or separately. Partial R^2 displays the coefficient of partial determination for the two predictors. The t-stat provided is that of the coefficient of the predictor in the joint linear regression. Higher t-stats indicate higher significance.

Predictors	R^2	Partial R^2	t-stat
Joint	0.34	NA	NA
Hydrophobicity	0.33	0.20	16.6
tRNA Abundance	0.18	0.01	2.5

While this multivariate linear regression approach is able to incorporate all the DMS data simultaneously, unlike the stratified approach above, it too has limitations. Notably, it assumes a linear relationship between the predictors and frameshifting and that the errors in each prediction have identical Gaussian distributions. Although these assumptions are not strictly fulfilled, the agreement between the stratified and regression approaches implies that hydrophobicity, not tRNA abundance or translation rate, is the dominant factor in controlling PRF.

4.3 Discussion

The interactions between the transcript and translation machinery that stimulate -1PRF have been studied in considerable detail [63, 101]. However, our recent investigations of the alphavirus structural polyprotein revealed that -1PRF is also sensitive to conformational transitions in the nascent chain [34]. The net efficiency of -1PRF in this system is therefore modulated by the interplay between structural features within both the transcript and nascent chain. To map these effectors, we measured the effects of 4,530 mutations on the efficiency of -1PRF in the context of a reporter containing a large fragment of the SINV structural polyprotein. This reporter retains all the native features that direct the targeting, processing, and cotranslational folding of the nascent chain (including the signal peptide and post-translational modification sites) as well as the features in the transcript that stimulate -1PRF (slip-site and stem-loop). Retaining these features allowed us to identify and map effectors in a manner that cannot be achieved using a conventional dual-luciferase reporter system bearing an N-terminal fusion domain). Our observed trends validate numerous expectations about the stimulatory RNA elements. For instance, all 128 single, double, and triple mutants that alter the heptanucleotide slip-site reduce -1PRF to some extent (Fig. 4.2 A,B). Additionally, the mutagenic trends downstream from the slip-site are entirely consistent with previous characterizations of the stimulatory stem-loop (Fig. 4.2C) [77]. Interestingly, the latter measurements also show that -1PRF can be enhanced by mutations that strengthen or extend this stem-loop (Fig. 4.2A). The fact that the stability of this stem-loop has not evolved to maximize -1PRF provides a clear example of how SINV has evolved to maintain a specific frameshifting efficiency.

We recently showed that ribosomal frameshifting in the SINV structural polyprotein is sensitive to mutations that perturb the translocon-mediated membrane integration of TM2 [34]. Considering the marginal hydrophobicity of this segment, we initially expected mutagenic trends within this region to reveal a simple relationship between the topological energetics of TM2 and the efficiency of -1PRF. However, the observed mutagenic trends cannot be explained by hydrophobicity alone. Instead, our DMS and molecular dynamics data suggest that stimulatory pulling forces are specifically generated by the transfer of the C-terminal portion of TM2 from the translocon to membrane. At this final stage of translocation, these residues interact with the lumenal edge of the lateral gate, which pulls the loop between TMs 2 and 3 into the protein conducting channel of the translocon (Fig. 4.5). This loop is wedged between TM2 and the translocon at the point of frameshifting, and the observed mutagenic patterns within this region do not simply track with hydrophobicity or codon usage (Fig. 4.3,4.5). This observation suggests this segment may form specific interactions that influence membrane integration. Nevertheless, it is unclear whether mutations within this loop alter -1PRF by reducing the force on the nascent chain or by delaying the transmission of the force to the peptidyl transfer center until after the ribosome passes the slip-site. Additional investigations are needed to explore the network of interactions between the nascent chain, translocon, and lipid bilayer that mediate this type of mechanochemical feedback [50]).

The ribosomal exit tunnel generally suppresses protein structure formation until the nascent chain clears the ribosome [120, 121]. We were therefore surprised to find that -1PRF is highly sensitive to mutations within the region encoding the portion of the nascent chain that occupies the exit tunnel during frameshifting (Fig. 4.3). Nevertheless, a recent investigation of the sequence constraints in other PRF motifs found that frameshifting is often enhanced by rare codons upstream of the slip-site [109]. An analysis of the effects of codon-level substitutions suggests changes in -1PRF may be related to the effects of mutations on the relative abundance of decoding tRNAs (Fig. 4.6), and CGMD simulations suggest stimulatory forces are sensitive to the rate of translation (Fig. 4.7). These observations strongly suggest translation kinetics are critical for the coupling between cotranslational folding and -1PRF. However, it remains unclear how much these substitutions actually impact the physiological rate of translation. Furthermore, the estimated impact of translation kinetics on pulling forces appears to be relatively modest in relation to the magnitude of their effects of -1PRF. Given these caveats, it may be that these mutations impact -1PRF in more than one way. For instance, the native codon bias in this region may also control the frequency of ribosomal collisions, which were recently found to stimulate -1PRF [122]. Modifications to the structure of the nascent chain could also potentially be relevant to -1PRF. This region encodes TM3, and it was recently found that hydrophobic helices exhibit a tendency to form helical structure within the exit tunnel [123]. Furthermore, native -1PRF levels are only maintained when mutations introduce other hydrophobic side chains within this segment (Fig. 4.3). Thus, the formation of helical structure within the exit tunnel could potentially also contribute to this mechanochemical coupling. Taken together, it seems the mechanistic basis of the mutagenic effects within this region are undoubtedly complex, and additional investigations are needed to tease apart each of these potential variables.

Together, our results reveal that the net efficiency of -1PRF in the alphavirus structural polyprotein arises from the coupling between multiple structural effectors. While this motif features the canonical slip-site and stem loop architecture found in most -1PRF sites, its efficiency is rendered tunable by mutations that alter mechanochemical transitions within the nascent chain and/ or the translation kinetics upstream of the slip-site. Our DMS data demonstrate that this degeneracy effectively expands the pool of genetic variations that are capable of adjusting the efficiency of -1PRF; an important consideration for viral evolution. Though this investigation focuses on a mechanochemical transition within the ribosome-translocon complex, it seems likely that many other classes of conformational transitions within the nascent chain may also be capable of stimulating -1PRF. Further investigations are needed to evaluate whether similar mechanisms are operative within other transcripts.

4.4 Methodology

Computational Predictions of Topological Energetics

The effects of mutations on the free energy associated with the transfer of TM2 from the translocon to the ER membrane were estimated using the ΔG predictor [88] using a series of individual predictions for TM2 variants assuming a WT sequence of LTPYALAPNAVIPTSLALLCCVR.

Coarse-grained simulations of polyprotein translation

Coarse-grained simulations use a previously developed and validated methodology [14, 15, 34]. Simulations are based on coarsened representations of the ribosome exit tunnel, Sec translocon, and nascent chain. The nascent chain is represented as a polymer of beads, each of which represents three amino acids. Each bead has a hydrophobicity and charge derived from its three constituent amino acids. The solvent and lipid bilayer are modeled implicitly. The lateral gate of the Sec translocon stochastically switches between open and closed conformations with probability dependent on the free energy difference between the two conformations. The structure of the ribosome and Sec translocon are based on cryo-EM structures [124], and aside from the opening/closing of the lateral gate, are fixed in place during simulation.

Model parameterization is unchanged from previously published work [14, 34]. Beads are added to the nascent chain at a rate of 5 amino acids per second unless otherwise specified. Integration is carried out using overdamped Langevin dynamics with a timestep of 300 ns and a diffusion coefficient of 253 nm^2/s. Translation starts 33 amino acids before the N-terminus of TM1 of the E2 protein (P663).

To measure pulling forces, translation is halted when the ribosome occupies the slip site. The dynamics of the nascent chain are then allowed to evolve for three seconds, during which time the forces exerted on the N-terminal bead are measured every 3 ms. Because the exit tunnel is truncated in the CGMD model, we add 27 amino acids to the index of the final bead to account for the unmodeled nascent chain in the omitted part of the exit tunnel. All single mutants from positions 725 to 754 are simulated. Each mutant was simulated in 100 independent runs in order to ensure the estimate for the mean pulling force has low statistical error. Data from trajectories in which TM1 does not adopt the correct N_{out} topology were not included in the analysis.

The ratios of topomeric isomers were obtained independently from the force-measuring simulations. To observe the final topology of TM2 in the membrane, the polyprotein was translated from amino acid P663 to I837 without pausing. Restraints were placed on TM1 to ensure it adopts the correct topology. No restraints were placed on TM2. Upon translation of the final bead, the nascent chain was released from the ribosome and simulation continued. After 3 seconds, the topology of the protein was recorded. Simulations were repeated 100 times to obtain the probability of TM2 integrating per mutant.

Atomistic simulations of polyprotein TM2 in the translocon

An atomistic model was built of polyprotein TM2 in the Sec translocon using the cryo-EM structure with PDB code of 6ITC as an initial template [119]. This structure captures a polypeptide entering a lipid bilayer through the lateral gate of SecY. The translocating peptide sequence in the structure was replaced with the polyprotein sequence using MODELLER [125], aligning the TM2 helix in polyprotein with the helical pro-OmpA signal sequence that is exposed to the lipid bilayer in the cryo-EM structure. Specifically, residue M1 of the peptide was replaced with residue E723 of the polyprotein. Two alternative structures were generated by shifting the mapping of TM2 amino acids two amino acids forwards. Specifically, residue M1 of the peptide was replaced with either residues L725 or P727. Unresolved amino acids in the translocating peptide were modeled in, also using MODELLER. The antibodies and GFP present in the cryo-EM structure were removed. The resulting structure was then solvated with TIP3P water and POPC lipids, filling a simulation box that extended 10 Å beyond the protein complex. The structure was energetically locally minimized, and then the alpha carbons in SecA more than 15 Å from the translocating peptide were fixed in place with 1 kcal/mol/$Å^2$ harmonic restraints. Molecular dynamics simulations were then run for 150 ns with a 1.5 fs timestep at 300K and 1 atm using Desmond [55]. Root mean square fluctuations of the amino acid alpha carbons were calculated using the final 100 ns of the trajectory.

Experimental methodology is available in the original publication on which this chapter is based[126].

Chapter 5

RESIDUE-BY-RESIDUE ANALYSIS OF COTRANSLATIONAL MEMBRANE PROTEIN INTEGRATION IN VIVO

Adapted from:

Felix Nicolaus, Ane Metola, Daphne Mermans, Amanda Liljenström, Ajda Krč, Salmo Mohammed Abdullahi, Matthew H. Zimmer, Thomas F. Miller III, and Gunnar von Heijne. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *eLife*, 10:1–16, 2021. ISSN 2050084X. doi: 10.7554/eLife.64302.

We follow the cotranslational biosynthesis of three multispanning *Escherichia coli* inner membrane proteins in vivo using high-resolution force profile analysis. The force profiles show that the nascent chain is subjected to rapidly varying pulling forces during translation and reveal unexpected complexities in the membrane integration process. We find that an N-terminal cytoplasmic domain can fold in the ribosome exit tunnel before membrane integration starts, that charged residues and membrane-interacting segments such as re-entrant loops and surface helices flanking a transmembrane helix (TMH) can advance or delay membrane integration, and that point mutations in an upstream TMH can affect the pulling forces generated by downstream TMHs in a highly position-dependent manner, suggestive of residue-specific interactions between TMHs during the integration process. Our results support the sliding model of translocon-mediated membrane protein integration, in which hydrophobic segments are continually exposed to the lipid bilayer during their passage through the SecYEG translocon.

5.1 Introduction

Most integral membrane proteins are cotranslationally integrated into their target membrane with the help of translocons such as bacterial SecYEG and YidC, and the eukaryotic Sec61 and EMC complexes [127, 128]. While the energetics of translocon-mediated integration of a transmembrane α-helix (TMH) is reasonably well understood [88], the actual integration process is not, other than in general terms. We have shown that force profile analysis (FPA), a method in which a translational arrest peptide (AP) engineered into a target protein serves as a sensor to measure the force exerted on a nascent polypeptide chain during translation, can be used to follow the cotranslational folding of soluble proteins and the membrane integration of a model TMH [4, 8, 129]. Here, we have applied FPA and coarse-grained molecular dynamics (CGMD) simulations to follow the cotranslational membrane integration of three multispanning *Escherichia coli* inner membrane proteins of increasing complexity (EmrE, GlpG, BtuC), providing the first residue-by-residue data on membrane protein integration in vivo.

5.2 Results

Force profile analysis

FPA takes advantage of the ability of APs to bind in the upper parts of the ribosome exit tunnel and thereby pause translation when their last codon is in the ribosomal A-site [130]. The duration of an AP-induced pause is reduced in proportion to pulling forces exerted on the nascent chain [2, 131], that is, APs can act as force sensors and can be tuned by mutation to react to different force levels [33]. In an FPA experiment, a series of constructs is made in which a force-generating sequence element (e.g., a TMH) is placed an increasing number of residues away from an AP (reflected in N, the number of residues from the start of the protein to the end of the AP), which in turn is followed by a C-terminal tail (Fig. 5.1A). In constructs where the TMH engages in an interaction that generates a strong enough pulling force (F) on the nascent chain at the point when the ribosome reaches the last codon of the AP, pausing will be prevented and mostly full-length protein will be produced during a short pulse with [^{35}S]-Met (Fig. 1B, middle). In contrast, in constructs where little force is exerted on the AP, pausing will be efficient and more of the arrested form of the protein will be produced (Fig. 5.1B, left and right). The fraction full-length protein produced, $f_{FL} = I_{FL}/(I_{FL} + I_A)$, where $I_F L$ and I_A are the intensities of the bands representing the full-length (FL) and arrested (A) species on an SDS-PAGE gel (Fig.5.1C), can therefore be used as a proxy for F in a given construct [15, 49, 131]. A plot of $f_{F}L$ versus N, a force profile (FP), thus can provide a detailed picture of the cotranslational process in question, as reflected in the variation in the force exerted on the nascent chain during translation. FPs can be recorded with up to single-residue resolution by increasing N in steps of one residue (corresponding to a lengthening of the nascent chain by ~3 Å).

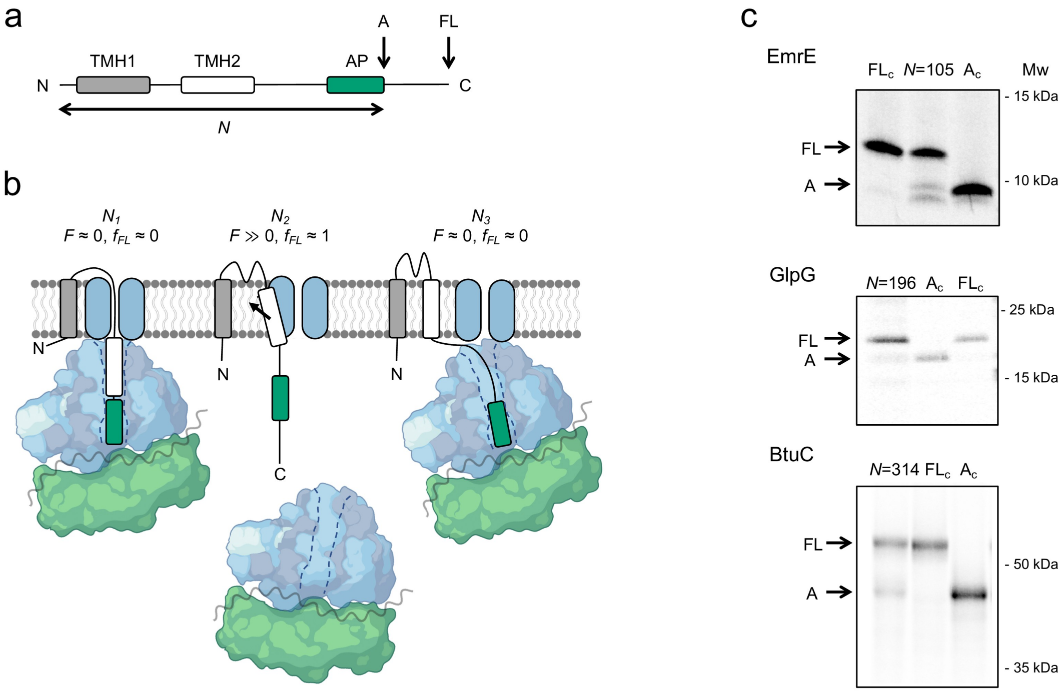


Figure 5.1: A) Basic construct. Arrested (A) and full-length (FL) products are indicated. B) At construct length N_1, TMH2 has not yet entered the SecYEG channel and no pulling force F is generated. At N_2, TMH2 is integrating into the membrane and $F \gg 0$. At N_3, TMH2 is already integrated and $F \approx 0$. C) SDS-PAGE gels showing A and FL products for [^{35}S]-Met labeled and immunoprecipitated EmrE(C$_{\text{out}}$) (N = 105), GlpG (N = 196), and BtuC (N = 314). Control constructs A_C and FL_C have, respectively, a stop codon and an inactivating Ala codon replacing the last Pro codon in the arrest peptide (AP). The band just below the A band in the EmrE(C$_{\text{out}}$) (N = 105) lane most likely represents ribosomes stacked behind the AP-stalled ribosomes [132] and is not included in the calculation of f_{FL}.

EmrE: 4 TMHs, 110 residues

We chose EmrE as an example of a small, relatively simple 4-TMH protein. EmrE is a dual-topology protein, that is, the monomers integrate into the inner membrane in a 50–50 mixture of N$_{\text{in}}$-C$_{\text{in}}$ and N$_{\text{out}}$-C$_{\text{out}}$ topologies; two oppositely oriented monomers then assemble into an antiparallel dimer [133, 134]. To avoid potential complications caused by the dual topology, we used EmrE(C$_{\text{out}}$), a mutant version that adopts the N_{out}-C_{out} topology [134], and further used the relatively weak SecM(*Ec*) AP [4] and included an human influenza hemagglutinin (HA) tag for immunoprecipitation (Fig. 5.2A). A series of EmrE(C_{out})-AP constructs was used to obtain the FP shown in Fig. 5.2B (orange curve), at 2–5 residues resolution. Also shown is an FP derived from a CGMD simulation (CGMD-FP, gray).

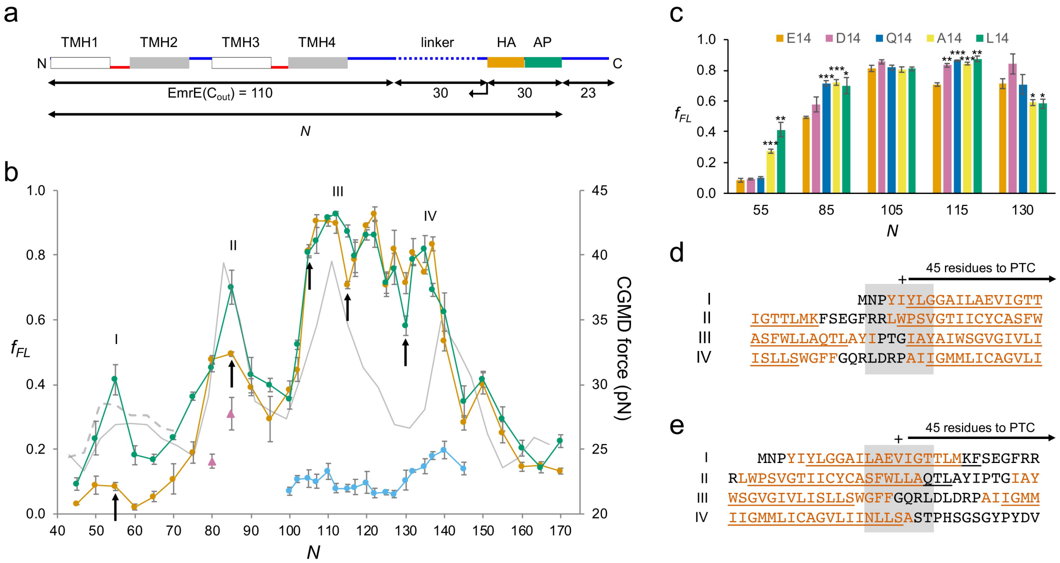


Figure 5.2: A) Construct design. EmrE(C_{out}) is shortened from the C-terminal end of the LepB-derived linker (dotted), as indicated by the arrow. Cytoplasmic (red) and periplasmic (blue) loops, and lengths of full-length EmrE(C_{out}), LepB-derived linker, HA tag + arrest peptide (AP), and C-terminal tail, are indicated. Since the 30-residue HA + AP segment is constant in all constructs, the force profile (FP) reflects nascent chain interactions occurring mainly outside the ribosome exit tunnel. B) FPs for EmrE(C_{out}) (orange), EmrE(C_{out},E14L) (green), EmrE(C_{out}) with SecM(*Ec*-sup1) AP (blue), EmrE(C_{out}, I37,I38→NN) (magenta triangles), and coarse-grained molecular dynamics (CGMD-FP) calculated with a -100 mV membrane potential (gray). C) Effects of mutations in E14 on f_FL values for the N values are indicated by arrows in B. p-values (two-sided t-test): *p < 0.05; **p < 0.01; ***p < 0.001. D,E) Sequences corresponding to peaks I-IV aligned from their N_{start} D) and N_{end} E) values. The + sign indicates 45 residues from the peptidyl transferase center (PTC). Hydrophobic transmembrane helix (TMH) segments are shown in orange and transmembrane α-helices underlined (PDB: 3B5D). Error bars in B and C indicate SEM values.

We have previously shown that a model TMH composed of Ala and Leu residues generates a peak in an FP recorded with the SecM(*Ec*) AP that reaches half-maximal amplitude (N_{start}) when the N-terminal end of the TMH is ~45 residues away from the peptidyl transferase center (PTC) [4], and a recent real-time FRET study of cotranslational membrane integration found that the N-terminal end of the first

TMH in a protein reaches the vicinity of the SecYEG translocon when it is 40–50 residues away from the PTC [135]. For EmrE(C_{out}) TMH1, this would correspond to constructs with $N \approx 50$. However, the f_{FL} values are hardly above background in this region of the FP. Due to the functionally important E14 residue, TMH1 is only marginally hydrophobic and does not become firmly embedded in the membrane until the protein dimerizes [136]. To ascertain whether the lack of a peak in the FP corresponding to the membrane integration of TMH1 is because of its low hydrophobicity, we mutated E14 to L. Indeed, in the FP obtained for EmrE(C_{out},E14L) (Fig. 5.2B, green curve), a clear peak appears at the expected chain length $N_{start} \approx 50$ residues. Mutation E14A yields an f_{FL} value intermediate between EmrE(C_{out},E14L) and EmrE(C_{out}) at $N = 55$ (Fig. 5.2C), while f_{FL} for the mutants EmrE(C_{out},C_{out}) and EmrE(C_{out},E14Q) is the same as for EmrE(C_{out}).

Peak II has $N_{start} \approx 76$, corresponding to a situation where the N-terminal end of TMH2 is ~45 residues from the PTC (Fig. 5.2D). The double mutation I37I38→NN in TMH2 reduces f_{FL} at $N = 80$ and 85 (magenta triangles), as expected. Unexpectedly, however, the E14L, E14A, and E14Q (but not the E14D) mutations in TMH1 increase f_{FL} at N = 85 (Fig. 5.2C), showing that a negatively charged residue (D or E) in position 14 in TMH1 specifically reduces the pulling force generated by TMH2 at $N = 85$, that is, when about one-half of TMH2 has integrated into the membrane. Likewise, f_{FL} values at $N = 115$ and 130 (but not at $N = 105$, included as a negative control) are specifically affected by mutations in E14: at $N = 115$ (one-half of TMH3 integrated), all four mutations in position 14 increase f_{FL} relative to E14, while at $N = 130$ (beginning of TMH4 integration) the E14A and E14L mutations decrease f_{FL} (Fig. 5.2C). FPA thus reveals long-range effects of mutations in E14 on three specific steps in the membrane integration of the downstream TMHs. This implies that TMH1 remains in the vicinity of the translocon and that E14 makes specific interactions with residues in the TMH2–TMH4 region during the membrane integration process. Further studies will be required to pinpoint these interactions and understand the role played by the slow dynamics of TMH1 integration [136].

Peak III has $N_{start} \approx 102$ residues, with the N-terminal end of TMH3 ~45 residues from the PTC (Fig. 5.2D). Peak IV is difficult to locate precisely in the FP because f_{FL} values are high throughout the TMH3–TMH4 region, but is seen at $N_{start} \approx 132$ residues when the strong SecM(*Ec*-sup1) AP [22] is used (blue curve), again with the N-terminal end of TMH4 ~45 residues from the PTC (Fig. 5.2D). As shown in Fig. 5.2E, the TMHs cease generating a pulling force when their C-terminal ends are ~45 residues away from the PTC, indicating that they are fully integrated at this point.

GlpG: 6 TMHs, 276 residues

We next studied GlpG, a medium-sized monomeric 6-TMH rhomboid protease with an ~60 residue cytoplasmic N-terminal domain (NTD) (Fig. 5.3A)[137, 138], a protein that allows us to follow the cotranslational folding of a soluble domain and integration of a membrane domain in the same experiment.

The FP is shown in Fig. 5.3B (orange curve). It was obtained at 5-residue resolution, except for the portion N = 168–224, which we measured with single-residue resolution. Peak I, at $N_{\text{start}} \approx 84$ residues, is conspicuously close to what would be expected for the folding of the NTD from previous studies of cotranslational folding of small globular domains in the ribosome exit tunnel [129]. To verify that the peak indeed represents folding of the NTD, we recorded an FP for the NTD by in vitro transcription-translation in the PURE system[139] and further made a destabilizing point mutation (F16E) in the core of the NTD (Fig. 5.3C). The FP obtained in vitro (magenta) overlaps peak I in the in vivo FP, and the mutation strongly reduces f_{FL} values for peak I both in vivo (green) and in vitro (black). Given that the NTD has a relative contact order of 15% and is predicted to fold on the ms time scale [140] while the elongation cycle on the ribosome takes ~100 ms/codon [141], the NTD has ample time to equilibrate between the unfolded and accessible folded states at each elongation step [142]. We conclude that the ~60 residue NTD folds inside the ribosome exit tunnel when its C-terminal end is 25–30 residues from the PTC, well before synthesis of the membrane domain has commenced.

Peaks II–VII in the FP correspond reasonably well to the CGMD-FP (gray) profiles. The unexpectedly low N_{start} value for peak III seems to be caused by an upstream periplasmic surface helix (Fig. 5.3F) (see below). Likewise, peak VI-a likely reflects the membrane integration of a hydrophobic, membrane-associated cytoplasmic segment located just upstream of TMH5. In contrast, the unexpectedly high N_{start} value for peak IV indicates that integration of TMH3 commences only when its N-terminal end is ~52 residues away from the PTC, possibly because of the tight spacing between TMH2 and TMH3.

As peak III saturates at $f_{FL} \approx 0.9$ over a rather wide range, we sought a more detailed view by using the strong SecM(Ec-Sup1) AP (Yap and Bernstein, 2009) (Fig. 5.3B,D, blue) and the medium-strong SecM(*Ec-Ms*) AP (Fig. 5.3D, green)[21].

The SecM(*Ec*-Sup1) FP allows a precise determination of $N_{\text{max}} = 200$, at which point the middle of TMH2 (L155) is located 45 residues from the PTC (Fig. 5.3F). The SecM(*Ec-Ms*) FP reveals additional detail: peak III is now seen to be composed of three subpeaks III-a, III-b, and III-c. III-a has $N_{\text{start}} = 182$, coinciding with the N-terminal end of the periplasmic surface helix reaching 45 resides away from the PTC. For III-b, $N_{\text{start}} \approx 190$, with the N-terminal end of TMH2 ~45 residues from the PTC. The major subpeak III-c at $N \approx 197$–204 finally corresponds well to the peak seen in the SecM(*Ec*-Sup1) and the CGMD FPs, and therefore represents the membrane insertion of the most hydrophobic part of TMH2. Taken together, subpeaks III-b and III-c are reminiscent of the biphasic pulling force pattern previously recorded for a model hydrophobic transmembrane segment using the medium-strong SecM(*Ms*) AP [4], which is closely related to the SecM(*Ec-Ms*) AP used here.

We further recorded a SecM(*Ec-Ms*) FP (magenta) for the triple mutation Y138, F139,L143 → NNN (Fig. 5.3E, sticks) that renders the periplasmic surface helix less hydrophobic: the mutation strongly reduces the amplitude of peak III-a, has only a small effect on peak III-b, and both reduces the amplitude and shifts N_{start} and N_{max} for peak III-c by approximately four residues (Fig. 5.3D,F). Thus, the periplasmic surface helix engages in hydrophobic interactions already during its passage through the translocon, presumably by sliding along a partly open lateral gate [33]. It also adds to the force generated by the membrane integration of TMH2, possibly by partitioning into the periplasmic leaflet of the inner membrane at approximately the same time that TMH2 enters the translocon.

BtuC: 10 TMHs, 326 residues

Finally, we studied BtuC, a vitamin B12 transporter with 10 TMHs, as an example of a large, multispanning protein with a complex fold [143]. In order to improve expression, we added the N-terminal part of LepB to the BtuC constructs (Fig. 5.4A) and used a LepB antiserum for immunoprecipitation. The N_{out}-C_{in} orientation of LepB TMH1 ensures that the N_{in}-C_{in} topology of BtuC will be maintained.

We identified 11 peaks in the FP (Fig. 5.4B, orange), one more than could be accounted for by the 10 TMHs. Since it was not possible to provide an unequivocal match between the BtuC FP and the CGMD-FP, we did two sets of controls. First, we chose constructs at or near peaks in the FP and CGMD-FP and mutated multiple hydrophobic residues (Leu, Ile, Val, Met) located 40–50 residues from the PTC to less hydrophobic Ala residues. The mutations caused significant drops in fFL ($p < 0.01$, two-sided t-test), except for construct N = 191 that is mutated at the extreme N-terminus of TMH5. The mutation data allowed us to identify the membrane integration of TMHs 1, 2, 3, 4, 5, 7, 8, 9, and 10 with peaks I, II, III, IV, V, VIII, IX, X, and XI, respectively; the overlapping peaks VIII and IX appear to represent the concerted integration of the closely spaced TMH7 and TMH8. However, peak II (corresponding to TMH2) is shifted to unexpectedly high, and peaks V (corresponding to TMH5), X (corresponding to TMH9), and XI (corresponding to TMH10) to unexpectedly low, N_{start} values (Fig. 5.4C). To confirm these assignments, we obtained FPs for the isolated TMH2 (dashed green), TMH8 (dashed light blue), and TMH10 (dashed pink) sequences (Fig. 5.4B) by introducing them into the periplasmic domain of LepB such that they maintained their natural N_{out}-C_{in} orientation (Fig. 5.4D); the FPs for the individual TMHs overlap the corresponding peaks II, IX, and XI in the full FP. Likewise, an FP obtained for a construct lacking TMH1-TMH4 overlaps the full FP, except that peak V is shifted to a higher N_{start} value, more in line with the peak seen in the CGMD-FP. The low N_{start} value for the N_{in}-C_{out}-oriented TMH5 in full-length BtuC may result from an early interaction between a positively charged patch (RFARRHLSTSR) just upstream of TMH5 and negatively charged lipid headgroups, while the low N_{start} values for peaks X and XI are likely caused by the short upstream hydrophobic segments LCGL and LAAALEL (Fig. 5.4C,F), similar to peak III in GlpG.

Remarkably, the N-terminal end of the isolated TMH2 is ~45 residues away from the PTC at N_{start}, suggesting that upstream sequence elements present in full-length BtuC delay the integration of TMH2 by ~10 residues (compare II* and II in Fig. 5.4C). The most conspicuous feature in the upstream region of TMH2 is the presence of three positively charged Arg residues, an uncommon occurrence in a periplasmic loop (Heijne, 1986). Indeed, when these residues are replaced by uncharged Gln residues in LepB-BtuC, peak II (dashed black in Fig. 5.4B,E) becomes almost identical to the FP for the isolated TMH2; a similar behavior is seen when the CGMD-FP simulation is run without an electrical membrane potential (Fig. 5.4E). Upstream positively charged residues thus delay the membrane integration of the N_{out}-oriented TMH2, possibly because of the energetic cost of translocating them against the membrane potential [8], or because they are temporarily retained in the negatively charged exit tunnel[135].

Neither peak VI nor VII seems to represent the integration of TMH6, but instead flanks the location expected from the CGMD-FP and HP and apparently corresponds, respectively, to the membrane insertion of a short periplasmic re-entrant helix and of a short cytoplasmic surface helix (Fig. 5.4C,H). The FP for the isolated TMH6 (Fig. 5.4B,G, dashed dark blue) peaks in the location expected from the CGMD-FP, between peaks VI and VII, and the FP for the isolated TMH5-6 part that includes the re-entrant helix but lacks the downstream surface helix is intermediate between the LepB-BtuC and the TMH6 FPs (Fig. 5.4G, dashed green). Thus, the membrane interactions of the periplasmic re-entrant helix and the cytoplasmic surface helix exert a strong effect on the membrane integration of the intervening TMH6.

5.3 Discussion

A detailed view of the cotranslational integration of three multispanning membrane proteins provided here shows that translocating nascent chains experience a distinct transition to a more hydrophobic environment at a distance of ~45 residues from the PTC, generating an oscillating force on the nascent chain that is ultimately transmitted to the PTC and varies in step with the appearance of each TMH in the vicinity of the SecYEG translocon channel. It seems likely that such oscillations can have multiple effects on the translation of membrane proteins, as recently demonstrated for ribosomal frameshifting [34], and may affect protein quality control [144].

Notably, TMHs also stop generating a force on the nascent chain when their C-terminal end reaches ~45 residues from the PTC, irrespective of whether their orientation is N_{out}-C_{in} or N_{in}-C_{out}. This is in agreement with the 'sliding' model of TMH integration [145], which posits that N_{out}-C_{in} TMHs have continuous lipid contact as they slide across the membrane along the open lateral gate in the SecYEG translocon, while N_{in}-C_{out} TMHs first partition into the cytoplasmic interface region of the membrane as they exit the ribosome (and therefore generate less pulling force than N_{out}-C_{in} TMHs [93] and only insert across the membrane as their polar C-terminal flanking region translocates through the central translocon channel. In both cases, the TMHs are embedded in the membrane (albeit in perpendicular orientations) when their C-terminal end is ~45 residues from the PTC. In the sliding model, the translocon channel serves as a conduit for polar nascent chain segments while hydrophobic segments are always in contact with surrounding lipid, similar to what has been proposed for the YidC/Oxa1 translocon family [146]. The lateral gate region in the SecYEG translocon thus in a certain sense mimics the water–bilayer interface environment[147].

We also find that the cytoplasmic NTD in GlpG folds already in the ribosome exit tunnel, before the first TMH has been synthesized. Further, the FPs for EmrE, GlpG, and BtuC to a first approximation match those predicted by CGMD calculations, but uncover a much richer picture of the membrane integration process where charged residues and membrane-interacting segments such as re-entrant loops and surface helices flanking a TMH show prominent interactions with the translocon and surrounding lipid. Finally, point mutations in EmrE TMH1 affect the pulling force generated by downstream TMHs in a highly position-dependent manner, suggestive of residue-specific interactions between TMHs during the membrane integration process. Complementing in vitro unfolding/folding studies [148, 149], real-time FRET analyses [135], chemical crosslinking [146], structure determination [150], and computational modeling [151], high-resolution in vivo FPA can thus help identify the molecular interactions underlying cotranslational membrane protein biogenesis with up to single-residue precision.

5.4 Methodology

Computer simulations of cotranslational membrane integration were carried out using a previously developed and validated CGMD model in which nascent proteins are mapped onto CG beads representing three amino acids [14, 15]. The nascent protein interacts with the Sec translocon and the ribosome via pairwise interactions that depend on the hydrophobicity and charge of the beads of the nascent protein. The interaction parameters are unchanged from previous work [14]. The lateral gate of the translocon switches between the open and closed conformations with probability dependent on the difference in free energy between the two conformations. The structures of the ribosome and translocon are based on cryo-EM structures and, aside from the lateral gate of the translocon, are fixed in place during the simulations. The lipid bilayer and cytosol are modeled implicitly. The positions of the nascent protein beads are evolved using overdamped Langevin dynamics with a timestep of 300 ns and a diffusion coefficient of 253 nm^2/s. Membrane potentials are included by adding an electrostatic energy term to the simulations, as previously described [15].

To simulate protein translation, new amino acids are added to the nascent chain at a rate of five amino acids per second. Simulations of EmrE, GlpG, and BtuC begin with 12 amino acids translated. Translation continues until the nascent protein reaches the desired length, at which point translation is halted and forces on the C-terminus of the nascent chain are measured every 3 ms for 6 s. This methodology has been found to accurately reproduce experimental FPs[15]. Forces are measured starting at a nascent protein length of 18 amino acids for EmrE and BtuC, and 70 for GlpG. The computational force profile (CGMD-FP) is then obtained by measuring the forces at lengths incremented by four amino acids. Simulations at different lengths are performed independently and repeated 100 times. Because the ribosomal exit tunnel is truncated in the CGMD model, a shift in the protein index is required to compare simulated and experimental results. Shifts of -12, -5, and -5 residues are used for EmrE, GlpG, and BtuC CGMD-FPs, respectively. The shifts are estimated by aligning the computational and experimental FPs and are in line with what is expected given the length of the truncated exit tunnel. Variation in the shift may reflect different degrees of compaction of the nascent chain. Although previous work provides a framework to estimate the experimentally observed fraction full length from simulated forces given a specific AP [15, 51], forces are reported directly to facilitate comparison between experiments performed with different APs.

Experimental methodology is available in the publication on which this chapter is based[9].

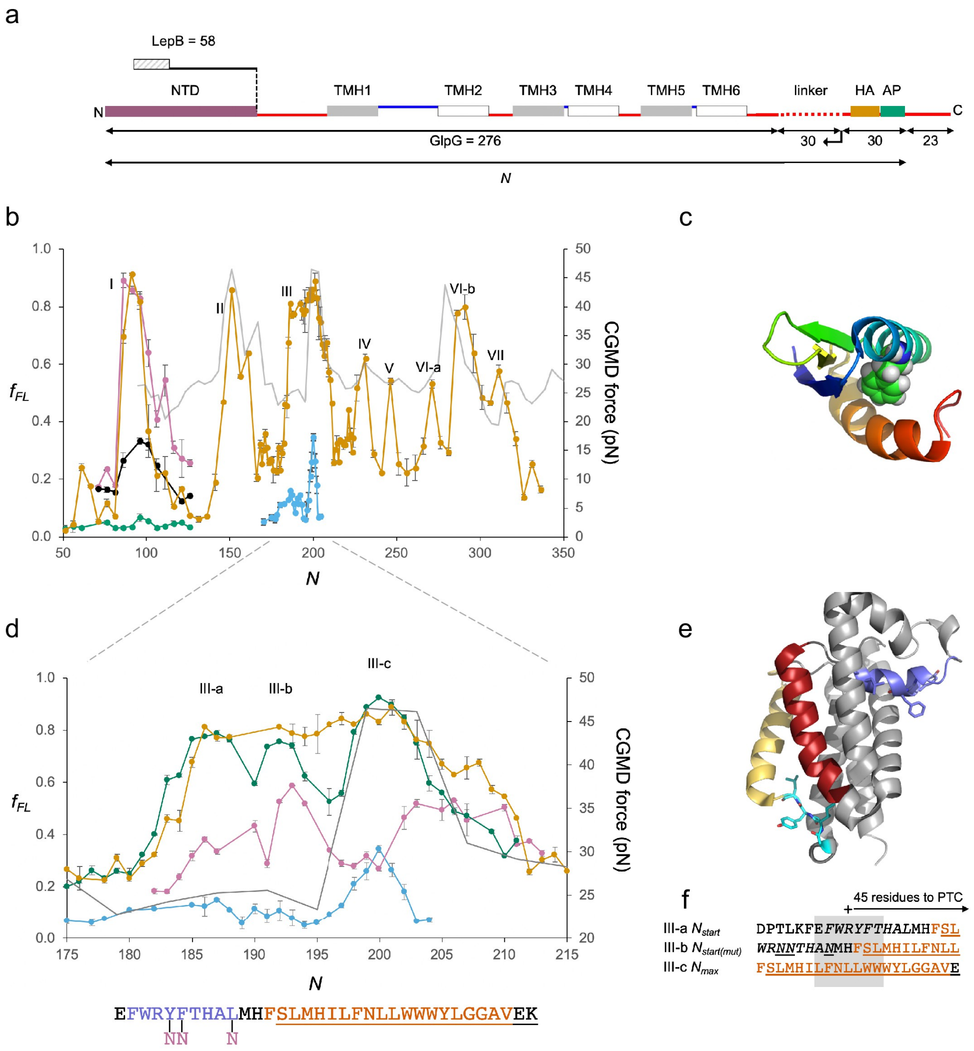


Figure 5.3: A) Construct design. The N-terminal LepB fusion is indicated. B) Force profiles (FPs) for GlpG and LepB-GlpG (N = 131–224) (orange), NTD(F16E) (green), in vitro translated NTD (magenta), and NTD(F16E) (black), LepB-GlpG with SecM(*Ec*-Sup1) AP (blue), and coarse-grained molecular dynamics (CGMD)-FP calculated with a -100 mV membrane potential (gray). Error bars indicate SEM values. C) NTD (PDB ID: 2LEP), with F16 in spacefill. D) Enlarged FPs for LepB-GlpG with SecM(*Ec*) AP (orange), SecM(*Ec*-Ms) AP (green), SecM(*Ec*-sup1) AP (blue), and GlpG(Y138,F139,L143→NNN) with SecM(*Ec*-Ms) AP (magenta). CGMD-FP in gray. E) Structure of GlpG with the periplasmic surface helix in blue, TMH2 in red, the membrane-associated cytoplasmic segment in cyan, and TMH5 in yellow. F) LepB-GlpG peak III-a and III-c sequences aligned, respectively, from their N_{start} and N_{max} values, and the mutant LepB-GlpG(Y138,F39,L143→NNN) peak III-c sequence aligned from its N_{max} value. Hydrophobic transmembrane helix (TMH) segments are shown in orange and transmembrane α-helices (PDB: 2IC8) underlined. The periplasmic surface helix is italicized.

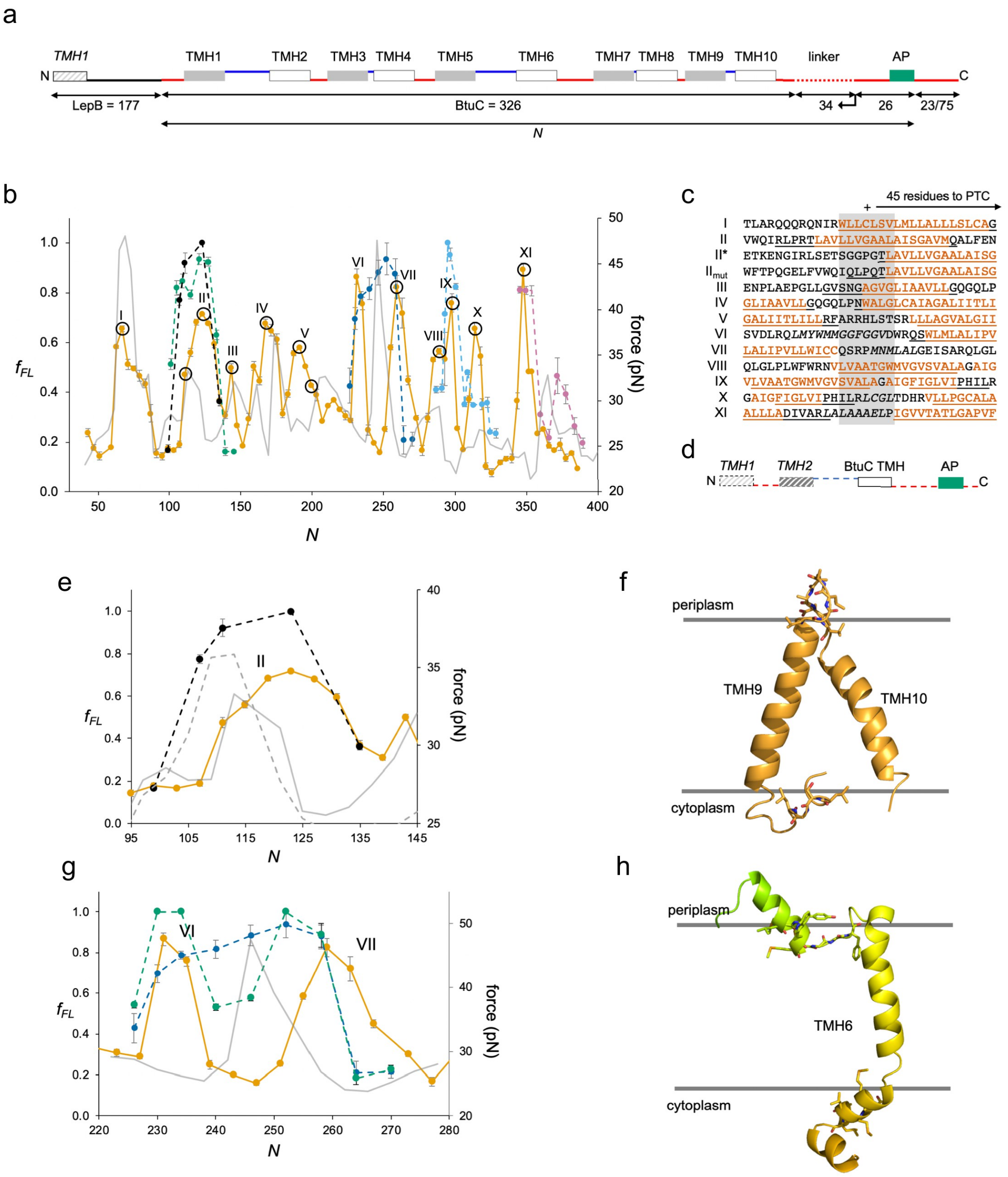


Figure 5.4: A) Construct design. B) FPs for BtuC (orange), BtuC-TMH2 (green), BtuC(R47,R56,R59 → QQQ) (black), BtuC-TMH6 (dark blue), BtuC-TMH8 (blue), BtuC-TMH10 (pink), and CGMD-FP calculated with a -100 mV membrane potential (gray). Error bars indicate SEM values. C) Sequences corresponding to peaks I–XI aligned from their N_{start} values. Hydrophobic TMH segments are shown in orange and membrane-embedded α-helices are underlined. Re-entrant loops and surface helices are italicized. D) Construct design for obtaining FPs of isolated N_{out}-oriented BtuC TMHs. Dashed segments are derived from LepB. E) Enlarged FPs for BtuC (orange) and (R47,R56,R59 → QQQ) (black), with CGMD-FPs calculated with (gray) and without (dashed gray) a -100 mV potential. F) BtuC TMH9-TMH10 (PDB ID: 2QI9). G) Enlarged FPs for BtuC (orange), isolated TMH6 (residues 187–206; blue), and isolated TMH5-6 (residues 138–206; green). The CGMD-FP is in gray. H) Structure of TMH6 including the upstream periplasmic re-entrant helix and the downstream cytoplasmic surface helix

BIBLIOGRAPHY

[1] Hitoshi Nakatogawa and Koreaki Ito. The ribosomal exit tunnel functions as a discriminating gate. *Cell*, 108(5):629–636, 2002.

[2] Daniel H. Goldman, Christian M. Kaiser, Anthony Milin, Maurizio Righini, Ignacio Tinoco, and Carlos Bustamante. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*, 348(6233):457–460, 4 2015.

[3] Hitoshi Nakatogawa and Koreaki Ito. Secretion monitor, secM, undergoes self-translation arrest in the cytosol. *Molecular Cell*, 7(1):185–192, 2001.

[4] Nurzian Ismail, Rickard Hedman, Nina Schiller, and Gunnar Von Heijne. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nature Structural and Molecular Biology*, 19(10):1018–1022, 10 2012.

[5] Florian Cymer and Gunnar von Heijne. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proceedings of the National Academy of Sciences*, 110(36):14640–14645, 2013.

[6] Ola B. Nilsson, Rickard Hedman, Jacopo Marino, Stephan Wickles, Lukas Bischoff, Magnus Johansson, Annika Müller-Lucks, Fabio Trovato, Joseph D. Puglisi, Edward P. O'Brien, Roland Beckmann, and Gunnar von Heijne. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports*, 12(10):1533–1540, 2015.

[7] Ola B Nilsson, Adrian A Nickson, Jeffrey J Hollins, Stephan Wickles, Annette Steward, Roland Beckmann, Gunnar von Heijne, and Jane Clarke. Cotranslational folding of spectrin domains via partially structured states. *Nature Structural and Molecular Biology*, 24(3), 2017.

[8] Nurzian Ismail, Rickard Hedman, Martin Lindén, and Gunnar von Heijne. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nature Structural and Molecular Biology*, 22(2):145–9, 1 2015.

[9] Felix Nicolaus, Ane Metola, Daphne Mermans, Amanda Liljenström, Ajda Krč, Salmo Mohammed Abdullahi, Matthew H. Zimmer, Thomas F. Miller III, and Gunnar von Heijne. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *eLife*, 10:1–16, 2021.

[10] Benjamin Fritch, Andrey Kosolapov, Phillip Hudson, Daniel A. Nissley, H. Lee Woodcock, Carol Deutsch, and Edward P. O'Brien. Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *Journal of the American Chemical Society*, 140(15):5077–5087, 2018.

[11] Yang Jiang and Edward P. O'Brien. Mechanical Forces Have a Range of Effects on the Rate of Ribosome Catalyzed Peptidyl Transfer Depending on Direction. *The Journal of Physical Chemistry B*, 2021.

[12] Mian Zhou, Jinhu Guo, Joonseok Cha, Michael Chae, She Chen, Jose M Barral, Matthew S Sachs, and Yi Liu. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, 495(7439):111–5, 2013.

[13] Bin Zhang and Thomas F. Miller. Long-Timescale Dynamics and Regulation of Sec-Facilitated Protein Translocation. *Cell Reports*, 2(4):927–937, 10 2012.

[14] Michiel J.M. Niesen, Connie Y. Wang, Reid C. Van Lehn, and Thomas F. Miller. Structurally detailed coarse-grained model for Sec-facilitated co-translational protein translocation and membrane integration. *PLoS Computational Biology*, 13(3):1–26, 3 2017.

[15] Michiel J.M. Niesen, Annika Müller-Lucks, Rickard Hedman, Gunnar von Heijne, and Thomas F. Miller. Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon. *Biophysical Journal*, 115(10):1885–1894, 11 2018.

[16] Koreaki Ito and Shinobu Chiba. Biological significance of nascent polypeptides that stall the ribosome. In *Regulatory Nascent Polypeptides*, pages 3–20. Springer Japan, Tokyo, 2014.

[17] Daniel N. Wilson, Stefan Arenz, and Roland Beckmann. Translation regulation via nascent polypeptide-mediated ribosome stalling. *Current Opinion in Structural Biology*, 37:123–133, 2016.

[18] Daniel H Goldman, Christian M Kaiser, Maurizio Righini, and Carlos Bustamante. Mechanical force release nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*, 716(2013):457–460, 2014.

[19] D. Oliver, J. Norman, and S. Sarker. Regulation of Escherichia coli secA by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. *Journal of Bacteriology*, 180(19):5240–5242, 1998.

[20] Xiuqi Chen, Nandakumar Rajasekaran, Kaixian Liu, and Christian M. Kaiser. Synthesis runs counter to directional folding of a nascent protein domain. *Nature Communications*, 11(1), 2020.

[21] José Arcadio Farías-Rico, Sara Kathrin Goetz, Jacopo Marino, and Gunnar von Heijne. Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Letters*, 591(1):155–163, 1 2017.

[22] Mee Ngan Yap and Harris D. Bernstein. The Plasticity of a Translation Arrest Motif Yields Insights into Nascent Polypeptide Recognition inside the Ribosome Tunnel. *Molecular Cell*, 34(2):201–211, 4 2009.

[23] Mee Ngan Yap and Harris D. Bernstein. The translational regulatory function of SecM requires the precise timing of membrane targeting. *Molecular Microbiology*, 81(2):540–553, 2011.

[24] Jun Zhang, Xijiang Pan, Kaige Yan, Shan Sun, Ning Gao, and Sen Fang Sui. Mechanisms of ribosome stalling by SecM at multiple elongation steps. *eLife*, 4:1–25, 2015.

[25] Shashi Bhushan, Thomas Hoffmann, Birgit Seidelt, Jens Frauenfeld, Thorsten Mielke, Otto Berninghausen, Daniel N. Wilson, and Roland Beckmann. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biology*, 9(1), 2011.

[26] T. Martin Schmeing, Kevin S. Huang, Scott A. Strobel, and Thomas A. Steitz. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*, 438(7067):520–524, 2005.

[27] James Gumbart, Eduard Schreiner, Daniel N. Wilson, Roland Beckmann, and Klaus Schulten. Mechanisms of SecM-mediated stalling in the ribosome. *Biophysical Journal*, 103(2):331–341, 2012.

[28] Stefan Arenz, Lars V. Bock, Michael Graf, C. Axel Innis, Roland Beckmann, Helmut Grubmüller, Andrea C. Vaiana, and Daniel N. Wilson. A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. *Nature Communications*, 7(May):12026, 2016.

[29] Birgit Seidelt, C Axel Innis, Daniel N. Wilson, Marco Gartmann, J Armache, E Villa, Leonardo Trabuco, Thomas Becker, Thorsten Mielke, Klaus Schulten, Thomas A Steitz, and Roland Beckmann. Structural Insight into Nascent Polypeptide Chain–Mediated Translational Stalling. *Science*, 326(5958):1410–1412, 2009.

[30] Ting Su, Jingdong Cheng, Daniel Sohmen, Rickard Hedman, Otto Berninghausen, Gunnar von Heijne, Daniel N. Wilson, and Roland Beckmann. The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife*, 6:1–17, 2017.

[31] Hazel A. Bracken and Cheryl A. Woolhead. Increased freedom of movement in the nascent chain results in dynamic changes in the structure of the SecM arrest motif. *Bioscience Reports*, 39(December 2018):1–15, 2019.

[32] Daniel Sohmen, Shinobu Chiba, Naomi Shimokawa-Chiba, C Axel Innis, Otto Berninghausen, Roland Beckmann, Koreaki Ito, and Daniel N. Wilson.

Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. *Nature Communications*, 6:6941, 2015.

[33] Florian Cymer, Rickard Hedman, Nurzian Ismail, and Gunnar Von Heijne. Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *Journal of Biological Chemistry*, 290(16):10208–10215, 4 2015.

[34] Haley R. Harrington, Matthew H. Zimmer, Laura M. Chamness, Veronica Nash, Wesley D. Penn, Thomas F. Miller III, Suchetana Mukhopadhyay, Jonathan P. Schlebach, Thomas F. Miller, Suchetana Mukhopadhyay, and Jonathan P. Schlebach. Cotranslational folding stimulates programmed ribosomal frameshifting in the alphavirus structural polyprotein. *Journal of Biological Chemistry*, 295(20):6798–6808, 5 2020.

[35] Albert Tsai, Guy Kornberg, Magnus Johansson, Jin Chen, and Joseph D. Puglisi. The dynamics of SecM-induced translational stalling. *Cell Reports*, 7:1521–1533, 2014.

[36] Yury S. Polikanov, Thomas A. Steitz, and C. Axel Innis. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nature Structural and Molecular Biology*, 21(9):787–793, 2014.

[37] Phillip J. Elms, John D. Chodera, Carlos Bustamante, and Susan Marqusee. The molten globule state is unusually deformable under mechanical force. *Proceedings of the National Academy of Sciences*, 109(10):3796–3801, 2012.

[38] Michael T. Woodside and Steven M. Block. Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annual Review of Biophysics*, 43(1):19–39, 2014.

[39] David E. Shaw, J. P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, pages 41–53, 11 2014.

[40] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, 105(41):15755–15760, 2008.

[41] Jakob T. Bullerjahn, Sebastian Sturm, and Klaus Kroy. Theory of rapid force spectroscopy. *Nature Communications*, 5:1–10, 2014.

[42] Riti Gupta, Dmitri Toptygin, and Christian M. Kaiser. The SecA motor generates mechanical force during protein translocation. *Nature Communications*, 11(1):1–11, 2020.

[43] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[44] G I Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–27, 1978.

[45] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters*, 96(10):1–4, 2006.

[46] Gaurav Arya. Models for recovering the energy landscape of conformational transitions from single-molecule pulling experiments. *Molecular Simulation*, 42(13):1102–1115, 2016.

[47] Eiji Ishii, Shinobu Chiba, Narimasa Hashimoto, Seiji Kojima, Michio Homma, Koreaki Ito, Yoshinori Akiyama, Hiroyuki Mori, and Linda L. Randall. Nascent chain-monitored remodeling of the Sec machinery for salinity adaptation of marine bacteria. *Proceedings of the National Academy of Science*, 112(40):E5513–E5522, 2015.

[48] Olga K. Dudko, Jérôme Mathé, Attila Szabo, Amit Meller, and Gerhard Hummer. Extracting kinetics from single-molecule force spectroscopy: nanopore unzipping of DNA hairpins. *Biophysical Journal*, 92(June):4188–4195, 2007.

[49] Sarah E. Leininger, Fabio Trovato, Daniel A. Nissley, and Edward P. O'Brien. Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proceedings of the National Academy of Sciences*, 116(12):201813003, 2019.

[50] Sarah E. Leininger, Karthik Narayan, Carol Deutsch, and Edward P. O'Brien. Mechanochemistry in Translation. *Biochemistry*, 58(47):4657–4666, 11 2019.

[51] Pengfei Tian, Annette Steward, Renuka Kudva, Ting Su, Patrick J. Shilling, Adrian A. Nickson, Jeffrey J. Hollins, Roland Beckmann, Gunnar Von Heijne, Jane Clarke, and Robert B. Best. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proceedings of the National Academy of Sciences*, 115(48):E11284–E11293, 11 2018.

[52] Christian M. Kaiser and Kaixian Liu. Folding up and Moving on—Nascent Protein Folding on the Ribosome. *Journal of Molecular Biology*, 430(22):4580–4591, 2018.

[53] Kakoli Mitra, Christiane Schaffitzel, Felcy Fabiola, Michael S. Chapman, Nenad Ban, and Joachim Frank. Elongation Arrest by SecM via a Cascade of Ribosomal RNA Rearrangements. *Molecular Cell*, 22(4):533–543, 2006.

[54] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics*, 78(8):1950–1958, 2010.

[55] Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 1(November), 2006.

[56] Neva Caliskan, Frank Peske, and Marina V. Rodnina. Changed in translation: MRNA recoding by -1 programmed ribosomal frameshifting. *Trends in Biochemical Sciences*, 40(5):265–274, 5 2015.

[57] Matthew T.J. Halma, Dustin B. Ritchie, Tonia R. Cappellano, Krishna Neupane, and Michael T. Woodside. Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proceedings of the National Academy of Sciences of the United States of America*, 116(39):19500–19505, 9 2019.

[58] Olivier Namy, Stephen J. Moran, David I. Stuart, Robert J.C. Gilbert, and Ian Brierley. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090):244–247, 5 2006.

[59] Young Chan Kim, Alexander W. Tarr, and Christopher N. Penfold. Colicin import into E. coli cells: A model system for insights into the import mechanisms of bacteriocins. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8):1717–1731, 2014.

[60] Neva Caliskan, Vladimir I. Katunin, Riccardo Belardinelli, Frank Peske, and Marina V. Rodnina. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell*, 157(7):1619–1631, 6 2014.

[61] Jin Chen, Alexey Petrov, Magnus Johansson, Albert Tsai, Seán E. O'leary, and Joseph D. Puglisi. Dynamic pathways of -1 translational frameshifting. *Nature*, 512(7514):328–332, 8 2014.

[62] Junhong Choi, Sinéad O'Loughlin, John F. Atkins, and Joseph D. Puglisi. The energy landscape of -1 ribosomal frameshifting. *Science Advances*, 6(1), 1 2020.

[63] Natalia Korniy, Ekaterina Samatova, Maria M. Anokhina, Frank Peske, and Marina V. Rodnina. Mechanisms and biomedical implications of –1 programmed ribosome frameshifting on viral and bacterial mRNAs, 7 2019.

[64] Lars V. Bock, Neva Caliskan, Natalia Korniy, Frank Peske, Marina V. Rodnina, and Helmut Grubmueller. Thermodynamic Control of Ribosomal Frameshifting. *Biophysical Journal*, 116(3):361a, 2019.

[65] Ashton Trey Belew, Arturas Meskauskas, Sharmishtha Musalgaonkar, Vivek M. Advani, Sergey O. Sulima, Wojciech K. Kasprzak, Bruce A. Shapiro, and Jonathan D. Dinman. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature*, 512(7514):265–269, 8 2014.

[66] Sawsan Napthine, Roger Ling, Leanne K. Finch, Joshua D. Jones, Susanne Bell, Ian Brierley, and Andrew E. Firth. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature Communications*, 8:1–11, 6 2017.

[67] Xinlu Wang, Yifang Xuan, Yuling Han, Xiang Ding, Kai Ye, Fuquan Yang, Pu Gao, Stephen P. Goff, and Guangxia Gao. Regulation of HIV-1 Gag-Pol Expression by Shiftless, an Inhibitor of Programmed -1 Ribosomal Frameshifting. *Cell*, 176(3):625–635, 1 2019.

[68] Jolene Ramsey and Suchetana Mukhopadhyay. Disentangling the frames, the state of research on the alphavirus 6K and TF proteins. *Viruses*, 9(8):1–21, 2 2017.

[69] Rebecca S. Brown, Judy J. Wan, and Margaret Kielian. The alphavirus exit pathway: What we know and what we wish we knew, 2 2018.

[70] M Lobigs, H X Zhao, and H Garoff. Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. *Journal of Virology*, 64(9):4346–4355, 1990.

[71] Matthew Mulvey and Dennis T. Brown. Assembly of the Sindbis virus spike protein complex. *Virology*, 219(1):125–132, 5 1996.

[72] Andrew E. Firth, Betty Y.W. Chung, Marina N. Fleeton, and John F. Atkins. Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virology Journal*, 5, 2008.

[73] D. Hallengard, M. Kakoulidou, A. Lulla, B. M. Kummerer, D. X. Johansson, M. Mutso, V. Lulla, J. K. Fazakerley, P. Roques, R. Le Grand, A. Merits, and P. Liljestrom. Novel Attenuated Chikungunya Vaccine Candidates Elicit Protective Immunity in C57BL/6 mice. *Journal of Virology*, 88(5):2858–2866, 3 2014.

[74] Adam Taylor, Julian V. Melton, Lara J. Herrero, Bastian Thaa, Liis Karo-Astover, Peter W. Gage, Michelle A. Nelson, Kuo-Ching Sheng, Brett A. Lidbury, Gary D. Ewart, Gerald M. McInerney, Andres Merits, and Suresh Mahalingam. Effects of an In-Frame Deletion of the 6k Gene Locus from the Genome of Ross River Virus . *Journal of Virology*, 90(8):4150–4159, 4 2016.

[75] J. E. Snyder, K. A. Kulcsar, K. L. W. Schultz, C. P. Riley, J. T. Neary, S. Marr, J. Jose, D. E. Griffin, and R. J. Kuhn. Functional Characterization of the Alphavirus TF Protein. *Journal of Virology*, 87(15):8511–8523, 8 2013.

[76] K. J. Rogers, S. Jones-Burrage, W. Maury, and S. Mukhopadhyay. TF protein of Sindbis virus antagonizes host type I interferon responses in a palmitoylation-dependent manner. *Virology*, 542:63–70, 3 2020.

[77] Betty Y.W. Chung, Andrew E. Firth, and John F. Atkins. Frameshifting in Alphaviruses: A Diversity of 3' Stimulatory Structures. *Journal of Molecular Biology*, 397(2):448–456, 3 2010.

[78] Joseph A. Kendra, Vivek M. Advani, Bin Chen, Joseph W. Briggs, Jinyi Zhu, Hannah J. Bress, Sushrut M. Pathy, and Jonathan D. Dinman. Functional and structural characterization of the chikungunya virus translational recoding signals. *Journal of Biological Chemistry*, 293(45):17536–17545, 2018.

[79] A. J. Snyder, K. J. Sokoloski, and S. Mukhopadhyay. Mutating Conserved Cysteines in the Alphavirus E2 Glycoprotein Causes Virus-Specific Assembly Defects. *Journal of Virology*, 86(6):3100–3111, 3 2012.

[80] Jolene Ramsey, Emily C. Renzi, Randy J. Arnold, Jonathan C. Trinidad, and Suchetana Mukhopadhyay. Palmitoylation of Sindbis Virus TF Protein Regulates Its Plasma Membrane Localization and Subsequent Incorporation into Virions. *Journal of Virology*, 91(3), 2 2017.

[81] Kerstin Gaedigk-Nitschko and Milton J. Schlesinger. The sindbis virus 6K protein can be detected in virions and is acylated with fatty acids. *Virology*, 175(1):274–281, 1990.

[82] Eugenia G. Politis, Amy F. Roth, and Nicholas G. Davis. Transmembrane topology of the protein palmitoyl transferase Akr1. *Journal of Biological Chemistry*, 280(11):10156–10163, 3 2005.

[83] P Liljeström and H Garoff. Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *Journal of Virology*, 65(1):147–154, 1991.

[84] R. Holland Cheng, Richard J. Kuhn, Norman H. Olson, Michael G. Rossmannˆ Hok-Kin Choi, Thomas J. Smith, and Timothy S. Baker. Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell*, 80(4):621–630, 2 1995.

[85] Lihong Chen, Ming Wang, Dongjie Zhu, Zhenzhao Sun, Jun Ma, Jinglin Wang, Lingfei Kong, Shida Wang, Zaisi Liu, Lili Wei, Yuwen He, Jingfei Wang, and Xinzheng Zhang. Implication for alphavirus host-cell entry and assembly indicated by a 3.5Å resolution cryo-EM structure. *Nature Communications*, 9(1), 12 2018.

[86] Guro Gafvelin and Gunnar von Heijne. Topological "frustration" in multi-spanning E. coli inner membrane proteins. *Cell*, 77(3):401–412, 5 1994.

[87] Travis R. Ruch and Carolyn E. Machamer. The coronavirus E protein: Assembly and beyond, 3 2012.

[88] Tara Hessa, Nadja M. Meindl-Beinker, Andreas Bernsel, Hyun Kim, Yoko Sato, Mirjam Lerch-Bader, Ingmarie Nilsson, Stephen H. White, and Gunnar Von Heijne. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450(7172):1026–1030, 12 2007.

[89] Tara Hessa, Hyun Kim, Karl Bihlmaier, Carolina Lundin, Jorrit Boekel, Helena Andersson, IngMarie Marie Nilsson, Stephen H. White, and Gunnar von Heijne. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433(7024):377–381, 1 2005.

[90] L Ivanova and M J Schlesinger. Site-directed mutations in the Sindbis virus E2 glycoprotein identify palmitoylation sites and affect virus budding. *Journal of Virology*, 67(5):2546–2551, 1993.

[91] Christine Ryan, Lidia Ivanova, and Milton J. Schlesinger. Effects of site-directed mutations of transmembrane cysteines in sindbis virus E1 and E2 glycoproteins on palmitylation and virus replication. *Virology*, 249(1):62–67, 9 1998.

[92] Hunsang Lee, Jisoo Min, Gunnar von Heijne, and Hyun Kim. Glycosylatable GFP as a compartment-specific membrane topology reporter. *Biochemical and Biophysical Research Communications*, 427(4):780–784, 11 2012.

[93] Florian Cymer, Nurzian Ismail, and Gunnar Von Heijne. Weak pulling forces exerted on Nin-orientated transmembrane segments during co-translational insertion into the inner membrane of Escherichia coli. *FEBS Letters*, 588(10):1930–1934, 5 2014.

[94] Florian Cymer and Gunnar Von Heijne. Cotranslational folding of membrane proteins probed by arrest-peptide- mediated force measurements. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36):14640–14645, 9 2013.

[95] Reid C. Van Lehn, Bin Zhang, and Thomas F. Miller. Regulation of multi-spanning membrane protein topology via post-translational annealing. *eLife*, 4(September2015):1–23, 9 2015.

[96] Kent E.S. Matlack, Benjamin Misselwitz, Kathrin Plath, and Tom A. Rapoport. BIP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane [2], 5 1999.

[97] Jeffrey L. Brodsky, Jennifer Goeckeler, and Randy Schekman. BiP and Sec63p are required for both co- and posttranslational protein translocation into the yeast endoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America*, 92(21):9643–9646, 10 1995.

[98] Kristi L. Muldoon-Jacobs and Jonathan D. Dinman. Specific effects of ribosome-tethered molecular chaperones on programmed -1 ribosomal frameshifting. *Eukaryotic Cell*, 5(4):762–770, 4 2006.

[99] Rebecca M. Voorhees, Israel S. Fernández, Sjors H.W. Scheres, and Ramanujan S. Hegde. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell*, 157(7):1632–1643, 6 2014.

[100] Stephen Wu, Panagiotis Angelikopoulos, Gerardo Tauriello, Costas Papadimitriou, and Petros Koumoutsakos. Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models. *Journal of Chemical Physics*, 145(24), 12 2016.

[101] John F. Atkins, Gary Loughran, Pramod R. Bhatt, Andrew E. Firth, and Pavel V. Baranov. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Research*, 44(15):7007–7078, 9 2016.

[102] Wesley D. Penn, Haley R. Harrington, Jonathan P. Schlebach, and Suchetana Mukhopadhyay. Regulators of Viral Frameshifting: More Than RNA Influences Translation Events, 9 2020.

[103] Tyler Jacks, Hiten D. Madhani, Frank R. Masiarz, and Harold E. Varmus. Signals for ribosomal frameshifting in the rous sarcoma virus gag-pol region. *Cell*, 55(3):447–458, 11 1988.

[104] H Reil, H Kollmus, U H Weidle, and H Hauser. A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *Journal of Virology*, 67(9):5579–5584, 1993.

[105] Guillaume Stahl, Gregory P. McCarty, and Philip J. Farabaugh. Ribosome structure: Revisiting the connection between translational accuracy and unconventional decoding. *Trends in Biochemical Sciences*, 27(4):178–183, 4 2002.

[106] David P. Giedroc and Peter V. Cornish. Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Research*, 139(2):193–208, 2 2009.

[107] Kathryn D. Mouzakis, Andrew L. Lang, Kirk A. Vander Meulen, Preston D. Easterday, and Samuel E. Butcher. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Research*, 41(3):1901–1913, 2 2013.

[108] Pramod R. Bhatt, Alain Scaiola, Gary Loughran, Marc Leibundgut, Annika Kratzel, Angus McMillan, Kate M. O' Connor, Jeffrey W. Bode, Volker Thiel, John F. Atkins, and Nenad Ban. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *bioRxiv*, pages 1–22, 2020.

[109] Martin Mikl, Yitzhak Pilpel, and Eran Segal. High-throughput interrogation of programmed ribosomal frameshifting in human cells. *Nature Communications*, 11(1), 2020.

[110] Katrina M. Kutchko, Emily A. Madden, Clayton Morrison, Kenneth S. Plante, Wes Sanders, Heather A. Vincent, Marta C.Cruz Cisneros, Kristin M. Long, Nathaniel J. Moorman, Mark T. Heise, and Alain Laederach. Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Research*, 46(7):3657–3670, 4 2018.

[111] Michiel J.M. Niesen, Matthew H. Zimmer, and Thomas F. Miller III. Dynamics of Co-translational Membrane Protein Integration and Translocation via the Sec Translocon. *Journal of the American Chemical Society*, 142(12):5449–5460, 2020.

[112] Kenneth A. Matreyek, Jason J. Stephany, and Douglas M. Fowler. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Research*, 45(11):1–12, 2017.

[113] Melissa A. Chiasson, Nathan J. Rollins, Jason J. Stephany, Katherine A. Sitko, Kenneth A. Matreyek, Marta Verby, Song Sun, Frederick P. Roth, Daniel De Sloover, Debora S. Marks, Allan E. Rettie, and Douglas M. Fowler. Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *eLife*, 9:1–25, 9 2020.

[114] Wesley D. Penn, Andrew G. McKee, Charles P. Kuntz, Hope Woods, Veronica Nash, Timothy C. Gruenhagen, Francis J. Roushar, Mahesh Chandak, Chris Hemmerich, Douglas B. Rusch, Jens Meiler, and Jonathan P. Schlebach. Probing biophysical sequence constraints within the transmembrane domains of rhodopsin by deep mutational scanning. *Science Advances*, 6(10), 2020.

[115] Ian Brierley, Alison J. Jenner, and Stephen C. Inglis. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *Journal of Molecular Biology*, 227(2):463–479, 9 1992.

[116] Shannon Yan, Jin Der Wen, Carlos Bustamante, and Ignacio Tinoco. Ribosome Excursions during mRNA Translocation Mediate Broad Branching of Frameshift Pathways. *Cell*, 160(5):870–881, 2 2015.

[117] C. Preston Moon and Karen G. Fleming. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences of the United States of America*, 108(25):10174–10177, 6 2011.

[118] Michiel J.M. Niesen, Stephen S. Marshall, Thomas F. Miller, and William M. Clemons. Improving membrane protein expression by optimizing integration efficiency. *Journal of Biological Chemistry*, 292(22):19537–19545, 2017.

[119] Chengying Ma, Xiaofei Wu, Dongjie Sun, Eunyong Park, Marco A. Catipovic, Tom A. Rapoport, Ning Gao, and Long Li. Structure of the substrate-engaged SecA-SecY protein translocation machine. *Nature Communications*, 10(1):1–9, 12 2019.

[120] Renuka Kudva, Pengfei Tian, Fátima Pardo-Avila, Marta Carroni, Robert B. Best, Harris D. Bernstein, and Gunnar von Heijne. The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding. *eLife*, 7:1–15, 2018.

[121] Avi J. Samelson, Madeleine K. Jensen, Randy A. Soto, Jamie H.D. Cate, and Susan Marqusee. Quantitative determination of ribosome nascent chain stability. *Proceedings of the National Academy of Sciences of the United States of America*, 113(47):13402–13407, 11 2016.

[122] Angela M. Smith, Michael S. Costello, Andrew H. Kettring, Robert J. Wingo, and Sean D. Moore. Ribosome collisions alter frameshifting at translational reprogramming motifs in bacterial mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 116(43):21769–21779, 10 2019.

[123] Manuel Bañó-Polo, Carlos Baeza-Delgado, Silvia Tamborero, Anthony Hazel, Brayan Grau, Ing Marie Nilsson, Paul Whitley, James C. Gumbart, Gunnar von Heijne, and Ismael Mingarro. Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nature Communications*, 9(1), 12 2018.

[124] Rebecca M Voorhees, Israel S Ferná Ndez, Sjors H W Scheres, and Ramanujan S Hegde. Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. *Cell*, 157:1632–1643, 2014.

[125] Benjamin Webb and Andrej Sali. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, 54(1):1–5, 6 2016.

[126] Patrick J Carmody, Matthew H. Zimmer, Charles P Kuntz, Haley R Harrington, Kate E Duckworth, Wesley D Penn, Suchetana Mukhopadhyay, Thomas F Miller III, and Jonathan P Schlebach. Coordination of -1 Programmed Ribosomal Frameshifting by Transcript and Nascent Chain Features Revealed by Deep Mutational Scanning. *bioRxiv*, 3 2021.

[127] Tom A. Rapoport, Long Li, and Eunyong Park. Structural and mechanistic insights into protein translocation. *Annual Review of Cell and Developmental Biology*, 33:369–390, 10 2017.

[128] Patrick J. Chitwood, Szymon Juszkiewicz, Alina Guna, Sichen Shao, and Ramanujan S. Hegde. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*, 175(6):1507–1519, 11 2018.

[129] José Arcadio Farías-Rico, Frida Ruud Selin, Ioanna Myronidi, Marie Frühauf, and Gunnar Von Heijne. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40):E9280–E9287, 10 2018.

[130] Koreaki Ito and Shinobu Chiba. Arrest peptides: Cis-acting modulators of translation, 6 2013.

[131] Grant Kemp, Ola B. Nilsson, Ola B. Nilsson, Pengfei Tian, Robert B. Best, Gunnar Von Heijne, and Gunnar Von Heijne. Cotranslational folding cooperativity of contiguous domains of a-spectrin. *Proceedings of the National Academy of Sciences of the United States of America*, 117(25):14119–14126, 6 2020.

[132] Luigi Notari, Markel Martínez-Carranza, José Arcadio Farías-Rico, Pål Stenmark, and Gunnar von Heijne. Cotranslational Folding of a Pentarepeat β-Helix Protein. *Journal of Molecular Biology*, 430(24):5196–5206, 12 2018.

[133] Yen Ju Chen, Owen Pornillos, Samantha Lieu, Che Ma, Andy P. Chen, and Geoffrey Chang. X-ray structure of EmrE supports dual topology model. *Proceedings of the National Academy of Sciences of the United States of America*, 104(48):18999–19004, 11 2007.

[134] Mikaela Rapp, Susanna Seppälä, Erik Granseth, and Gunnar Von Heijne. Emulating membrane protein evolution by rational design. *Science*, 315(5816):1282–1284, 3 2007.

[135] Evan Mercier, Wolfgang Wintermeyer, and Marina V Rodnina. Cotranslational insertion and topogenesis of bacterial membrane proteins monitored in real time. *The EMBO Journal*, 39(15), 8 2020.

[136] Maximilian Seurig, Moira Ek, Gunnar von Heijne, and Nir Fluman. Dynamic membrane topology in an unassembled membrane protein. *Nature Chemical Biology*, 15(10):945–948, 10 2019.

[137] Allison R. Sherratt, David R. Blais, Houman Ghasriani, John Paul Pezacki, and Natalie K. Goto. Activity-based protein profiling of the escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry*, 51(39):7794–7803, 10 2012.

[138] Yongcheng Wang, Yingjiu Zhang, and Ya Ha. Crystal structure of a rhomboid family intramembrane protease. *Nature*, 444(7116):179–183, 11 2006.

[139] Yoshihiro Shimizu, Takashi Kanamori, and Takuya Ueda. Protein synthesis by pure translation systems. *Methods*, 36(3):299–304, 7 2005.

[140] Kevin W. Plaxco, Kim T. Simons, and David Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277(4):985–994, 4 1998.

[141] R. Young and H. Bremer. Polypeptide chain elongation rate in Escherichia coli B/r as a function of growth rate. *Biochemical Journal*, 160(2):185–194, 1976.

[142] Grant Kemp, Renuka Kudva, Andrés de la Rosa, and Gunnar von Heijne. Force-Profile Analysis of the Cotranslational Folding of HemK and Filamin Domains: Comparison of Biochemical and Biophysical Folding Assays. *Journal of Molecular Biology*, 431(6):1308–1314, 3 2019.

[143] Rikki N. Hvorup, Birke A. Goetz, Martina Niederer, Kaspar Hollenstein, Eduardo Perozo, and Kaspar P. Locher. Asymmetry in the structure of the ABC transporter - Binding protein complex BtuCD-BtuF. *Science*, 317(5843):1387–1390, 9 2007.

[144] Ramya Lakshminarayan, Ben P. Phillips, Imogen L. Binnian, Natalia Gomez-Navarro, Norberto Escudero-Urquijo, Alan J. Warren, and Elizabeth A. Miller. Pre-emptive Quality Control of a Misfolded Membrane Protein by Ribosome-Driven Effects. *Current Biology*, 30(5):854–864, 3 2020.

[145] Florian Cymer, Gunnar Von Heijne, and Stephen H. White. Mechanisms of integral membrane protein insertion and folding. *Journal of Molecular Biology*, 427(5):999–1022, 3 2015.

[146] Haoze He, Andreas Kuhn, and Ross E. Dalbey. Tracking the Stepwise Movement of a Membrane-inserting Protein In Vivo. *Journal of Molecular Biology*, 432(2):484–496, 1 2020.

[147] Dagan C. Marx and Karen G. Fleming. Local Bilayer Hydrophobicity Modulates Membrane Protein Stability. *Journal of the American Chemical Society*, 143(2):764–772, 1 2021.

[148] Hao Yu, Matthew G.W. Siewny, Devin T. Edwards, Aric W. Sanders, and Thomas T. Perkins. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science*, 355(6328):945–950, 3 2017.

[149] Hyun Kyu Choi, Duyoung Min, Hyunook Kang, Min Ju Shon, Sang Hyun Rah, Hak Chan Kim, Hawoong Jeong, Hee Jung Choi, James U. Bowie, and Tae Young Yoon. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science*, 366(6469):1150–1156, 11 2019.

[150] Lukas Kater, Benedikt Frieg, Otto Berninghausen, Holger Gohlke, Roland Beckmann, and Alexej Kedrov. Partially inserted nascent chain unzips the lateral gate of the Sec translocon. *EMBO reports*, 20(10), 10 2019.

[151] Wei Lu, Nicholas P. Schafer, and Peter G. Wolynes. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nature Communications*, 9(1), 12 2018.