Institute of Physics,

Polish Academy of Sciences

INFLUENCE OF THE RIBOSOME ON PROTEIN EJECTION AND FOLDING

A Dissertation in

Physics

by

Quyen V. Vu

Supervisor: Prof. Dr hab. Mai Suan Li

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

2023

DECLARATION

I, Quyen V. Vu, declare that this thesis entitled “Influence of the ribosome on protein ejection and folding” and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at the Institute of Physics, Polish Academy of Sciences.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this Institute or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given or cited. Except for such quotations, this thesis is entirely my work.
- I have acknowledged all primary sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clearly what was done by others and what I have contributed.

Warsaw, 26 May 2023

Quyen V. Vu

ABSTRACT

Proteins are synthesized by macromolecular machines called ribosomes, which are found in cells across all species, from bacteria to humans. They perform various tasks necessary to support life. To carry out their functions, many proteins must first self-assemble into a specific configuration known as the native state. The process of a protein attaining its native state is termed protein folding. The folding of proteins in isolation has been extensively studied for over a half-century. However, within cells, proteins are translated by the ribosome based on information contained in an mRNA sequence and emerge through the exit tunnel to the cytosol after synthesis. Proteins can acquire tertiary structure at any stage: during their biosynthesis, as they are ejected through the ribosome's exit tunnel, or posttranslationally – after their release from the ribosome. Indeed, several computational and experimental studies have shown that proteins can start to fold while they are still being synthesized by the ribosome. This phenomenon, known as cotranslational folding, is mediated by the spatial constraints of the ribosomal exit tunnel as well as the interactions between the nascent chain and the ribosome surface. These factors can potentially impact the kinetics and pathways of protein folding. Therefore, gaining a comprehensive understanding of protein behavior during their early stages of existence is of utmost importance and remains a significant focus of ongoing research.

This thesis contains three computational studies related to protein ejection and folding on the ribosome. The introduction to the ribosome and protein folding on the ribosome is summarized in Chapter 1. Chapter 2 describes the computational methods with a focus on the computational modeling and analyses used in the research presented in this dissertation.

In Chapter 3, the ejection process of nascent protein out of the ribosome exit tunnel is described. This process has not been studied before as it is believed to be fast, show little variation between proteins, and have no biological significance. Using a combination of multiscale modeling, and ribosome profiling experimental data, we find a greater than 1000-fold variation in ejection times. Nascent proteins enriched in negatively charged residues near their C-terminus eject the fastest, while nascent proteins enriched in positively charged residues tend to eject much more slowly. More work is required to pull slowly ejecting proteins out of the exit

tunnel than quickly ejecting proteins, according to all-atom steered molecular dynamics simulations. An energetic decomposition reveals that the slow ejection is due to the strong attractive electrostatic interactions between the nascent chain and the negatively charged ribosomal-RNA lining the exit tunnel, while the quick ejection of proteins is due to their repulsive electrostatic interactions with the exit tunnel. Ribosome profiling data from Escherichia coli reveals that the presence of slowly ejecting sequences correlates with ribosomes spending more time at stop codons. This indicates that the ejection process might delay ribosome recycling and could influence the cotranslational behavior of proteins.

Chapter 4 presents the results of the all-atom simulations of hydrophobic interactions in the presence and absence of the ribosome. Interactions between the ribosome and nascent protein can destabilize folded domains in the ribosome exit tunnel's vestibule, the last 3 nm of the exit tunnel where tertiary folding can occur. Here, we test if the contribution to this destabilization is the weakening of the hydrophobic association, which is the driving force for protein folding. The potential-of-mean force between two methane molecules along the center line of the ribosome exit tunnel and in bulk solution was calculated. The results indicate that the associated methanes are half as stable in the ribosome's vestibule as compared to bulk solution, demonstrating that the hydrophobic effect is weakened by the presence of the ribosome. We demonstrate that the weakening of the hydrophobic effect is due to the increased ordering of water molecules in the presence of the ribosome. These findings mean that nascent proteins pass through a ribosome vestibule environment that can destabilize folded structures. This, in turn, can potentially impact cotranslational protein folding pathways, as well as their energetics and kinetics.

In Chapter 5, the influence of protein synthesis and posttranslational folding on protein folding efficiency is described and compared to the folding from denatured states in bulk solution. To make this comparison, coarse-grained molecular dynamics simulations were performed for dihydrofolate reductase (DHFR), type III chloramphenicol acetyltransferase (CAT-III), and D-alanine–D-alanine ligase B (DDLB) proteins. The results indicate that the influence of ribosomes on folding efficiency depends on the protein size and complexity. For small, simple folds (DHFR), the ribosome facilitates efficient folding by preventing misfolding. However, for larger, more complex proteins (CAT-III and DDLB), the ribosome may not

promote folding and may contribute to intermediate misfolds during translation. Additionally, it was found that the folding efficiency correlates with the presence of tertiary structural elements known as entanglements in the native structure.

Finally, Chapter 6 summarizes the conclusions that can be drawn from this work and directions for future research.

Streszczenie

Białka są syntetyzowane przez wielkocząsteczkowe maszyny zwane rybosomami, które znajdują się w komórkach wszystkich gatunków, od bakterii po ludzi. Wykonują one różne zadania niezbędne do podtrzymania życia, ale aby pełnić swoje funkcje, wiele białek musi najpierw samo przybrać specyficzną strukturę znaną jako stan natywny, a proces osiągania go nazywany jest zwijaniem białek. Zwijanie izolowanych białek jest badane od ponad pół wieku, jednak w komórkach białka są tłumaczone przez rybosom na podstawie informacji zawartych w sekwencji mRNA i po syntezie wychodzą przez tunel wyjściowy do cytozolu. Białka mogą uzyskać strukturę trzeciorzędową na dowolnym etapie: podczas ich biosyntezy, gdy są uwalniane przez tunel wyjściowy rybosomu lub potranslacyjnie - po ich uwolnieniu z rybosomu. Różne badania obliczeniowe i eksperymentalne wykazały, że białka mogą zacząć się zwijać, gdy są nadal syntetyzowane przez rybosom. W zjawisku tym, znanym jako kotranslacyjne zwijanie, pośredniczą ograniczenia przestrzenne rybosomalnego tunelu wyjściowego, a także oddziaływania między powstającym łańcuchem a powierzchnią rybosomu. Czynniki te mogą potencjalnie wpływać na kinetykę i ścieżki zwijania białek, dlatego też tak ważne jest zrozumienie zachowania białek na wczesnych etapach ich istnienia.

Praca doktorska zawiera trzy projekty obliczeniowe opisujące zwijanie białek w rybosomie. Opis rybosomu i procesu zwijania białek w rybosomie jest podsumowany w rozdziale 1., z kolei rozdział 2. opisuje metody obliczeniowe, skupiając się na modelowaniu molekularnym i analizach używanych w badaniach przedstawionych w tej dysertacji. W rozdziale 3. opisano proces uwalniania powstającego białka z tunelu wyjściowego rybosomu. Ten proces nie był wcześniej badany, ponieważ uważano, że jest szybki, wykazuje jedynie niewielkie zmiany między białkami i nie ma znaczenia biologicznego. Wykorzystując kombinację modelowania wieloskalowego i analizy profilowania rybosomów, znaleźliśmy ponad 1000-krotną różnicę w czasach uwalniania białek z rybosomu. Powstające białka wzbogacone w reszty o ładunku ujemnym w pobliżu ich C-końca są uwalniane najszybciej, podczas gdy białka wzbogacone w reszty o ładunku dodatnim mają tendencję do znacznie

wolniejszego uwalniania z rybosomu. Pełnoatomowe symulacje sterowanej dynamiki molekularnej wykazały, że wymagane jest włożenie wyższej pracy, aby wyciągnąć białka powoli uwalniane z tunelu wyjściowego niż te uwalniane szybko. Natomiast dekompozycja członów energii ujawniła, że powolne uwalnianie spowodowane jest silnymi przyciągającymi oddziaływaniami elektrostatycznymi pomiędzy powstającym łańcuchem a ujemnie naładowanym kanałem rybosomu z związanym RNA, podczas gdy szybkie uwalnianie białek spowodowane jest ich odpychającymi oddziaływaniami elektrostatycznymi z tunelem wyjściowym. Dane z profilowania rybosomów z *Escherichia coli* pokazują, że obecność sekwencji białek, które są uwalniane powoli koreluje z dłuższym czasem spędzanym przez rybosomy na kodonach stop, co wskazuje, że proces uwalniania może opóźniać recykling rybosomu.

Rozdział 4. przedstawia wyniki symulacji pełnoatomowych dotyczących oddziaływań hydrofobowych w przypadku obecności rybosomu oraz jego braku (w roztworze). Badania wykazały, że oddziaływania między rybosomem a powstającym białkiem w przedsionku tunelu wyjściowego rybosomu (ostatnie 3 nm tunelu wyjściowego, gdzie może nastąpić zwijanie struktur trzeciorzędowych) mogą destabilizować powstające domeny. Za pomocą obliczeń potencjału średniej siły pomiędzy dwoma cząstkami metanu wzdłuż linii środkowej tunelu wyjściowego rybosomu i w roztworze prawdziliśmy, czy do tej destabilizacji przyczynia się osłabienie asocjacji hydrofobowej, która jest siłą napędową zwijania białek. Wyniki wskazują, że związane cząsteczki metanu są dwa razy mniej stabilne w przedsionku rybosomu w porównaniu z warunkami w roztworze, co dowodzi, że efekt hydrofobowy jest osłabiony przez obecność rybosomu. Dodatkowo stwierdziliśmy, że osłabienie efektu hydrofobowego wynika z większego uporządkowania cząsteczek wody w obecności rybosomu. Te odkrycia oznaczają, że powstające białka przechodzą przez środowisko przedsionka rybosomu, które może destabilizować zwijające się struktury, a to z kolei może potencjalnie wpływać na ścieżki zwijania białek kotranslacyjnych, a także na ich energetykę i kinetykę.

W Rozdziale 5. opisano i porównano wpływ syntezy białek i zwijania posttranslacyjnego na efektywność zwijania ze stanów zdenaturowanych w stosunku do zwijania w roztworze. Gruboziarniste symulacje dynamiki molekularnej zostały użyte do porównania, jak reduktaza

dihydrofolianowa (DHFR), acetylotransferaza chloramfenikolowa typu III (CAT-III) i ligasa B D-alaniny–D-alaniny zwijają się podczas i po syntezie na rybosomie, w porównaniu do zwijania ze stanu rozwiniętego w roztworze. Wyniki wskazują, że wpływ rybosomów na efektywność zwijania białek zależy od ich wielkości i złożoności. Dla małych, prostych struktur(DHFR), rybosom ułatwia efektywne zwijanie, zapobiegając nieprawidłowemu zwijaniu, jednak dla większych, bardziej złożonych białek (CAT-III i DDLB), rybosom może nie sprzyjać zwijaniu i może przyczyniać się do powstawania nieprawidłowo zwiniętych struktur podczas translacji. Dodatkowo stwierdzono, że efektywność zwijania koreluje z zaplątaniem obecnym w strukturze natywnej.

Rozdział 6. podsumowuje wnioski z mojej pracy i kierunki przyszłych badań.

Publications used in this thesis

1. Nissley, D. A.; **Vu, Q. V.;** Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J. Am. Chem. Soc.* 2020, *142* (13), 6103–6110.
2. **Vu, Q. V.;** Jiang, Y.; Li, M. S.; O'Brien, E. P. The Driving Force for Co-Translational Protein Folding Is Weaker in the Ribosome Vestibule Due to Greater Water Ordering. *Chem. Sci.* 2021, *12* (35), 11851–11857.
3. **Vu, Q. V**.; Nissley, D. A., Jiang, Y.; O'Brien, E. P.; Li, M. S. Is Posttranslational Folding More Efficient Than Refolding from a Denatured State: A Computational Study. *J. Phys. Chem. B* 2023, 127 (21), 4761–4774.

Other publications:

1. Leininger, S. E.; Rodriguez, J.; **Vu, Q. V.**; Jiang, Y.; Li, M. S.; Deutsch, C.; O'Brien, E. P. Ribosome Elongation Kinetics of Consecutively Charged Residues Are Coupled to Electrostatic Force. *Biochemistry* 2021, *60* (43), 3223–3235.
2. Halder, R.; Nissley, D. A.; Sitarik, I.; Jiang, Y.; Rao, Y.; **Vu, Q. V.;** Li, M. S.; Pritchard, J. R.; O'Brien, E. P. How soluble misfolded proteins bypass chaperones at the molecular level. *Nat. Commun.* 2023 (accepted).
3. Dang, L. P.; Nissley, D. A.; Sitarik, I.; **Vu, Q. V.;** Jiang, Y.; Li, M. S.; O'Brien, E. P. Synonymous Mutations Can Alter Protein Dimerization through Localized Interface Misfolding Involving Self-Entanglements. *bioRxiv* **2021**, 2021.10.26.465867. https://doi.org/10.1101/2021.10.26.465867.
4. **Vu, Q. V.;** Sitarik, I.; Jiang, Y.; Yadav, D.; Sharma, P.; Fried, S. D.; Li, M. S.; O'Brien, E. P. A Newly Identified Class of Protein Misfolding in All-Atom Folding Simulations Consistent with Limited Proteolysis Mass Spectrometry. bioRxiv 2022. https://doi.org/10.1101/2022.07.19.500586.

TABLE OF CONTENTS

List of Figures xvii

List of Tables xix

Abbreviation xx

ACKNOWLEDGEMENTS xxii

Chapter 1. Introduction 1

1.1 Protein and the folding problem 1

1.2 Hydrophobic effect – the driving force for protein folding 3

1.3 Ribosome 4

1.4 Ribosome exit tunnel 7

1.5 Protein folding on the ribosome 8

1.5.1 Some proteins fold in the exit tunnel 9

1.5.2 Ribosome destabilizes folded domains 10

1.5.3 The folding kinetics of proteins are slower on the ribosome 11

1.5.4 Folding pathways of proteins on and off the ribosome 12

1.6 Thesis objectives 13

Chapter 2. Computational background 14

2.1 Molecular dynamics simulation 14

2.2 All-atom modeling 16

2.3 Coarse-grained modeling 17

2.4 All-atom modeling of 50S *E. coli* ribosome 19

2.5 Coarse-grained modeling of 50S *E. coli* ribosome 20

2.6 Steered molecular dynamics simulation 21

2.7 Umbrella sampling simulation 23

2.8 Entropy-enthalpy decomposition 25

2.9 Calculation of water tetrahedral order parameters 25

2.10 Calculation of fraction of native contacts, Q. 26

2.11 Estimating the folding time of slow-folding proteins with a large proportion of unfolded trajectories 26

2.12 Definition of the progress variable ς used to monitor the sequence of pairs of native secondary structure elements formed during the folding process 27

2.13 Identifying entanglement and the changes in entanglement 28

Chapter 3. Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling 31

3.1 Introduction 31

3.2 Publication 35

3.2.1 Abstract 35

3.2.2 Author contribution statements 36

3.2.3 Publication 42

Chapter 4. The Driving Force for Cotranslational Protein Folding is Weaker In the Ribosome Vestibule due to Greater Water Ordering 52

4.1 Introduction 52

4.2 Publication 54

4.2.1 Abstract 54

4.2.2 Author contribution statements 56

4.2.3 Publication 60

Chapter 5. Is Posttranslational Folding More Efficient Than Refolding from a Denatured State: A Computational Study 69

5.1 Introduction 69

5.2 Publication 71

5.2.1 Abstract 71

5.2.2 Author contribution statements 73

5.2.3 Publication 78

Chapter 6. Conclusions and future directions 94

6.1 Conclusions 94

6.2 Future directions 95

References.. 97

LIST OF FIGURES

Figure 1.1: Four distinct levels of protein structure organization. 1

Figure 1.2: The bacterial ribosome (70S, PDB ID: 4v9d[27]) consists of two subunits: the small subunit (SSU, 30S) and the large subunit (LSU, 50S). 5

Figure 1.3: Geometry and the electric potential of ribosome exit tunnel in prokaryotic and eukaryotic. 7

Figure 2.1: (a) An all-atom model and (b) C_α coarse-grained model of Dihydrofolate reductase (DHFR) proteins. 16

Figure 2.2: Surface representation of the large subunit (the 50S) of the *E. coli* ribosome and the simulated region containing the exit tunnel. 20

Figure 2.3: A truncated coarse-grained representation of the ribosome exit tunnel and surface used in all synthesis and ejection simulations 21

Figure 2.4: (Left) Schematic of SMD simulations of pulling protein from the ribosome exit tunnel. (Right) Force-displacement profile from SMD simulations. 22

Figure 2.5: Schematic of umbrella sampling to calculate the potential of mean force along the reaction coordinate. 24

Figure 2.6: Visualizing lasso-entanglement. 28

Figure 3.1: Coarse-grained simulations of nascent protein synthesis and ejection. 31

Figure 3.2: All-atom steered molecular dynamics simulation of pulling of a nascent protein (yellow) from the ribosome exit tunnel (cyan). 33

Figure 4.1: (a) Cross-section of the 50S subunit of *E.coli* highlighting the ribosome (gray), exit tunnel (black), and the last 3nm of the exit tunnel known as the ‘vestibule’ (red). (b) The portion of the ribosome exit tunnel used in the simulation. 53

Figure 5.1: Crystal structures of DHFR, CAT-III, and DDLB proteins with domain-based coloring 70

LIST OF TABLES

Table 1.1: Comparison of bacterial and eukaryotic ribosomes..... 6

ABBREVIATION

aa	Amino acid
ACE	Acetyl capping group
CAT-III	Type III chloramphenicol acetyltransferase protein
CI	Confident interval
CG	Coarse-grained
CTD	C-terminal domain
Da	Dalton (unit)
DDLB	D-alanine–D-alanine ligase B protein
DHFR	Dihydrofolate reductase protein
E. coli	*Escherichia coli*
FRET	Förster resonance energy transfer
LSU	The large subunit of the ribosome
MD	Molecular dynamics
mRNA	Messenger RNA
NTD	N-terminal domain
PDB	Protein Data Bank
PTC	Peptidyl-transferase center
RNA	Ribonucleic acid
RNC	Ribosome nascent protein complex

S	Svedberg (unit)
SecM	Secretion monitor protein
SSU	The small subunit of the ribosome
SMD	Steered molecular dynamics
tRNA	Transfer RNA
US	Umbrella sampling
UI	Umbrella integration
VMD	Visualize Molecular Dynamics (software)
WHAM	The weighted histogram analysis method

ACKNOWLEDGEMENTS

There are many people I need to thank for bringing the past five years. I am deeply grateful to my supervisors, Prof. Mai Suan Li and Prof. Edward P. O'Brien (Penn State University), for their outstanding guidance, support, and encouragement throughout my Ph.D. journey. They have been exemplary mentors who inspired me to pursue scientific excellence and challenged me to grow as a researcher and a person. They have taught me how to approach research problems and how to write high-quality research papers. They also treated me with kindness and care like family. I am honored to have worked with them and learned from their expertise and experience.

I would also like to thank Li's wife, Nguyen Thi Ngoc Tam for her support. She has been taking care of me like a mother when I was in Poland. She always made me feel welcome and comfortable in her home and helped me with many practical matters. She is a wonderful person who has shown me great kindness and generosity.

I would like to express my gratitude to Prof. Trinh Xuan Hoang (Vietnam Academy of Science and Technology) for his invaluable support in initiating my Ph.D. journey in Poland. He helped me to connect with my current supervisor. Without his assistance and encouragement, I would not have been able to pursue my academic goals in this field.

My collaborators from Penn State University, led by Ed O'Brien, have been instrumental in advancing my research and enriching my learning experience. I am especially grateful to Yang, Dan, Ian, Viraj, and Sarah for their valuable contributions, feedback, and expertise. It was a pleasure and an honor to work with such a talented and dedicated team.

I would like to express my sincere gratitude to the members of the theoretical department of the Institute of Physics of the Polish Academy of Sciences and Prof. Li's group members for their valuable friendship and support. They have enriched my life with enjoyable moments in the lab and outside. I am fortunate to have such wonderful colleagues and friends.

I would like to thank Prof. Anna Niedźwiecka, Prof. Bartosz Różycki, and Dr. Paweł Krupa for their valuable and insightful discussions throughout the past few years concerning my

research. They have always been approachable and supportive, providing me with guidance whenever I had questions.

I am grateful for my friends and would like to extend my thanks to each of them. Nguyen, your consistent support in providing me with an open door when I needed a break from my research is deeply appreciated. Your willingness to listen to my stories and share intriguing things with me has meant a lot. Thuan, Hoang, and Khuong, I am also grateful for your efforts in keeping me updated on what's happening in our hometown and for providing me with valuable assistance. Your help has been invaluable to me.

Most importantly, I owe a special debt of gratitude to my family for their unconditional love, support, and sacrifice. I thank my parents for giving me the opportunity to pursue my education and for always believing in me. I thank my brother Quang Vu and my sister Quynh Vu and Ngoc Uyen for their care and encouragement. I thank my niece Thanh Ngan for bringing joy and happiness to our family. I would like to honor the memory of my grandfather, my father-in-law, and my friend Minh Chang, who provided unwavering support throughout my Ph.D. journey, but sadly passed away before witnessing its completion. They remain in my heart and mind, and I believe they are proud of me from above.

Last but not least, I would like to express my deepest appreciation to my wife Thu Thao, who has been my best friend, partner, and soulmate throughout this journey. She has been a constant source of love, strength, and motivation. She has supported me in every possible way, emotionally, financially, and practically. She has endured many hardships and sacrifices for me to pursue my dream. She has also shared with me many moments of joy and success.

DEDICATION

This dissertation is dedicated to my parents, Ben Nguyen and Quan Vu, and my wife Thu Thao.

Chapter 1. INTRODUCTION

1.1 PROTEIN AND THE FOLDING PROBLEM

After billions of years of evolution, proteins have emerged as the most complex structures known to science. These remarkable macromolecules are comprised of only twenty canonical amino acids, each with distinct chemical properties. The canonical amino acids are further categorized into several groups based on the chemical characteristics of their side chains. These groups include positively charged amino acids (Arg, Lys, and His), negatively charged amino acids (Asp, Glu), uncharged polar amino acids (Asn, Gln, Ser, Thr, and Tyr), and nonpolar amino acids (Ala, Gly, Val, Leu, Ile, Pro, Phe, Met, Trp, and Cys). A polypeptide chain is

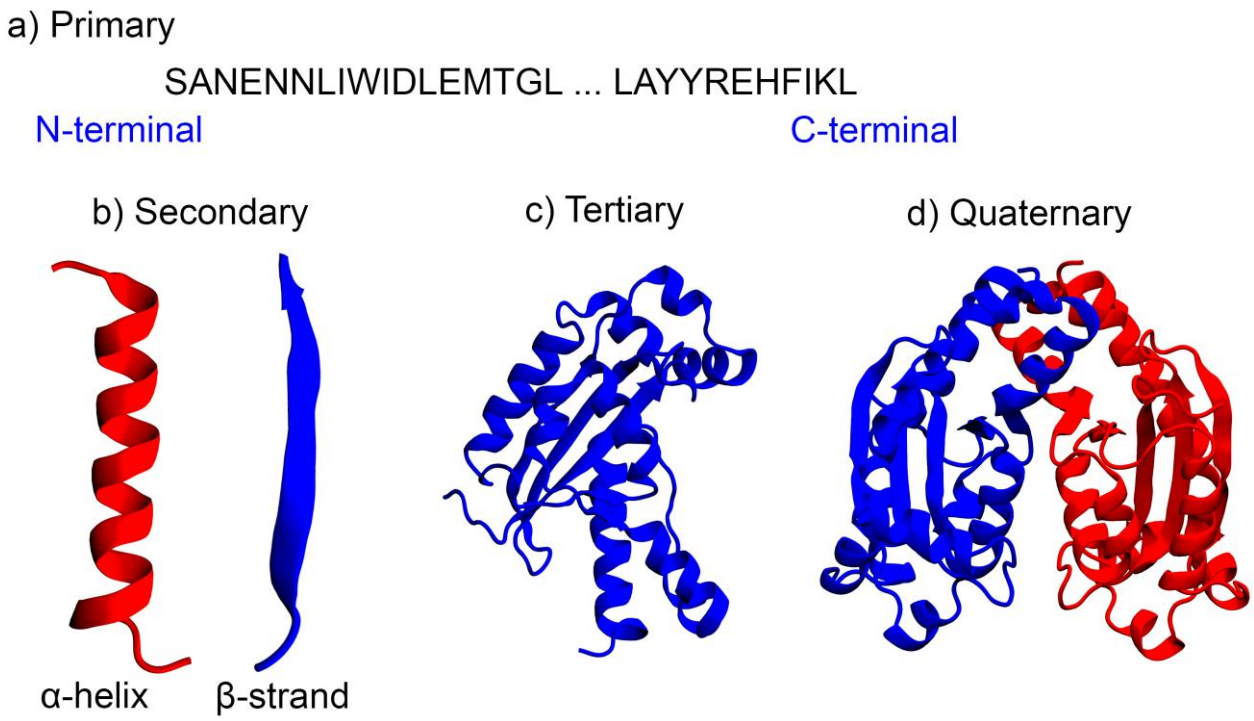


Figure 1.1: Four distinct levels of protein structure organization. a) Primary structure is a sequence of amino acids from N- to C-terminus. b) Secondary structure (two most common types of secondary structure: α-helix and β-strand are presented). c) Tertiary structure. d) Quaternary structure is formed by the complex of two monomers (blue and red). Panels c and d are generated from PDB ID: 1YTA.

formed by the covalent bond (peptide bond) between amino acids. The order of amino acids within the chain determines the structure of the protein[1] and ultimately dictates its function.

Proteins exhibit four levels of structural organization. The primary structure refers to the linear sequence of amino acids in a polypeptide chain from the N-terminus to the C-terminus (Figure 1.1a). The secondary structure of a protein is characterized by the local spatial arrangement of the polypeptide chain, which is stabilized by hydrogen bonds in the peptide backbone. The most common types of secondary structures are α-helix and β-strand (Figure 1.1b). The three-dimensional arrangement of a single polypeptide chain, as dictated by the interactions between its side chains, is referred to as the tertiary structure (Figure 1.1c). When a protein is composed of multiple polypeptide chains, the complete structure is designated as the quaternary structure (Figure 1.1d).

Proteins serve various functions in supporting life, including as building blocks for tissues and catalytic and signaling agents. To carry out their functions, many proteins must self-assemble into specific structures (known as native states), and misfolding can lead to disruptive diseases[2–4]. Protein folding, the process by which proteins attain their native state, has been extensively studied for over half a century. How do proteins fold? A famous thought experiment proposed in 1969 by Cyrus Levinthal[5] is that if the folding is the process of sampling all possible configurations, then for a simple protein of 100 amino acids where each can have three configurations, there would be a total of 3^{100} states to sample. Suppose the time to sample each configuration is about 10^{-15} s (timescale of bond rotation). In that case, it will take about 10^{25} years to sample all possible configurations, which is much longer than the universe's age (approximately 13 billion years). However, in reality, such small proteins can fold rapidly within microseconds[6]. This puzzle is known as the "Levinthal paradox". Nowadays, it is widely accepted that the protein folding kinetics can be described by the funnel theory proposed by Wolynes, Onuchic, and Dill[7,8]. According to the funnel picture[7,8], this process involves a downhill conformational search toward the native state, which has the global free energy minimum. The folding process is complex[5,9] and may expose the misfold[10]. Proteins that fail to fold into their native state are either aggregated or tagged for degradation. Since Anfinsen's experiment about 60 years ago[11], it has been observed that ribonuclease can spontaneously self-assemble into its native structure. This observation has been replicated with many other proteins as well. As a result, it is widely accepted that thermodynamics determines the native

conformation of a protein, and the native state of a protein is determined by its amino acid sequence[1]. However, efficient reversible folding and unfolding in solution is generally observed only for small proteins (up to 100 amino acids[12]), such as single-domain proteins, while multidomain proteins (account for 30–40% in prokaryotic and up to 75% in eukaryotic cells[13]) tend to misfold and form insoluble aggregates[14,15] (these larger proteins usually require external factors to fold). Fortunately, life has evolved various mechanisms to assist proteins fold[16]. *In vivo*, the folding process is assisted by other proteins or molecular machinery such as chaperones[17–19], and the ribosome[20–23]. The function of a protein depends on its structure, which is determined by its folding. Therefore, understanding the protein folding process is crucial and remains a significant area of research in biophysics and biochemistry.

1.2 HYDROPHOBIC EFFECT – THE DRIVING FORCE FOR PROTEIN FOLDING

The hydrophobic effect is a phenomenon that describes the tendency of nonpolar molecules to aggregate in a water solution. An example of this is that oil and water do not mix. Water is a unique solvent because of its polar nature. It boils at 373 K and freezes at 273 K; these temperatures are higher than other molecules with similar molecular weight. This suggests there are some strong bonding networks among water molecules. In fact, a water molecule has a partial negative charge on the oxygen atom and a partial positive charge on the hydrogen atoms. The polarization feature of water molecules allows them to form hydrogen bonds with other molecules with opposite charges, such as the oxygen atoms of other water molecules. Strong hydrogen network dominants the solvent properties of water.

On the other hand, a hydrophobic molecule cannot form hydrogen bonds with water because it has no charge or polarity. When a hydrophobic molecule is transferred into a water solvent, it disrupts the hydrogen-bonding network of water molecules. To minimize this disruption, water molecules form a cage-like structure around the hydrophobic molecule, isolating it from the rest of the solution. This process reduces the entropy of the water molecules at the interface region. However, it is favorable in terms of the system's free energy because it preserves the number of hydrogen bonds in the solution. When two or more hydrophobic molecules are present in water, they tend to aggregate together. This is because of clustering,

they reduce the surface area exposed to water and thus decrease the number of water molecules that need to form cages around them (release some freeze water molecules at the solvation shell to bulk). This increases the entropy of the water molecules in the solution and lowers the system's free energy. Therefore, hydrophobic aggregation is thermodynamically driven by entropy rather than enthalpy. The hydrophobic effect is considered the primary driving force for the folding of globular proteins. It results in the burial of the hydrophobic residues in the core of the protein (minimizing the loss of hydrogen bonds) and the hydrophilic residues at the surface (which can form hydrogen bonds with water). The hydrophobic effect also reduces the entropy loss of water molecules that would otherwise form ordered cages around the nonpolar groups.

The hydrophobic effect is not the only force involved in protein folding[24], as other interactions, such as hydrogen bonds, electrostatic interactions, disulfide bonds, and metal coordination, also play important roles. However, the hydrophobic effect is considered the dominant factor (contributes around 60% of protein stability[25,26]) that guides protein folding and provides thermodynamic stability to proteins.

1.3 RIBOSOME

Ribosomes were first discovered by George E. Palade in 1955 using an electron microscope, for which he shared the Nobel Prize in Physiology and Medicine in 1974. The detailed structure and mechanism of ribosomes were later revealed by the experimental work of Venkatraman Ramakrishnan, Thomas Steitz, and Ada Yonath, who jointly won the Nobel Prize in Chemistry in 2009. For the story behind the discovery of the ribosome's structure and how science happens, I recommend that readers check out the book "Gene Machine: The Race to Decipher the Secrets of the Ribosome" by Venkatraman Ramakrishnan. We know ribosomes are complex and essential molecular machines in all cells responsible for protein synthesis from messenger RNA molecules. They comprise ribosomal RNA (rRNA) and more than 50 different ribosomal proteins. They function to translate the genetic code in messenger RNA (mRNA) into a specific order of amino acids, which then form functional proteins. Ribosome ensures that the sequences are built in the correct order.

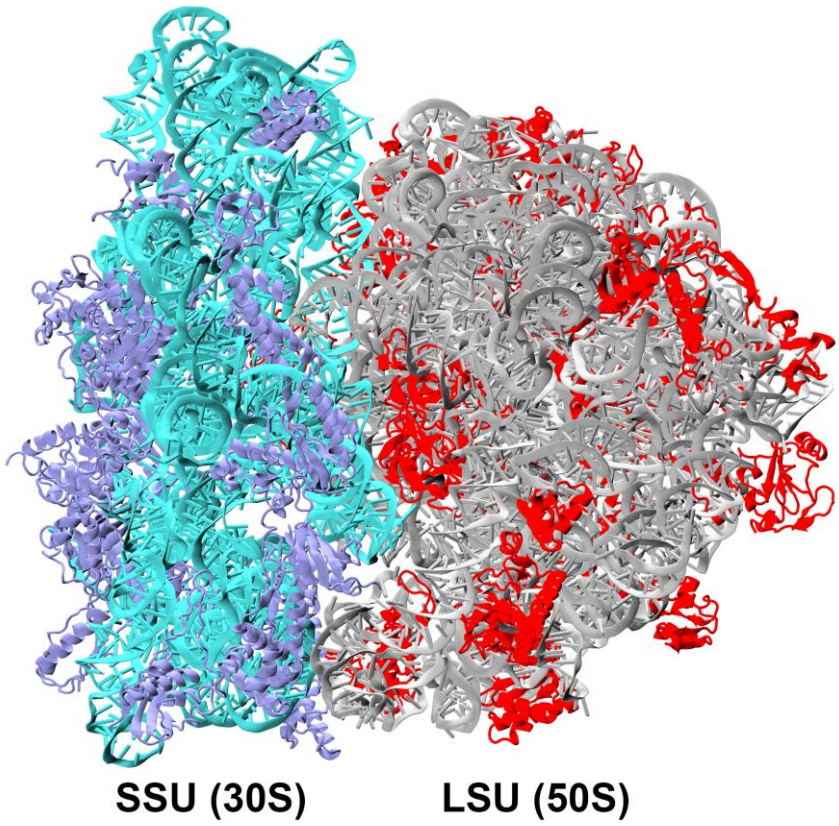


Figure 1.2: The bacterial ribosome (70S, PDB ID: 4v9d[27]) consists of two subunits: the small subunit (SSU, 30S) and the large subunit (LSU, 50S). The SSU and LSU comprise ribosomal RNA (rRNA) and ribosomal proteins. The rRNA and proteins of the SSU are colored in cyan and ice-blue, respectively, while those of the LSU are colored in silver and red, respectively.

Ribosomes consist of two subunits (Figure 1.2): a large subunit (LSU) and a small subunit (SSU). Each subunit contains one or more rRNA molecules and many ribosomal proteins (Table 1.1). These subunits work together: the small subunit provides a framework for tRNA, binds to mRNA, and decodes the genetic code it carries, while the large subunit catalyzes the formation of peptide bonds between the amino acids in the growing polypeptide chain. The peptide bond formation is catalyzed by the peptidyl-transferase center (PTC), and the emerging nascent proteins exit the ribosome through the exit tunnel located in the LSU. The size and composition of ribosomes differ between different organisms[28]. Bacteria and other prokaryotes have smaller ribosomes called 70S ribosomes, which comprise a small subunit (30S) and a large subunit (50S). Animals and other eukaryotes have larger ribosomes called 80S ribosomes, which

consist of a small subunit (40S) and a large subunit (60S). The Archaeal ribosomes are similar to the bacteria ribosome in general dimensions (70S ribosome).

Table 1.1: Comparison of bacterial and eukaryotic ribosomes.

Organism	Ribosome	Subunit	Component	
			Ribosomal RNAs	Number of ribosomal proteins
Bacteria	70S (weight: ~ 2.5 MDa)	50S	23S and 5S	31
		30S	16S	21
Eukaryote	80S (weight: ~ 4.2 MDa)	60S	28S, 5.8S and 5S	49
		40S	18S	33

Understanding how ribosomes work is crucial for elucidating the molecular mechanisms of gene expression, protein folding, cellular regulation, and evolution. Moreover, ribosomes are essential targets for many antibiotics that inhibit bacterial protein synthesis and treat infections[29,30].

1.4 RIBOSOME EXIT TUNNEL

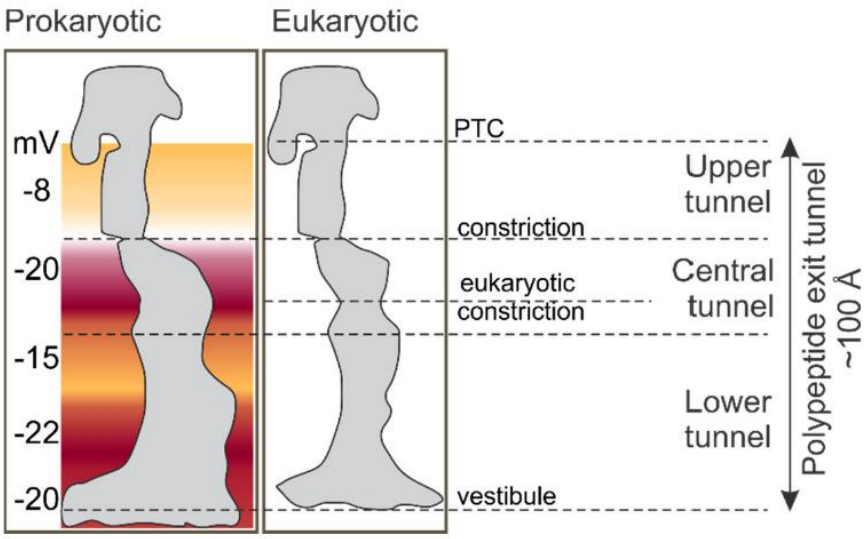


Figure 1.3: Geometry and the electric potential of ribosome exit tunnel in prokaryotic and eukaryotic. Figure adapted from Liutkute *et al.*[31]

The ribosome exit tunnel is located in the large subunit of the ribosome and spans from the peptidyl transferase center (PTC) to the outer ribosome surface. The shape of the exit tunnel is about 100 Å – 120 Å in length (depending on where the open end is defined) and varies between 10 and 20 Å in diameter[32–34], providing a confined space where the nascent chain begins to fold. Residues lining the exit tunnel are highly conserved in the zone proximal to the PTC[35]. The exit tunnel is not straight, it is bent and has a constriction site at ~30 Å from PTC in prokaryotic cells and an additional constriction site formed by ribosomal protein uL4 in the eukaryotic ribosome (Figure 1.3). The final 20 Å of the tunnel is known as the vestibule and is wider than the rest of the tunnel. The vestibule region of the bacterial tunnel is wider than the eukaryotic tunnel, composed of ribosomal proteins uL23 and uL24 in bacteria and an additional ribosomal protein eL39 in eukaryotes[35]. The exit tunnel can accumulate a segment of ~30 amino acids in an extended conformation and a domain with the size of about 60 amino acids in the helix conformation[19,22,36].

The ribosome exit tunnel is primarily composed of RNA (23S in bacteria and 28S in eukaryotes), a highly charged density biomolecule, creating a distinct electrostatic environment[37]. On average, the tunnel exhibits a more negative charge and is quite heterogeneous than the cellular matrix[37]. The geometry and composition of the tunnel potentially impact the translation dynamics[38,39] and protein folding[40–43]. Notably, the electrostatic nature of the exit tunnel allows it to interact with various protein sequences, leading to stalled translation, such as those seen with tnaC and SecM sequences[44–46]. Translation can only resume if sufficient force is applied to dislodge these sequences[47,48]. This phenomenon has been utilized *in vitro* to track the location of protein folding on the ribosome[42,49]. As proteins fold, they generate an entropic force transmitted back to the PTC site via the protein backbone[50]. In addition, Lucent *et al.* utilized molecular dynamics simulations and demonstrated that the ribosome exit tunnel exhibits increased ordering and reduces the rotational entropy of water[51]. This makes the exit tunnel to be a unique environment, and the chemical heterogeneity of the exit tunnel is vital to regulate downstream processes such as the protein elongation[52], potentially impacting the translation[39,53], and regulating the early event of protein folding such as protein ejection and cotranslational folding.

1.5 PROTEIN FOLDING ON THE RIBOSOME

Proteins can acquire tertiary structure at any stage: during their biosynthesis, as they are ejected through the ribosome's exit tunnel, or posttranslationally – after their release from the ribosome. Cotranslational folding, the concomitant acquisition of stable tertiary structure by nascent protein segments during protein synthesis, occurs both *in vitro* and *in vivo*. This process is critical in ensuring proteins' proper folding and function in cells. During translation, the nascent protein first passes through a tunnel within the large ribosomal subunit before emerging at the solvent side. Indeed, many proteins fold cotranslationally[54–61] as they begin to emerge from the exit tunnel and acquire tertiary structure before their synthesis is complete.

It is important to note that the rate of protein synthesis is typically slower than the folding of small-domain proteins, with small single-domain globular proteins able to fold spontaneously within microseconds to hours[6]. In bacteria, for example, the rate of protein synthesis is

approximately 15-20 amino acids per second[62], taking roughly 5 seconds to synthesize a small single protein domain of 100 residues. Cotranslational folding may be even more critical in eukaryotes[20], where the translation rate is slower, ranging from 3-4 amino acids per second[62], and the average size of proteins is larger, with the median protein length in eukaryotic cells being 361 residues[63]. While small proteins may have sufficient time to adopt preferred conformations or even fold to the native state in the ribosome exit tunnel or surface during synthesis[49], larger proteins may not attain their stable native conformation cotranslationally. However, they can still achieve some collapsed states and only fold posttranslationally once they have left the tunnel[21,64–67]. It has been estimated that one-third of *E. coli* proteins fold cotranslationally[68].

The interactions between ribosomes and nascent proteins can perturb the folding process in terms of proteins' folding kinetics and self-assembly pathways. Recent experimental findings have highlighted the importance of protein synthesis and cotranslational folding, indicating that one-third of *E. coli* proteins cannot refold in bulk solution after being completely unfolded by denaturants[69]. This suggests that cotranslational folding is critical to their ability to reach their native state[70].

Overall, the ribosome is not only responsible for protein synthesis but also plays an essential role in protein folding. Cotranslational folding represents a vital aspect of the intricate and tightly regulated process by which cells produce functional proteins. This process has evolved as a means for the cell to maintain proteostasis by mitigating the risk of misfolding and aggregation. Therefore, understanding the mechanisms and dynamics of cotranslational folding, which involves protein folding under the influence of ribosomes, is essential for gaining insights into the folding and function of proteins and developing new strategies to prevent misfolding protein diseases.

1.5.1 *Some proteins fold in the exit tunnel*

The ribosomal exit tunnel plays a crucial role in the folding of proteins during their biosynthesis[40,71]. This narrow channel extends from the peptidyl transferase center (PTC) of the ribosome to its outer surface, with a width that ranges from 10 Å at the constriction site to 20 Å in the vestibule. The exit tunnel restricts the ability of proteins to self-interact and form tertiary

structures. However, it has been found that a simple structure motif like α-helix is allowed to form in the upper tunnel.

Computational studies using a simple cylinder geometry to model the ribosomal exit tunnel have shown that a small helix can form in the upper region of the tunnel, approximately 20-30 Å from the PTC, and is stabilized entropically by the ribosome[72]. This conclusion is supported by other studies that visualize nascent proteins using cryo-EM, showing that α-helices can form in both the upper and lower regions of the tunnel[42,73]. Furthermore, fluorescence resonance energy transfer (FRET) studies have shown that transmembrane segments can also form α-helices within the exit tunnel in the proximity of the PTC[74].

Recent studies have shown that the ribosomal exit tunnel can also support the folding of larger domains at the vestibule region, located approximately 80 Å away from the PTC at the end of the tunnel. This region is wider than the rest of the tunnel, with a diameter of about 20 Å, allowing many domains to fold. Computational simulations have predicted that an 80-residue protein domain can fold in the ribosome vestibule[36], and subsequent studies using arrested peptide essays combined with molecular dynamics simulations have verified that an entire ADR1a Zinc-Finger domain can fold into a native structure deep inside the exit tunnel[42]. Similarly, the SecM arrest peptide has been used as a force sensor to probe the cotranslational folding of nine small protein domains (<70 residues) of various topologies, including α-helices or β-sheets. The study has shown that these domains can fold in the first 80 Å of the exit tunnel, indicating that these protein domains initiate folding while still inside the exit tunnel[49].

Larger or multi-domain proteins can only begin to fold once they have left the exit tunnel[64–67]. It is because the space available to the nascent protein abruptly expands once the protein reaches the ribosome's surface. For example, the N-terminal domain of HemK can form a compact, intermediate state deep inside the tunnel, but the native fold is attained only upon leaving the ribosome[75].

1.5.2 *Ribosome destabilizes folded domains*

Nascent proteins can acquire secondary and some limited tertiary structures before emerging from the ribosome exit tunnel[42,49,76,77]. These early structures might be essential

elements in forming the native state. Several experiments have indicated that the folded domains in the presence of ribosomes are less stable than those without ribosomes. For example, Samelson *et al.* employed pulse proteolysis to determine the thermodynamic stability of DHFR, RNase H, and Barnase proteins tethered to the ribosome at various linker lengths and compare them to the stability of the isolated protein[78]. They found that the ribosome destabilizes the compact form of proteins, resulting in a destabilizing effect of up to 2 kcal/mol on the polypeptide chain. This destabilization decreases as the distance from the peptidyl transferase center increases. Another study used a single-molecule optical tweezer to investigate the folding of a five-domain elongation factor G (EF-G) protein. The results showed that domain III of EF-G still unfolded even though it had emerged from the ribosome exit tunnel[79].

Thus, the ribosome may contribute an extra layer to regulate the protein folding process by preventing the formation of partially folded states until the protein has fully emerged from the ribosome.

1.5.3 *The folding kinetics of proteins are slower on the ribosome*

In terms of kinetics, laser optical tweezer experiments measure the folding rate of protein bound to ribosome showing that protein folds slower on the ribosome at various linker lengths compared to folding in bulk solution. A pioneer work by Kaiser *et al.* used single-molecule experiments on an arrested ribosome have revealed that due to the interaction with the ribosome surface, T4-lysozyme's folding rate is significantly slower near the ribosome surface, even after it has emerged from the ribosome exit tunnel, as compared to when it folds in free solution. By extending the linker length between protein and PTC, the folding rate approaches its bulk value[80]. Increasing the salt concentration increases the protein folding rate on stalled ribosomes. However, they did not observe that salt concentration affects the folding rate of free T4 lysozyme, suggesting that the electrostatic interactions between the nascent protein and the negatively charged ribosome surface are responsible for this deceleration in the folding rate[80]. Also, using an optical tweezer, Liu *et al.* showed that the ribosome modulates the apparent folding rate of elongation factor G[81]. The authors found that in the ribosome-nascent chain complex, there is an optimal value of linker length at which the apparent folding rate equals

folding in bulk solution. Below this value, the disordered nascent polypeptide interacts with the ribosome, effectively slowing its folding rate. At lengths beyond this optimal length, additional emerged portions of the neighbor domain become available to interact with the G-domain and also disfavor folding.

Ribosomes can also delay the formation of cotranslational intermediates at the emerging N-terminus of the multidomain calcium-binding protein, disfavor the formation of misfolded intermediates, and increase the rate of their unfolding to maintain a folding-competent nascent polypeptide[82]. Delaying the compaction of nascent chains could ensure that folding into stable conformations does not occur before the entire sequence is fully accessible, thus promoting the correct folding of the nascent protein.

1.5.4 *Folding pathways of proteins on and off the ribosome*

Interactions between the ribosome and the protein complicate the protein folding problem. This suggests that ribosomes actively anticipate the protein folding process. Consequently, whether protein pathways are conserved on and off the ribosome is unclear. The main question remains: to what extent does the ribosome help proteins fold? As mentioned above, the ribosome can destabilize nascent protein folds and delay folding until the entire domain is exposed; thus, ribosomes can alter protein folding pathways. Various studies showed that ribosomes assist proteins in folding efficiently. For example, O'Brien *et al.* utilized a coarse-grained model to simulate the cotranslational folding of protein G on an arrested ribosome. They found that the dominant folding pathways changed on the ribosome and that the number of unique pathways decreased by 28% on the ribosome[36]. Tanaka et al. used coarse-grained molecular simulation to study the role of the ribosome in guiding SufI multi-domain protein folding, finding that folding on the ribosome is more efficient than refolding[83]. Dabrowski-Tumanski *et al.* computationally studied a deeply knotted protein and found that the ribosome plays a crucial role in knot formation[84].

On the other hand, several other studies found that the folding pathways on and off the ribosome are robust. For example, structure-based models in combination with an arrest-peptide assay and cryo-EM experiments indicate that the folding of titin I27 is conserved on and off

ribosome[76]. Similarly, experiments and molecular simulations of src SH3 show that its folding pathways are the same on and off ribosome[85]. Given the relative paucity of experimental and computational data on the differences between folding on and off the ribosome for large proteins, we believe the influence of the ribosome on protein folding mechanisms remains an open question.

1.6 THESIS OBJECTIVES

The main goal of this thesis is to investigate the influences of ribosomes on proteins at their early stages of existence. This thesis contributes to understanding the interaction between the ribosome and the nascent protein in several ways. We performed a multiscale study of how the electrostatic interactions affect the protein ejection from the ribosome after the protein synthesis, which has not been explored before. Understanding the molecular mechanisms that weaken the hydrophobic interaction, a driving force for protein folding, in the ribosomal vestibule. This explains the experimental observation that the folded domain is less stable and the folding kinetic is slower on the ribosome than the bulk solution. To explore the role of protein synthesis and posttranslational folding on protein folding and compare it to folding in bulk. Finally, the main findings of this work are summarized, and some unresolved questions and directions for future research are highlighted.

Chapter 2. COMPUTATIONAL BACKGROUND

2.1 MOLECULAR DYNAMICS SIMULATION

Molecular dynamics (MD) simulation is a powerful computational tool to study the behavior of the interacting system over time. Depending on the research question and the available computational resources, different models (levels of detail) can characterize the system of interest and their environment. In MD simulations, forces are calculated at every step to integrate the equations of motion, allowing us to observe the system's evolution over time. By recording the positions and velocities of the system, we obtain the phase space, which allows us to calculate physical properties. By simulating the motion and interactions of individual particles, MD allows scientists to investigate the properties of systems at the atomic and molecular scale, including the behavior of biological molecules such as proteins and nucleic acids. MD has become an essential tool in many fields, including chemistry, physics, materials science, and biophysics, and has played a crucial role in advancing our understanding of the behavior of matter at the atomic level. MD provides an interface between the theory and experiment and sometimes is a so-called *in-silico* experiment.

The primary justification of the MD method is based on the ergodicity hypothesis: ensemble averages are equal to the time averages of the system taken over a long time interval (eq. (2.1)). To my knowledge, this assumption has not been proven yet. Hence, by performing MD simulation for a sufficiently long timescale, any physical properties of the system can be obtained via the time average from the simulated trajectory, and we can conclude the ensemble properties.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(t) dt \tag{2.1}$$

In eq. (2.1), notations $\langle A \rangle_{ensemble}$, $\langle A \rangle_{time}$ are an ensemble and time average, A is any macroscopic quantity of the system.

To perform MD simulations, we need software to numerically solve the equations of motion and a force field that defines how particles in the system interact. For the former

component, it is pretty convenient nowadays that many software packages have been designed to perform these tasks efficiently, including open-source (free software) and commercial software, such as GROMACS[86], AMBER[87], CHARMM[88], and OpenMM[89], etc. As for the latter component, the force field is a set of parameters and equations that define how particles interact, including the strengths and types of interactions such as bonds, angles, torsions, non-bonded interactions (van der Waals forces, and electrostatic interactions), and possibly other terms. The choice of force field depends on the desired model resolution. Two popular models used in the biophysics community are coarse-grained and all-atom models (Figure 2.1). These models have been extensively used in this dissertation to study various problems on the ribosome. Coarse-grained models simplify the molecular structure by grouping several atoms into larger units called beads and have an effective mass, charge, and other properties represented for a group of atoms. Coarse-grained models reduce the degrees of freedom and allow longer time scales and larger system sizes to be simulated[90]. However, they also lose some information about atomic details. All-atom models explicitly represent every atom in the system, with realistic masses, charges, and interaction potentials. All-atom models accurately describe the molecular structure and dynamics and can capture subtle effects such as hydrogen bonding or atomic conformational changes. However, they also require more computational power and memory and limit the time scales and system sizes that can be simulated. Both coarse-grained and all-atom models have advantages and disadvantages depending on the application. Therefore, choosing an appropriate model that balances accuracy and efficiency for a given problem is essential. Sometimes, hybrid models that combine coarse-grained and all-atom representations can also be used to achieve a multiscale simulation[91]. In the following sections, these models will be briefly described.

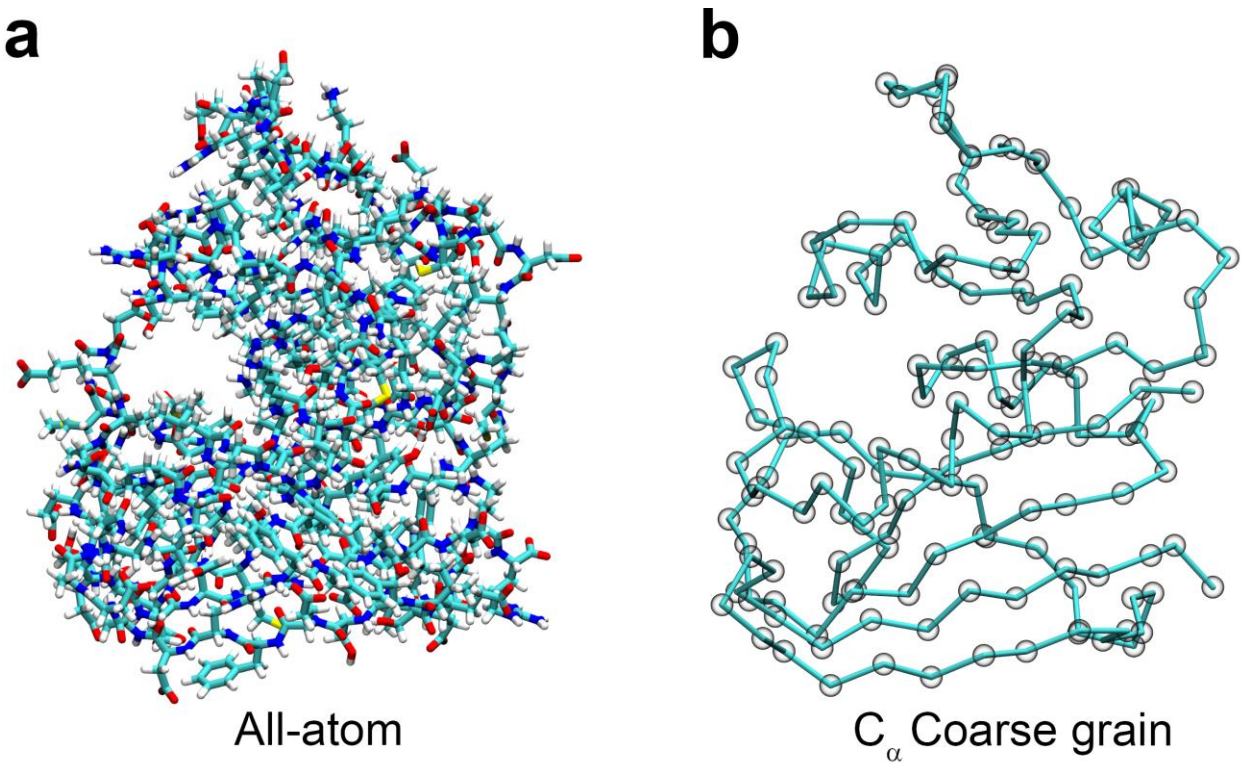


Figure 2.1: (a) An all-atom model and (b) C_α coarse-grained model of Dihydrofolate reductase (DHFR) proteins.

2.2 ALL-ATOM MODELING

The all-atom model explicitly represents the atomic nuclei, including solvent and ions (Figure 2.1a), and employs an empirical potential energy function, commonly known as a "force field" to model the system. Many different all-atom force fields have been developed to study biomolecules, and the most commonly used included AMBER[92–98], CHARMM[99–103], GROMOS[104,105], OPLS[106], etc. Different force fields may have different levels of accuracy and applicability depending on the system being studied. Some force fields are specifically designed for specific molecules or materials, while others aim for broader coverage. In this dissertation, we used AMBER99SB[107] to model the ribosome and protein.

The functional form of AMBER99SB force field:

$$E = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{n=1}^{3} V_n[1 + cos(n\omega - \gamma_n)] + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.2)$$

The first two terms of the eq. (2.2) describe the bonded potential between two and three particles, which are modeled using harmonic functions with force constants k_b, k_θ and equilibrium values of r_0, θ_0, respectively. The third term represents the dihedral potential between four points, where V_n is the dihedral force constant, n is dihedral periodicity and γ_n is a phase of the dihedral angle. The final term describes the non-bonded potentials, including the van der Waals interaction represented by the Lennard-Jones 6-12 function and the electrostatic interactions modeled by Coulombic interactions.

2.3 COARSE-GRAINED MODELING

A coarse-grained model is a simplified representation of a complex system that aims to capture its essential features while discarding irrelevant details[90] (Figure 2.1b). The idea behind coarse-graining is to reduce the degrees of freedom in a system by grouping atoms or molecules into larger units, such as beads or segments. This simplification allows for computationally feasible simulations and can provide insight into the system's behavior over longer timescales than an all-atom simulation.

Coarse-grained models are typically parameterized to reproduce experimental data or data from more detailed simulations. They can study a wide range of phenomena, such as protein folding[108,109], membrane structure[110], and phase separation of biomolecules[111–114], etc. While coarse-grained models are inherently less accurate than more detailed models, they can provide a valuable and efficient tool for understanding complex systems and designing new materials with desired properties.

In our structural-based coarse-grained model, each residue is represented by one interaction site centered on the C_α atom[115–117]. The potential energy for a given configuration of the C_α coarse-grained model is calculated using the following equation:

$$
\begin{aligned}
E = & \sum_i k_\mathrm{b}(r_i - r_0)^2 \\
& + \sum_i -\frac{1}{\gamma}\ln\left\{\exp[-\gamma(k_\alpha(\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha)] + \exp\left[-\gamma k_\beta\left(\theta_i - \theta_\beta\right)^2\right]\right\} \\
& + \sum_i \sum_j^4 k_{\mathrm{D}_j}\left(1 + \cos\left[j\varphi_i - \delta_j\right]\right) + \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0\varepsilon_\mathrm{r} r_{ij}} \exp\left[-\frac{r_{ij}}{l_\mathrm{D}}\right] \\
& + \sum_{ij\,\in\,\{\mathrm{NC}\}} \epsilon_{ij}^{\mathrm{NC}}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{ij\notin\{\mathrm{NC}\}} \epsilon_{ij}^{\mathrm{NN}}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]
\end{aligned}
\tag{2.3}
$$

The equation (2.3) presented here describes the summation of potential energy contributions from various interactions. These include the contributions from Cα – Cα bonds, bond angles, dihedral angles, electrostatic interactions, Lennard-Jones-like native interactions, and repulsive non-native interactions. Specifically, the bond potential between two adjacent interaction sites is modeled by a harmonic potential with a bond force constant $k_b = 50\frac{kcal}{mol\times Å^2}$, an equilibrium bond length r_0 of 3.81 Å, and a pseudo bond length r_i for the i^{th} bond. The angle potential is modeled by a double-well potential, which describes bond angles associated with both α-helix and β-sheet conformations[118]. Constants of the double-well angle potential include $\gamma = 0.1\frac{\mathrm{mol}}{\mathrm{kcal}}, k_\alpha = 106.4\frac{\mathrm{kcal}}{\mathrm{mol\times rad^2}}, \theta_\alpha = 1.6\ \mathrm{rad}, \varepsilon_\alpha = 4.3\frac{\mathrm{kcal}}{\mathrm{mol}}, k_\beta = 26.3\frac{\mathrm{kcal}}{\mathrm{mol\times rad^2}}, \theta_\beta =$ 2.27 rad. k_{D_j} and δ_j are the dihedral force constant and the phase at periodicity j, respectively. φ_i is the i^{th} pseudo dihedral angle. Electrostatics are treated using the Debye-Hückel theory with a Debye length l_D of 10 Å and a dielectric constant of 78.5. Lysine and arginine C_α sites are assigned $q = +e$, glutamic acid and aspartic acid are assigned $q = -e$, and all other interaction sites are uncharged[119]. The contribution from native interactions is computed using the 12-10-6

potential of Karanicolas and Brooks[120], with the depth of the energy minimum for a native contact $\epsilon_{ij}^{\mathrm{NC}} = n_{ij}\epsilon_{\mathrm{HB}} + \eta\epsilon_{ij}$, where ϵ_{HB} and ϵ_{ij} represent energetic contributions arising from hydrogen bonding and van der Waals contacts between residues i and j identified from the crystal structure of the protein, respectively. n_{ij} is the number of hydrogen bonds formed between residues i and j and $\epsilon_{\mathrm{HB}} = 0.75\,\frac{\mathrm{kcal}}{\mathrm{mol}}$. The value of ϵ_{ij} is set based on the Betancourt-Thirumalai pairwise potential[121], while the scaling factor η is determined for each protein based on a previously published training set[122] to reproduce realistic protein stabilities for different structural classes. Collision diameters σ_{ij} between Cα interaction sites involved in native contacts are set equal to the distance between the Cα of the corresponding residues in the crystal structure divided by $2^{\frac{1}{6}}$. For non-native interactions (last term), $\varepsilon_{ij}^{\mathrm{NN}} = 1.32 \times 10^{-4}\,\frac{\mathrm{kcal}}{\mathrm{mol}}$, and σ_{ij} is set to the average of the radii of the residues involved[120]. NC and NN stand for native contact and non-native contact respectively.

2.4 ALL-ATOM MODELING OF 50S *E. COLI* RIBOSOME

The large subunit of the ribosome is a complex structure consisting of several megadaltons of RNA and protein. Due to the computational cost and time required for an all-atom simulation of the entire subunit, we focused on simulating the structure around the ribosome exit tunnel (red region in Figure 2.2), which is the primary focus of this thesis. This approach involves cropping the subunit to reduce computational cost while preserving the physical properties of the ribosome exit tunnel.

To achieve this, we aligned the 50S subunit of the *E. coli* ribosome (PDB ID: 3R8T) with the long axis of the exit tunnel, which is defined as the vector between peptidyl transferase center (atom N6 of nucleotide A2602, blue sphere in Figure 2.2) and the Cβ atom of Ala50 in ribosomal protein L24, which protrude to the open end of the tunnel, along the x-axis of the simulation coordinate system. Subsequently, we cropped the ribosome to form a rectangular box (white color in Figure 2.2) around the exit tunnel such that the minimum distance along the y- and z-axis between the tunnel wall and the removed part is about 3 nm.

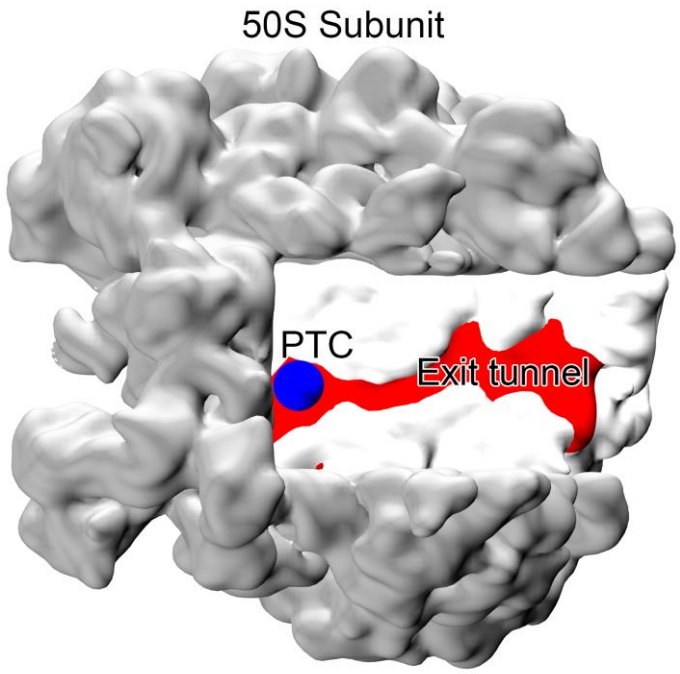


Figure 2.2: Surface representation of the large subunit (the 50S) of the *E. coli* ribosome and the simulated region containing the exit tunnel. The large subunit of the *E. coli* ribosome (PDB ID: 3R8T) is shown in gray, with the simulated region inside the black rectangle. The exit tunnel (red) is where nascent proteins are transported, and the peptidyl transferase center (PTC) site is highlighted as a blue sphere.

2.5 COARSE-GRAINED MODELING OF 50S *E. COLI* RIBOSOME

The structure of the 50S ribosome contained in PDB ID: 3R8T was reduced to a cutout of the exit tunnel and surface near the exit tunnel opening (Figure 2.3). The entire 50S structure was initially subjected to coarse-graining, utilizing a three/four-point RNA model and the protein's Cα model. In this model, nucleotides containing pyrimidines and purines were represented by 3 and 4 interaction sites[119], respectively. These interaction sites were characterized by a negative charge of $q = -1e$ located at the phosphate position, one at the centroid of the ribose ring, and one at the centroid of each conjugated ring in the base. The origin of the simulation coordinate system $(0, 0, 0)$ is placed at the position of the N6 atom of A2602, and the positive x-axis points from this origin towards the exit tunnel opening. The positive x-axis, therefore, lies along the long axis of the ribosome exit tunnel. Only ribosome interaction sites within 30 Å of the nascent chain or with an x-coordinate greater than 60 Å were retained for computational efficiency.

Additionally, residues with an x-coordinate greater than 60 Å but with zero solvent accessible surface area were removed, utilizing the COOR SURF functionality of CHARMM with RPROBE = 1.8 Å. This probe size was selected to be smaller than the smallest nascent chain interaction site, thereby removing only those ribosome sites that cannot interact with the nascent chain. Furthermore, an 18-residue loop of ribosomal protein L24, extending over the exit tunnel opening (blue color in Figure 2.3), was allowed to fluctuate in the model, and the rest of the truncated ribosomal atoms were made rigid.

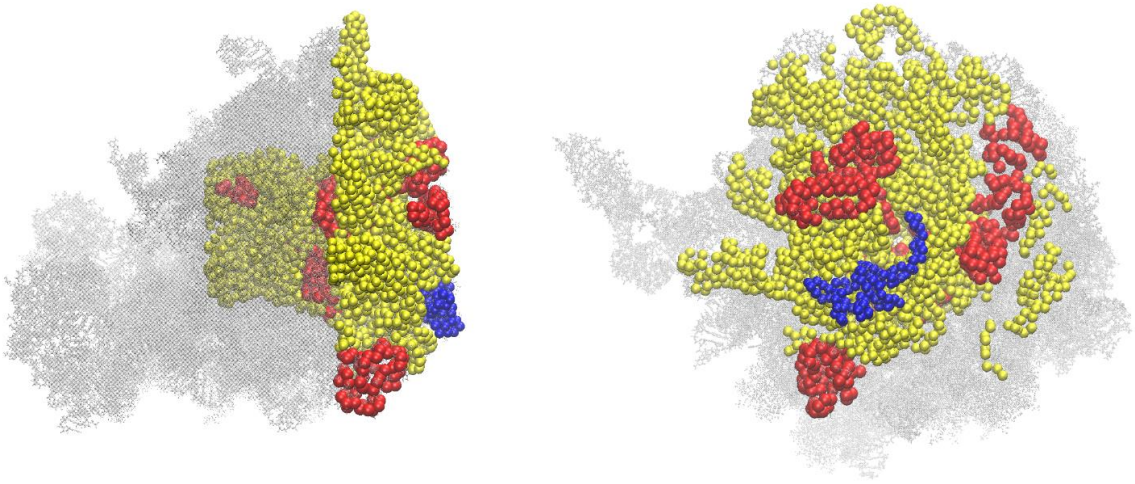


Figure 2.3: A truncated coarse-grained representation of the ribosome exit tunnel and surface used in all synthesis and ejection simulations (viewed from the side and top-down perspectives). The model was superimposed onto the entire 50S ribosome subunit (PDB ID: 3R8T) in gray. Ribosomal RNA, ribosomal proteins (excluding L24), and the L24 protein are colored yellow, red, and blue, respectively.

2.6 STEERED MOLECULAR DYNAMICS SIMULATION

Steered molecular dynamics simulation (SMD), first proposed by Grubmuller and coworkers[123], is a powerful technique in computational biophysics that allows us to study how biomolecules respond to external forces. SMD mimics single-molecule force spectroscopy experiments, such as atomic force microscopy[124] (AFM), laser optical tweezers[125], and magnetic tweezers[126]. SMD can reveal necessary information about processes such as protein unfolding[127,128], ligand binding[123], and conformational changes[127].

In SMD simulations, an external force is applied to a dummy atom along the pulling direction. The dummy atom is attached to a part of the system of interest (usually called the 'SMD atom') by a virtual spring with a constant k. There are two popular methods of SMD simulation: applying a constant external force or applying an external force to pull the dummy atom at a constant velocity $\vec{v}$.

In constant force SMD simulations, a constant external force is applied to a specific atom or region of interest within the biomolecular system. The applied force can be achieved by directly applying a force to the atom or employing a virtual spring. As the constant force is applied, the system responds by undergoing conformational changes, stretching, or unfolding.

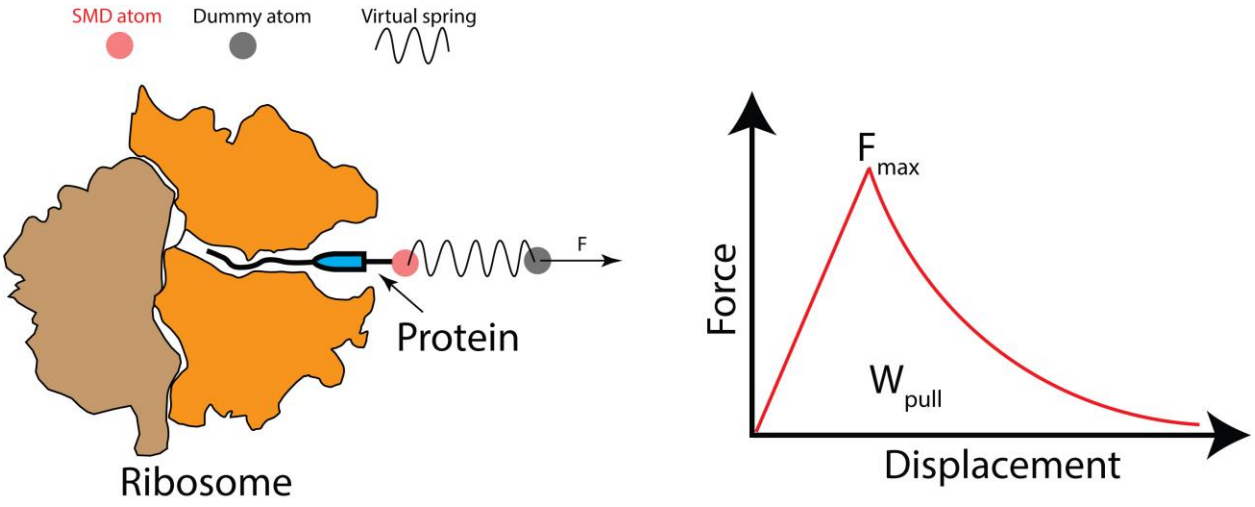


Figure 2.4: (Left) Schematic of SMD simulations of pulling protein from the ribosome exit tunnel. (Right) Force-displacement profile from SMD simulations.

Constant velocity SMD simulations, on the other hand, involve applying a force to a dummy atom that is connected to the region of interest. The applied force drags the dummy atom moving with a constant speed along the pulling direction. The force experienced by the system between the dummy atom and the SMD atom is measured by:

$$\vec{F} = -\nabla U \qquad (2.4)$$

$$U = \frac{1}{2}k[vt - (\vec{r} - \vec{r_0})\vec{n}]^2$$

where, $k, v, t, \vec{r}, \overrightarrow{r_0}$ and $\vec{n}$ are spring constant, pulling velocity, time, the actual position of the SMD atom, the initial position of the SMD atom, and the pulling direction.

By recording the position and force experienced by the SMD atom over time, we can obtain valuable data, such as force-displacement profiles or force-time profiles, that can be used to characterize the mechanical stability of the system. From these profiles, we can obtain quantitative information about the response of the biomolecule to the applied force, including the strength of interatomic interactions indicated by the rupture forces (the maximum force in the force-displacement/time profile) and the work applied to the system by the external force[129]. The pulling work can be calculated from the force-displacement profile as follows:

$$W_{pull} = \int F dx = \sum_{i=0}^{N-1} \left(\frac{F_i + F_{i+1}}{2} \right) \times (x_{i+1} - x_i) \tag{2.5}$$

Here, N is the number of frames, F_i and x_i refer to the pulling force and position of the SMD atom at frame i.

These techniques have proven to be powerful tools for studying the mechanical stability, unfolding pathways, and the response of biomolecules under controlled external forces, aiding in designing novel therapeutic strategies and elucidating fundamental biological processes.

2.7 UMBRELLA SAMPLING SIMULATION

Umbrella sampling is a method used to calculate the potential of mean force along the predefined reaction coordinate ξ. The umbrella sampling method was introduced by Torrie and Valleau in 1977 to improve sampling efficiency[130] and become a widely adapted method for enhanced sampling in biomolecular research. The primary goal of umbrella sampling is to overcome the limitations of conventional molecular dynamics simulations, which often limit simulation time and struggle to explore rare events due to the high energy barriers.

The general concept of umbrella sampling involves dividing the reaction coordinate into small regions or windows (Figure 2.5a). Then each window is simulated independently (Figure

2.5b), with an additional bias potential applied to ensure that the system adequately samples that particular region.

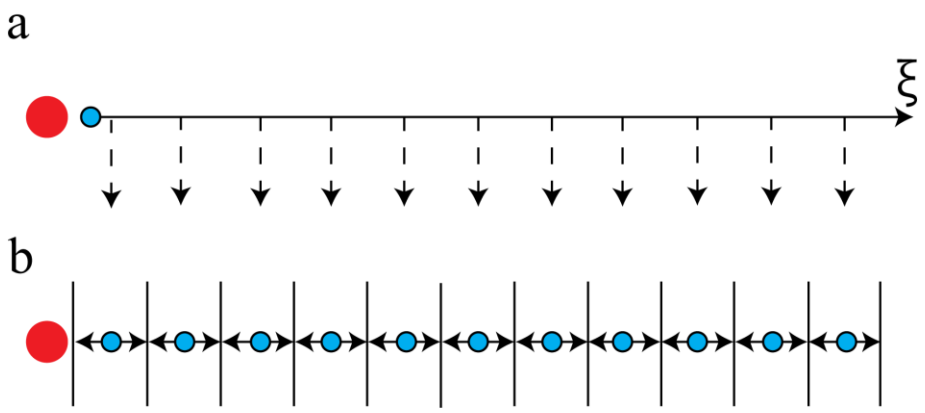


Figure 2.5: Schematic of umbrella sampling to calculate the potential of mean force along the reaction coordinate. (a) The reaction coordinate is divided into small regions, and (b) each region is sampled independently.

The most commonly used bias potential is the harmonic potential due to its simplicity:

$$\omega_i(\xi) = \frac{1}{2}k(\xi - \xi_i^{ref})^2 \tag{2.6}$$

Here, k is the force constant of the bias potential and ξ_i^{ref} is the reference center of window i. Through a series of simulations performed across the various windows, umbrella sampling generates an ensemble of configurations statistically representative of the entire reaction coordinate. By analyzing the resulting data from multiple windows using advanced statistical techniques, *i.e.*, WHAM[131,132] (Weighted Histogram Analysis Method) or UI[133] (umbrella integration), one can reconstruct the free energy profile or potential of mean force along the reaction coordinate.

2.8 ENTROPY-ENTHALPY DECOMPOSITION

To calculate the entropy contribution to the free energy at temperature T, we performed two more sets of umbrella sampling to obtain the free energy at temperatures $T + \Delta T$ and $T - \Delta T$ and then utilized the finite difference temperature[134] of the free energy at each inter solute separation r:

$$-T\Delta S(r) = T\frac{\Delta G(r, T + \Delta T) - \Delta G(r, T - \Delta T)}{2\Delta T} \tag{2.7}$$

The enthalpy component is:

$$\Delta H(r, T) = \Delta G(r, T) + T\Delta S(r) \tag{2.8}$$

2.9 CALCULATION OF WATER TETRAHEDRAL ORDER PARAMETERS

The tetrahedral orientational (q) and translational (S_k) order parameters[135–137] were used to estimate the structural ordering of water.

The orientational order parameter measures how far the directions of the surrounding four nearest neighbors are from a tetrahedral arrangement. Here, we used the rescaled equation suggested by *Errington* and *Debenedetti*[136] :

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos\psi_{jk} + \frac{1}{3}\right)^2 \tag{2.9}$$

The rescaled version of q is defined in a way such that if the molecules are in a random arrangement, then the six angles associated with the center molecules are independent, thus $\langle q\rangle = 0$. In the case of a perfect tetrahedral network, $\cos\psi_{jk} = -\frac{1}{3}$, then $\langle q\rangle = 1$.

The translational order parameter S_k:

$$S_k = 1 - \frac{1}{3}\sum_{k=1}^{4}\frac{(r_k - \bar{r})^2}{4\bar{r}^2} \tag{2.10}$$

S_k measures the variance of the radial distances between central water oxygen and the four nearest neighbors' water oxygen, r_k is the radial distance from the central oxygen atom to the k^{th} peripheral oxygen atom and $\bar{r}$ is the mean value of four radial distances. S_k increases when the local tetrahedral order increases and reaches a maximum value of 1 for a perfect tetrahedron arrangement.

2.10 CALCULATION OF FRACTION OF NATIVE CONTACTS, Q.

Two residues are considered to form a native contact if their C_α atoms are less than 8 Å apart in the crystal structure. To account for thermal fluctuations in contact distances during simulation, a flexibility parameter $\Delta = 1.2$ was used: a native contact between two residues is classified to be formed in a current frame of the simulated trajectory if their distance is shorter than 1.2 times the distance in the crystal structure. Only contacts between pair of residues i and j both within secondary structural elements as identified by STRIDE[138] and satisfy the criterion $|i - j| > 3$, where i and j are the residue indices, were considered. Any secondary segment that is shorter than 4 residues was excluded from the analysis.

2.11 ESTIMATING THE FOLDING TIME OF SLOW-FOLDING PROTEINS WITH A LARGE PROPORTION OF UNFOLDED TRAJECTORIES

Usually, the folding time of protein will be reported as the median folding time. However, when the portion of folded trajectories is less than 50 % of total trajectories, it is not possible to estimate the folding time as the median first passage time. Therefore, we consider the three-state folding kinetics with parallel pathways. State A folds rapidly to the native state N at the rate k_1, and state B folds slowly to the native state with a much smaller rate k_2 ($k_1 \gg k_2$), and there is no interconversion between A and B. We have a set of ordinary differential equations respecting the rate of changing portion of states A and B:

$$\begin{cases} \frac{d[A]}{dt} = -k_1[A] \\ \frac{d[B]}{dt} = -k_2[B] \end{cases} \quad (2.11)$$

where [A] and [B] are the portion of non-native states A and B. The portion (survival probability) of non-native states at time t: $S_{\mathrm{U}}(t) = [A](t) + [B](t) = c_1 \exp(-k_1 t) + c_2 \exp(-k_2 t)$, where c_1, c_2 are arbitrary constants. The initial condition that at time $t = 0$, the survival probability of non-native state $= 1$, we have $S_U(t = 0) = c_1 + c_2 = 1$, this yields: $c_2 = 1 - c_1$.

Hence, we computed the survival probability of the unfolded state as a function of time from simulations, and the resulting time series were then fit to the double-exponential equation:

$$S_{\mathrm{U}}(t) = c_1 \exp(-k_1 t) + (1 - c_1) \exp(-k_2 t) \quad (2.12)$$

c_1, k_1, k_2 are fitting parameters. The time constants of the two kinetic phases: $\tau_1 = \frac{1}{k_1}, \tau_2 = \frac{1}{k_2}$ with the larger of these two times determining the overall timescale of the folding process, $\tau_2 \gg \tau_1$. To estimate the uncertainty of the folding time when fitting to double-exponential folding kinetics, we apply bootstrap resampling by randomly selecting trajectories from the list of simulations.

2.12 Definition of the progress variable ς used to monitor the sequence of pairs of native secondary structure elements formed during the folding process

Protein folding occurs hierarchically, with secondary structural elements first forming individually, then cooperatively coalescing into the tertiary structure. Hence, we characterize protein folding process as the temporal sequence of the formation of stable pairs of native secondary structural elements. To account for the significant variation in folding times among different trajectories, we monitored the folding process as a function of a progressive variable[139], ς, defined as:

$$\varsigma = \langle \frac{t_{\text{pair},i}}{t_{\text{fold},i}} \rangle \tag{2.13}$$

Where: $< \cdots >$ indicates the average over all folded trajectories, and $t_{\text{pair},i}$ and $t_{\text{fold},i}$ are the folding time of the pair and the whole protein folding time of folded trajectory i. With this definition, we have $0 \leq \varsigma \leq 1$, with $\varsigma = 0$ meaning that the pair under study folds at the start of the simulation, and $\varsigma = 1$ indicates the pair folds as the last step in the folding process. To determine the sequence of pairs of the secondary structure formation, we consider a pair between two secondary structure elements with more than one native contact. A pair is considered folded if its fraction of native contacts is larger than the threshold determined from native simulations.

2.13 IDENTIFYING ENTANGLEMENT AND THE CHANGES IN ENTANGLEMENT

Entanglement is defined by the presence of two structural components (Figure 2.6a): a loop formed by a protein backbone segment closed by a non-covalent native contact and another protein segment threaded through and around this loop, sometimes multiple times. We used the numerically invariant linking numbers[140,141] to identify lasso-like entanglements, which describe the linking between a closed loop and an open segment in three-dimensional space (Figure 2.6).

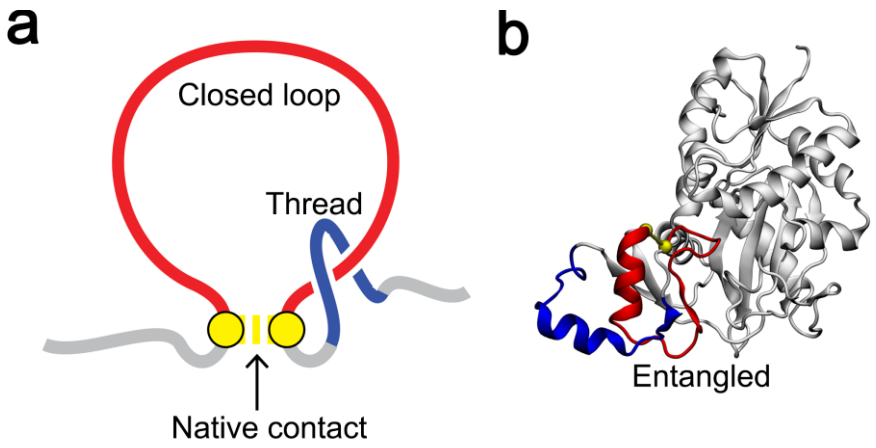


Figure 2.6: Visualizing lasso-entanglement. (a) An illustration of two geometric elements that compose an entanglement: the closed loop is colored in red, and the threading segment is in blue. (b) A misfolded entangled state from the protein D-alanine – D-alanine Ligase B (DDLB) with the closed loop and crossing section of the threading segment colored in red and blue, respectively.

For a given structure of an N-residue protein, with a native contact present at residues (i,j), the coordinates $\boldsymbol{R}_l$ and the gradient $\mathrm{d}\boldsymbol{R}_l$ of the point l on the curves were first calculated as:

$$\begin{cases} \boldsymbol{R}_l = \dfrac{1}{2}(\boldsymbol{r}_l + \boldsymbol{r}_{l+1}) \\ d\boldsymbol{R}_l = \boldsymbol{r}_{l+1} - \boldsymbol{r}_l \end{cases} \tag{2.14}$$

where $\boldsymbol{r}_l$ is the coordinates of the C_α atom in residue l. The linking numbers between N-tail, $g_{\mathrm{N}}(i,j)$, and C-tail, $g_{\mathrm{C}}(i,j)$ and the closed loop that is closed by native contact between residues i and j was calculated as:

$$\begin{cases} g_{\mathrm{N}}(i,j) = \dfrac{1}{4\pi}\displaystyle\sum_{m=6}^{i-5}\sum_{n=i}^{j-1} \frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3} \cdot (\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \\ g_{\mathrm{C}}(i,j) = \dfrac{1}{4\pi}\displaystyle\sum_{m=i}^{j-1}\sum_{n=j+4}^{N-6} \frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3} \cdot (\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \end{cases} \tag{2.15}$$

The total linking number for a native contact (i,j) is estimated as:

$$g(i,j) = \mathrm{round}[g_N(i,j)] + \mathrm{round}[g_C(i,j)] \tag{2.16}$$

Comparing the absolute value of the total linking number for a native contact (i,j) to that of a reference state allows us to detect a gain or loss of linking between the backbone trace loop and the terminal open curves and any switches in chirality[142].

The degree of entanglement G is defined as the fraction of native contacts change entanglement and is time-dependent:

$$G(t) = \frac{1}{M}\sum_{(i,j)} \Theta\left[(i,j) \in NC \bigcap g(i,j,t) \neq g^{\mathrm{native}}(i,j)\right] \tag{2.17}$$

where (i,j) is the native contact in the crystal structure; *NC* is the set of native contacts formed in the current structure at time t; $g(i,j,t)$ and $g^{\mathrm{native}}(i,j)$ are, respectively, the total linking number of the contact (i,j) at time t, and native structures estimated using eq. (2.16). M is the

total number of native contacts in the native structure, and Θ is a Heaviside step function, which equals 1 if the condition is true and 0 if the condition is false.

The difference between $g(i,j,t)$ and $G(t)$ is that $g(i,j,t)$ characterized the entanglement in a given structure of the contact (i,j) at time t, while $G(t)$ provided information about the total number of contacts that changed the entanglement at time t.

Chapter 3. ELECTROSTATIC INTERACTIONS GOVERN EXTREME NASCENT PROTEIN EJECTION TIMES FROM RIBOSOMES AND CAN DELAY RIBOSOME RECYCLING

3.1 INTRODUCTION

Ribosome synthesizes protein vectorially from the N-terminus to the C-terminus along the mRNA template. After reaching the stop codon in mRNA, the covalent bond between the nascent protein and tRNA breaks, and the nascent protein ejects from the ribosome exit tunnel. However, the ejection process has not been experimentally characterized. This is likely due to the belief that the process is rapid, shows slight variation between proteins, and has no biological significance. Furthermore, the complex nature of the ribosome and the rapid timescales associated with nascent chain ejection pose significant challenges to sample preparation and experimental measurement. Recent molecular dynamics simulations study proposed that the physicochemical properties of the exit tunnel can regulate the nascent protein exit and ion flux[143].

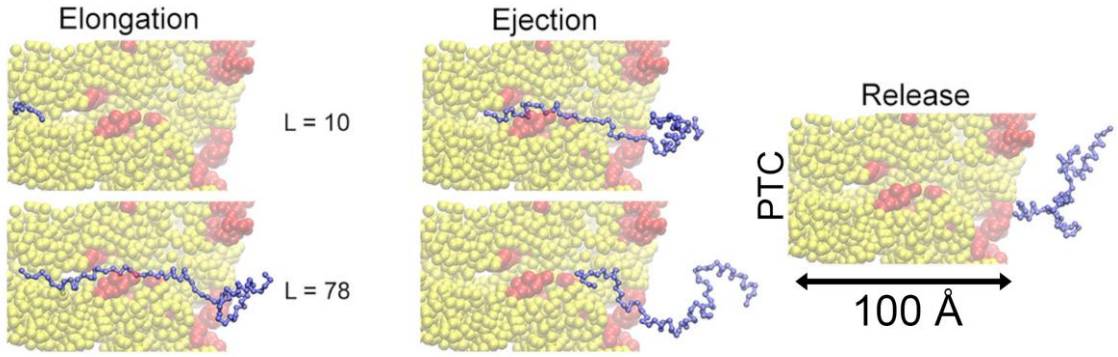


Figure 3.1: Coarse-grained simulations of nascent protein synthesis and ejection. Coarse-grained simulations begin with the elongation phase, during which the protein (blue) is synthesized on the ribosome (rRNA and ribosomal protein of ribosome are colored yellow and red, respectively). Once the full-length protein is synthesized, ejection occurs. Ejection is complete once the C-terminal residue is about 100 Å from the ribosome's peptidyl transferase center (PTC).

In this study, we aimed to investigate the ejection times of proteins from the ribosome exit tunnel (middle panel in Figure 3.1). To accomplish this, we utilized simulations of 122 full-length *E. coli* proteins using a coarse-grained model of the ribosome nascent chain complex (see section 2.3). These proteins were chosen to represent the size and structural class distributions of the *E. coli* cytosolic proteome. Each protein was subjected to 50 independent simulations, and the ejection time was measured as the duration from breaking the bond between the C-terminal residue and P-site tRNA to the point at which it reached the end of the exit tunnel. Our findings revealed that the ejection times of nascent proteins ranged 242-fold, meaning some proteins eject very slowly.

Furthermore, the proteins at the extremes of this distribution had markedly different electrostatic characteristics in their last 30 residues located in the exit tunnel. Specifically, proteins with many positive charges in their C-terminus ejected much more slowly than those with negative charges in the same region. Therefore, we hypothesized that electrostatic properties of the nascent protein C-terminal segment are responsible for extremely fast or slow ejection times.

To test this hypothesis, we performed simulations with the removal of negative or positive charges of amino acids in the last 30 residues for fast or slow-ejecting proteins, respectively. Our results indicated that removing negative charges from fast ejectors slowed the ejection process by 5-98%, while removing positive charges from slow ejecting proteins sped up the ejection time by 48-99%. These results support the hypothesis that electrostatic interactions are the primary factor governing proteins' extremely fast or slow ejection times.

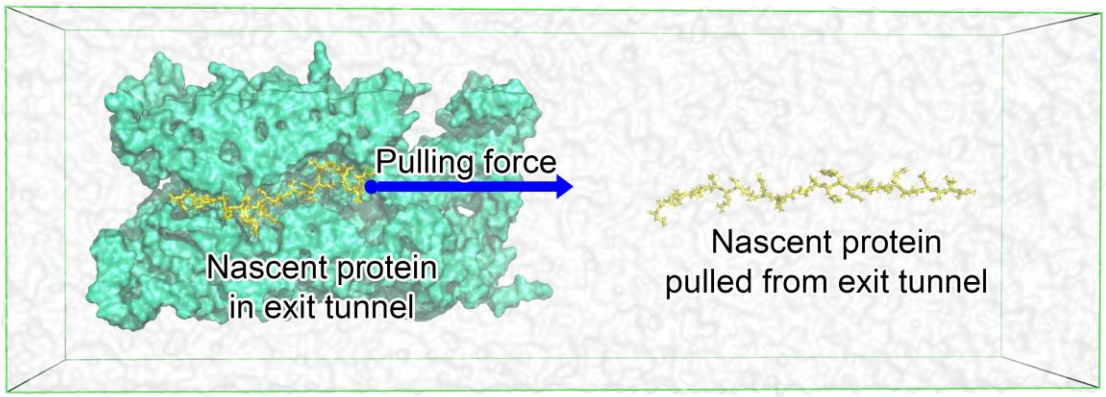


Figure 3.2: All-atom steered molecular dynamics simulation of pulling of a nascent protein (yellow) from the ribosome exit tunnel (cyan).

Coarse-grained models allow larger systems to be simulated over longer timescales than all-atom models but neglect atomic details that may impact results. However, with existing computational facilities, conventional unrestrained all-atom molecular dynamics simulations cannot simulate the complete ejection process of nascent proteins from the ribosome. To test the robustness of conclusions from coarse-grained models, we conducted constant velocity steered molecular dynamics simulations, in which an external force was applied to the N-terminus of the protein to pull it from the ribosome exit tunnel. The last 30 C-terminal residues of each coarse-grained trajectory were back-mapped to all-atom resolution and subjected to SMD simulations. We found that slowly ejecting proteins require 28% more work on average to be extracted from the exit tunnel than quickly ejecting proteins, and this difference is statistically significant. Slowly ejecting proteins interact more strongly with the ribosome exit tunnel than quickly ejecting proteins, with electrostatic interactions being the dominant force. The consistency of all-atom and coarse-grained results indicates that electrostatic interactions between nascent proteins and the ribosome govern ejection times.

This study also raises the biological question of whether the broad range of ejection times has any downstream consequences. To test this, we performed the ribosome profiling analysis, which shows that slow-ejecting proteins cause a significant enrichment of ribosome density at their stop codons, which suggests that the ribosome spends more time on stop codons when a slowly ejecting sequence is present compared to when a quickly ejecting sequence is present.

This finding suggests that slow-ejecting proteins can delay the ribosome recycling stage of protein synthesis and perturb the translation cycle.

3.2 PUBLICATION

3.2.1 *Abstract*

The ejection of nascent proteins out of the ribosome exit tunnel after their covalent bond to transfer-RNA has been broken has not been experimentally studied due to challenges in sample preparation. Here, we investigate this process using a combination of multi-scale modeling, ribosome profiling, and gene ontology analyses. Simulating the ejection of a representative set of 122 *E. coli* proteins we find a greater than thousand-fold variation in ejection times. Nascent proteins enriched in negatively charged residues near their C-terminus eject fastest, while nascent chains enriched in positively charged residues tend to eject much more slowly. More work is required to pull slow-ejecting proteins out of the exit tunnel than fast-ejecting proteins, according to all-atom simulations. An energetic decomposition reveals that this is due, for slow ejecting proteins, to the strong attractive electrostatic interactions between the nascent chain and negatively charged ribosomal-RNA lining the exit tunnel, and for fast ejecting proteins it is due to their repulsive electrostatic interactions with the exit tunnel. Ribosome profiling data from *E. coli* reveals that the presence of slow ejecting sequences correlates with ribosomes spending more time at stop codons, indicating the ejection process might delay ribosome recycling. Proteins that have the highest positive charge density at their C-terminus are overwhelmingly ribosomal proteins, suggesting the possibility that this sequence feature may aid in the co-translational assembly of ribosomes by delaying release of nascent ribosomal proteins into the cytosol. Thus, nascent chain ejection times from the ribosome can vary greatly between proteins due to differential electrostatic interactions, can influence ribosome recycling, and could be particularly relevant to the synthesis and co-translational behavior of some proteins.

DOI: 10.1021/jacs.9b12264

Publication source: https://pubs.acs.org/doi/10.1021/jacs.9b12264

3.2.2 *Author contribution statements*

Daniel A. Nissley
Department of Statistics
University of Oxford
Oxford, OX1 3LB, United Kingdom

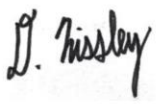
STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was to perform coarse-grain model simulations, analyze the results, and help write the manuscript along with my co-authors.

Oxford, 23 May 2023



Daniel A. Nissley

Quyen V. Vu
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

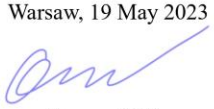
STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was performing all-atom simulations, analyzing the results, preparing figures, and participating in writing the manuscript.

Warsaw, 19 May 2023



Quyen V. Vu

Nabeel Ahmed
Senior Scientist, Bioinformatics
Moderna Therapeutics
200 Technology Square
Cambridge MA, 02739

moderna

STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was processing the ribosome profiling data, analyzing the results and editing the manuscript.

NJ, USA, 23 May 2023

Nabeel Ahmed

Yang Jiang
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802

STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was analyzing the results, preparing figures, and writing the manuscript.

PA USA, 22 May 2023

Yang Jiang

Yang Jiang

Prof. Mai Suan Li, PhD
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was analyzing and interpreting the results, and helping to write the manuscript.

Warsaw, 26 May 2023

Mai Suan Li

Prof. Edward P. O'Brien, PhD
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802
E-mail: epo2@psu.edu
Tel: 1-814-867-5100

STATEMENT

I declare that I am the co-author of the publication:

- Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling**. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110.

My contribution was designing the research methodology, interpreting the data, writing the manuscript, and supervising the overall project.

PA USA, 26 May 2023

Edward P. O'Brien

3.2.3 *Publication*

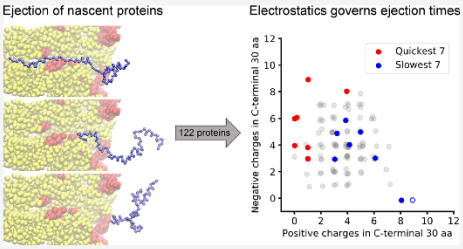
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling

Daniel A. Nissley, Quyen V. Vu, Fabio Trovato, Nabeel Ahmed, Yang Jiang, Mai Suan Li, and Edward P. O'Brien*

Cite This: J. Am. Chem. Soc. 2020, 142, 6103–6110
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The ejection of nascent proteins out of the ribosome exit tunnel, after their covalent bond to transfer-RNA has been broken, has not been experimentally studied due to challenges in sample preparation. Here, we investigate this process using a combination of multiscale modeling, ribosome profiling, and gene ontology analyses. Simulating the ejection of a representative set of 122 *E. coli* proteins we find a greater than 1000-fold variation in ejection times. Nascent proteins enriched in negatively charged residues near their C-terminus eject the fastest, while nascent chains enriched in positively charged residues tend to eject much more slowly. More work is required to pull slowly ejecting proteins out of the exit tunnel than quickly ejecting proteins, according to all-atom simulations. An energetic decomposition reveals, for slowly ejecting proteins, that this is due to the strong attractive electrostatic interactions between the nascent chain and the negatively charged ribosomal-RNA lining the exit tunnel, and for quickly ejecting proteins, it is due to their repulsive electrostatic interactions with the exit tunnel. Ribosome profiling data from *E. coli* reveals that the presence of slowly ejecting sequences correlates with ribosomes spending more time at stop codons, indicating that the ejection process might delay ribosome recycling. Proteins that have the highest positive charge density at their C-terminus are overwhelmingly ribosomal proteins, suggesting the possibility that this sequence feature may aid in the cotranslational assembly of ribosomes by delaying the release of nascent ribosomal proteins into the cytosol. Thus, nascent chain ejection times from the ribosome can vary greatly between proteins due to differential electrostatic interactions, can influence ribosome recycling, and could be particularly relevant to the synthesis and cotranslational behavior of some proteins.



INTRODUCTION

Translation is the process by which a protein is synthesized from an mRNA and is carried out by the ribosome molecular machine. The four phases of translation (initiation, elongation, termination, and ribosome recycling) are areas of intense research due to the essential role of protein synthesis in life. Each phase is composed of multiple steps, many of which have been characterized in terms of the structures adopted by the molecules involved, the mechanisms of conformational and chemical transitions, and the rates associated with these transitions.[1] Translation termination in *E. coli*, for example, consists of some four steps: the binding of a release factor to a stop codon in the A-site of the ribosome, the hydrolysis of the covalent bond connecting the C-terminus of the nascent chain to the P-site tRNA, the ejection of the nascent protein out of the exit tunnel, and the dissociation of the release factor from the ribosome (Figure 1). While rates for release factor binding and hydrolysis have been measured,[2−6] the diffusion of the nascent chain out of the exit tunnel has not been experimentally characterized due to challenges with sample preparation. Additionally, the presumably fast time scales of nascent chain ejection make experimental measurement a challenge.

The ribosome exit tunnel is composed of both rRNA and ribosomal protein and is therefore a chemically heterogeneous environment through which nascent proteins pass into the cytosol. At approximately 10 nm long and roughly 1.5 nm in diameter, the interactions between the exit tunnel and nascent chains can exhibit the full range of intermolecular forces, including charge−charge, hydrogen bonding, and hydrophobic interactions. Highly attractive forces can exist between some regions of the exit tunnel and some nascent peptide sequences.[7] Indeed, peptide sequences known as stalling sequences have evolved to take advantage of these interactions and bind to the tunnel wall so tightly that they drastically slow

Received: December 2, 2019
Published: March 6, 2020

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

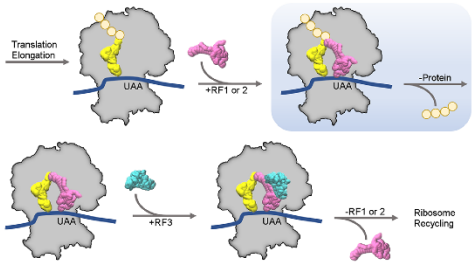


Figure 1. Translation termination in *E. coli*. Translation termination begins after translation elongation ends when a stop codon (e.g., UAA) enters the ribosome's A-site (ribosome shown as a gray outline). Release factor (RF) 1 or 2 (magenta) binds the stop codon in the A-site and catalyzes the hydrolysis of the peptidyl-tRNA bond between the nascent protein (orange spheres) and P-site tRNA (yellow). The nascent protein then diffuses out of the exit tunnel of the ribosome, which is around 10 nm in length. Following ejection, RF3 (cyan) catalyzes the release of RF1 or 2, allowing translation to proceed to the final phase, ribosome recycling. This study focuses on the second panel in this figure, highlighted in light blue. P-site tRNA and RF1 structures are generated from PDB ID 3OSK. The RF3 structure is generated from PDB ID 2H5E.

translation elongation,[8,9] the biological benefit of which is to regulate downstream protein synthesis. When stretches of positively charged residues are present in the exit tunnel, they slow protein elongation under both in vitro[10] and in vivo conditions. Since nascent protein elongation and nascent protein ejection both involve the passage of nascent protein segments through the exit tunnel, the interactions that can be large and impactful during elongation also have the potential to be important during ejection.

In this study, we use a combination of coarse-grained and all-atom simulations, ribosome profiling data, and gene ontology analysis to estimate the relative range of ejection time scales that can occur across the cytosolic proteome of *E. coli*, determine the intermolecular forces that give rise to the extremes of ejection times, find experimental evidence that slowly ejecting sequences can delay later stages of translation, and identify nascent ribosomal proteins as some of the slowest ejecting proteins from the ribosome.

■ SIMULATION METHODS

Single-Domain Protein Selection and Model Building. The database from which the 122 proteins were selected contains 1014 cytosolic protein structures, 598 of which were single-domain proteins and the rest multidomain proteins.[11] A domain in this database is classified as either α or β if more than 70% of its residues identified by STRIDE[12] to be in secondary structural elements were in α-helixes or β-strands, respectively. Domains that simultaneously had α-helical and β-strand content greater than 30% were classified as α/β.[11] Given the sequence length distribution of these proteins, we determined that we could feasibly simulate the synthesis and ejection of 122 proteins in total. To maintain the ratio of single- to multidomain proteins in the database, we selected 72 single-domain proteins and 50 multidomain proteins. Of the 598 single-domain proteins in the database there are 250 α, 55 β, and 293 α/β domains; this ratio of structural classes was reproduced in the subset of 72 single-domain proteins by randomly selecting 30 α, 7 β, and 35 α/β proteins (Tables S1 and S2). PDB files for each single-domain protein were retrieved from the Protein Data Bank,[13] and their corresponding mRNA sequences (NCBI assembly eschColi_K12) were retrieved using the University of California Santa Cruz microbe table browser (http://microbes.ucsc.edu/). Randomly selected PDBs from the database were accepted only if the crystallized sequence had no amino acid mutations in comparison to the amino acid sequence which would result from the translation of the eschColi_K12 mRNA. However, small sections of amino acids (12 or less) or small numbers of heavy atoms (less than 10) that were not resolved in the experimental structure were rebuilt on the basis of the reference genome sequence and minimized in CHARMM.[14]

Multidomain Protein Selection and Model Building. Fifty multidomain proteins were selected randomly from the same previously published database of *E. coli* globular proteins from which single-domain proteins were selected.[11] These multidomain proteins are listed in Table S3. The amino acid sequence of each PDB was aligned to the translated sequence of the corresponding gene in NCBI assembly eschColi_K12, and missing residues and domains were identified. Some PDBs contained large missing sections; to fill in these regions with reasonable structures, PDBs representing the same gene product were used to reconstruct missing sections after structural alignment in VMD.[15] PREDATOR[16] and IUPRED[17] were used to predict whether those residues not resolved in any other PDB were intrinsically unstructured, in which case they were rebuilt and minimized in CHARMM rather than templated using other structures. When homologous structures from *E. coli* were not available, homologous structures from other organisms were used as a template for the protein model, provided the sequence similarity was greater than 30% and the backbone RMSD between the regions common to the two structures was $\lesssim$2 Å.

Reconstructing missing domains or sections of the multidomain proteins in this way resulted in models that still had, in some cases, sections of missing atoms or mismatched amino acids relative to the consensus mRNA sequence. All proteins were therefore subjected to a rebuilding phase to add missing atoms and correct mutations or sequence mismatches (Table S4). Because the reconstructed segments were generated in an extended conformation, minimization was performed in vacuo for 200 steps. This short minimization was sufficient to resolve steric clashes. For proteins with short stretches of missing residues (less than 10), this minimized configuration was accepted as the final atomistic model. If a protein contained one or more long stretches of missing residues (more than 10) or disconnected domains, then the minimized protein structure was subjected to additional dynamics at 310 K. In this phase, the reconstructed atoms within each templated domain were left free to move, thereby allowing the structure to locally equilibrate. The smallest domain in each protein was also left unrestrained in order to allow it to reorient with respect to all other domains into a favorable conformation. All other atoms were either held fixed or harmonically restrained to the experimentally solved structure with a force constant of 1 kcal/(mol·Å^2). All reconstructions, minimizations, and molecular dynamics simulations were performed using CHARMM with the par27 force field.[14] The minimized structures were solvated in TIP3P[18] water and 150 mM NaCl, gradually heated to 310 K for 100 ps, and then equilibrated for 1.5 ns at the same temperature. Production runs had different durations, spanning from 20 to 50 ns. Langevin dynamics with a friction coefficient of 1.0 ps^{-1} and a time step of 1.5 fs were used. For each protein, the conformation with the lowest potential energy was selected as the final atomistic model. We emphasize that the purpose of the molecular dynamics simulations was not to thoroughly explore the conformational space of the multidomain proteins but rather to provide reasonable atomistic conformations for building coarse-grained models.

Domains in the multidomain proteins were initially defined according to CATH,[19] which is also used in the original database.[11] Domain residue numberings were shifted to match the translated sequence and modeling of the missing atoms performed (Table S4). The final domain definitions reported in Table S3 include residues and domains that were modeled as described in Table S4.

Coarse-Grained Force Field and Model Construction. The potential energy for a given configuration of the C_α coarse-grained model is calculated using the equation

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

$$E = \sum_i k_b(r_i - r_0)^2 + \sum_i \sum_{j=1}^{4} k_{\varphi,ij}(1 + \cos[j\varphi_i - \delta_{ij}])$$
$$+ \sum_i -\frac{1}{\gamma}\ln\{\exp[-\gamma(k_\alpha(\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha)] + \exp[-\gamma k_\beta(\theta_i - \theta_\beta)^2]\}$$
$$+ \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \exp\left|-\frac{r_{ij}}{l_D}\right| + \sum_{ij\in\{NC\}} \epsilon_{ij}^{NC}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$
$$+ \sum_{ij\notin\{NC\}} \epsilon_{ij}^{NN}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

The terms in this equation represent, from left to right, summations over the contributions from C_α–C_α bonds, dihedral angles, bond angles, electrostatic interactions, Lennard-Jones-like native interactions, and repulsive non-native interactions to the total potential energy. The bond, dihedral, and angle terms have been described in detail elsewhere.[20,21] Electrostatics are treated using Debye–Hückel theory with a Debye length, l_D, of 10 Å and a dielectric of 78.5; lysine and arginine C_α sites are assigned $q = +e$, glutamic acid and aspartic acid are assigned $q = -e$, and all other interaction sites are uncharged.[22] The contribution from native interactions is computed using the 12–10–6 potential of Karanicolas and Brooks.[20] The value of ϵ_{ij}^{NC}, which sets the depth of the energy minimum for a native contact, is calculated as $\epsilon_{ij}^{NC} = n_{ij}\epsilon_{HB} + \eta\epsilon_{ij}$. Here, ϵ_{HB}, and ϵ_{ij} represent energy contributions arising from hydrogen bonding and van der Waals contacts between residues i and j identified from the all-atom structure of the protein, respectively. n_{ij} is the number of hydrogen bonds formed between residues i and j and $\epsilon_{HB} = 0.75$ kcal/mol. The value of ϵ_{ij} is set on the basis of the Betancourt–Thirumalai pairwise potential.[23] The scaling factor η is determined for each of our 122 proteins (Tables S2 and S3) based on a previously published training set[24] to reproduce realistic protein stabilities for different structural classes (Table S5, Table S6, and Supplementary Methods). A single value of η is applied to all native contacts for a given single-domain protein and for each individual domain and interface in multidomain proteins. Collision diameters, σ_{ij}, between C_α interactions sites involved in native contacts are set equal to the distance between the C_α of the corresponding residues in the crystal structure divided by $2^{1/6}$. For non-native interactions, ϵ_{ij}^{NN} is set to 0.000132 kcal/mol and σ_{ij} is computed as previously reported.[20]

Simulations of Nascent Protein Synthesis and Ejection. The coarse-grained model of each protein in the single- and multidomain protein data sets was synthesized starting from a single residue (see the two exceptions below) using a modified version of a previously published protocol on a coarse-grained representation of the 50S *E. coli* ribosome (details of the ribosome model can be found in Supplementary Methods).[25] The dwell time at a particular nascent chain length was randomly selected from an exponential distribution with a mean equal to the average decoding time of the codon in the A-site. Average decoding times are taken from the Fluitt-Viljoen model[26] and scaled to reproduce an overall average of 12.6 ns (840 000 integration time steps of 0.015 ps duration; see Table S7) based on a previously published training set.[24] A planar restraint in the yz plane through the point (58, 0, 0) Å is used to prevent the nascent chain from contacting the underside of the ribosome cutout. Fifty trajectories were run for each of the 122 proteins in the data set. After synthesis was completed for a given trajectory, the harmonic restraint on the C-terminal bead to model the covalent bond between the nascent protein and the P-site tRNA was removed. Simulations of termination were run until the C-terminal residue of each trajectory reached an x coordinate of 100 Å or greater, indicating that the protein exited the tunnel. Ejection times are calculated as the time between when the C-terminal harmonic restraint is removed and when the C-terminal residue reaches an x coordinate of ≥100 Å. Two proteins (PDB IDs 2KFW and 3GN5) became stalled in the exit tunnel when synthesis was begun from a single residue; synthesis for these two proteins was therefore initiated from a nascent chain length of 50 residues. One protein (PDB ID 4DCM) did not eject from the exit tunnel in 27 of 50 trajectories during 25 days of CPU time when its wild-type C-terminal charges were used; the ejection time for this protein is therefore reported as a lower bound. Mean ejection times for all 122 proteins are listed in Table S8.

All-Atom Steered Molecular Dynamics Simulations. The 50S subunit of the *E. coli* ribosome (PDB ID 3R8T) was aligned with the long axis of the exit tunnel, defined to be between atom N6 of nucleotide A2602 and the C_β atom of Ala50 in ribosomal protein L24, along the x axis of the simulation coordinate system. The ribosome was then cropped to form a rectangular box around the exit tunnel with dimensions of 13.10590 × 8.44869 × 8.18680 nm^3. Coarse-grained structures of the C-terminal 30 aa of nascent proteins from the final time step of synthesis simulations were backmapped to atomistic resolution for use as starting structures. The first step in backmapping is the insertion of coarse-grained sites representing amino acid side chains near their corresponding C_α beads followed by energy minimization in the C_α side-chain model force field[27] with all C_α positions restrained. Backbone and side-chain all-atom structures were then rebuilt using Prodart2[28] and Pulchra,[29] respectively, on the minimized C_α side-chain model. The final backmapped structure was obtained after energy minimization within the generalized Born (GB) implicit water environment.[30] The N-terminus of the segment was capped by the N-terminal acetyl capping group (ACE) and the atomistic protein structure inserted into the atomistic exit tunnel structure.

A simulation box was constructed with a minimum of 1 nm between the edge of the cropped ribosome and the periodic boundary wall in all dimensions and then extended 15 nm in the positive x dimension to accommodate the nascent protein when fully extracted from the exit tunnel at the end of the steered molecular dynamics simulation. The system was neutralized with Na^+ before adding 5 mM $MgCl_2$ and 100 mM NaCl. Next, the system was minimized in the gas phase with the steepest-descent algorithm. Harmonic restraints on all C_α atoms of the nascent peptide and all heavy atoms of the ribosome with a force constant of 1000 kJ/(mol·nm^2) were employed to prevent the nascent protein from moving during minimization. The system was then equilibrated in the gas phase for 300 ps to allow ions to rapidly find binding sites on the ribosome, with harmonic restraints again applied to C_α atoms of the nascent chain and all heavy atoms of the ribosome.

The cropped ribosome and nascent protein were then solvated and equilibrated. First, 1 ns of dynamics was carried out in the NVT ensemble, followed by 1 ns of dynamics in the NPT ensemble, with the temperature and pressure held at 310 K and 1 atm, respectively. To allow the nascent protein and the ribosome exit tunnel to reach equilibrium in the all-atom model, we performed a second NPT simulation for 10 ns with harmonic restraints applied to P and C_α atoms of the ribosome that were more than 28 Å from the x axis and all C_α atoms of the nascent protein. The center of mass of the N-terminal ACE residue was then pulled from the exit tunnel with a cantilever speed of 0.25, 1, or 5 nm/ns and a spring constant of 600 kJ/(mol·nm^2). All simulations were carried out with GROMACS 2018[31] using the AMBER99SB[32] force field and the TIP3P[18] water model. The particle mesh Ewald method[33] was used to calculate the long-range electrostatic interactions beyond 1.2 nm. Lennard-Jones interactions were calculated within a distance of 1.2 nm. The Nose–Hoover thermostat[34,35] and Parrinello–Rahman barostat[36] were employed to maintain the temperature and pressure at 310 K and 1 atm, respectively. The LINCS algorithm[37] was used to constrain all bonds, and the integration time step was set to 2 fs.

Simulations were carried out using this protocol for 5 quickly ejecting proteins (PDB IDs 1FM0, 1Q5X, 1T8K, 2KFW, and 3BMB) and 5 slowly ejecting proteins (PDB IDs 1AH9, 1JW2, 2JO6, 2PTH, and 3IV5) using 21 different initial configurations from the coarse-grained synthesis simulations for each different protein.

■ RESULTS AND DISCUSSION

To estimate the range of nascent chain ejection times across different proteins, we simulated the synthesis and ejection of

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103–6110

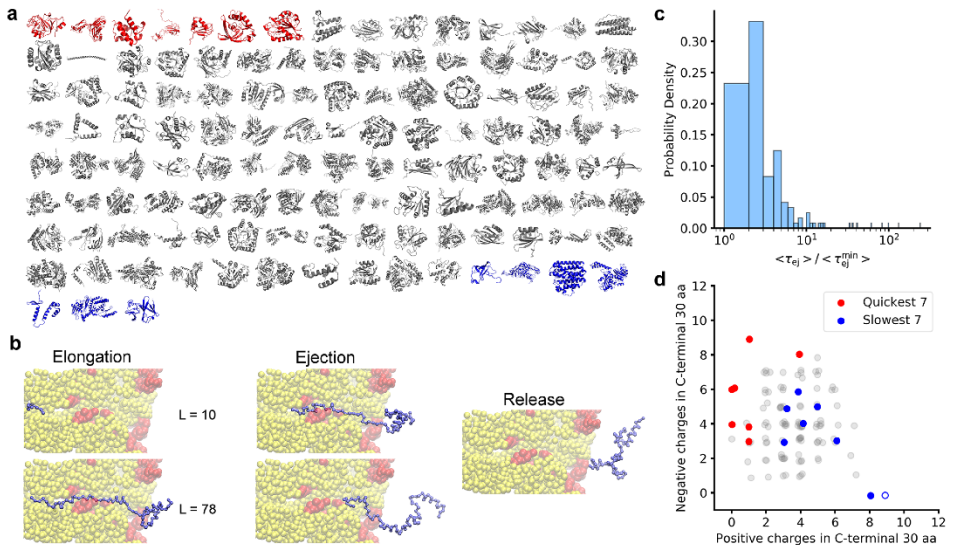


Figure 2. The 242-fold variation in ejection times is related to the presence of charged residues in the C-termini of proteins. (a) The set of 122 *E. coli* proteins that were simulated is shown from top left to bottom right by increasing ejection time. The top 5% fastest and slowest ejecting proteins are colored red and blue, respectively, while the middle 90% are colored gray. (b) Coarse-grained simulations begin with the elongation phase, during which the protein (blue) is synthesized on the ribosome (rRNA and protein colored yellow and red, respectively). Once the full-length protein is synthesized, ejection occurs. Ejection is complete once the C-terminal residue is 100 Å from the peptidyl transferase center of the ribosome. (c) Distribution of mean ejection times, $\langle \tau_{ej} \rangle$, for 121 of the 122 proteins shown in (a) (excluding 4DCM), normalized by the factor $\langle \tau_{ej}^{min} \rangle$ which is the smallest $\langle \tau_{ej} \rangle$ found in the set of proteins. (d) Number of positive and negative charges in each protein's C-terminal 30 residues. The fastest and slowest ejectors (bottom and top 5% of the distribution in panel c) are colored red and blue, respectively, and exist as separate, nonoverlapping populations along these metrics. Values from other proteins are displayed in transparent gray. Random noise (jitter) has been added to minimize overlapping points. As discussed in the main text, the single unfilled blue data point at 9 positive charges and zero negative charges is for PDB ID 4DCM, for which an exact ejection time could not be calculated because not all of its trajectories were released from the ribosome in the simulation. Therefore, this protein was excluded from the distribution in panel c.

122 full-length *E. coli* proteins using a coarse-grained representation of the ribosome nascent chain complex (Figure 2a,b).[25,27,38,39] This set of 122 proteins is representative of the globular *E. coli* cytosolic proteome as a whole because it reproduces the proteome-wide protein size and structural class distributions (Figure S1, Table S1, and Simulation Methods). Fifty statistically independent synthesis and ejection trajectories were run for each protein. We find that the mean protein ejection time, defined as the average time it takes for the C-terminal nascent chain residue to reach the end of the exit tunnel after the bond between the protein and P-site tRNA is broken, varies 242-fold across 121 of these proteins (Figure 2c). We observed that the proteins at the extremes of this ejection time distribution, that is, those proteins in the top and bottom 5%, have markedly different electrostatic characteristics in their C-termini (Figure 2d), the last 30 residues of which are in the exit tunnel. Quickly ejecting proteins tend to have abundant negatively charged residues and few positively charged residues in their C-terminal 30 residues (red dots in Figure 2d). In contrast, slowly ejecting proteins tend to have fewer negatively charged residues and more positively charged residues (blue dots in Figure 2d). We note the exceptional case of the protein with PDB ID 4DCM (unfilled blue point in Figure 2d), which is the 122nd protein in our set. (Note that complete protein names are provided in Tables S2 and S3.) While complete ejection occurred for all other proteins, only 23 out of the 50 simulation trajectories of 4DCM were fully ejected from the exit tunnel. Under a conservative estimate, this protein's average ejection time is 7031-fold slower than the fastest ejecting protein in our data set. Consistent with electrostatics being important, 4DCM also has the greatest positive charge density in our set of proteins. These results indicate that there is a 3-order-of-magnitude spread in ejection times across *E. coli* cytosolic proteins and suggest that the very fast ejectors are fast because they are electrostatically repelled by the exit tunnel, which is lined with negatively charged rRNA, while the very slow ejectors are slow because they are electrostatically attracted to the exit tunnel wall.

To test this electrostatic hypothesis within our coarse-grained model, we set to zero all negative or positive charges of amino acids in the C-terminal 30 residues of the quickly or slowly ejecting proteins, respectively. All other interactions and charges involving the ribosome and nascent chain remained the same. Rerunning the ejection simulations for these sequences, we find that removing negative charges from the set of quick ejectors slowed the ejection process by 5−98% (average 44%) and removing positive charges from the set of slowly ejecting proteins sped up the ejection process by 48−99% (average 82%, 4DCM results excluded) (Table 1). These results are consistent with the hypothesis that electrostatic interactions are a causal factor in influencing extremely fast or extremely slow ejection times out of the ribosome tunnel.

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

Table 1. Ejection Times upon Neutralization of C-Terminal Positive or Negative Residues

quickly ejecting peptides			
PDB ID	wild type (ns)	no (−) charges (ns)[a]	% change
1Q5X	0.31	0.47	51.9
3M7M	0.31	0.41	30.8
1T8K	0.32	0.42	31.9
2KFW	0.32	0.41	27.3
1FM0	0.36	0.38	4.55
3BMB	0.37	0.57	51.8
1AG9	0.39	0.76	95.8
2HGK	0.40	0.65	62.3
1FJJ	0.41	0.44	8.06
2HO9	0.41	0.60	43.8
1SVT	0.42	0.50	19.4
1SG5	0.45	0.88	98.0
slowly ejecting peptides			
PDB ID	wild type (ns)	no (+) charges (ns)[a]	% change
4IM7	3.92	0.69	−82.3
1JW2	4.39	1.94	−55.9
2PTH	4.75	0.86	−81.9
1D2F	5.02	0.64	−87.2
3OFO	10.22	5.28	−48.3
1AH9	11.28	0.59	−94.8
1NG9	12.66	2.31	−81.8
1RQJ	18.80	0.84	−95.5
1T4B	25.66	2.58	−90.0
3IV5	30.47	0.45	−98.5
1U0B	40.75	1.04	−97.5
2JO6	74.73	23.31	−68.8
4DCM	>2170	0.54	−100.0

[a]Columns labeled "no (−) charges" and "no (+) charges" are ejection times from simulations in which negative or positive charges in C-terminal 30 aa, respectively, are made electrically neutral.

While coarse-grained models can simulate larger systems for longer times in comparison to all-atom simulations, they leave out atomic details that have the potential to influence these results. It is currently not possible to simulate the complete ejection process of nascent chains from ribosomes using unrestrained all-atom molecular dynamics simulations. Therefore, to qualitatively test the robustness of the conclusions from our coarse-grained model, we carried out nonequilibrium all-atom steered molecular dynamics simulations in which the nascent protein is pulled from the ribosome exit tunnel using an external pulling force applied to the N-terminus of the protein (Figure 3a). If the coarse-grained model results are correct, then we predict that it will be harder to pull (as measured by the nonequilibrium work) the slowly ejecting chains out of the exit tunnel as compared to the quickly ejecting chains due to differential electrostatic interactions with the ribosome exit tunnel. Twenty-one independent trajectories were run for each of five quickly ejecting and five slowly ejecting proteins drawn from the bottom and top 10% of the distribution of ejection times, respectively. A cantilever speed of 0.25 nm/ns was used for all simulations. In these all-atom simulations, we find that the slowly ejecting nascent proteins require 28% (95% CI [19%, 36%] computed from bootstrapping; $p < 1 \times 10^{-8}$ computed from the permutation test) more work on average to be extracted from the exit tunnel than quickly ejecting nascent proteins (Figure 3b,c). Decomposing

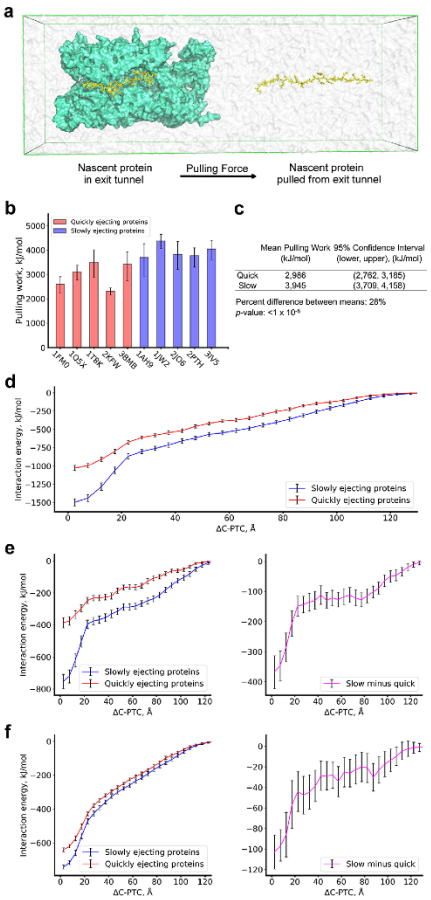


Figure 3. Slowly ejecting proteins are more electrostatically attracted to the ribosome exit tunnel. (a) Initial (left) and final (right) conformations from all-atom, steered molecular dynamics simulations of the extraction of a nascent protein (yellow) from the ribosome (cyan). (b) Mean pulling work required to extract 10 different nascent proteins from the ribosome exit tunnel from 21 statistically independent simulations per protein. Error bars are 95% confidence intervals calculated by bootstrapping. (c) Results from the statistical comparison between the overall means of the slowly and quickly ejecting sets. Confidence intervals are calculated as in (b). The p value is estimated using a permutation test. (d) Total interaction energy between the ribosome and nascent protein as a function ΔC − PTC, the distance between the C_α atom of the C-terminal residue of the nascent protein and the N6 atom of nucleotide A2602 in the peptidyl transferase center of the ribosome. (e) Electrostatic contribution to the total interaction energy (left) and the difference between the slowly and quickly ejecting data set mean electrostatic interaction energies (right). (f) The same as in (d) but for the van der Waals interaction energy. These results were obtained using a cantilever speed of 0.25 nm/ns.

the intermolecular interactions in these simulations, we find that slowly ejecting nascent proteins have stronger interactions

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

with the ribosome tunnel wall than quickly ejecting proteins (Figure 3d–f), with the majority of this energy difference due to electrostatic rather than van der Waals interactions (Figure 3e,f). Qualitatively equivalent results are obtained when the cantilever speed is increased 4- or 20-fold to 1 or 5 nm/ns, respectively (Figures S2 and S3). Thus, the all-atom results and coarse-grained results are consistent, lending further support to the hypothesis that electrostatic interactions between the nascent chain and ribosome govern the extremes of nascent chain ejection times.

An important biological question is whether there are any downstream consequences of this broad range of ejection times. We hypothesized that the slowest ejecting sequences might delay the onset of the next and final step of translation (Figure 1), ribosome recycling, during which molecular factors interact with the ribosome to aid the dissociation of the small and large ribosomal subunits. This hypothesis predicts that ribosomes will dwell for longer at stop codons when a slowly ejecting sequence is present compared to when a quickly ejecting sequence is present. To test this hypothesis, we analyzed ribosome profiling data from *E. coli*[40] of those cytosolic proteins that have the highest charge density at their C-terminus. Ribosome profiling is an experimental technique that measures a signal, called the "reads", that is proportional to the number of ribosomes sitting at a particular codon position on the various cellular copies of an mRNA transcript.[41] As such, the greater the normalized ribosome density at a codon, the longer the ribosome spent at that codon position. The normalized ribosome density at a codon position is the number of reads at that codon divided by the average number of reads per codon arising from the coding sequence of the transcript. Therefore, our hypothesis predicts that there will be greater ribosome density at the stop codon for proteins that have the highest number of positive charges in their C-terminus compared to those that have high negative charge density. Therefore, we restricted our analysis to high-coverage transcripts (Supplementary Methods) encoding proteins with either ≥8 positive and ≤2 negative residues in their 30 C-terminal residues, which we predict to be slowly ejecting proteins (n = 22 proteins), or with ≥8 negative residues and ≤2 positively charged residues (n = 22 proteins) in their 30 C-terminal residues, which we predict to be quickly ejecting proteins. We could not include all of the fastest and slowest ejecting proteins from our simulations in this analysis because the read coverage of their transcripts was very sparse in the ribosome profiling data, meaning that their signal-to-noise ratio is too low to be useful. However, two proteins (PDB IDs 1T8K and 3IV5) for which we simulated ejection times did have sufficient read coverage and are included in this analysis. We observe that the putative slow ejectors have on average 3.3-fold higher ribosome density at the stop codon compared to the fast ejectors (Figure 4a,b; median ribosome densities across fast and slow ejector sets are, respectively, 0.248 and 0.812; the difference between medians is significant based on the Mann–Whitney U Test, p = 0.011). These results are consistent with the hypothesis that the slowest ejecting nascent chains tend to delay ribosome recycling.

To further explore the potential biological ramifications of very fast or slow ejection times, we carried out a gene ontology analysis to determine whether putative slowly and quickly ejecting proteins are more likely than random chance to be associated with particular cellular or biochemical processes (Supplementary Methods). The putative quickly ejecting

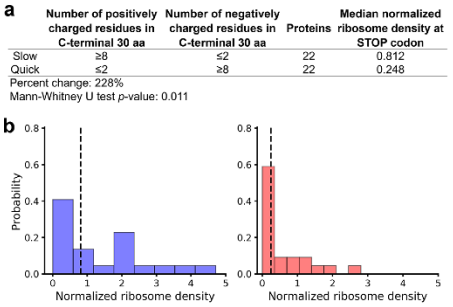


Figure 4. Presence of slowly ejecting proteins in ribosome-nascent chain complexes correlated with longer ribosome dwell times at the stop codon. (a) Proteins with ≥8 positive and ≤2 negative residues in their C-terminus as well as proteins with ≥8 negative and ≤2 positive residues in their C-terminus were selected from the *E. coli* ribosome profiling data from ref 40. A total of 22 proteins fit into each category. The median normalized ribosome density is higher for proteins enriched in positive charge at the C-terminus (p value 0.011, Mann–Whitney U test). (b) Histograms of normalized ribosome density at the stop codon for the subsets of proteins enriched in positive (left, blue histogram) or negative (right, red histogram) amino acids.

proteins exhibit no significant relationship to any particular biological processes. However, 14 of the 22 potentially slowly ejecting proteins are associated with translation, and 13 are ribosomal subunit proteins. Two hypotheses can explain this observation. First, slow ejection increases the time a nascent protein is available for cotranslational assembly,[42−44] suggesting that these highly positively charged C-terminal segments might have evolved to aid in the efficient cotranslational assembly of ribosomes in the *E. coli* cytosol. Second, ribosomal proteins may have evolved positively charged segments solely to aid in their interactions with rRNA in the context of a fully assembled ribosomal subunit, with their slow ejection times being a biologically irrelevant consequence of this fact. Indeed, each of the 13 ribosomal proteins identified by this analysis is in contact with rRNA based on the analysis of a crystal structure of the *E. coli* ribosome in the nonrotated conformation (PDB ID 4V9D). It will be an interesting area of future research to test these distinct hypotheses.

■ CONCLUSIONS

Our results indicate that nascent protein ejection times are very broad, that the extremes are primarily driven by interactions of the high charge density of either positive or negative residues near the nascent protein's C-terminus with the negatively charged ribosome exit tunnel, and that very slowly ejecting chains can delay ribosome recycling. The fact that ribosomal proteins have some of the most highly positively charged C-termini across the *E. coli* proteome suggests the intriguing possibility that their charge density did not evolve just to strengthen their binding affinity for rRNA but could also be beneficial by making them slow ejectors, thereby affording more time for potential cotranslational assembly processes to occur. While we have demonstrated that electrostatics are essential for extreme ejection times by running simulations without the charges present in the nascent chain C-termini, other factors must also play a role in determining ejection times. As can be seen in Figure 2d, some

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

proteins with typical ejection times (gray dots) have a similar number of positive charges in the C-terminus as proteins with very slow ejection times. We speculate that other factors that influence the ejection time could include the backbone structural propensity and the size of the amino acids in the protein sequence. Helical backbone preferences and large amino acids are more likely to sterically clash with the walls of the exit tunnel, while extended strand backbone preferences and small amino acids might make diffusion out of the tunnel sterically easier. This study is the first to our knowledge to provide evidence that the seemingly mundane act of diffusion of nascent proteins out of the exit tunnel can vary greatly between proteins, have downstream cellular consequences, and might be particularly biologically relevant to the synthesis of ribosomal proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.9b12264.

Full details of coarse-grained model parametrization and simulations, additional all-atom steered molecular dynamics results with different cantilever speeds, and GO pathway analysis details (PDF)

Fast ejection (AVI)

Intermediate ejection (AVI)

Slow ejection (AVI)

■ AUTHOR INFORMATION

Corresponding Author

Edward P. O'Brien − *Department of Chemistry, Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, and Institute for Computational and Data Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0001-9809-3273; Email: epo2@psu.edu

Authors

Daniel A. Nissley − *Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States*

Quyen V. Vu − *Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland;* orcid.org/0000-0002-9863-0486

Fabio Trovato − *Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States*

Nabeel Ahmed − *Department of Chemistry and Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, United States*

Yang Jiang − *Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0003-1100-9177

Mai Suan Li − *Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland; Institute for Computational Sciences and Technology, Ho Chi Minh City, Vietnam;* orcid.org/0000-0001-7021-7916

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.9b12264

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.P.O. acknowledges support from the National Science Foundation (MCB-1553291) for the simulation component and (ABI-1759860) for the bioinformatics component of this study, as well as the National Institutes of Health (R35-GM124818). Portions of numerical computations and data analysis in this work have been carried out on the CyberLAMP cluster, which is supported by NSF-MRI-1626251 and operated by the Institute for CyberScience at The Pennsylvania State University. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. M.S.L. acknowledges that this work was supported by Narodowe Centrum Nauki (grant no. 2015/19/B/ST4/02721), PLGrid Infrastructure in Poland, and the Department of Science and Technology, Ho Chi Minh City, Vietnam.

■ REFERENCES

(1) Rodnina, M. V. Translation in Prokaryotes. *Cold Spring Harbor Perspect. Biol.* **2018**, *10*, No. a032664.

(2) Trappl, K.; Mathew, M. A.; Joseph, S. Thermodynamic and Kinetic Insights into Stop Codon Recognition by Release Factor 1. *PLoS One* **2014**, *9*, e94058.

(3) Pierson, W. E.; et al. Uniformity of peptide release is maintained by methylation of release factors. *Cell Rep.* **2016**, *17*, 11−18.

(4) Zavialov, A. V.; Mora, L.; Buckingham, R. H.; Ehrenberg, M. Release of Peptide Promoted by the GGQ Motif of Class 1 Release Factors Regulates the GTPase Activity of RF3. *Mol. Cell* **2002**, *10*, 789−798.

(5) Kuhlenkoetter, S.; Wintermeyer, W.; Rodnina, M. V. Different substrate-dependent transition states in the active site of the ribosome. *Nature* **2011**, *476*, 351−355.

(6) Shoemaker, C. J.; Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E1392−E1398.

(7) Petrone, P. M.; Snow, C. D.; Lucent, D.; Pande, V. S. Side-chain recognition and gating in the ribosome exit tunnel. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16549−16554.

(8) Murakami, A.; Nakatogawa, H.; Ito, K. Translation arrest of SecM is essential for the basal and regulated expression of SecA. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12330−12335.

(9) Gumbart, J.; Schreiner, E.; Wilson, D. N.; Beckmann, R.; Schulten, K. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys. J.* **2012**, *103*, 331−341.

(10) Lu, J.; Deutsch, C. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *J. Mol. Biol.* **2008**, *384*, 73−86.

(11) Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P. In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E132−E140.

(12) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566−579.

(13) Berman, H. M.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(14) Brooks, B. R.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(15) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(16) Frishman, D.; Argos, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng., Des. Sel.* **1996**, *9*, 133−142.

(17) Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433−3434.

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

(18) Jorgensen, W. L.; et al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(19) Sillitoe, I.; et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res.* **2013**, *41*, D490−D498.
(20) Karanicolas, J.; Brooks, C. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **2002**, *11*, 2351−2361.
(21) Best, R. B.; Chen, Y. G.; Hummer, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of Arc repressor. *Structure* **2005**, *13*, 1755−1763.
(22) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Trigger factor slows Co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J. Am. Chem. Soc.* **2012**, *134*, 10920−10932.
(23) Betancourt, M. R.; Thirumalai, D. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361−369.
(24) Leininger, S. E.; Trovato, F.; Nissley, D. A.; O'Brien, E. P. Domain topology, stability, and translation speed determine mechanical force generation on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 5523−5532.
(25) Nissley, D. A.; O'Brien, E. P. Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome. *J. Phys. Chem. B* **2018**, *122*, 9927−9937.
(26) Fluitt, A.; Pienaar, E.; Viljoen, H. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* **2007**, *31*, 335−346.
(27) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13403−13408.
(28) Moore, B. L.; Kelley, L. A.; Barber, J.; Murray, J. W.; Macdonald, J. T. High-Quality Protein Backbone Reconstruction from Alpha Carbons Using Gaussian Mixture Models. *J. Comput. Chem.* **2013**, *34*, 1881−1889.
(29) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29*, 1460.
(30) Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275−291.
(31) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E.; et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(32) Hornak, V.; et al. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(33) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.
(34) Nosé, S. A unified formulation of the constant temperature molecular dynamics. *J. Chem. Phys.* **1984**, *81*, 511−519.
(35) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(36) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(38) Fritch, B.; et al. Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J. Am. Chem. Soc.* **2018**, *140*, 5077−5087.
(39) O'Brien, E. P.; Vendruscolo, M.; Dobson, C. M. Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.* **2012**, *3*, 868.
(40) Mohammad, F.; Green, R.; Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife* **2019**, *8*, 1−25.
(41) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R. S.; Weissman, J. S. Genome-Wide Analysis of in Vivo Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (Washington, DC, U. S.)* **2009**, *324*, 218−224.
(42) Kamenova, I.; et al. Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* **2019**, *10*, 1740.
(43) Shiber, A.; et al. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* **2018**, *561*, 268.
(44) Natan, E.; Wells, J. N.; Teichmann, S. A.; Marsh, J. A. Regulation, evolution and consequences of cotranslational protein complex assembly. *Curr. Opin. Struct. Biol.* **2017**, *42*, 90−97.

https://dx.doi.org/10.1021/jacs.9b12264
J. Am. Chem. Soc. 2020, 142, 6103−6110

Chapter 4. THE DRIVING FORCE FOR COTRANSLATIONAL PROTEIN FOLDING IS WEAKER IN THE RIBOSOME VESTIBULE DUE TO GREATER WATER ORDERING

4.1 INTRODUCTION

Cotranslational protein folding is the concomitant folding of a protein with its synthesis by the ribosome. During translation, the nascent protein emerges into the ribosome exit tunnel– the first microenvironment with which the nascent protein interacts. The first location where tertiary protein folding can occur is the ribosome vestibule, the last 3 nm of the ribosome exit tunnel (Figure 4.1a). Experiments and simulations have indicated that many domains can fold on the ribosome vestibule[36,42,49,144]. This process is of great interest to the scientific community because how a protein folds during its early stages can significantly impact its fate within the cell[145].

Experiments have observed that individual domains fold in the vestibule are often less stable than the same domain outside the exit tunnel when measured on translationally arrested ribosomes. Even just outside the vestibule, the native state is often less stable than in bulk solution[78,146,147]. Single-molecule laser optical tweezer experiments[80,81,148] have found that the folding process for two different proteins on stalled ribosomes becomes slower the closer the domain is to the ribosome's outer surface, with the trend line suggesting folding is slower still in the vestibule. Increasing the salt concentration leads to an enhanced rate of protein folding on the ribosome. However, this effect is not observed to a significant extent when considering the folding rate of isolated proteins. These findings suggest that electrostatic interactions with the ribosome surface contribute to the observed deceleration in folding rate[80]. In addition, the hydrophobic effect is the primary driving force of protein folding[24,149], and changes in salt concentration can also change the strength of the hydrophobic effect[150]. This suggests the possibility that the hydrophobic effect can be weakened in the presence of the ribosome.

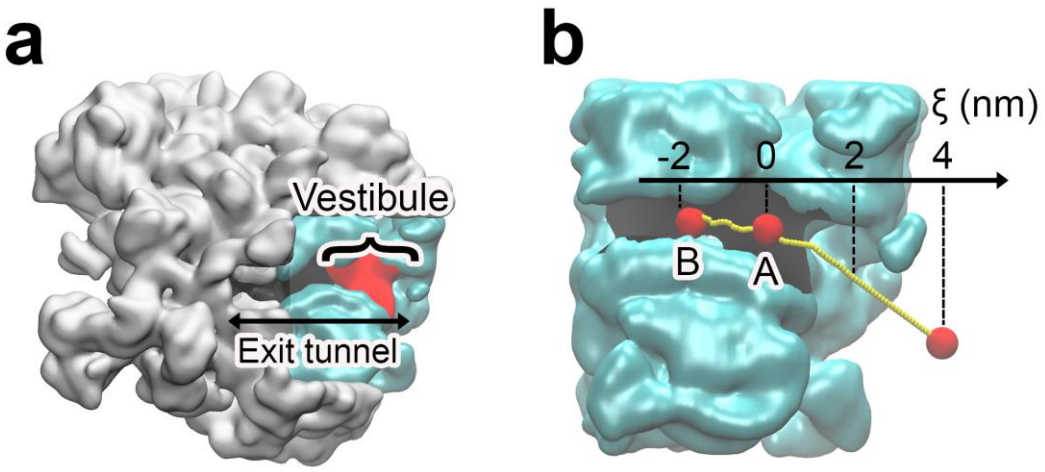


Figure 4.1: (a) Cross-section of the 50S subunit of *E.coli* highlighting the ribosome (gray), exit tunnel (black), and the last 3nm of the exit tunnel known as the 'vestibule' (red). (b) The portion of the ribosome exit tunnel used in the simulation. The center line of the exit tunnel is represented as a yellow dotted line, and the locations of points A and B (where we calculate the association) are highlighted.

In this work, we tested the novel hypothesis that the environment around the ribosome weakens the hydrophobic effect, thereby contributing to decreased protein stability and slowing folding. To do this, we used the physical chemistry approach and calculated the potential of mean force between two methanes (hydrophobic molecule) in the ribosome exit tunnel (Figure 4.1b) and bulk solution, as well as compared thermodynamic and water structure properties. Our key findings are:

1. Near the ribosome the contact minimum between two methane molecules is half as stable as in bulk solution, demonstrating that the hydrophobic effect is weakened in the presence of the ribosome.
2. Thermodynamic decomposition[151] and structural analyses[135–137] reveal that the weakening of the hydrophobic effect is due to the increased ordering of water molecules in the presence of the ribosome. Specifically, increased water ordering reduces the entropy gain of water released from the first-solvation shell upon association of the two hydrophobic groups. Hence, the driving force for the hydrophobic association is weakened.

3. Finally, we examine the implications of this finding for cotranslational protein folding by estimating how much this effect destabilizes a domain's folded state. The hydrophobic effect contributes about 60% to the free energy difference between the folded and unfolded state[25,26]. Therefore, we estimate that the free energy of protein stability is decreased by 60%×0.5=30%. For a typical protein of 80 residues that can fold in the vestibule[49,144] and has a free energy of stability of -25 kJ/mol in bulk[152], the stability of the folded state will be decreased by around -7.5 kJ/mol in the ribosome vestibule due to the weakening of the hydrophobic effect.

These results are significant because they identify a hitherto unknown effect of the ribosome on the primary driving force for protein folding, identify the molecular mechanism by which this occurs, and provide an explanation for several experimental observations. The results have broad implications for protein folding assembly and the cotranslational processing of nascent proteins by chaperones and enzymes.

4.2 PUBLICATION

4.2.1 *Abstract*

Interactions between the ribosome and nascent chain can destabilize folded domains in the ribosome exit tunnel's vestibule, the last 3 nm of the exit tunnel where tertiary folding can occur. Here, we test if a contribution to this destabilization is a weakening of hydrophobic association, the driving force for protein folding. Using all-atom molecular dynamics simulations, we calculate the potential-of-mean force between two methane molecules along the center line of the ribosome exit tunnel and in bulk solution. Associated methanes, we find, are half as stable in the ribosome's vestibule as compared to bulk solution, demonstrating that the hydrophobic effect is weakened by the presence of the ribosome. This decreased stability arises from a decrease in the amount of water entropy gained upon the association of the methanes. And this decreased entropy gain originates from water molecules being more ordered in the vestibule as compared to bulk solution. Therefore, the hydrophobic effect is weaker in the vestibule because waters released from the first solvation shell of methanes upon association do

not gain as much entropy in the vestibule as they do upon release in bulk solution. These findings mean that nascent proteins pass through a ribosome vestibule environment that can destabilize folded structures, which has the potential to influence co-translational protein folding pathways, energetics, and kinetics.

DOI: 10.1039/D1SC01008E

Publication source: https://pubs.rsc.org/en/content/articlelanding/2021/sc/d1sc01008e

4.2.2 *Author contribution statements*

Quyen V. Vu
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

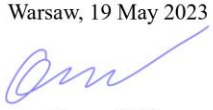
STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **The Driving Force for Co-Translational Protein Folding Is Weaker in the Ribosome Vestibule Due to Greater Water Ordering.** *Chem. Sci.* **2021**, *12* (35), 11851–11857.

My contribution was performing simulations, analyzing the results, preparing figures, and participating in writing the manuscript.

Warsaw, 19 May 2023



Quyen V. Vu

Yang Jiang
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802

PENNSTATE
1855

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **The Driving Force for Co-Translational Protein Folding Is Weaker in the Ribosome Vestibule Due to Greater Water Ordering.** *Chem. Sci.* **2021**, *12* (35), 11851–11857.

My contribution was analyzing the results, preparing figures, and writing the manuscript.

PA USA, 22 May 2023

Yang Jiang

Yang Jiang

Prof. Mai Suan Li, PhD
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **The Driving Force for Co-Translational Protein Folding Is Weaker in the Ribosome Vestibule Due to Greater Water Ordering.** *Chem. Sci.* **2021**, *12* (35), 11851–11857.

I was responsible for interpreting the data, writing the manuscript, and supervising the project.

Warsaw, 26 May 2023

Mai Suan Li

Prof. Edward P. O'Brien, PhD
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802
E-mail: epo2@psu.edu
Tel: 1-814-867-5100

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V.; Jiang, Y.; Li, M. S.; O'Brien, E. P. **The Driving Force for Co-Translational Protein Folding Is Weaker in the Ribosome Vestibule Due to Greater Water Ordering.** *Chem. Sci.* **2021**, *12* (35), 11851–11857.

My contribution consisted of designing the research methodology, interpreting the data, writing the manuscript, and supervising the overall project.

PA USA, 26 May 2023

Edward P. O'Brien

4.2.3 *Publication*

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2021, **12**, 11851

All publication charges for this article have been paid for by the Royal Society of Chemistry

The driving force for co-translational protein folding is weaker in the ribosome vestibule due to greater water ordering†

Quyen V. Vu, ‡[a] Yang Jiang, ‡[b] Mai Suan Li *[ac] and Edward P. O'Brien*[bde]

Received 19th February 2021
Accepted 2nd August 2021

DOI: 10.1039/d1sc01008e

rsc.li/chemical-science

Interactions between the ribosome and nascent chain can destabilize folded domains in the ribosome exit tunnel's vestibule, the last 3 nm of the exit tunnel where tertiary folding can occur. Here, we test if a contribution to this destabilization is a weakening of hydrophobic association, the driving force for protein folding. Using all-atom molecular dynamics simulations, we calculate the potential-of-mean force between two methane molecules along the center line of the ribosome exit tunnel and in bulk solution. Associated methanes, we find, are half as stable in the ribosome's vestibule as compared to bulk solution, demonstrating that the hydrophobic effect is weakened by the presence of the ribosome. This decreased stability arises from a decrease in the amount of water entropy gained upon the association of the methanes. And this decreased entropy gain originates from water molecules being more ordered in the vestibule as compared to bulk solution. Therefore, the hydrophobic effect is weaker in the vestibule because waters released from the first solvation shell of methanes upon association do not gain as much entropy in the vestibule as they do upon release in bulk solution. These findings mean that nascent proteins pass through a ribosome vestibule environment that can destabilize folded structures, which has the potential to influence co-translational protein folding pathways, energetics, and kinetics.

Introduction

The association of hydrophobic side chains is the primary driving force for protein folding.[1,2] The first location that tertiary protein folding can occur is in the ribosome vestibule, corresponding to the last 3 nm of the ribosome exit tunnel (red region in Fig. 1a). The nascent polypeptide chain passes through the 10 nm exit tunnel that is lined with ribosomal proteins and RNA, and out into the cellular milieu. Simulations first predicted,[3,4] and experiments later verified,[5,6] that many domains are sterically permitted to fold in the ribosome vestibule because the vestibule is wider than the rest of the tunnel (diameter is about 3 nm as compared to 1.5 nm).

A variety of changes in folding thermodynamics and kinetics occur as a domain passes through the vestibule and outside the exit tunnel, and some can be protein specific. While co-translational folding can occur in the vestibule according to computer simulations[4] and cryo-EM structures,[7] NMR experiments,[8] and fraction-full-length protein profiles, which are proportional to force,[9] have found that, with the exception of ADR1 protein,[10] individual domains in the vestibule are often less stable as compared to the same domain outside the exit tunnel when measured on translationally arrested ribosomes. Even just outside the vestibule the native state is often less stable than in bulk solution.[11,12]

In terms of kinetics, single molecule laser optical tweezer experiments[13–15] have found that the folding process for two different proteins on stalled ribosomes becomes slower the closer the domain is to the ribosome's outer surface, with the trend line suggesting folding is slower still in the vestibule. Indeed, a number of computer simulations of co-translational folding find slower folding rates near the outer ribosome surface and in the vestibule.[4,16] Additionally, increasing the salt concentration in solution was found to accelerate domain folding just outside the vestibule indicating electrostatic interactions play a role in this slowdown.[13] Changes in salt concentration can also change the strength of the hydrophobic effect,[17]

[a]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland. E-mail: masli@ifpan.edu.pl*

[b]*Department of Chemistry, Penn State University, University Park, Pennsylvania, USA. E-mail: epo2@psu.edu*

[c]*Institute for Computational Sciences and Technology, Quang Trung Software City, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam*

[d]*Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Penn State University, University Park, Pennsylvania, USA*

[e]*Institute for Computational and Data Sciences, Penn State University, University Park, Pennsylvania, USA*

† Electronic supplementary information (ESI) available: Full details of all-atom molecular dynamics simulations, setup of umbrella sampling simulations, procedure of enthalpy–entropy decomposition, entropy of water and tetrahedral parameters calculations. See DOI: 10.1039/d1sc01008e

‡ These authors contributed equally.

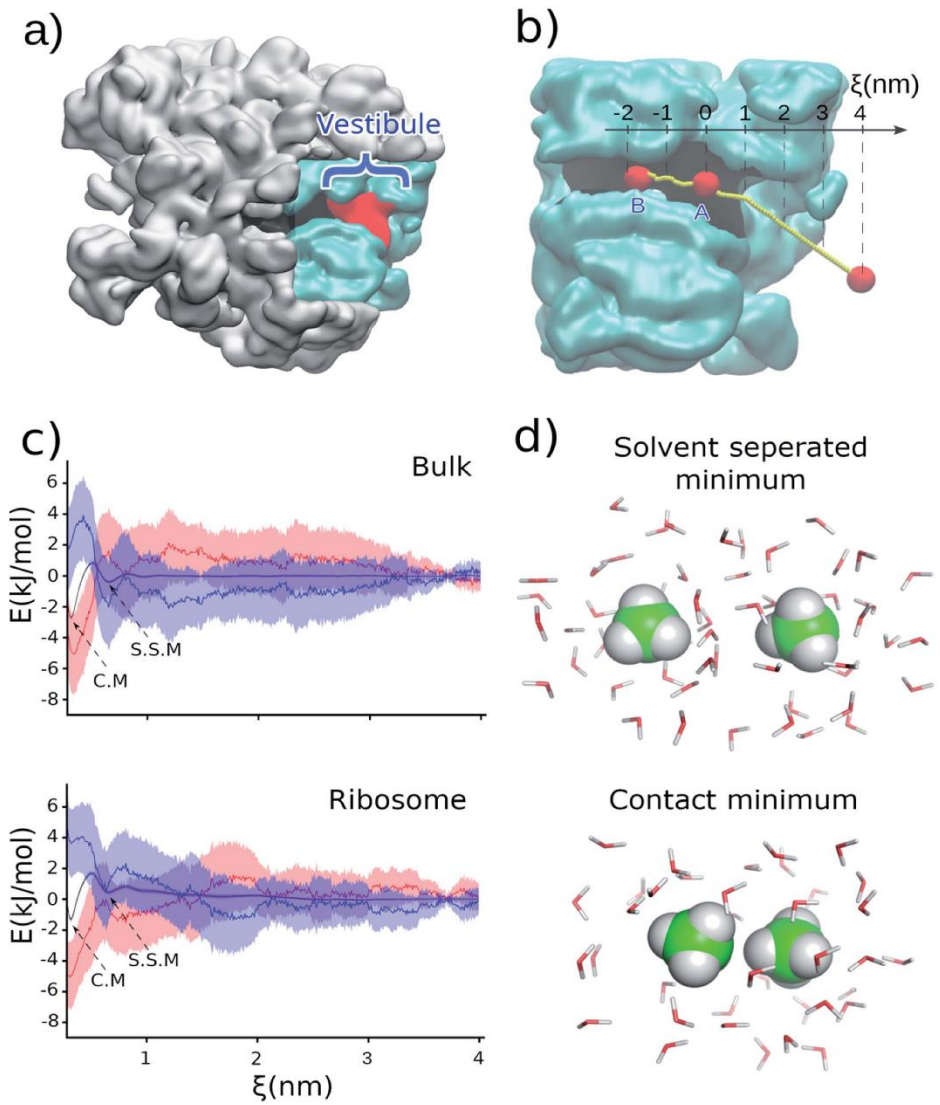


Fig. 1 (a) Cross-section of the 50S subunit of *E. coli* highlighting the ribosome (gray), exit tunnel (black), and last 3 nm of the exit tunnel known as the 'vestibule' (red) where tertiary folding can occur. (b) The portion of the ribosome exit tunnel used in the simulations. The center-line of the exit tunnel is represented as a yellow dotted line. (c) Potential of mean force (ΔG, black), enthalpy (ΔH, blue), and negative entropy term ($-T\Delta S$, red) in bulk (upper) and in the ribosome exit tunnel (lower) between two methanes with one methane at point *A*. The shaded regions present 95% confident intervals calculated from bootstrapping. (d) A snapshot of methanes in solvent-separated minimum and in contact minimum configurations.

with higher salt concentration leading to increased water density around the component ions compared to pure water.[18] This suggests the possibility that variation in the hydrophobic effect could arise in the exit tunnel vestibule due to the electrostatic environment it creates.

In this study, we examine whether there is a decrease in the affinity of hydrophobes for one another – a classic measure of the strength of the hydrophobic effect – in the ribosome's vestibule.

To test this hypothesis we carried out classical, all-atom molecular dynamics simulations of the association of two

Table 1 Free energy, enthalpy, and entropy at contact minimum and solvent separated minimum (95% confidence interval about the mean is reported in parentheses)

		Contact minimum (kJ mol^{-1})			Solvent-separated minimum (kJ mol^{-1})			Contact minimum minus solvent-separated minimum (kJ mol^{-1})		
System		ΔG	ΔH	$T\Delta S$	ΔG	ΔH	$T\Delta S$	$\Delta\Delta G$	$\Delta\Delta H$	$T\Delta\Delta S$
Point *A*	Bulk	−2.71	2.09	4.80	−0.40	−1.72	−1.32	−2.31	3.81	6.12
		(−2.82, −2.61)	(−0.51, 4.72)	(2.20, 7.42)	(−0.49, −0.29)	(−4.56, 1.25)	(−4.12, 1.58)	(−2.37, −2.26)	(2.88, 4.77)	(5.23, 7.07)
	Ribosome	−1.31	3.58	4.90	0.43	1.05	0.62	−1.74	2.54	4.28
		(−1.47, −1.17)	(1.32, 5.78)	(2.63, 7.07)	(0.30, 0.56)	(−1.45, 3.50)	(−1.78, 3.13)	(−1.78, −1.70)	(1.09, 3.81)	(2.83, 5.55)
Point *B*	Bulk	−3.40	—	—	−0.37	—	—	−3.03	—	—
		(−3.60, −3.10)			(−0.49, −0.18)			(−3.16, −2.82)		
	Ribosome	−2.35	—	—	0.09	—	—	−2.44	—	—
		(−3.07, −1.82)			(−0.31, 0.45)			(−3.05, −2.01)		

hydrophobic molecules both in the presence and absence of the *Escherichia coli* ribosome at 310 K, the optimal growth temperature of this organism. We calculate the potential of mean force between two methanes (CH_4) along the center line of the ribosome exit tunnel (Fig. 1b). We study methane because it is a model compound for the alanine side chain. Methanes are also closely chemically related to methyl moieties (CH_2) – the most common building block for more complex hydrophobic molecules. The transfer free energy of hydrophobic molecules is directly proportional to the number of methyl's, suggesting our results for methane will be relevant to larger hydrophobic side chains. Additionally, the small number of degrees of freedom of methane means that we can obtain precise statistics in the simulations. The center line is the line along the exit tunnel that is maximally separated from all ribosomal atoms (yellow line in Fig. 1b). Since the hydrophobic effect is water mediated, we calculate association along this center line so that the methanes are always solvated (see radial distribution functions in Fig. S1†), and do not come into direct contact with the exit tunnel walls.

Results and discussion

Ribosome reduces the hydrophobic driving force for protein folding

Holding one methane fixed at position A (labelled in Fig. 1b), which is about 2.5 nm into the exit tunnel, we find that the potential of mean force between this methane and a methane brought along the center line exhibits a solvent separated minimum (labeled 'S.S.M.' in Fig. 1c) and contact minimum (labeled 'C.M.' in Fig. 1c). Since the ribosome exit tunnel is 1.5 nm in diameter, on average, we focus on testing for changes in the hydrophobic effect at distances less than this. Therefore, we examine free energy differences between the contact minimum and solvent-separated minimum. We find the contact minimum is 1.74 kJ mol^{-1} (95% confidence interval (CI): [1.70, 1.78] kJ mol^{-1}, calculated from bootstrapping) more stable than the solvent separated minimum (Table 1). Carrying out this simulation in bulk solution (*i.e.*, without the ribosome) along the same spatial path shown in Fig. 1b, the contact minimum is 2.31 kJ mol^{-1} (95% CI: [2.26, 2.37], bootstrapping) more stable than the solvent separated minimum (Table 1). Thus, the presence of the ribosome vestibule decreases the stability of the associated methanes (p-value $< 1 \times 10^{-6}$, one-sided permutation test).

To test if this conclusion is robust at different positions along the center line of the tunnel, we carried out the same simulations and analyses but with the methanes associated approximately 2 nm further inside the exit tunnel (point B in Fig. 1b – lower region of the ribosome exit tunnel, where helix formation has been experimentally observed to occur). We find that although the relative stabilities are different as compared to point A (Table 1 and Fig. S2 in ESI Text†), which is to be expected in the heterogeneous environment along the exit tunnel, it is still the case that the presence of the ribosome leads to a less stable associated state relative to the solvent separated minimum (−2.44 kJ mol^{-1}, 95% CI [−3.05, −2.01] in the ribosome *versus* −3.03 kJ mol^{-1}, 95% CI [−3.16, −2.82] in bulk). These results are in agreement with an earlier study on cylindrical confinement.[19] We conclude from these data that the presence of the ribosome decreases the affinity of hydrophobic molecules for one another in the exit tunnel where co-translational tertiary protein folding can occur.

We note that because we are projecting the non-linear path of the methane (yellow line in Fig. 1b) onto the linear reaction coordinate ξ (black line in Fig. 1b) this leads to the situation that in bulk solution the difference in contact *versus* solvent separated minimum stabilities depend on whether they are computed using the path to point *A* or *B* (Fig. 1b and Table 1). The difference in the curvature of the path near point *A* and *B* leads to the projection of different probability densities onto to ξ. This is acceptable, however, because we are only interpreting the thermodynamic properties along the same path – *e.g.*, association at point *A* in bulk *versus* association at point *A* in the ribosome exit tunnel.

To understand why the hydrophobic effect is weakened by the ribosome we calculated the entropy and enthalpy of association at position *A* at 310 K using data from multiple temperatures (see Methods†). We find that upon going from the solvent-separated minimum to the contact minimum there is no statistically significant difference in the enthalpy of

Table 2 Translational, rotational, and total entropy of water in different regions of exit tunnel (95% confidence interval about the mean is reported in parentheses)

	System		
Property	Bulk	Point *A*	Point *B*
Translational entropy (J K^{-1} mol^{-1})	56.04 (55.72, 56.22)	47.89 (47.42, 48.48)	44.53 (43.08, 45.64)
Rotational entropy (J K^{-1} mol^{-1})	13.15 (13.03, 13.24)	13.13 (12.40, 13.89)	12.69 (12.52, 12.93)
Total entropy (J K^{-1} mol^{-1})	69.19 (68.75, 69.46)	61.02 (60.18, 62.37)	57.22 (56.01, 58.26)

association ($\Delta\Delta H$) calculated in bulk solution (3.81 kJ mol^{-1}, 95% CI: [2.88, 4.77]) *versus* in the presence of the ribosome (2.54 kJ mol^{-1}, 95% CI: [1.09, 3.81], Table 1). The *p*-value is 0.08 (one-sided permutation test) for the difference between these two values. Thus, changes in enthalpy of association do not cause the change in stability of hydrophobic association in the

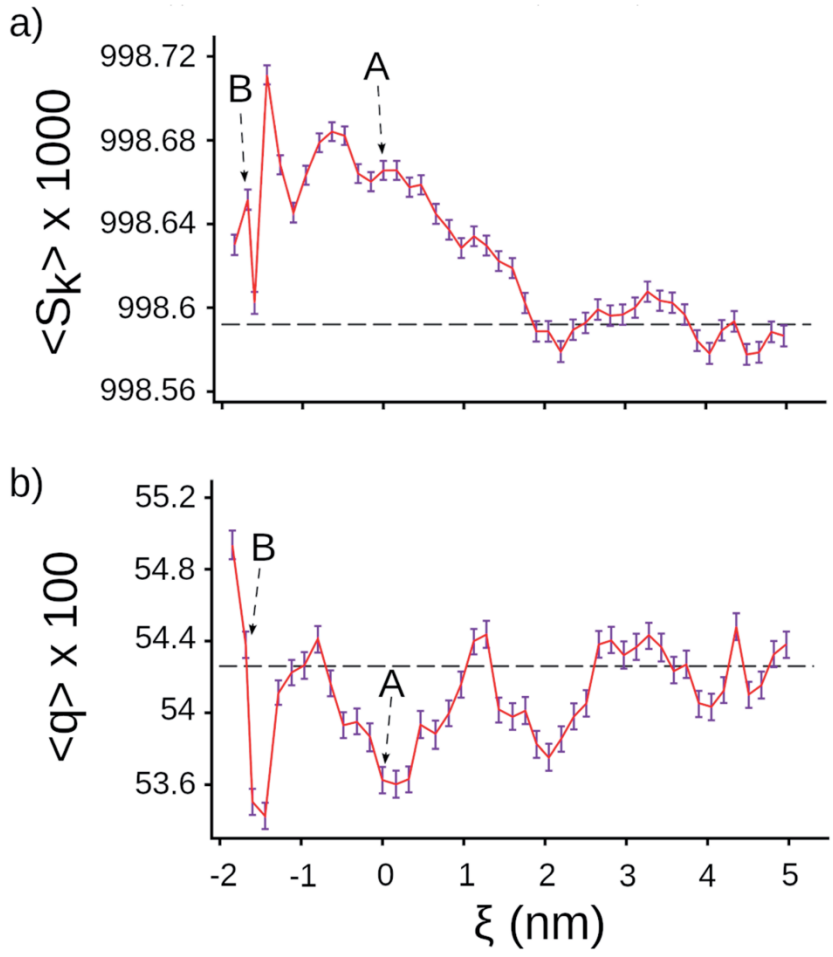


Fig. 2 Tetrahedral parameters for water molecules along the center line of the ribosome. (a) Distance order parameter S_k (eqn (S4)†), (b) orientational order parameter *q* (eqn (S5)†). The horizontal lines are the average value calculated for water molecules in bulk, the error bar presents 95% confident intervals calculated from bootstrapping.

vestibule. We do find a difference, however, in the entropy of association. In bulk solution the entropic term ($T\Delta\Delta S$) is 6.12 kJ mol^{-1} (95% CI [5.23, 7.07]), while on the ribosome it is 4.28 kJ mol^{-1} (95% CI [2.83, 5.55]). The difference between these two values is statistically significant (p-value = 0.02, one-sided permutation test). Both entropy terms are positive, meaning that there is a gain in entropy upon association of the two methanes. However, the gain in entropy is smaller in the presence of the ribosome (4.28 kJ mol^{-1} *versus* 6.12 kJ mol^{-1}). Thus, the ribosome-induced weakening of hydrophobic association arises from a smaller gain in entropy upon going from the solvent separated configuration of the methanes to the associated state.

The ribosome exit tunnel has more ordered solvent compared to bulk solution

Previous studies have demonstrated that the entropy gain upon the association of hydrophobes arises from the release of several ordered water molecules from the first solvation shell of the methanes, and their subsequent gain in rotational and translational entropy.[20] This suggests the hypothesis that water molecules are more ordered in the exit tunnel as compared to bulk solution, resulting in a smaller gain in entropy when waters are released upon methane association. Indeed, an earlier simulation study observed greater water ordering in the exit tunnel and reduced rotational entropy.[21] We tested this hypothesis in two ways. First, we computed the entropy of water in the region around positions *A* and *B* in the absence of methanes, as well as in bulk, using the two-phase thermodynamic method[22–24] (see Methods in ESI Text†). We find the total entropy of water decreases from 69.19 J K^{-1} mol^{-1} (95% CI: [68.75, 69.46]) in bulk solution, to the smaller values of 61.02 (95% CI: [60.18, 62.37]) and 57.22 (95% CI: [56.01, 58.26]) J K^{-1} mol^{-1}, respectively, at points *A* and *B* (see Table 2). And that this decrease arises from a large decrease in water's translational entropy and a smaller decrease in waters rotational entropy (Table 2). (Note well, since the TIP3P water molecules in our simulations are rigid the vibrational entropy is zero and not reported in Table 2.) Thus, water molecules have less translational and rotational entropy in the vestibule.

Next, we tested whether we could detect signatures of greater water ordering in the exit tunnel by using the tetrahedral orientational (q) and translational (S_k) order parameters.[25–27] These two metrics measure two different aspects of how closely five water molecules are to forming a tetrahedron, which is the minimum potential energy structure. We computed q and S_k at each point along the center line by selecting the water molecule that was closest to that point and its four nearest-neighbor water molecules (see Methods in ESI Text†). We find that S_k is higher in the exit tunnel than in bulk (Fig. 2a), indicating that the water molecules adopt a more tetrahedral structure in terms of their distances from the central water molecule. The orientational parameter q, however, fluctuates above and below the bulk value, indicating the ribosome distorts the water cluster angular configuration to be more or less tetrahedral at different points along the tunnel (Fig. 2b). The angular degrees-of-freedom of the tetrahedron are softer than the distance degrees-of-freedom, meaning that it takes more energy to change the distances than the angles. Taken together, these results demonstrate that water molecules are more ordered in the exit tunnel and have decreased entropy. They also indicate that the smaller entropy gain upon association of methanes arises from the fact that the newly liberated waters are released into an environment where the water molecules are more ordered and have less entropy.

Ribosome sites that could influence the hydrophobic effect

Negatively charged residues and groups, such as the phosphates of RNA, lead to more water ordering than positively charged residues, while polar and non-polar residues cause the least changes in water structure.[28] For small hydrophobic residues water's tetrahedral structure in the first solvation shell is equivalent to bulk water.[29] Thus, greater water ordering in the exit tunnel is due in part to charged amino acids and the phosphate groups of 23S RNA. For completeness, we have identified all charged and polar residues within 1.5 nm of points *A* and *B* and report them in Table 3.

Estimated effect on co-translational protein folding

We can estimate how much this weakening of the hydrophobic effect will affect the stability of a typical protein domain. We first note that the stability of the contact minimum is half of its bulk value in the exit tunnel (−1.31 kJ mol^{-1} 95% CI [−1.47, −1.17] *versus* −2.71 kJ mol^{-1} 95% CI [−2.82, −2.61]). Next, we note that the hydrophobic effect contributes 60%, on average, to the free energy difference between the folded and unfolded states.[30,31] Therefore, the weakening of the hydrophobic effect will decrease the folded state stability by around 30% (=60%/2). A typical 80 residue protein (which can fold in the ribosome vestibule[3,6]) has a free energy of stability of −25 kJ mol^{-1} in bulk solution.[32] Hence, the stability of folded state is decreased by around −7.5 kJ mol^{-1} (=−25 kJ mol^{-1} × 0.5 × 0.6) due to the reduction of the hydrophobic effect in the exit tunnel. While

Table 3 Distance between methane molecules at the contact minimum at points *A* and *B* (Fig. 1b) and residues lining the ribosome exit tunnel. Residue indices are followed PDBID: 3R8T

	Point *A*			Point *B*		
Index	Residue	Chain	Distance (Å)	Residue	Chain	Distance (Å)
1	A507	A (23S)	11.4	LYS83	S (L22)	6.77
2	A508	A (23S)	12.0	A471	A (23S)	7.95
3	C1335	A (23S)	12.0	GLN72	T (L23)	8.55
4	G1334	A (23S)	13.0	ARG84	S (L22)	10.0
5	HIS70	T (L23)	13.1	A472	A (23S)	10.1
6	A1322	A (23S)	13.6	A470	A (23S)	10.7
7	A492	A (23S)	13.9	G1259	A (23S)	12.4
8	A91	A (23S)	14.1	C461	A (23S)	12.6
9	C1319	A (23S)	14.4	A1322	A (23S)	13.2
10	U92	A (23S)	14.4	A508	A (23S)	14.7
11	A1336	A (23S)	15.0	U1258	A (23S)	14.9

a rough estimate, it suggests that destabilization on the order of 10's of kJ mol^{-1} is possible.

Conclusions

In summary, the hydrophobic effect is weaker in the ribosome vestibule because water molecules are more ordered than in bulk. This greater ordering decreases the entropy gain of water molecules released from the first hydration shell of hydrophobic moieties, thereby weakening the entropic driving force for hydrophobic association.

More broadly, understanding the folding of proteins at their earliest stage of existence, *i.e.*, during synthesis, is an area of intense research efforts because what happens during this crucial period can influence the fate of a protein in a cell.[33] Studies have found that for some proteins their folding pathways can differ from that of bulk solution[34–36] due to the N- to C-terminal synthesis of proteins, interactions between the nascent chain and ribosome,[11,13] and the speed of protein synthesis.[37] To this list, our results indicate that the weakening of the hydrophobic effect – the primary driving force of protein folding – is also likely to influence nascent protein folding. The hydrophobic effect is still present in the vestibule, and hence our results are consistent with observations that protein domains do fold in the exit tunnel. However, our results indicate the structures that do form will not be as thermodynamically stable as in bulk solution. This decreased stability in the vestibule has the potential to slow down the rate of protein co-translational folding, and may be a mechanism contributing to slower folding and decreased stability.

In addition, post-translation protein folding may be aided by the presence of the ribosome[38–40] potentially through outer ribosome surface interactions that are not accessible to nascent chains emerging from the exit tunnel. All of this points to a rich set of scenarios of the role of the ribosome on co- and post-translational protein folding, and the role of solvent in mediating nascent protein behavior.

Data availability

All the data that support the findings of this study are available from the corresponding authors upon reasonable request.

Author contributions

E. P. O. designed the research. Q. V. V. and Y. J. conducted the simulations. All authors analyzed the results, wrote and reviewed the article.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

EPO acknowledges support from the National Science Foundation (MCB-1553291) as well as the National Institutes of Health (R35-GM124818). Portions of numerical computations in this work have been carried out on the XSEDE supercomputer, which is supported by MCB-160069. MSL acknowledges that this work was supported by the National Science Centre, Poland (grant 2019/35/B/ST4/02086), PLGrid Infrastructure in Poland, and the Department of Science and Technology, Ho Chi Minh city, Vietnam (grant 07/2019/HĐ-KHCNTT).

Notes and references

1 K. A. Dill, *Biochemistry*, 1990, **29**, 7133–7155.
2 M. Compiani and E. Capriotti, *Biochemistry*, 2013, **52**, 8601–8624.
3 E. P. O'Brien, S. T. D. Hsu, J. Christodoulou, M. Vendruscolo and C. M. Dobson, *J. Am. Chem. Soc.*, 2010, **132**, 16928–16937.
4 E. P. O'Brien, J. Christodoulou, M. Vendruscolo and C. M. Dobson, *J. Am. Chem. Soc.*, 2011, **133**, 513–526.
5 O. B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, J. D. Puglisi, E. P. O'Brien, R. Beckmann and G. von Heijne, *Cell Rep.*, 2015, **12**, 1533–1540.
6 J. Marino, G. Von Heijne and R. Beckmann, *FEBS Lett.*, 2016, **590**, 655–660.
7 O. B. Nilsson, A. A. Nickson, J. J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. Von Heijne and J. Clarke, *Nat. Struct. Mol. Biol.*, 2017, **24**, 221–225.
8 L. D. Cabrita, A. M. E. Cassaignau, H. M. M. Launay, C. A. Waudby, T. Wlodarski, C. Camilloni, M. E. Karyadi, A. L. Robertson, X. Wang, A. S. Wentink, L. S. Goodsell, C. A. Woolhead, M. Vendruscolo, C. M. Dobson and J. Christodoulou, *Nat. Struct. Mol. Biol.*, 2016, **23**, 278–285.
9 G. Kemp, R. Kudva, A. de la Rosa and G. von Heijne, *J. Mol. Biol.*, 2019, **431**, 1308–1314.
10 F. Wruck, P. Tian, R. Kudva, R. B. Best, G. von Heijne, S. J. Tans and A. Katranidis, *Commun. Biol.*, 2021, **4**, 1–8.
11 A. J. Samelson, M. K. Jensen, R. A. Soto, J. H. D. Cate and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 13402–13407.
12 M. K. Jensen, A. J. Samelson, A. Steward, J. Clarke and S. Marqusee, *J. Biol. Chem.*, 2020, **295**, 11410–11417.
13 C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco and C. Bustamante, *Science*, 2011, **334**, 1723–1727.
14 K. Liu, J. E. Rehfus, E. Mattson and C. M. Kaiser, *Protein Sci.*, 2017, **26**, 1439–1451.
15 K. Liu, K. Maciuba and C. M. Kaiser, *Mol. Cell*, 2019, **74**, 310–319.
16 P. Tian, A. Steward, R. Kudva, T. Su, P. J. Shilling, A. A. Nickson, J. J. Hollins, R. Beckmann, G. Von Heijne, J. Clarke and R. B. Best, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E11284–E11293.
17 T. Ghosh, A. Kalra and S. Garde, *J. Phys. Chem. B*, 2005, **109**, 642–651.
18 R. M. Leberman and A. M. Soper, *Nature*, 1995, **378**, 364–366.
19 S. Vaitheeswaran and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17636–17641.
20 D. E. Smith, L. Zhang and A. D. J. Haymet, *J. Am. Chem. Soc.*, 1992, **114**, 5875–5876.

21 D. Lucent, C. D. Snow, C. E. Aitken and V. S. Pande, *PLoS Comput. Biol.*, 2010, **6**, e1000963.
22 S. T. Lin, M. Blanco and W. A. Goddard, *J. Chem. Phys.*, 2003, **119**, 11792–11805.
23 S. T. Lin, P. K. Maiti and W. A. Goddard, *J. Phys. Chem. B*, 2010, **114**, 8191–8198.
24 T. A. Pascal, S. T. Lin and W. A. Goddard, *Phys. Chem. Chem. Phys.*, 2011, **13**, 169–181.
25 P. Chau and A. J. Hardwick, *Mol. Phys.*, 1998, **93**, 511–518.
26 J. R. Errington and P. G. Debenedetti, *Nature*, 2001, **409**, 318–321.
27 E. Duboué-Dijon and D. Laage, *J. Phys. Chem. B*, 2015, **119**, 8406–8418.
28 B. Qiao, F. Jiménez-Ángeles, T. D. Nguyen and M. O. De La Cruz, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 19274–19281.
29 T. Hajari and S. Bandyopadhyay, *J. Chem. Phys.*, 2017, **146**, 225104.
30 C. N. Pace, H. Fu, K. L. Fryar, J. Landua, S. R. Trevino, B. A. Shirley, M. M. N. Hendricks, S. Iimura, K. Gajiwala, J. M. Scholtz and G. R. Grimsley, *J. Mol. Biol.*, 2011, **408**, 514–528.
31 C. N. Pace, B. A. Shirley, M. McNutt and K. Gajiwala, *FASEB J.*, 1996, **10**, 75–83.
32 D. De Sancho, U. Doshi and V. Munoz, *J. Am. Chem. Soc.*, 2009, **131**, 2074–2075.
33 A. K. Sharma and E. P. O'Brien, *Curr. Opin. Struct. Biol.*, 2018, **49**, 94–103.
34 O. B. Nilsson, A. A. Nickson, J. J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. Von Heijne and J. Clarke, *Nat. Struct. Mol. Biol.*, 2017, **24**, 221–225.
35 A. J. Samelson, E. Bolin, S. M. Costello, A. K. Sharma, E. P. O'Brien and S. Marqusee, *Sci. Adv.*, 2018, **4**, eaas9098.
36 M. Liutkute, E. Samatova and M. V. Rodnina, *Biomolecules*, 2020, **10**, 97.
37 F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M. V. Rodnina and A. A. Komar, *Mol. Cell*, 2016, **61**, 341–351.
38 B. Das, S. Chattopadhyay and C. Das Gupta, *Biochem. Biophys. Res. Commun.*, 1992, **183**, 774–780.
39 S. Chattopadhyay, S. Pal, D. Pal, D. Sarkar, S. Chandra and C. Das Gupta, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 1999, **1429**, 293–298.
40 A. Basu, D. Samanta, D. Das, S. Chowdhury, A. Bhattacharya, J. Ghosh, A. Das and C. DasGupta, *Biochem. Biophys. Res. Commun.*, 2008, **366**, 598–603.

Chapter 5. IS POSTTRANSLATIONAL FOLDING MORE EFFICIENT THAN REFOLDING FROM A DENATURED STATE: A COMPUTATIONAL STUDY

5.1 INTRODUCTION

Protein folding is a fundamental biological process, and the folding of isolated proteins has been studied extensively for over 50 years. *In vivo*, proteins are synthesized by the ribosomes during the nonequilibrium translation process. It has been shown that many proteins fold cotranslationally as they begin to emerge from the exit tunnel and acquire tertiary structure before their synthesis is complete[54–61]. During protein synthesis, the ribosome confines nascent proteins within a narrow region of the exit tunnel, which restricts their ability to self-interact and form tertiary structures. Consequently, the folding mechanisms of proteins may differ on and off the ribosome.

Experimental and computational studies have investigated the folding of a few proteins on and off the ribosome[76,78,80,83–85]. The evidence suggests that the ribosome's role in protein folding is protein-specific. For example, previous studies on titin I27 and src SH3 indicate that their folding pathways are the same on and off the ribosome[76,85]. On the other hand, coarse-grained molecular simulations find that folding in the presence of ribosome is more efficient for multi-domain protein SufI and deeply knotted protein Tp0624 compared to the absence of ribosome[83,84]. *In vivo*, nascent proteins diffuse into the cytosol after synthesis; if folding is not completed on the ribosome, it may complete posttranslationally. Hence, the ribosome may only influence the formation of intermediate states, which nonetheless can change the outcome of folding[148,153]. Therefore, the influence of the ribosome on the folding of proteins remains unclear due to the relative paucity of experimental and computational data.

In this study, we conducted coarse-grained molecular dynamics simulations of protein synthesis and post-translational folding and protein refolding from a denatured state to investigate the ribosome's influence on protein folding mechanisms. We focused on three *E. coli* enzyme proteins: dihydrofolate reductase (DHFR), type III chloramphenicol acetyltransferase (CAT-III), and D-alanine–D-alanine ligase B (DDLB). Our simulations were analyzed using

various folding analyses, including a new entanglement parameter to decipher the differences in folding on and off the ribosome.

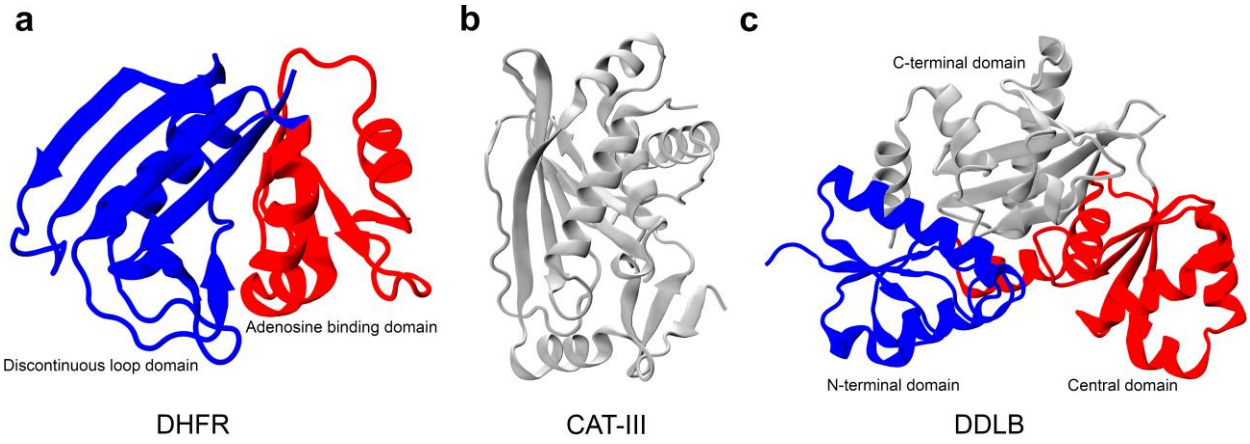


Figure 5.1: Crystal structures of DHFR, CAT-III, and DDLB proteins with domain-based coloring (a) Crystal structure of DHFR the discontinuous loop and the adenosine binding domains are shown in blue and red, respectively. (b) CAT-III is a single-domain protein shown in grey, and (c) DDLB protein with the N-terminal, central, and C-terminal domains shown in blue, red, and grey, respectively.

Our key findings suggest that the ribosome's influence on protein folding mechanisms varies depending on the size and complexity of the protein. DHFR folds more efficiently due to protein synthesis, while the ribosome does not promote the folding of CAT-III and DDLB and may contribute to the formation of intermediate misfolded states during translation. These misfolded states persist after translation and do not convert to the native state over a long period.

We also found that the sequence of secondary structure formations significantly differed for DHFR, while CAT-III and DDLB were robust on and off the ribosome. Additionally, we hypothesized that the native topologies of CAT-III and DDLB may lead to a large proportion of misfolding due to the presence of native entanglement. Recent research has predicted a link between misfolding involving a change in entanglement status and long-lived misfolded states[116,117]. Our analysis shows that DHFR does not contain any entanglement in its native structure, while CAT-III and DDLB have many native entanglements.

Considering the existence of entanglement, we combine the fraction of native contacts and the degree of entanglement to characterize the protein folding pathways. Our findings

indicate that protein synthesis assists the folding of DHFR by avoiding non-native entangled states compared to refolding from the unfolded ensemble. Conversely, non-native entangled states act as a kinetic trap in both refolding and posttranslational folding of CAT-III and DDLB.

Our study highlights the complex interplay between the ribosome and protein folding and provides insight into the mechanisms of protein folding on and off the ribosome.

5.2 PUBLICATION

5.2.1 *Abstract*

The folding of proteins into their native conformation is a complex process that has been extensively studied over the past half-century. The ribosome, the molecular machine responsible for protein synthesis, is known to interact with nascent proteins, adding further complexity to the protein folding landscape. Consequently, it is unclear whether the folding pathways of proteins are conserved on and off the ribosome. The main question remains: to what extent does the ribosome help proteins fold? To address this question, we used coarse-grained molecular dynamics simulations to compare the mechanisms by which the proteins dihydrofolate reductase, type III chloramphenicol acetyltransferase, and D-alanine–D-alanine ligase B fold during and after vectorial synthesis on the ribosome to folding from the full-length unfolded state in bulk solution. Our results reveal that the influence of the ribosome on protein folding mechanisms varies depending on the size and complexity of the protein. Specifically, for a small protein with a simple fold, the ribosome facilitates efficient folding by helping the nascent protein avoid misfolded conformations. However, for larger and more complex proteins, the ribosome does not promote folding and may contribute to the formation of intermediate misfolded states cotranslationally. These misfolded states persist post-translationally and do not convert to the native state during the 6 μs runtime of our coarse-grain simulations. Overall, our study highlights the complex interplay between the ribosome and protein folding and provides insight into the mechanisms of protein folding on and off the ribosome.

DOI: 10.1021/acs.jpcb.3c01694

Publication source: https://pubs.acs.org/doi/10.1021/acs.jpcb.3c01694

5.2.2 *Author contribution statements*

Quyen V. Vu
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

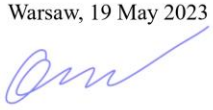
STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V., Nissley, D. A., Jiang, Y., O'Brien, E. P., Li, M. S. **Is Posttranslational Folding More Efficient Than Refolding from a Denatured State : A Computational Study.** *J. Phys. Chem. B* **2023**, 127 (21), 4761–4774.

My contribution was performing simulations, analyzing the results, preparing figures, interpreting the results, and participating in writing the manuscript.

Warsaw, 19 May 2023



Quyen V. Vu

Daniel A. Nissley
Department of Statistics
University of Oxford
Oxford, OX1 3LB, United Kingdom

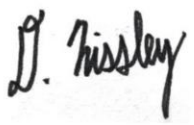
STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V; Nissley, D. A.; Jiang, Y.; O'Brien, E. P.; Li, M. S. **Is Posttranslational Folding More Efficient Than Refolding from a Denatured State: A Computational Study.** *J. Phys. Chem. B* **2023**, 127 (21), 4761–4774.

My contribution was analyzing and interpreting the results and helping to write the manuscript.

Oxford, 23 May 2023



Daniel A. Nissley

Yang Jiang
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V., Nissley, D. A., Jiang, Y., O'Brien, E. P., Li, M. S. **Is Posttranslational Folding More Efficient Than Refolding from a Denatured State : A Computational Study.** *J. Phys. Chem. B* **2023**, 127 (21), 4761–4774.

My contribution was writing computer code, analyzing the results, interpreting the results, and writing the manuscript.

PA USA, 22 May 2023

Yang Jiang

Yang Jiang

Prof. Edward P. O'Brien, PhD
Department of Chemistry
The Pennsylvania State University
104 Chemistry Building
University Park PA, 16802
E-mail: epo2@psu.edu
Tel: 1-814-867-5100

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V., Nissley, D. A., Jiang, Y., O'Brien, E. P., Li, M. S. **Is Posttranslational Folding More Efficient Than Refolding from a Denatured State : A Computational Study.** *J. Phys. Chem. B* **2023**, 127 (21), 4761–4774.

My contribution consisted of designing the research, interpreting the data, and writing the manuscript.

PA USA, 26 May 2023

Edward P. O'Brien

Prof. Mai Suan Li, PhD
Division of Theoretical Physics
Institute of Physics, Polish Academy of Sciences
Al. Lotnikow 32/46
02-668 Warsaw, Poland

STATEMENT

I declare that I am the co-author of the publication:

- Vu, Q. V., Nissley, D. A., Jiang, Y., O'Brien, E. P., Li, M. S. **Is Posttranslational Folding More Efficient Than Refolding from a Denatured State : A Computational Study.** *J. Phys. Chem. B* **2023**, 127 (21), 4761–4774.

My contribution consisted of designing the research, interpreting the data, writing the manuscript, and supervising the overall project.

Warsaw, 26 May 2023

Mai Suan Li

5.2.3 *Publication*

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Is Posttranslational Folding More Efficient Than Refolding from a Denatured State: A Computational Study

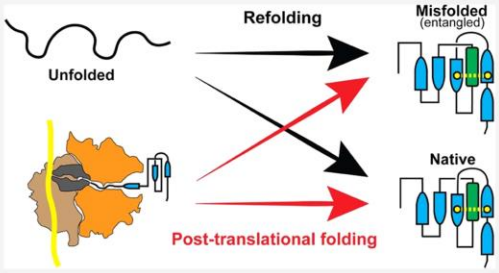
Quyen V. Vu, Daniel A. Nissley, Yang Jiang, Edward P. O'Brien,* and Mai Suan Li*

Cite This: *J. Phys. Chem. B* 2023, 127, 4761−4774

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The folding of proteins into their native conformation is a complex process that has been extensively studied over the past half-century. The ribosome, the molecular machine responsible for protein synthesis, is known to interact with nascent proteins, adding further complexity to the protein folding landscape. Consequently, it is unclear whether the folding pathways of proteins are conserved on and off the ribosome. The main question remains: to what extent does the ribosome help proteins fold? To address this question, we used coarse-grained molecular dynamics simulations to compare the mechanisms by which the proteins dihydrofolate reductase, type III chloramphenicol acetyltransferase, and D-alanine−D-alanine ligase B fold during and after vectorial synthesis on the ribosome to folding from the full-length unfolded state in bulk solution. Our results reveal that the influence of the ribosome on protein folding mechanisms varies depending on the size and complexity of the protein. Specifically, for a small protein with a simple fold, the ribosome facilitates efficient folding by helping the nascent protein avoid misfolded conformations. However, for larger and more complex proteins, the ribosome does not promote folding and may contribute to the formation of intermediate misfolded states cotranslationally. These misfolded states persist posttranslationally and do not convert to the native state during the 6 μs runtime of our coarse-grain simulations. Overall, our study highlights the complex interplay between the ribosome and protein folding and provides insight into the mechanisms of protein folding on and off the ribosome.



INTRODUCTION

Proteins are synthesized by ribosomes during the nonequilibrium process of translation and must fold to a specific native state, dictated by their amino acid sequence, to function. During translation, proteins are synthesized vectorially from N- to C-terminus based on an mRNA template. The nascent protein is initially confined to the ribosome exit tunnel, an ∼10 nm long tunnel with a diameter of 1−2 nm that can accommodate approximately 30 amino acids of the elongating protein.[1,2] Due to its dimensions, the exit tunnel restricts the ability of the protein to self-interact and form a tertiary structure. However, many proteins fold cotranslationally[3−6] as they begin to emerge from the exit tunnel and acquire a tertiary structure before their synthesis is complete. Though some small domains can fold inside the exit tunnel,[3−5] most proteins can only begin to fold once they have left the exit tunnel.[7−10] The nonequilibrium nature of protein synthesis means that the ability of a protein to fold cotranslationally can depend on the speed at which amino acids are added to the growing nascent chain.[11,12] Refolding of a protein from its full-length denatured state, however, allows all segments of the protein to simultaneously fold without the restriction of the exit tunnel or the influence of translation kinetics. Bulk refolding thus presents the opportunity for the formation of a vast number of non-native contacts between amino acids. In general, cotranslational folding is thought to be a beneficial process that aids in the efficient folding of complex proteomes.[13−15] The importance of cotranslational folding is highlighted by the recent experimental finding that one-third of *Escherichia coli* (*E. coli*) proteins are not able to refold in bulk solution after complete unfolding,[16] suggesting that cotranslational folding is critical to their ability to reach their native state.

The folding of a small number of proteins has been experimentally and computationally studied on and off the ribosome.[17−22] Evidence so far suggests that the role of the ribosome in folding is protein-specific. For example, structure-based models in combination with an arrest-peptide assay and cryo-EM experiments indicate that the folding of titin I27 is conserved on and off the ribosome.[21] Similarly, experiments and molecular simulations of src SH3 show that its folding

Received: March 13, 2023
Revised: May 4, 2023
Published: May 18, 2023

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

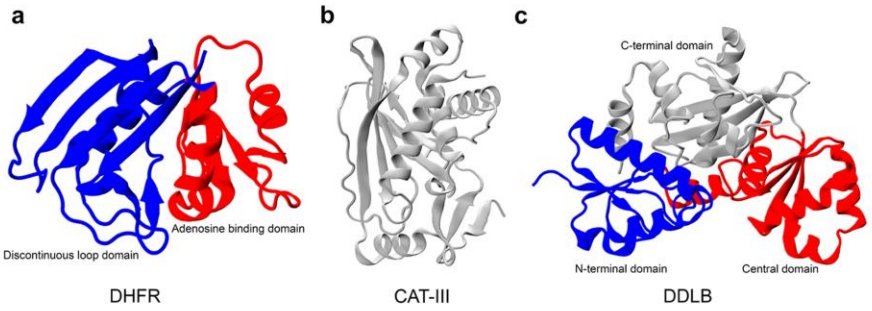


Figure 1. Crystal structures of three proteins in this study with domain-based coloring. (a) Crystal structure of DHFR (PDB ID: 4KJK); the discontinuous loop and the adenosine binding domains are shown in blue and red, respectively. (b) CAT-III is a single domain protein, which is shown in gray, and (c) DDLB protein (PDB ID: 4C5C), with the N-terminal, central, and C-terminal domains are shown in blue, red, and gray, respectively.

pathways are the same on and off the ribosome.[22] On the other hand, Tanaka et al. used coarse-grained molecular simulation to study the role of the ribosome in guiding multidomain protein folding, finding that folding on the ribosome is more efficient compared to refolding.[18] Dabrowski-Tumanski et al. computationally studied a deeply knotted protein and found that the ribosome plays a key role in knot formation.[20] In terms of kinetics, single-molecule laser optical tweezer experiments have found that the arrested ribosome nascent chain complexes have reduced protein folding rates compared to folding in bulk.[17,23] These studies mostly focus on small proteins (~100 residues) folding on translationally arrested ribosomes. In vivo, many nascent proteins diffuse into the cytosol after synthesis; if folding is not completed on the ribosome, it may complete posttranslationally. Hence, the ribosome may only influence the formation of intermediate states, which nonetheless can change the outcome of folding.[24,25] Given the relative paucity of experimental and computational data on the differences between folding on and off the ribosome for large proteins, we believe the influence of the ribosome on protein folding mechanisms remains an open question.

Performing all-atom folding simulations for large proteins is computationally infeasible. In this study, we, therefore, utilize a topology-based coarse-grained model to simulate the refolding in bulk solution as well as the co- and posttranslational folding of three *E. coli* enzymes (Figure 1): (i) dihydrofolate reductase (DHFR, 159 residues, PDB ID: 4KJK[26]), (ii) type III chloramphenicol acetyltransferase (CAT-III, 213 residues, PDB ID: 3CLA[27]), and (*iii*) D-alanine−D-alanine ligase B (DDLB, 306 residues, PDB ID: 4C5C[28]). DHFR, the smallest of the three, is composed of two domains.[29,30] The adenosine binding domain (ABD) consists of residues 38−106, and the discontinuous loop domain (DLD) comprises residues 1−37 and 107−159 (Figures 1a and S1). DHFR catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate and has been a target enzyme of antifolate drugs.[31] The native structure of CAT-III is composed of eight β-sheets and five α-helices (Figures 1b and S1); CAT-III is responsible for the high level of bacterial resistance to chloramphenicol.[32] Finally, DDLB is a three-domain protein composed of an N-terminal domain (residues 1−85), central domain (residues 86−180), and C-terminal domain (residues 181−306), each of which is classified as α/β. At the secondary structure level, DDLB contains 10 β-sheets and 11 α-helices (Figures 1c and S1) and is an essential enzyme for the proper synthesis and maintenance of the bacterial cell wall.[33]

In this work, we apply multiple order parameters for protein folding, including the recently described entanglement parameter G, to investigate differences in folding on and off the ribosome. We find that while the ribosome assists the folding of DHFR, it does not promote the folding of CAT-III and DDLB, both of which contain a native entanglement. Our results support a mechanism by which the ribosome may promote the formation of intermediate misfolded states with non-native entanglements; these intermediates are kinetically trapped and persist for long time scales posttranslationally.

MATERIALS AND METHODS

Simulation Details and Construction of Coarse-Grain Model. We employ a previously published Gō-based coarse-grain model[11,34] in which each amino acid is represented by a single interaction site placed at the C_α atom with a specific van der Waals radius for each amino acid; ribosomal RNA is represented as three or four beads per nucleotide, with one bead located at the phosphate position, another at the centroid of the ribose ring, and one at the centroid of each conjugated ring in the base (one bead for pyrimidine nucleobases and two beads for purine nucleobases). The potential energy of a configuration in this model is computed by the equation

$$
\begin{aligned}
E = {} & \sum_i k_b (r_i - r_0)^2 + \sum_i \sum_{j=1}^{4} k_{\varphi,ij}[1 + \cos(j\varphi_i - \delta_{ij})] \\
& + \sum_i -\frac{1}{\gamma} \ln\{\exp[-\gamma(k_\alpha(\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha)] \\
& + \exp[-\gamma k_\beta(\theta_i - \theta_\beta)^2]\} + \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \exp\left[-\frac{r_{ij}}{l_D}\right] \\
& + \sum_{ij \in \{NC\}} \varepsilon_{ij}^{NC}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{ij \notin \{NC\}} \varepsilon_{ij}^{NN}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (1)
\end{aligned}
$$

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

The potential energy of a given conformation is calculated as a sum of the contributions from bonds, dihedral angles, bond angles, electrostatic interactions, Lennard-Jones-like native interactions, and repulsive non-native interactions. Model parameters are described in the previous studies.[11,34] Parameters for three proteins in this study were taken from the previous work.[11]

In posttranslational folding simulations, we first performed continuous synthesis using the wild-type mRNA sequences, which are presented in Table S1. Synthesis simulations were conducted using a previously described protocol,[11,35] with a cutout of the ribosome exit tunnel and surface. Codon-specific translation times were obtained from a previous study[11] (Supplementary Table 8 of ref 11). Once the protein sequence was fully synthesized, the covalent bond between the C-terminal site and the peptidyl transferase center (PTC) was cleaved and the protein was allowed to diffuse through the ribosome exit tunnel. Protein dissociation from the ribosome was defined as the point at which the position of the C-terminal residue was greater than 20 Å from the ribosome surface. At this point, the ribosome was removed and the left protein was able to undergo posttranslational folding in the absence of the ribosome.

The refolding simulations were initiated from the unfolded state, characterized by a low fraction of native contacts, Q value. Initial conformations for refolding simulations were generated by heating the native state of the protein to 1000 K for 15 ns. The final conformation from heating was then temperature-quenched at 310 K to initialize refolding. All simulations were carried out using a Langevin thermostat at a temperature of 310 K, with a time step of 15 fs and a friction coefficient of 0.050 ps^{-1}. All simulations were carried out using OpenMM 7.7.[36]

In order to characterize protein folding, we conducted 200 statistically independent folding trajectories for each protein under investigation (100 trajectories of refolding and 100 trajectories of posttranslational folding). Each trajectory lasted for 6 μs, which corresponds to a real-time duration of approximately 24 seconds based on the relative acceleration of folding in these coarse-grain models relative to real time scales.[11,34] For CAT-III and DDLB, which had a high prevalence of misfolded trajectories, we extended the simulation time to 30 and 15 μs, respectively, in order to determine if the proteins would eventually fold correctly in a longer time scale.

Calculation of the Fraction of Native Contacts, *Q*, and Its Usage to Determine Folded Trajectories. Two residues are considered to form a native contact if their α carbons are less than 8 Å apart in the crystal structure. To account for thermal fluctuations in contact distances during simulation, a flexibility parameter $\Delta = 1.2$ was used: a native contact between two residues is classified to be formed in a current frame of the simulated trajectory if their distance is shorter than 1.2 times the distance in the crystal structure. The fraction of native contacts, Q, was calculated for each protein during their posttranslational folding or refolding simulations. Only contacts between pair of residues i and j both within secondary structural elements as identified by STRIDE[37] and satisfying the criterion $|i - j| > 3$, where i and j are the residue indices, were considered; we excluded any secondary segment that is shorter than four residues from the analysis. To determine when a given trajectory of a protein is folded, we first characterized the fraction of native contact, Q, of each protein's native state by performing ten 1.5 μs coarse-grained simulations at 310 K initialized from the native-state coordinates. The threshold for protein folding during refolding or posttranslational folding simulations, $Q_{threshold}$, was determined as $Q_{threshold} = \langle Q^{NS}_{mode} \rangle - 3\sigma$, where $\langle Q^{NS}_{mode} \rangle$ is the average Q_{mode} over all 15 ns windows of the ten 1.5 μs native-state simulations and σ is the standard deviation of $\langle Q^{NS}_{mode} \rangle$. To determine when folding occurred during refolding or posttranslational folding simulations, the mode of the Q values over a sliding 15 ns window was compared to the $Q_{threshold}$. A given trajectory is defined as folded if during its time evolution, $Q^{15\text{-ns}}_{mode} \geq Q_{threshold}$, the folding time is the first time that the above condition is met.[35,38] The threshold value of Q for each protein is presented in Table 1.

Table 1. Threshold Value of *Q* of Three Proteins Computed from 10 Native-State Simulations Used to Determine if a Given Trajectory of Protein Folds

protein	$Q_{threshold} = \langle Q^{NS}_{mode} \rangle - 3\sigma$
DHFR	0.9221
CAT-III	0.9269
DDLB	0.9521

Estimating the Folding Time of Slow-Folding Proteins with a Large Proportion of Unfolded Trajectories. When the portions of folded trajectories are less than 50% of total trajectories, it is not possible to estimate the folding time as the median first passage time.

We consider three-state folding kinetics with parallel pathways. State A folds rapidly to the native state N at the rate k_1, and state B folds slowly to the native state with a much smaller rate k_2 ($k_1 \gg k_2$), and there is no interconversion between A and B. We have a set of ordinary differential equations respecting the rate of changing portion of states A and B

$$\begin{cases} \dfrac{d[A]}{dt} = -k_1[A] \\ \dfrac{d[B]}{dt} = -k_2[B] \end{cases} \tag{2}$$

where [A] and [B] are the portion of non-native states A and B. The portion (survival probability) of non-native states at time t: $S_U(t) = [A](t) + [B](t) = c_1 \exp(-k_1 t) + c_2 \exp(-k_2 t)$, where c_1 and c_2 are arbitrary constants. The initial condition that at time $t = 0$, the survival probability of non-native state = 1, we have $S_U\ (t = 0) = c_1 + c_2 = 1$, this yields: $c_2 = 1 - c_1$.

Hence, we computed the survival probability of the unfolded state as a function of time from simulations, and the resulting time series were then fit to the double-exponential equation

$$S_U(t) = c_1 \exp(-k_1 t) + (1 - c_1) \exp(-k_2 t) \tag{3}$$

c_1, k_1, and k_2 are the fitting parameters. The time constants of the two kinetic phases are $\tau_1 = \frac{1}{k_1}$, $\tau_2 = \frac{1}{k_2}$, with the larger of these two times determining the overall time scale of the folding process, $\tau_2 \gg \tau_1$. To estimate the uncertainty of the folding time when fitting to double-exponential folding kinetics, we apply bootstrap resampling by randomly selecting trajectories from the list of simulations. We only consider the random sample with the coefficient of determination $R^2 > 0.9$.

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

This procedure was applied to estimate the folding time of CAT-III.

Definition of the Progress Variable ς and Use to Monitor the Sequence of Pairs of Native Secondary Structure Elements Formed during the Folding Process. To account for the significant variation in folding times among different trajectories, we monitored folding pathways as a function of a progressive variable,[39] ς, defined as

$$\varsigma = \left\langle \frac{t_{\mathrm{pair},i}}{t_{\mathrm{fold},i}} \right\rangle \tag{4}$$

where <···> indicates the average over all folded trajectories, and $t_{\mathrm{pair},i}$ and $t_{\mathrm{fold},i}$ are the folding time of pair and the whole protein folding time of the folded trajectory i, respectively. With this definition, we have $0 \leq \varsigma \leq 1$, $\varsigma = 0$, which means that the pair under studied folds at the start of the simulation, and $\varsigma = 1$ indicates the pair folds as the last step in the folding process. To determine the sequence of pairs of the secondary structure formation (defined in Figure S1 and Table S2), we consider a pair between two secondary structure elements that have more than one native contact. A pair is considered to be folded if its fraction of native contacts is larger than the threshold determined from native simulations. In our analysis of folding pathways, trajectories that did not fold within the 6 μs simulation duration were excluded.

Identifying Entanglement and the Changes in Entanglement. We use the approximation to the partial Gaussian double integration method proposed by Baiesi and co-workers[40] to calculate these partial linking numbers for a closed (loop) and open curve (termini). To identify lasso-like entanglements, we used the numerically invariant linking numbers,[41] which describe the linking between a closed loop and an open segment in a three-dimensional space. This procedure is a modified version of the original protocol proposed by Baiesi to detect entanglement in coarse-grain protein structures. The original protocol is not computationally efficient to analyze trajectories since for each pair of contact, we have to calculate the linking number for all possible combinations of loop and threading segments. In our modified protocol, we only have to calculate the linking number between the closed loop (closes by native contact) and two tails. The closed loop is composed of the peptide backbone connecting residues i and j that form a native contact. Outside this loop is an N-terminal segment composed of residues 5 through $i - 4$ and a C-terminal segment composed of residues $j + 4$ through $N - 5$, where we exclude the first five residues of the N-terminal curve, the last five residues of the C-terminal curve, and four residues before and after the native contact to eliminate the error introduced by both the high flexibility and contiguity of the termini and trivial entanglements in the local structure; this metric is similar to whGLN.[42] We can characterize the entanglement of each tail with the loop formed by the native contacts with two partial linking numbers denoted g_{N} and g_{C}. For a given structure of an N-residue protein, with a native contact present at residues (i, j), the coordinates $\boldsymbol{R}_l$ and the gradient $\mathrm{d}\boldsymbol{R}_l$ of the point l on the curves were calculated as

$$\begin{cases} \boldsymbol{R}_l = \dfrac{1}{2}(\boldsymbol{r}_l + \boldsymbol{r}_{l+1}) \\ \mathrm{d}\boldsymbol{R}_l = \boldsymbol{r}_{l+1} - \boldsymbol{r}_l \end{cases} \tag{5}$$

where $\boldsymbol{r}_l$ is the coordinates of the C_α atom in residue l. The linking numbers $g_{\mathrm{N}}(i, j)$ and $g_{\mathrm{C}}(i, j)$ were calculated as

$$\begin{cases} g_{\mathrm{N}}(i, j) = \dfrac{1}{4\pi} \displaystyle\sum_{m=6}^{i-5} \sum_{n=i}^{j-1} \frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3} \cdot (\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \\ g_{\mathrm{C}}(i,\ j) = \dfrac{1}{4\pi} \displaystyle\sum_{m=i}^{j-1} \sum_{n=j+4}^{N-6} \frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3} \cdot (\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \end{cases} \tag{6}$$

The total linking number for a native contact (i, j) is therefore estimated as

$$g(i, j) = \mathrm{round}[g_{\mathrm{N}}(i, j)] + \mathrm{round}[g_{\mathrm{C}}(i, j)] \tag{7}$$

Comparing the absolute value of the total linking number for a native contact (i, j) to that of a reference state allows us to detect a gain or loss of linking between the backbone trace loop and the terminal open curves as well as any switches in chirality. Therefore, there are six changes in linking cases we should consider when using this approach to quantify entanglement (see Supplementary Figure S1 and Table 1 of ref 43).

The degree of entanglement G is defined as

$$G(t) = \frac{1}{M} \sum_{(i,j)} \Theta[(i, j) \in \mathrm{NC} \cap g(i, j, t) \neq g^{\mathrm{native}}(i, j)] \tag{8}$$

where (i, j) is the native contact in the crystal structure; NC is the set of native contacts formed in the current structure at time t; and $g(i,j,t)$ and g^{native} (i,j) are, respectively, the total linking number of the contact (i, j) at time t and native structures estimated using eq 7. M is the total number of native contacts in the native structure and Θ is a Heaviside step function, equals 1 if the condition is true and equals 0 if the condition is false.

The difference between $g(i,j,t)$ and $G(t)$ is $g(i,j,t)$, which is characterized by the entanglement in a given structure of contact (i, j) at time t, while $G(t)$ provided information about the total number of contacts that changed the entanglement at time t.

Clustering and Coarse-Graining Conformational Space (Q, G). The projection of conformation space onto (Q,G) reveals intermediate states that may be hidden when projected onto Q alone, as two states can have the same value of Q but one may be entangled while the other is not. Entanglement can prevent a protein from reaching its native state, as the loop-threading segment is improperly organized. Entangled states thus can form kinetic traps with large energy barriers preventing progression to the folded state, as large sections of the protein must unfold to allow disentanglement. To derive the log probability surface as a function of (Q,G), we first combined (Q, G) data from refolding and posttranslational folding for each protein and applied the Min−Max algorithm[44] for normalization. K-mean++ clustering[45] was then utilized to identify microstates, with 200, 400, and 400 clusters (microstates) being used for DHFR, CAT-III, and DDLB, respectively. As k-mean++ is a distance-based clustering algorithm, the normalization of data was necessary to prevent one-order parameter from dominating the distance measure. The resulting clusters were further coarsened into a small number of metastable states using the PCCA+ algorithms[46] to facilitate the interpretation of the folding pathways. The

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

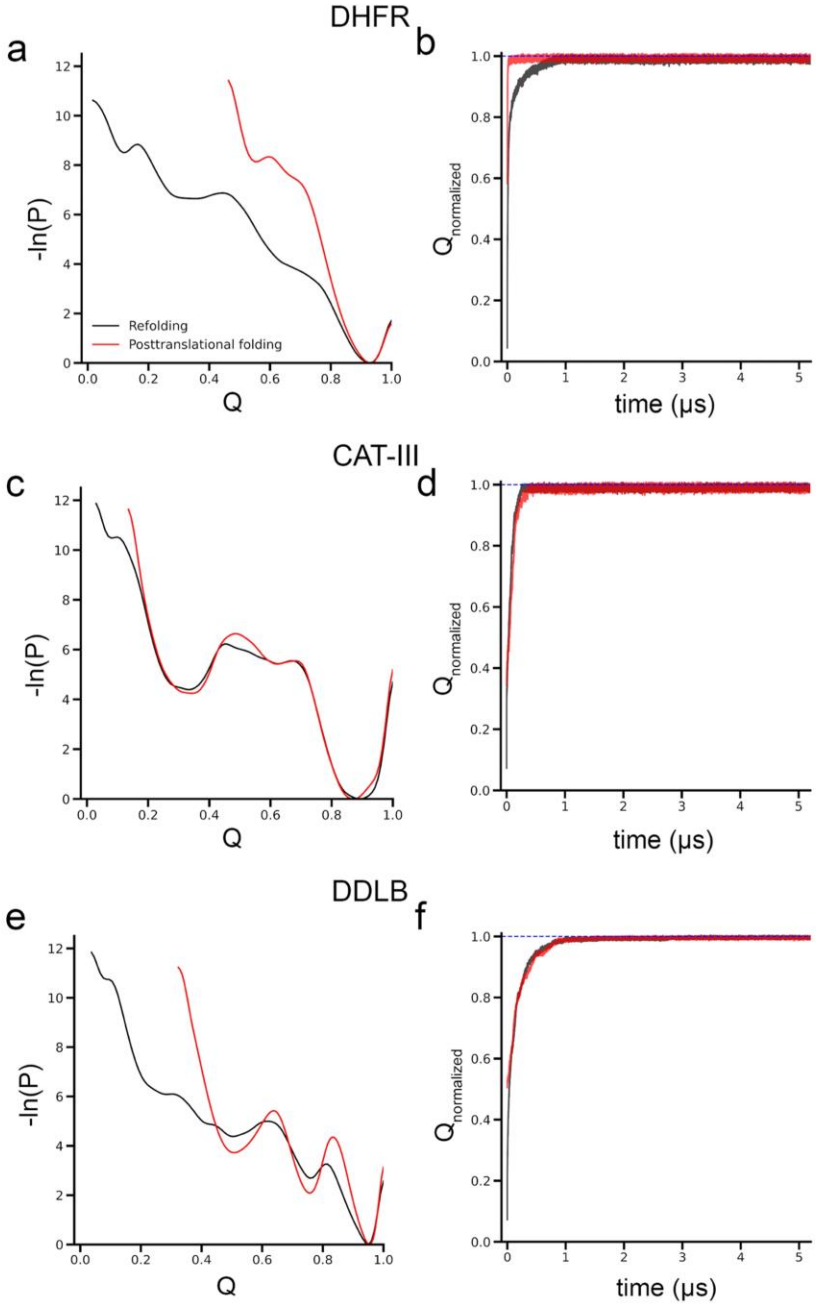


Figure 2. Influence of ribosomes on protein folding is protein-specific. Ribosome assists DHFR folding but does not promote CAT-III and DDLB folding. (a) Log probability landscape ($-\ln(P)$, where P is the probability of sampling a particular Q value) of DHFR, (b) average of the normalized fraction of native contacts, Q, of the folded trajectories as the function of time of DHFR; (c, d) same as in panels (a, b) but for CAT-III protein; and (e,f) same as in panels (a, c) but for DDLB. Refolding and posttranslational folding results are plotted in black and red colors, respectively. The blue-dashed lines in panels (b, d, f) indicate $Q_{normalized} = 1$.

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

number of metastable states was determined based on the presence of a gap in the eigenvalue spectrum of the transition probability matrix; 11, 14, and 13 metastable states were used for DHFR, CAT-III, and DDLB, respectively. Both the clustering and coarse-graining processes were performed by using the PyEmma[47] and Deeptime[48] packages.

Identify Folding Pathways along the Order Parameters (*Q*, *G*). To identify folding pathways from the simulated trajectories, the following procedure was followed:

(1) For each discrete trajectory, the starting state of the first frame is added to the pathway.
(2) The trajectory is then advanced, and the next state that differed from the last recorded state in the pathway was identified. If this state had not yet been recorded in the pathway, it was added to the pathway. If the state is already been recorded in the pathway, the pathway was truncated at the first instance of the recorded state and the trajectory was advanced from that point.
(3) Repeat Step (2) until the end of the trajectory is reached.

This process resulted in pathways that contained no loops, and only recorded the on-pathway states for each discrete trajectory. The distribution of distinct pathways and the probabilities of transitioning from one state to another was then estimated based on the pathways of all of the discrete trajectories. The initial, folded, and misfolded states (in the folding/misfolding pathways plots) are colored yellow, blue, and red, respectively. A state is considered misfolded if there is a trajectory that becomes trapped in that state, and there is no direct transition to the native state. The size of the nodes is proportional to the probability of the state appearing in the coarse-grained trajectories. The size of the edges connecting the nodes is proportional to the number of transitions between states, and the red number beside the edge is the total number of transitions observed in the coarse-grained trajectories.

Back-Mapping the Coarse-Grained Model to an All-Atom Model for Visualization. To backmap the coarse-grained model to all-atom representation, the first step was to add coarse-grained interaction sites that represent the side-chain center of mass near the corresponding C_α beads. Then, the orientation of the side-chain center of mass beads was optimized through energy minimization while restraining the C_α positions. Next, Pulchra software[49] was used to rebuild the nonhydrogen atoms of both the backbone and the side chain. Finally, additional energy minimization was performed in vacuum with position restraints applied to all C_α atoms to obtain the final all-atom structure.

RESULTS AND DISCUSSION

DHFR Folds More Efficiently due to Protein Synthesis. To understand the influence of protein synthesis and the ribosome on the folding of DHFR, we constructed a topology-based coarse-grain model (see the Materials and Methods section) and simulated its folding through two different processes. First, we simulated protein refolding starting from a thermally unfolded ensemble. Second, to probe its folding when synthesized by the ribosome, we simulated continuous synthesis and posttranslational folding. This model has been previously shown to reproduce the cotranslational folding of HemK N-terminal domain,[2] accurately predict changes in enzyme-specific activities,[11] and to predict misfolded conformations of GlpD that qualitatively agree with LiP-MS experiments.[35] To characterize the similarities and differences in how proteins reach the native state, we only analyzed the trajectories that resulted in successful folding.

We find that DHFR folds more efficiently when synthesized by the ribosome and undergoes posttranslational folding. However, when refolding from unfolded ensembles, some trajectories are trapped in misfolded states ($Q < Q_{threshold}$) during the 6 μs of simulation time. Specifically, DHFR rapidly transitions from the initial structural ensemble to the folded ensemble. Since these simulations are out-of-equilibrium, we cannot speak of free-energy landscapes, which are time-independent; instead, we compute log probability landscapes (Figure 2a), which are time-dependent. This nonequilibrium landscape perspective for refolding and posttranslational folding simulations reveals differences between the two processes. DHFR has a well-defined structure composed of two main subdomains: the adenosine binding subdomain (ABD, residues 38−106) and the discontinuous loop subdomain (DLD, residues 1−37 and 107−159) (Figure 1a). In posttranslational folding simulations, this protein samples a smaller region of Q and the ABD domain folds cotranslationally and has the native form ($Q_{ABD} = 0.98$; Figure S2) at the start of posttranslational simulations. The DLD domain, consisting of both the N-terminus outside of the ribosome exit tunnel and the C-terminus, which is still within the exit tunnel, has a lower degree of native contacts $Q_{DLD} = 0.27$ (Figure S2). As a result, at the start of the posttranslational simulation, the overall structure of DHFR has approximately 60% of its native contacts formed, and the protein simply rearranges the DLD domain into the correct registry when the C-terminus is released from the exit tunnel. All trajectories reach the folded state ($Q \geq Q_{threshold}$ or $Q_{normalized} \geq 1$) with a median folding time of 20.5 ns (95% confidence interval (CI) [18.5 ns, 24.8 ns], computed from bootstrapping). In contrast, refolding from the thermally unfolded ensemble involves initial conformations with a high degree of disorder ($Q < 0.1$ for both ABD and DLD domains; Figure S2), sampling a wider range of the log probability landscape (Figure 2a). Overall, the protein takes a longer time to reach the native state compared to posttranslational folding (Figure 2b), with a median folding time of 140.5 ns (95% CI [114.6 ns, 196.1 ns]) (Table 2). Only 92 (95% CI [86, 97]) trajectories fold out of 100 during the simulation. The difference between the median folding times is significant (p-value $< 1 \times 10^{-6}$, permutation test; Table 2), as well as the number of folded trajectories (p-value = 0.007; Table 2) between posttranslational folding and refolding. In both cases, the folding of DHFR proceeds with the ABD folding into its native form first, followed by the folding of the DLD (Figure S2). The folding of DLD is thus rate-limiting to the formation of the overall native structure.

Protein Synthesis Does Not Increase the Folding Efficiency of CAT-III and DDLB. Using the same simulation protocol as DHFR, we performed refolding and posttranslational folding for CAT-III and DDLB proteins. In contrast to DHFR, the folding dynamics and population of folded trajectories for CAT-III and DDLB are relatively insensitive to posttranslational folding versus refolding. Specifically, for CAT-III, the log probability landscape of CAT-III is almost identical between posttranslational folding and refolding (Figure 2c). The progress of normalized Q of the folded trajectories is similar (Figure 2d), and the difference in the number of folded trajectories is insignificant (p-value = 0.14;

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

Table 2. Folding Times and the Number of Folded Trajectories of Proteins in Refolding and Posttranslational Folding Simulations (95% Confidence Interval and *p*-Value Are Calculated from the Bootstrap Resampling and Permutation Test with 10^6 Iterations)

	refolding		posttranslational folding	
protein	# folded trajectories [95% CI]	folding time (ns) [95% CI]	# folded trajectories [95% CI]	folding time (ns) [95% CI]
DHFR	92 [86, 97]	140.5 [114.6, 196.1]	100 [100, 100]	20.5 [18.5, 24.8]
	p-value (folded trajectories) = 0.007			
	p-value (folding time) < 10^{-6}			
CAT-III	42 [32, 52]	2.3×10^5 [6.5×10^4, 1.7×10^{12}]	31 [22, 40]	2.05×10^5 [7.8×10^4, 1.6×10^{12}]
	p-value (folded trajectories) = 0.14			
	p-value (folding time) = 0.96			
DDLB	76 [67, 84]	522.5 [412.1, 712.2]	78 [70, 86]	426.3 [264.7, 690.9]
	p-value (folded trajectories) = 0.87			
	p-value (folding time) = 0.18			

Table 2). There are a large number of misfolded trajectories ($Q < Q_{threshold}$; Table 2) within the simulation time of 6 μs. The proportion of folded trajectories for CAT-III is less than 50%; we, therefore, estimated its folding time by fitting the survival probability of the unfolded state as a function of time to a three-state kinetic model (eq 3; see Materials and Methods section). There is no statistical difference in folding times for CAT-III between refolding (2.3×10^5 ns, 95% CI [6.5×10^4 ns, 1.7×10^{12} ns]) and posttranslational folding (2.05×10^5 ns, 95% CI [7.8×10^4 ns, 1.6×10^{12} ns]), *p*-value = 0.96 (Table 2).

In the case of DDLB, more than 50% of trajectories are folded; hence, the median folding time could be estimated. We find that the median folding time in refolding is 522.5 ns (95% CI [412.1 ns, 712.2 ns]), compared to the folding time in posttranslational folding, which is 426.3 ns ([264.7 ns, 690.9 ns]). We find that there is no difference in the median folding times or the number of folded trajectories between the refolding and posttranslational folding simulations (*p*-value = 0.87 for the number of folded trajectories and *p*-value = 0.18 for the median folding time comparisons; Figure 2f and Table 2). However, there are some observed differences: the log probability landscape in the posttranslational folding of DDLB sampled a smaller region along the Q coordinate, and the local minima were deeper compared to refolding (Figure 2e). This suggests that the cotranslational formation of native contacts may have occurred after translation.

To test the influence of simulation time on the results, the misfolded trajectories for CAT-III were extended to 30 μs and the misfolded trajectories for DDLB were extended to 15 μs. We find that only one additional trajectory each from the refolding and posttranslational folding simulations of CAT-III folds during this extended duration, at 15 and 29.2 μs, respectively. No misfolded trajectories of DDLB folded in either the refolding or posttranslational folding simulations. This suggests that these misfolded trajectories are kinetically trapped and unlikely to convert to the folded state at longer time scales—consistent with previously published results.[11]

Measuring the Folding Mechanisms of Proteins Using Progress Variable ς Reveals the Differences for DHFR and Remains Robust for CAT-III and DDLB. Protein folding is typically thought to occur in a hierarchical fashion, with secondary structural elements first forming individually and then cooperatively coalescing into tertiary structures. With this in mind, we characterize the folding process of DHFR, CAT-III, and DDLB as the temporal sequence of formation of their stable pairs of native secondary structural elements with the aid of a progress variable, ς (see the Materials and Methods section, eq 4). The value of ς is relative to the time of complete folding of the protein, with $\varsigma = 0$ indicating that the pair folds at the start of the simulation and $\varsigma = 1$ indicating the pair folds as the last step in the folding process. To simplify the analysis, we restrict ourselves to pairs of secondary structures that have more than one native contact, as described in the Materials and Methods section and Table S2.

Based on this analysis, we observe a significant difference in DHFR. In posttranslational folding, all pairs of the native secondary structural elements belonging to the ABD domain fold cotranslationally ($\varsigma \sim 0$), while in refolding, most of the pairs fold at the end of the folding process ($\varsigma \sim 1$) (Figure 3a and Table 3). This suggests that the vectorial synthesis from the N-terminus to the C-terminus prevents the spontaneous cotranslational folding of some β-sheets in the C-terminal (C1, C2) and that the complete folding of DHFR occurs immediately upon release of the C-terminal from the ribosomal exit tunnel. These observations are consistent with previous experimental studies that have found that the central domain (ABD) acts as an independent folding unit during translation, while the DLD domain folds posttranslationally.[30] For CAT-III, the sequence of secondary structure pair folding is similar in both refolding and posttranslational folding, with all pairs folding late during the folding process ($\varsigma \sim 1$; Figure 3b and Table 3). For DDLB, the overall folding order is similar, but some differences were observed, such as in posttranslational folding, four pairs in the center domain (C13, C19, C22, and C24) fold cotranslationally ($\varsigma = 0$), two pairs in the N-terminal domain (C7, C8) fold posttranslationally but before the complete folding occurs ($\varsigma \sim 0.65$; Figure 3c and Table 3), while these pairs fold at the last event in refolding. Thus, protein synthesis and posttranslational folding do not significantly perturb the folding mechanisms of CAT-III and DDLB.

Native Entanglements Exist in the Crystal Structure of CAT-III and DDLB Proteins. We hypothesized that there is something distinct about the native topologies of CAT-III and DDLB that leads to a large proportion of misfolding. Several recent papers have predicted a link between misfolding involving a change in the entanglement status and long-lived misfolded states,[11,35] including the failure to form native entanglements. Indeed, this is the molecular hypothesis explaining the observation that experimental folding rates of proteins decrease as the number of times the threading segment pierces the loop increases.[40] To further understand this phenomenon, we investigate whether entanglement may play a role here by calculating the degree of entanglement for these proteins using eq 7.

We find that the crystal structure of DHFR does not contain any entanglements. In contrast, CAT-III has 16 native entanglements, with 14 of them consisting of a loop located near the N-terminus and a threading segment at the C-terminus. The remaining two native entanglements have a loop located near the C-terminus and a threading segment at the N-terminus. Similarly, DDLB has 36 native entanglements, half of

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

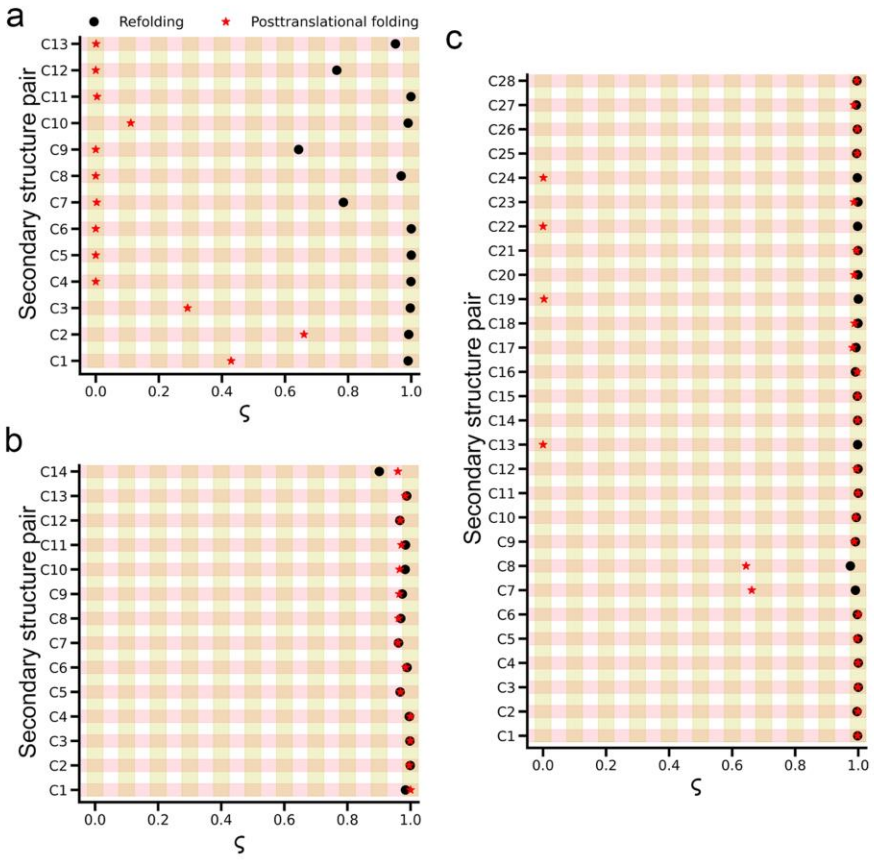


Figure 3. Comparisons of folding processes of DHFR, CAT-III, and DDLB in posttranslational folding versus refolding are shown as temporal sequences of secondary structure pairs formed over time with the aid of a progress variable ς. (a) Folding mechanism of DHFR is significantly different: all pairs in the ABD domain fold cotranslationally in posttranslational folding simulations, (b) CAT-III: there is no difference between posttranslational folding versus refolding, and (c) DDLB protein exhibits a small difference in four pairs of the center domain (C13, C19, C22, and C24) and two pairs (C7, C8) in the N-terminal domain. Refolding and posttranslational folding data are represented by black circles and red stars, respectively.

Table 3. Sequence of Native Secondary Structure Pair Formation during the Folding Process of DHFR, CAT-III, and DDLB Proteins[a]

protein	refolding	posttranslational folding
DHFR	C9 → C12 → C7 → (C8, C13) → (C1, C2, C3, C4, C5, C6, C10, C11)	(C4, C5, C6, C7, C8, C9, C11, C12, C13) → C10 → C3 → C1 → C2
CAT-III	C14 → (C5, C7, C8, C9, C12) → (C1, C2, C3, C4, C6, C10, C11, C13)	(C5, C7, C8, C9, C10, C11, C12, C14) → (C1, C2, C3, C4, C6, C13)
DDLB	(C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C27, C28)	(C13, C19, C22, C24) → (C7, C8) → (C1, C2, C3, C4, C5, C6, C9, C10, C11, C12, C14, C15, C16, C17, C18, C20, C21, C23, C25, C26, C27, C28)

[a]Pairs in parentheses represent secondary structures that are folded simultaneously.

which consist of a loop located closer to the N-terminus and a threading segment at the C-terminus, while the other half has a loop located closer to the C-terminus and a threading segment at the N-terminus. Representative examples of these entanglements are shown in Figure 4a,b for CAT-III and DDLB, respectively. Furthermore, proteins with an entanglement loop closer to the N-terminus were found to be folded more difficultly than the proteins with a loop closer to the C-terminus.[50] This explains why DHFR (without native entanglement) can fold easily and small portions of CAT-III trajectories (most entanglement loops are located near the N-terminus) folds in our simulation. This observation suggests that entanglement plays an important role in the proper folding of proteins.

Protein Synthesis Assists the Folding of DHFR by Avoiding Misfolded States with Non-Native Entangle-

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

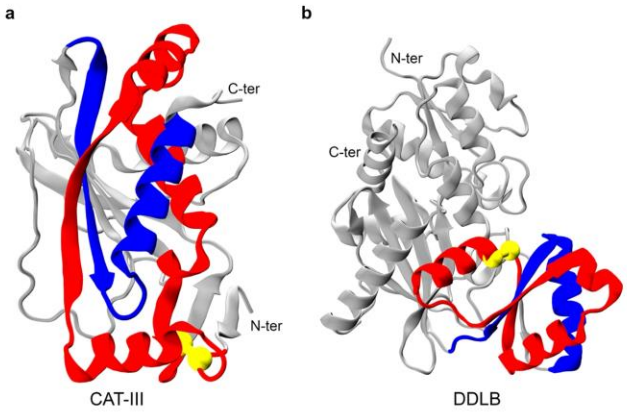


Figure 4. Example of native entanglements in the crystal structures of CAT-III and DDLB. The closed loop and crossing section of the threading segment of their entangled regions are colored red and blue, respectively. The loops are closed by noncovalent contacts between two residues (colored yellow), and the rest part of the protein is colored gray. (a) Representative native entanglement in CAT-III: the loop (colored red) is closed by a native contact between residues 8 and 77, and the threading segment consists of residues 177−208. (b) Representative native entanglement in DDLB: the loop (colored red) consists of residues 98−146, and the threading segment consists of residues 160−184.

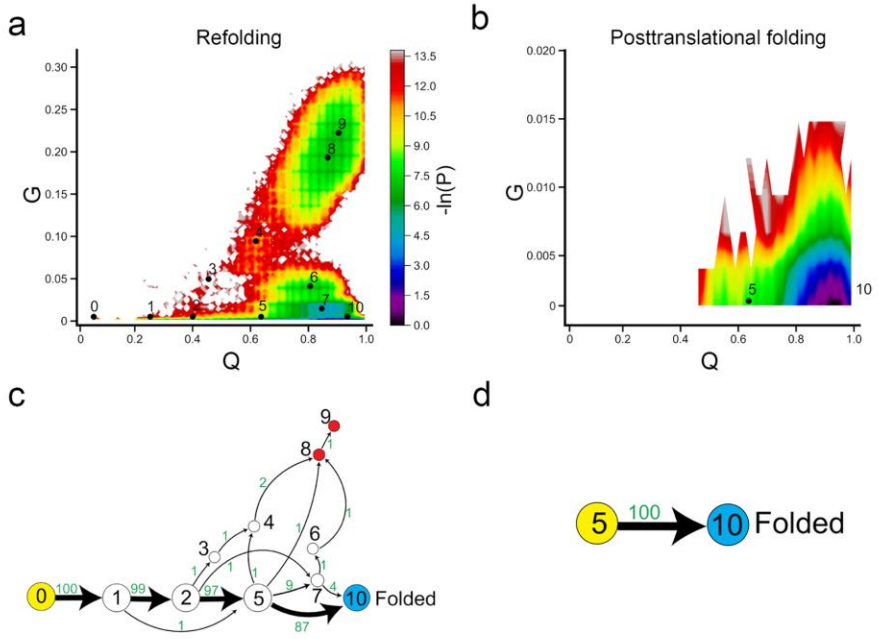


Figure 5. Ribosome helps DHFR fold more efficiently. (a, b) $-\ln(P)$ surface in refolding and posttranslational folding, respectively, where P is the probability of sampling particular Q and G values. The centers of metastable states and their corresponding indices are shown on top of the surface (black points). (c, d) Transition network from discrete trajectories of refolding and posttranslational folding simulations. The yellow, red, and sky blue nodes correspond to the initial, misfolded, and folded states, respectively. The black numbers on the nodes match the indices of metastable states in panels (a) and (b). The red numbers beside the edges indicate the number of direct transitions between states observed in discrete trajectories.

ments. Entanglement plays an important role in the proper folding of proteins. To further characterize the folding pathways of DHFR, we clustered the conformational space based along the order parameters Q and G and then assigned them to metastable states (see the Materials and Methods section). In posttranslational folding, DHFR can spontane-

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

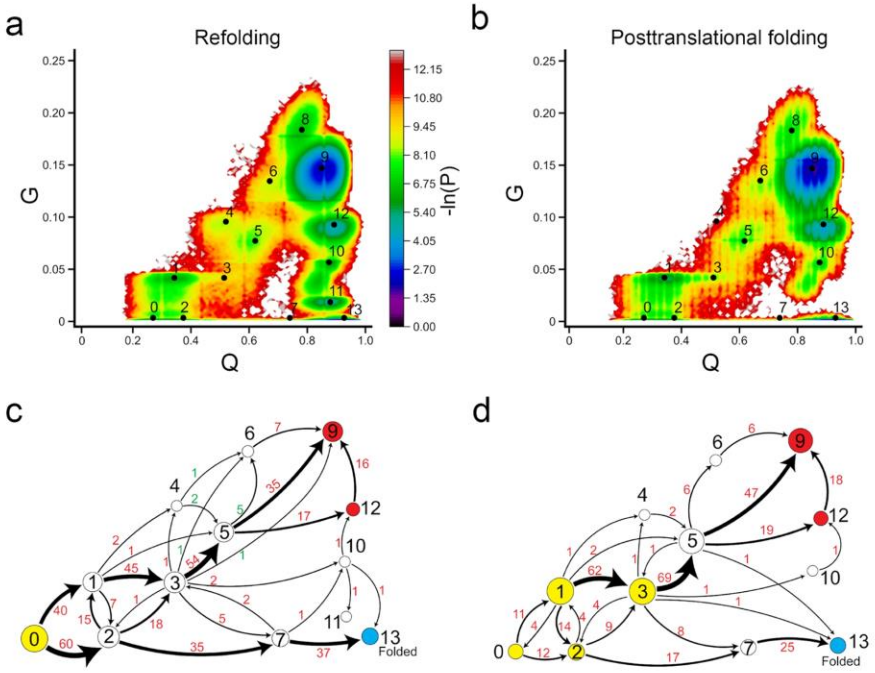


Figure 6. Protein synthesis does not increase the folding efficiency of CAT-III. (a, b) $-\ln(P)$ surface in refolding and posttranslational folding, respectively, where P is the probability of sampling particular Q and G values. The centers of metastable states and their corresponding indices are shown on top of the surface (black points). (c, d) Transition network from discrete trajectories of refolding and posttranslational folding simulations. The yellow, red, and sky blue nodes correspond to the initial, misfolded, and folded states, respectively. The black numbers on the nodes match the indices of metastable states in panels (a) and (b). The red numbers beside the edges indicate the number of direct transitions between states observed in discrete trajectories.

ously fold to its native state once the C-terminus is released from the ribosome. The two-dimensional log probability surface is concentrated in the region around the folded state (small G, high Q; Figure 5b), which is consistent with our 1D log probability landscape from the previous section. Specifically, the posttranslational folding simulations of DHFR only sample two states, 5 and 10 (the folded state). There are no misfolded trajectories in posttranslational folding, as all trajectories reach the folded state at the end of the simulation. The protein cotranslationally folds to the ensemble state 5, which has about 60% of native contacts formed (the fraction of native contacts with non-native entanglement is negligible, around 0.16%), and the folding process simply involves diffusion to the folded state (state 10). Folding network analyses reveal that 100% of folding pathways go straight from the initial state 5 to the folded state 10 (Figure 5d). There is no off-pathway state in the posttranslational folding of DHFR.

Refolding from the thermally unfolded ensemble is more complicated, compared to posttranslational folding. The $-\ln(P)$ surface has sampled a broad region in the non-native (low Q) or near-native (high Q) regions. We found that the population of DHFR refolding samples had a large number of entangled states, indicated by high values of G (Figures 5a and S3). The protein follows two parallel pathways to reach the native state: we find that the dominant pathway (*→ 5 → 10), which is the only pathway observed in posttranslational folding, accounts for 87% of the total trajectories in refolding simulation and a small portion (four trajectories, accounts for 4% of total trajectories) folds via intermediate state 7 (*→ 7 → 10). In addition, we find that 9% of trajectories become trapped in misfolded states (states 7−9). The broader $-\ln(P)$ surface in refolding is caused by a small number of misfolded trajectories. Five trajectories become trapped in state 7, three trajectories become trapped in state 8, and one trajectory becomes trapped in state 9. States 8 and 9 are off-pathway misfolded states, as we do not observe any folding events (conversion to the folded state 10) if the protein visits these states. When the protein samples the near-native state 7, only 40% of trajectories can fold successfully (* → 7 → 10/folded), while the remaining 60% fold to misfolded states (Figure 5c).

Non-Native Entangled States Act as a Kinetic Trap in Both Refolding and Posttranslational Folding of CAT-III and DDLB. In contrast to DHFR, it seems that the ribosome has less effect on the folding/misfolding mechanism of CAT-III. The conformational space is very similar between refolding and posttranslational folding, and these two processes share almost all of the observed states (Figure 6). This is reasonable as we have observed that when the protein synthesis is completed, there is a small portion of native contacts that have been formed in CAT-III and hence can be considered an unfolded state ($Q \sim 30\%$; Figure 2d). Therefore, when the protein dissociates from the ribosome and undergoes

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

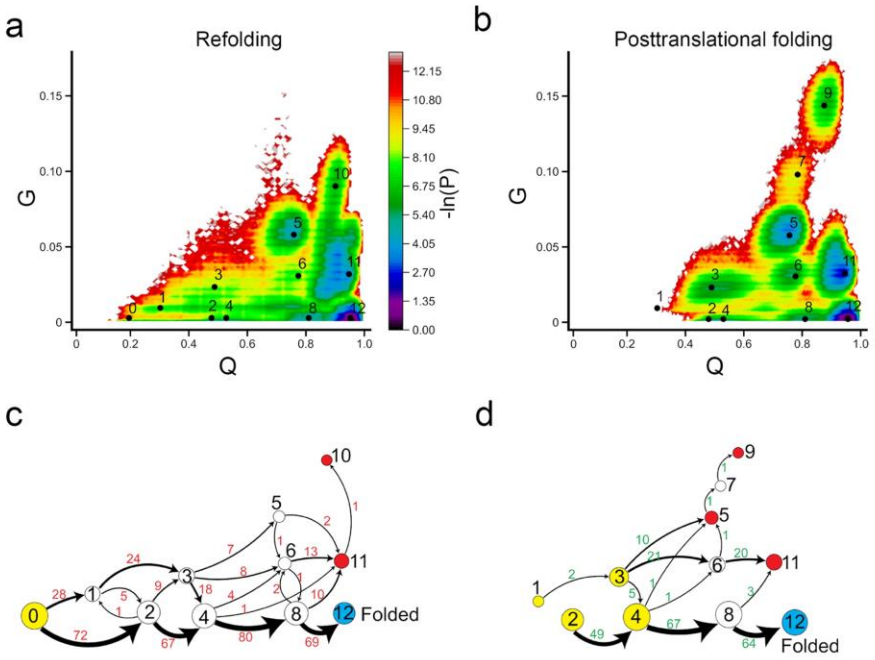


Figure 7. No difference in folding mechanisms of DDLB between refolding and posttranslational folding. (a, b) $-\ln(P)$ surface in refolding and posttranslational folding, respectively, where P is the probability of sampling particular Q and G values. The centers of metastable states and their corresponding indices are shown on top of the surface (black points). (c, d) Transition network from discrete trajectories of refolding and posttranslational folding simulations. The yellow, red, and sky blue nodes correspond to the initial, misfolded, and folded states, respectively. The black numbers on the nodes match the indices of metastable states in panels (a) and (b). The red numbers beside the edges indicate the number of direct transitions between states observed in discrete trajectories.

posttranslational folding, this process is similar to folding from unfolded ensembles.

There are two critical classes of intermediate states in the folding of CAT-III: state 1, which leads to misfolding when some native contacts change entanglement, and state 2, which leads to the native state without changing entanglement. In posttranslational folding, a large number of trajectories initiate in state 1 (68%, then transition to state 3) and state 3 (9%), with some portions of native contacts changing entanglement. These trajectories mainly end up in misfolded states (states 9 and 12). Only 27% (CI 95% [19%, 36%]) of total trajectories can fold to the native state. In refolding, the process starts in the fully unfolded state 0 and diversifies to state 1 (40%), where some contacts change entanglement and lead to further misfolding, and a larger number of trajectories go to state 2 (60%) and then fold correctly to the native state. This results in slightly more folded trajectories in refolding (38%, CI 95% [30%, 49%]) compared to posttranslational folding. Thus, protein synthesis and posttranslational folding do not increase the folding efficiency of CAT-III compared to refolding but rather cause the protein to partially fold into misfolded intermediate states.

States 9 and 12 are likely long-lived misfolded states, as even when we extended the simulation time to 30 μs, we did not observe any misfolded trajectories folding to the native state (when considering both Q and G parameters). All of these misfolded states are near-native (high Q) and have a large number of native contacts changing entanglement (Figure S4).

Similar to CAT-III, the ribosome does not aid in the proper folding of DDLB (Figure 7 and Table 4). Our simulations indicate that the overall $-\ln(P)$ surface is similar in refolding and posttranslational folding simulations. The dominant folding pathway is * → 2 → 4 → 8 → 12. In the posttranslational folding simulation, if the DDLB protein is in states 2 or 4 after protein synthesis (which occurs in 64% of trajectories), it has a high likelihood of successfully folding posttranslationally (2|4 → folded: 98.4%). On the other hand, if the protein is in states 1 or 3 after protein synthesis (36% of all trajectories in our simulations), it is likely to result in a misfolded state posttranslationally (1|3 → misfolded: 97.7%).

Table 4. Percentage of Folding Pathways of DDLB in Refolding and Posttranslational Folding Simulations

pathways	percent (%)	pathways	percent (%)
Refolding			
0 → 1	28	1 → misfolded	67.9
0 → 2	72	2 → folded	83.3
Posttranslational folding			
cotranslational folding → 1|3	36	1|3 → misfolded	97.7
cotranslational folding → 2|4	64	2|4 → folded	98.4

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

This has also been observed experimentally for other proteins.[13,24,25]

Analysis of refolding pathways shows a similar distribution to posttranslational folding, with two classes of folding: one leading to correct folding (69%) and the other leading to misfolding (31%). In refolding simulations, proteins that start in a fully unfolded state (state 0) diversify into the intermediate misfolded state 1 (28% of transitions, with a change in entanglement) and remain trapped in misfolded states (1 → misfolded: 67.9%), while those that sink to state 2 mainly transition to the native state 12 (2 → folded: 83.3%).

State 10 is observed in refolding simulations but not in posttranslational folding, while states 7 and 9 are observed in posttranslational folding but not in refolding. These differences are exhibited in a single misfolded trajectory. In both refolding and posttranslational folding, we did not observe the transition from the near-native state 11 to the native state 12.

Overall, protein synthesis does not increase the folding efficiency of CAT-III and DDLB; intermediate states with non-native entanglement form cotranslationally and persist posttranslationally, and these states act as kinetic traps in protein folding. It should be noted that this work uses a "structure-based" model of protein folding, which encodes that the native state is the global minimum of free energy in our simulations; hence, misfolded states (i.e., those observed for CAT-III and DDLB) are metastable states and kinetically trapped, meaning that they have high free-energy barriers separated from the native state, making them convert to the native state very slowly. One possible limitation of our approach is that the non-native entangled states that we observed can be artifacts of our coarse-grained model. However, in a recent study, we showed that non-native entangled states also occur in all-atom simulations of proteins,[43] suggesting that they are not model-dependent. Moreover, various recent studies have also reported a correlation between changes in entanglement and digestion patterns from Limited Proteolysis Mass Spectrometry.[11,35] Taken together, these results suggest that our coarse-grained model predictions are reliable.

■ CONCLUSIONS

Protein folding in vivo is not solely regulated by the ribosome. Various other proteins and folding factors, such as chaperones, play a critical role in the process.[51−53] In this study, we aimed to investigate the influence of the ribosome on protein folding alone. While it is commonly believed that the ribosome is generally effective in assisting protein folding to native conformations,[14,15,54,55] our data do not consistently support this assumption. We do find the ribosome increases the folding efficiency of DHFR, in which two domains ABD and DLD fold independently. The ribosome confines the DLD domain inside the exit tunnel, allowing the ABD domain to fold cotranslationally and without interference; then, the DLD domain arranges into the correct native topology once released from the ribosome. In contrast, during refolding, all segments of the protein are simultaneously folding, presenting the opportunity for the formation of several non-native contacts between amino acids, thus enhancing the probability of being trapped in entangled misfolded states. For CAT-III and DDLB, which contain native entanglements, we did not observe an improvement in folding efficiency due to the ribosome, and in some cases, the ribosome caused these proteins to form intermediate misfolded states during cotranslational synthesis and these misfolded states persisted posttranslationally.

In conclusion, our findings suggest that the effect of ribosomes on protein folding is protein-specific and cannot be described by a universal rule. In general, the ribosome does not have a significant influence on folding outcomes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.3c01694.

Secondary structure diagrams of DHFR, CAT-III, and DDLB; fraction of native contacts versus time of ABD and DLD domains of DHFR; representative structures of the metastable states from clustering of DHFR; representative structures of the metastable states from clustering of CAT-III; representative structures of the metastable states from clustering of DDLB; mRNA templates used in continuous synthesis simulations; and structural definitions from pairs of secondary structures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Edward P. O'Brien − *Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States; Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences and Institute for Computational and Data Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0001-9809-3273; Email: epo2@psu.edu

Mai Suan Li − *Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland; Institute for Computational Sciences and Technology, Ho Chi Minh City 700000, Vietnam;* orcid.org/0000-0001-7021-7916; Email: masli@ifpan.edu.pl

Authors

Quyen V. Vu − *Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland;* orcid.org/0000-0002-9863-0486

Daniel A. Nissley − *Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.;* orcid.org/0000-0003-0550-9394

Yang Jiang − *Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0003-1100-9177

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.3c01694

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.S.L. acknowledges that this work was supported by the National Science Centre, Poland (grant 2019/35/B/ST4/02086). E.P.O. acknowledges support from the National Science Foundation (MCB-1553291) as well as the National Institutes of Health (R35-GM124818). This research was supported in part by the TASK Supercomputer Center in

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

Gdansk and PLGrid Infrastructure in Poland (Prometheus and Ares supercomputers).

REFERENCES

(1) Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. The Geometry of the Ribosomal Polypeptide Exit Tunnel. *J. Mol. Biol.* **2006**, *360*, 893−906.

(2) Nissley, D. A.; O'Brien, E. P. Structural Origins of FRET-Observed Nascent Chain Compaction on the Ribosome. *J. Phys. Chem. B* **2018**, *122*, 9927−9937.

(3) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. New Scenarios of Protein Folding Can Occur on the Ribosome. *J. Am. Chem. Soc.* **2011**, *133*, 513−526.

(4) Marino, J.; Von Heijne, G.; Beckmann, R. Small Protein Domains Fold inside the Ribosome Exit Tunnel. *FEBS Lett.* **2016**, *590*, 655−660.

(5) Nilsson, O. B.; Hedman, R.; Marino, J.; Wickles, S.; Bischoff, L.; Johansson, M.; Müller-Lucks, A.; Trovato, F.; Puglisi, J. D.; O'Brien, E. P.; et al. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **2015**, *12*, 1533−1540.

(6) Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P. In Vivo Translation Rates Can Substantially Delay the Cotranslational Folding of the Escherichia Coli Cytosolic Proteome. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E132−E140.

(7) Cabrita, L. D.; Cassaignau, A. M. E.; Launay, H. M. M.; Christopher, A. A Structural Ensemble of a Ribosome-Nascent Chain Complex during Co-Translational Protein Folding. *Nat. Struct. Mol. Biol.* **2017**, *23*, 278−285.

(8) Goldman, D. H.; Kaiser, C. M.; Milin, A.; Righini, M.; Tinoco, I.; Bustamante, C. Mechanical Force Releases Nascent Chain-Mediated Ribosome Arrest in Vitro and in Vivo. *Science* **2015**, *348*, 457−460.

(9) Farías-Rico, J. A.; Selin, F. R.; Myronidi, I.; Frühauf, M.; Von Heijne, G. Effects of Protein Size, Thermodynamic Stability, and Net Charge on Cotranslational Folding on the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E9280−E9287.

(10) Frydman, J. Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annu. Rev. Biochem.* **2001**, *70*, 603−647.

(11) Jiang, Y.; Neti, S. S.; Sitarik, I.; Pradhan, P.; To, P.; Xia, Y.; Fried, S. D.; Booker, S. J.; O'Brien, E. P. How Synonymous Mutations Alter Enzyme Structure and Function over Long Timescales. *Nat. Chem.* **2023**, *15*, 308−318.

(12) Buhr, F.; Jha, S.; Thommen, M.; Mittelstaet, J.; Kutz, F.; Schwalbe, H.; Rodnina, M. V.; Komar, A. A. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol. Cell* **2016**, *61*, 341−351.

(13) Alexander, L. M.; Goldman, D. H.; Wee, L. M.; Bustamante, C. Non-Equilibrium Dynamics of a Nascent Polypeptide during Translation Suppress Its Misfolding. *Nat. Commun.* **2019**, *10*, No. 2709.

(14) Liutkute, M.; Samatova, E.; Rodnina, M. V. Cotranslational Folding of Proteins on the Ribosome. *Biomolecules* **2020**, *10*, 97.

(15) Netzer, W. J.; Hartl, F. U. Recombination of Protein Domains Facilitated by Co-Translational Folding in Eukaryotes. *Nature* **1997**, *388*, 343−349.

(16) To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D. Nonrefoldability Is Pervasive across the E. Coli Proteome. *J. Am. Chem. Soc.* **2021**, *143*, 11435−11448.

(17) Kaiser, C. M.; Goldman, D. H.; Chodera, J. D.; Tinoco, I.; Bustamante, C. The Ribosome Modulates Nascent Protein Folding. *Science* **2011**, *334*, 1723−1727.

(18) Tanaka, T.; Hori, N.; Takada, S. How Co-Translational Folding of Multi-Domain Protein Is Affected by Elongation Schedule: Molecular Simulations. *PLoS Comput. Biol.* **2015**, *11*, e1004356.

(19) Samelson, A. J.; Jensen, M. K.; Soto, R. A.; Cate, J. H. D.; Marqusee, S. Quantitative Determination of Ribosome Nascent Chain Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 13402−13407.

(20) Dabrowski-Tumanski, P.; Piejko, M.; Niewieczerzal, S.; Stasiak, A.; Sulkowska, J. I. Protein Knotting by Active Threading of Nascent Polypeptide Chain Exiting from the Ribosome Exit Channel. *J. Phys. Chem. B* **2018**, *122*, 11616−11625.

(21) Tian, P.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; von Heijne, G.; Clarke, J.; Best, R. B. Folding Pathway of an Ig Domain Is Conserved on and off the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E11284−E11293.

(22) Guinn, E. J.; Tian, P.; Shin, M.; Best, R. B.; Marqusee, S. A Small Single-Domain Protein Folds through the Same Pathway on and off the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 12206−12211.

(23) Liu, K.; Maciuba, K.; Kaiser, C. M. The Ribosome Cooperates with a Chaperone to Guide Multi-Domain Protein Folding. *Mol. Cell* **2019**, *74*, 310−319.e7.

(24) Plessa, E.; Chu, L. P.; Chan, S. H. S.; Thomas, O. L.; Cassaignau, A. M. E.; Waudby, C. A.; Christodoulou, J.; Cabrita, L. D. Nascent Chains Can Form Co-Translational Folding Intermediates That Promote Post-Translational Folding Outcomes in a Disease-Causing Protein. *Nat. Commun.* **2021**, *12*, No. 6447.

(25) Liu, K.; Rehfus, J. E.; Mattson, E.; Kaiser, C. M. The Ribosome Destabilizes Native and Non-Native Structures in a Nascent Multidomain Protein. *Protein Sci.* **2017**, *26*, 1439−1451.

(26) van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P. E.; Fraser, J. S. Automated Identification of Functional Dynamic Contact Networks from X-Ray Crystallography. *Nat. Methods* **2013**, *10*, 896−902.

(27) Leslie, A. G. W. Refined Crystal Structure of Type III Chloramphenicol Acetyltransferase at 1·75 Å Resolution. *J. Mol. Biol.* **1990**, *213*, 167−186.

(28) Batson, S.; De Chiara, C.; Majce, V.; Lloyd, A. J.; Gobec, S.; Rea, D.; Fülöp, V.; Thoroughgood, C. W.; Simmons, K. J.; Dowson, C. G.; et al. Inhibition of D-Ala:D-Ala Ligase through a Phosphorylated Form of the Antibiotic D-Cycloserine. *Nat. Commun.* **2017**, *8*, No. 1939.

(29) Arai, M.; Iwakura, M.; Matthews, C. R.; Bilsel, O. Microsecond Subdomain Folding in Dihydrofolate Reductase. *J. Mol. Biol.* **2011**, *410*, 329−342.

(30) Wales, T. E.; Pajak, A.; Roeselová, A.; Shivakumaraswamy, S.; Howell, S.; Hartl, F. U.; Engen, J. R.; Balchin, D. Resolving Chaperone-Assisted Protein Folding on the Ribosome at the Peptide Level. *bioRxiv* 2022. DOI: 10.1101/2022.09.23.509153.

(31) Liu, C. T.; Francis, K.; Layfield, J. P.; Huang, X.; Hammes-Schiffer, S.; Kohen, A.; Benkovic, S. J. Escherichia Coli Dihydrofolate Reductase Catalyzed Proton and Hydride Transfers: Temporal Order and the Roles of Asp27 and Tyr100. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 18231−18236.

(32) Day, P. J.; Shaw, W. V. Acetyl Coenzyme A Binding by Chloramphenicol Acetyltransferase. Hydrophobic Determinants of Recognition and Catalysis. *J. Biol. Chem.* **1992**, *267*, 5122−5127.

(33) al-Bar, O. A. M.; O'Connor, C. D.; Giles, I. G.; Akhtar, M. D-Alanine:D-Alanine Ligase of Escherichia Coli. Expression, Purification and Inhibitory Studies on the Cloned Enzyme. *Biochem. J.* **1992**, *282*, 747−752.

(34) Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J. Am. Chem. Soc.* **2020**, *142*, 6103−6110.

(35) Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P. Universal Protein Misfolding Intermediates Can Bypass the Proteostasis Network and Remain Soluble and Less Functional. *Nat. Commun.* **2022**, *13*, No. 3081.

(36) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(37) Frishman, D.; Argos, P. Knowledge-based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566−579.

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

(38) Halder, R.; Nissley, D. A.; Sitarik, I.; O'Brien, E. P. Subpopulations of Soluble, Misfolded Proteins Commonly Bypass Chaperones: How It Happens at the Molecular Level. *bioRxiv* 2021. DOI: 10.1101/2021.08.18.456736.
(39) Li, M. S.; Kouza, M.; Hu, C. K. Refolding upon Force Quench and Pathways of Mechanical and Thermal Unfolding of Ubiquitin. *Biophys. J.* **2007**, *92*, 547−561.
(40) Baiesi, M.; Orlandini, E.; Seno, F.; Trovato, A. Exploring the Correlation between the Folding Rates of Proteins and the Entanglement of Their Native States. *J. Phys. A Math. Theor.* **2017**, *50*, No. 504001.
(41) Kauffman, L. *Knots and Physics, XVI*; World Scientific Pub. Co.: Singapore, 1993.
(42) Niemyska, W.; Millett, K. C.; Sulkowska, J. I. GLN: A Method to Reveal Unique Properties of Lasso Type Topology in Proteins. *Sci. Rep.* **2020**, *10*, No. 15186.
(43) Vu, Q. V.; Sitarik, I.; Jiang, Y.; Yadav, D.; Sharma, P.; Fried, S. D.; Li, M. S.; O'Brien, E. P. A Newly Identified Class of Protein Misfolding in All-Atom Folding Simulations Consistent with Limited Proteolysis Mass Spectrometry. *bioRxiv* 2022. DOI: 10.1101/2022.07.19.500586.
(44) Visalakshi, N. K.; Thangavel, K. Impact of Normalization in Distributed K-Means Clustering. *Int. J. Soft Comput.* **2009**, *4*, 168−172.
(45) Macqueen, J. *Some Methods for Classification and Analysis of Multivariate Observations*, Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967; pp 281−297.
(46) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147−179.
(47) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(48) Hoffmann, M.; Scherer, M.; Hempel, T.; Mardt, A.; de Silva, B.; Husic, B. E.; Klus, S.; Wu, H.; Kutz, N.; Brunton, S. L.; Noé, F. Deeptime: A Python Library for Machine Learning Dynamical Models from Time Series Data. *Mach. Learn. Sci. Technol.* **2022**, *3*, No. 015009.
(49) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29*, 1460−1465.
(50) Baiesi, M.; Orlandini, E.; Seno, F.; Trovato, A. Sequence and Structural Patterns Detected in Entangled Proteins Reveal the Importance of Co-Translational Folding. *Sci. Rep.* **2019**, *9*, No. 8426.
(51) Willmund, F.; Del Alamo, M.; Pechmann, S.; Chen, T.; Albanèse, V.; Dammer, E. B.; Peng, J.; Frydman, J. The Cotranslational Function of Ribosome-Associated Hsp70 in Eukaryotic Protein Homeostasis. *Cell* **2013**, *152*, 196−209.
(52) Hartl, F. U.; Hayer-Hartl, M. Converging Concepts of Protein Folding in Vitro and in Vivo. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574−581.
(53) Kramer, G.; Boehringer, D.; Ban, N.; Bukau, B. The Ribosome as a Platform for Co-Translational Processing, Folding and Targeting of Newly Synthesized Proteins. *Nat. Struct. Mol. Biol.* **2009**, *16*, 589−597.
(54) Waudby, C. A.; Burridge, C.; Cabrita, L. D.; Christodoulou, J. Thermodynamics of Co-Translational Folding and Ribosome−Nascent Chain Interactions. *Curr. Opin. Struct. Biol.* **2022**, *74*, No. 102357.
(55) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Ulrich Hartl, F. Co-Translational Domain Folding as the Structural Basis for the Rapid de Novo Folding of Firefly Luciferase. *Nat. Struct. Biol.* **1999**, *6*, 697−705.

https://doi.org/10.1021/acs.jpcb.3c01694
J. Phys. Chem. B 2023, 127, 4761−4774

Chapter 6. CONCLUSIONS AND FUTURE DIRECTIONS

6.1 CONCLUSIONS

The research presented in this thesis aimed to gain fundamental insight into the process of protein ejection and folding on the ribosome using various theoretical and computational techniques. Using coarse-grained and all-atom MD simulation coupled with enhanced sampling techniques, we obtained the following results:

1. Ejection time spans at least two orders of magnitude, meaning some proteins eject very slowly.
2. Due to charged rRNA, electrostatic interactions are the primary driving force for very slow and very fast ejection.
3. Slow ejection can have the biological consequence of delaying later stages of protein translation.
4. Near the ribosome the contact minimum between two methane molecules is half as stable as compared to in bulk solution, demonstrating that the hydrophobic effect is weakened in the presence of the ribosome.
5. Thermodynamic decomposition and structural analyses reveal that the weakening of the hydrophobic effect is due to the increased ordering of water molecules in the presence of the ribosome. Specifically, increased water ordering reduces the entropy gain of water released from the first-solvation shell upon association of the two hydrophobic groups, weakening the driving force for the hydrophobic association.
6. It was shown that the protein stability, described by the differences in Gibbs free energies between the folded and the unfolded states, is decreased by 30% in the presence of the ribosome.
7. We showed that the influence of the ribosome on protein folding mechanisms varies depending on the size and complexity of the protein. DHFR folds more efficiently due to protein synthesis, while the ribosome does not promote the folding of CAT-III and DDLB and may contribute to the formation of intermediate misfolded states during

translation. These misfolded states persist after translation and do not convert to the native state over a long period.

8. Analyzing the folded trajectories from our simulations, we find that the sequence of secondary structure formations is significantly different for DHFR, while CAT-III and DDLB are robust on and off the ribosome.
9. The presence of native entanglement plays an essential role in the folding process of proteins. DHFR does not contain any entanglement in its native structure, while CAT-III and DDLB contain many entanglements in their native structure.
10. Considering the existence of entanglement, we find that protein synthesis assists the folding of DHFR by avoiding misfolded states with non-native entanglements compared to refolding from the unfolded ensemble. In contrast, these non-native entangled states act as a kinetic trap in both refolding and posttranslational folding of CAT-III and DDLB.

6.2 FUTURE DIRECTIONS

I utilized methane as a hydrophobic model to estimate that protein stability's free energy decreases by approximately 30% in the ribosomal vestibule. It is interesting and could significantly impact future studies of an actual protein folding on the ribosome. To this end, using all-atom molecular dynamics simulations, I plan to investigate the folding of small proteins like villin in the ribosome vestibule and in bulk solution. We expect that due to decreased hydrophobic effect in the ribosome exit tunnel, the folding in solution should be faster than on the ribosome.

Using the coarse-grained model, I have predicted the existence of misfolded states with non-native entanglement in CAT-III and DDLB. It should be helpful to confirm this conclusion in the all-atom model.

REFERENCES

(1) Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science (80-.).* **1973**, *181* (4096), 223–230. https://doi.org/10.1126/science.181.4096.223.

(2) Miller, S. B. M.; Mogk, A.; Bukau, B. Spatially Organized Aggregation of Misfolded Proteins as Cellular Stress Defense Strategy. *J. Mol. Biol.* **2015**, *427* (7), 1564–1574. https://doi.org/10.1016/j.jmb.2015.02.006.

(3) Sweeney, P.; Park, H.; Baumann, M.; Dunlop, J.; Frydman, J.; Kopito, R.; McCampbell, A.; Leblanc, G.; Venkateswaran, A.; Nurmi, A.; Hodgson, R. Protein Misfolding in Neurodegenerative Diseases: Implications and Strategies. *Transl. Neurodegener.* **2017**, *6* (1), 1–13. https://doi.org/10.1186/s40035-017-0077-5.

(4) Hartl, F. U. Protein Misfolding Diseases. *Annu. Rev. Biochem.* **2017**, *86* (1), 21–26. https://doi.org/10.1146/annurev-biochem-061516-044518.

(5) Levinthal, C. How to Fold Graciously. *Mössbauer Spectrosc. Biol. Syst. Proc.* **1969**, *24* (41), 22–24.

(6) Garbuzynskiy, S. O.; Ivankov, D. N.; Bogatyreva, N. S.; Finkelstein, A. V. Golden Triangle for Folding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (1), 147–150. https://doi.org/10.1073/pnas.1210180110.

(7) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4* (1), 10–19. https://doi.org/10.1038/nsb0197-10.

(8) Onuchic, J. N.; Wolynes, P. G. Theory of Protein Folding. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70–75. https://doi.org/10.1016/j.sbi.2004.01.009.

(9) Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys.* **1968**, *65*, 44–45.

(10) Luheshi, L. M.; Crowther, D. C.; Dobson, C. M. Protein Misfolding and Disease: From the Test Tube to the Organism. *Curr. Opin. Chem. Biol.* **2008**, *12* (1), 25–31. https://doi.org/10.1016/j.cbpa.2008.02.011.

(11) Haber, E.; Anfinsen, C. B. Regeneration of Enzyme Activity by Air Oxidation of Reduced Subtilisin-Modified Ribonuclease. *J. Biol. Chem.* **1961**, *236* (2), 422–424. https://doi.org/10.1016/S0021-9258(18)64379-0.

(12) Brockwell, D. J.; Radford, S. E. Intermediates: Ubiquitous Species on Folding Energy Landscapes? *Curr. Opin. Struct. Biol.* **2007**, *17* (1), 30–37.

https://doi.org/10.1016/j.sbi.2007.01.003.

(13) Ekman, D.; Björklund, Å. K.; Frey-Skött, J.; Elofsson, A. Multi-Domain Proteins in the Three Kingdoms of Life: Orphan Domains and Other Unassigned Regions. *J. Mol. Biol.* **2005**, *348* (1), 231–243. https://doi.org/10.1016/j.jmb.2005.02.007.

(14) Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. Protein Folding: Then and Now. *Arch. Biochem. Biophys.* **2008**, *469* (1), 4–19. https://doi.org/10.1016/j.abb.2007.05.014.

(15) Jahn, M.; Buchner, J.; Hugel, T.; Rief, M. Folding and Assembly of the Large Molecular Machine Hsp90 Studied in Single-Molecule Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (5), 1232–1237. https://doi.org/10.1073/pnas.1518827113.

(16) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In Vivo Aspects of Protein Folding and Quality Control. *Science (80-.).* **2016**, *353* (6294). https://doi.org/10.1126/science.aac4354.

(17) Willmund, F.; Del Alamo, M.; Pechmann, S.; Chen, T.; Albanèse, V.; Dammer, E. B.; Peng, J.; Frydman, J. The Cotranslational Function of Ribosome-Associated Hsp70 in Eukaryotic Protein Homeostasis. *Cell* **2013**, *152* (1–2), 196–209. https://doi.org/10.1016/j.cell.2012.12.001.

(18) Hartl, F. U.; Hayer-Hartl, M. Converging Concepts of Protein Folding in Vitro and in Vivo. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 574–581. https://doi.org/10.1038/nsmb.1591.

(19) Kramer, G.; Boehringer, D.; Ban, N.; Bukau, B. The Ribosome as a Platform for Co-Translational Processing, Folding and Targeting of Newly Synthesized Proteins. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 589–597. https://doi.org/10.1038/nsmb.1614.

(20) Netzer, W. J.; Hartl, F. U. Recombination of Protein Domains Facilitated by Co-Translational Folding in Eukaryotes. *Nature* **1997**, *388* (6640), 343–349. https://doi.org/10.1038/41024.

(21) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Ulrich Hartl, F. Co-Translational Domain Folding as the Structural Basis for the Rapid de Novo Folding of Firefly Luciferase. *Nat. Struct. Biol.* **1999**, *6* (7), 697–705. https://doi.org/10.1038/10754.

(22) Cassaignau, A. M. E.; Cabrita, L. D.; Christodoulou, J. How Does the Ribosome Fold the Proteome? *Annu. Rev. Biochem.* **2020**, *89*, 389–415. https://doi.org/10.1146/annurev-biochem-062917-012226.

(23) Pechmann, S.; Willmund, F.; Frydman, J. The Ribosome as a Hub for Protein Quality

Control. *Mol. Cell* **2013**, *49* (3), 411–421. https://doi.org/10.1016/j.molcel.2013.01.020.

(24) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29* (31), 7133–7155. https://doi.org/10.1021/bi00483a001.

(25) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Shirley, B. A.; Hendricks, M. M. N.; Iimura, S.; Gajiwala, K.; Scholtz, J. M.; Grimsley, G. R. Contribution of Hydrophobic Interactions to Protein Stability. *J. Mol. Biol.* **2011**, *408* (3), 514–528. https://doi.org/10.1016/j.jmb.2011.02.053.

(26) Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces Contributing to the Conformational Stability of Proteins. *FASEB J.* **1996**, *10* (1), 75–83. https://doi.org/10.1096/fasebj.10.1.8566551.

(27) Dunkle, J. A.; Wang, L.; Feldman, M. B.; Pulk, A.; Chen, V. B.; Kapral, G. J.; Noeske, J.; Richardson, J. S.; Blanchard, S. C.; Cate, J. H. D. Structures of the Bacterial Ribosome in Classical and Hybrid States of TRNA Binding. *Science (80-.).* **2011**, *332* (6032), 981–984. https://doi.org/10.1126/science.1202692.

(28) Alberts, B.; Johnson, A.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; Wilson, J., Hunt, T., Eds.; W.W. Norton & Company, 2017; Vol. 58. https://doi.org/10.1201/9781315735368.

(29) Yonath, A. Antibiotics Targeting Ribosomes: Resistance, Selectivity, Synergism, and Cellular Regulation. *Annu. Rev. Biochem.* **2005**, *74*, 649–679. https://doi.org/10.1146/annurev.biochem.74.082803.133130.

(30) Schlünzen, F.; Zarivach, R.; Harms, J.; Bashan, A.; Tocilj, A.; Albrecht, R.; Yonath, A.; Franceschi, F. Structural Basis for the Interaction of Antibiotics with the Peptidyl Transferase Centre in Eubacteria. *Nature* **2001**, *413* (6858), 814–821. https://doi.org/10.1038/35101544.

(31) Liutkute, M.; Samatova, E.; Rodnina, M. V. Cotranslational Folding of Proteins on the Ribosome. *Biomolecules* **2020**, *10* (1), 97. https://doi.org/10.3390/biom10010097.

(32) Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. The Geometry of the Ribosomal Polypeptide Exit Tunnel. *J. Mol. Biol.* **2006**, *360* (4), 893–906. https://doi.org/10.1016/j.jmb.2006.05.023.

(33) Yonath, A.; Leonard, K. R.; Wittmann, H. G. A Tunnel in the Large Ribosomal Subunit

Revealed by Three-Dimensional Image Reconstruction. *Science (80-.).* **1987**, *236* (4803), 813–816. https://doi.org/10.1126/science.3576200.

(34) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science (80-.).* **2000**, *289* (5481), 920–930. https://doi.org/10.1126/science.289.5481.920.

(35) Dao Duc, K.; Batra, S. S.; Bhattacharya, N.; Cate, J. H. D.; Song, Y. S. Differences in the Path to Exit the Ribosome across the Three Domains of Life. *Nucleic Acids Res.* **2019**, *47* (8), 4198–4210. https://doi.org/10.1093/nar/gkz106.

(36) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. New Scenarios of Protein Folding Can Occur on the Ribosome. *J. Am. Chem. Soc.* **2011**, *133* (3), 513–526. https://doi.org/10.1021/ja107863z.

(37) Lu, J.; Kobertz, W. R.; Deutsch, C. Mapping the Electrostatic Potential within the Ribosomal Exit Tunnel. *J. Mol. Biol.* **2007**, *371* (5), 1378–1391. https://doi.org/10.1016/j.jmb.2007.06.038.

(38) Choi, J.; Grosely, R.; Prabhakar, A.; Lapointe, C. P.; Wang, J.; Puglisi, J. D. How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. *Annu. Rev. Biochem.* **2015**, *87* (1), 421–449. https://doi.org/10.1146/annurev-biochem-060815-014818.

(39) Dao Duc, K.; Song, Y. S. *The Impact of Ribosomal Interference, Codon Usage, and Exit Tunnel Interactions on Translation Elongation Rate Variation*; 2018; Vol. 14. https://doi.org/10.1371/journal.pgen.1007166.

(40) Thommen, M.; Holtkamp, W.; Rodnina, M. V. Co-Translational Protein Folding: Progress and Methods. *Curr. Opin. Struct. Biol.* **2017**, *42*, 83–89. https://doi.org/10.1016/j.sbi.2016.11.020.

(41) Wruck, F.; Tian, P.; Kudva, R.; Best, R. B.; von Heijne, G.; Tans, S. J.; Katranidis, A. The Ribosome Modulates Folding inside the Ribosomal Exit Tunnel. *Commun. Biol.* **2021**, *4* (1), 1–8. https://doi.org/10.1038/s42003-021-02055-8.

(42) Nilsson, O. B.; Hedman, R.; Marino, J.; Wickles, S.; Bischoff, L.; Johansson, M.; Müller-Lucks, A.; Trovato, F.; Puglisi, J. D.; O'Brien, E. P.; Beckmann, R.; von Heijne, G. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Rep.* **2015**, *12* (10), 1533–1540. https://doi.org/10.1016/j.celrep.2015.07.065.

(43) Kudva, R.; Tian, P.; Pardo-Avila, F.; Carroni, M.; Best, R. B.; Bernstein, H. D.; von Heijne, G. The Shape of the Bacterial Ribosome Exit Tunnel Affects Cotranslational Protein Folding. *Elife* **2018**, *7*, 1–15. https://doi.org/10.7554/eLife.36326.

(44) Mankin, A. S. Nascent Peptide in the "birth Canal" of the Ribosome. *Trends Biochem. Sci.* **2006**, *31* (1), 11–13. https://doi.org/10.1016/j.tibs.2005.11.007.

(45) Murakami, A.; Nakatogawa, H.; Ito, K. Translation Arrest of SecM Is Essential for the Basal and Regulated Expression of SecA. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (33), 12330–12335. https://doi.org/10.1073/pnas.0404907101.

(46) Zhang, J.; Pan, X.; Yan, K.; Sun, S.; Gao, N.; Sui, S.-F. Mechanisms of Ribosome Stalling by SecM at Multiple Elongation Steps. *Elife* **2015**, *4*, 1–25. https://doi.org/10.7554/eLife.09684.

(47) E., B. M.; B., P. L.; B., O. D. Translocon "Pulling" of Nascent SecM Controls the Duration of Its Translational Pause and Secretion-Responsive SecA Regulation. *J. Bacteriol.* **2003**, *185* (22), 6719–6722. https://doi.org/10.1128/JB.185.22.6719-6722.2003.

(48) Ismail, N.; Hedman, R.; Schiller, N.; von Heijne, G. A Biphasic Pulling Force Acts on Transmembrane Helices during Translocon-Mediated Membrane Integration. *Nat. Struct. Mol. Biol.* **2012**, *19* (10), 1018–1022. https://doi.org/10.1038/nsmb.2376.

(49) Marino, J.; Von Heijne, G.; Beckmann, R. Small Protein Domains Fold inside the Ribosome Exit Tunnel. *FEBS Lett.* **2016**, *590* (5), 655–660. https://doi.org/10.1002/1873-3468.12098.

(50) Leininger, S. E.; Narayan, K.; Deutsch, C.; O'Brien, E. P. Mechanochemistry in Translation. *Biochemistry* **2019**, *58* (47), 4657–4666. https://doi.org/10.1021/acs.biochem.9b00260.

(51) Lucent, D.; Snow, C. D.; Aitken, C. E.; Pande, V. S. Non-Bulk-like Solvent Behavior in the Ribosome Exit Tunnel. *PLoS Comput. Biol.* **2010**, *6* (10), e1000963.

(52) Leininger, S. E.; Rodriguez, J.; Vu, Q. V.; Jiang, Y.; Li, M. S.; Deutsch, C.; O'Brien, E. P. Ribosome Elongation Kinetics of Consecutively Charged Residues Are Coupled to Electrostatic Force. *Biochemistry* **2021**, *60* (43), 3223–3235. https://doi.org/10.1021/acs.biochem.1c00507.

(53) Wilson, D. N.; Beckmann, R. The Ribosomal Tunnel as a Functional Environment for Nascent Polypeptide Folding and Translational Stalling. *Curr. Opin. Struct. Biol.* **2011**, *21*

(2), 274–282. https://doi.org/10.1016/j.sbi.2011.01.007.

(54) Rajasekaran, N.; Kaiser, C. M. Co-Translational Folding of Multi-Domain Proteins. *Front. Mol. Biosci.* **2022**, *9* (April), 1–9. https://doi.org/10.3389/fmolb.2022.869027.

(55) Eichmann, C.; Preissler, S.; Riek, R.; Deuerling, E. Cotranslational Structure Acquisition of Nascent Polypeptides Monitored by NMR Spectroscopy. **2010**, *107* (20), 9111–9116. https://doi.org/10.1073/pnas.0914300107.

(56) Han, Y.; David, A.; Liu, B.; Magadán, J. G.; Bennink, J. R.; Yewdell, J. W.; Qian, S. B. Monitoring Cotranslational Protein Folding in Mammalian Cells at Codon Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (31), 12467–12472. https://doi.org/10.1073/pnas.1208138109.

(57) Dobson, C. M.; Hsu, S. D.; Fucini, P.; Cabrita, L. D.; Christodoulou, J. Structure and Dynamics of a Ribosome-Bound Nascent Chain by NMR Spectroscopy. **2007**. https://doi.org/10.1073/pnas.0704664104.

(58) Frydman, J.; Nimmesgern, E.; Ohtsuka, K.; Hartl, F. U. Folding of Nascent Polypeptide Chains in a High Molecular Mass Assembly with Molecular Chaperones. *Nature* **1994**, *370* (6485), 111–117. https://doi.org/10.1038/370111a0.

(59) Bergman, L. W.; Kuehl, W. M. Formation of Intermolecular Disulfide Bonds on Nascent Immunoglobulin Polypeptides. *J. Biol. Chem.* **1979**, *254* (13), 5690–5694. https://doi.org/10.1016/s0021-9258(18)50469-5.

(60) Chen, W.; Helenius, J.; Braakman, I.; Helenius, A. Cotranslational Folding and Calnexin Binding during Glycoprotein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (14), 6229–6233. https://doi.org/10.1073/pnas.92.14.6229.

(61) Kolb, V. A.; Makeyev, E. V.; Spirin, A. S. Folding of Firefly Luciferase during Translation in a Cell-Free System. *EMBO J.* **1994**, *13* (15), 3631–3637. https://doi.org/10.1002/j.1460-2075.1994.tb06670.x.

(62) Fedyukina, D. V.; Cavagnero, S. Protein Folding at the Exit Tunnel. *Annu. Rev. Biophys.* **2011**, *40* (1), 337–359. https://doi.org/10.1146/annurev-biophys-042910-155338.

(63) Brocchieri, L.; Karlin, S. Protein Length in Eukaryotic and Prokaryotic Proteomes. *Nucleic Acids Res.* **2005**, *33* (10), 3390–3400. https://doi.org/10.1093/nar/gki615.

(64) Cabrita, L. D.; Cassaignau, A. M. E.; Launay, H. M. M.; Christopher, A. A Structural Ensemble of a Ribosome-Nascent Chain Complex during Co-Translational Protein

Folding. **2017**, *23* (4), 278–285. https://doi.org/10.1038/nsmb.3182.A.

(65) Goldman, D. H.; Kaiser, C. M.; Milin, A.; Righini, M.; Tinoco, I.; Bustamante, C. Mechanical Force Releases Nascent Chain-Mediated Ribosome Arrest in Vitro and in Vivo. *Science (80-.).* **2015**, *348* (6233), 457–460. https://doi.org/10.1126/science.1261909.

(66) Farías-Rico, J. A.; Selin, F. R.; Myronidi, I.; Frühauf, M.; Von Heijne, G. Effects of Protein Size, Thermodynamic Stability, and Net Charge on Cotranslational Folding on the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (40), E9280–E9287. https://doi.org/10.1073/pnas.1812756115.

(67) Frydman, J. Folding of Newly Translated Proteins In Vivo: The Role of Molecular Chaperones. *Annu. Rev. Biochem.* **2001**, *70* (1), 603–647. https://doi.org/10.1146/annurev.biochem.70.1.603.

(68) Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P. In Vivo Translation Rates Can Substantially Delay the Cotranslational Folding of the Escherichia Coli Cytosolic Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (2), E132-140. https://doi.org/10.1073/pnas.1213624110.

(69) To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D. Nonrefoldability Is Pervasive across the E. Coli Proteome. *J. Am. Chem. Soc.* **2021**, *143* (30), 11435–11448. https://doi.org/10.1021/jacs.1c03270.

(70) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In Vivo Aspects of Protein Folding and Quality Control. *Science (80-.).* **2016**, *353* (6294), aac4354. https://doi.org/10.1126/science.aac4354.

(71) Rodnina, M. V; Wintermeyer, W. Protein Elongation, Co-Translational Folding and Targeting. *J. Mol. Biol.* **2016**, *428* (10), 2165–2185. https://doi.org/10.1016/j.jmb.2016.03.022.

(72) Ziv, G.; Haran, G.; Thirumalai, D. Ribosome Exit Tunnel Can Entropically Stabilize α-Helices. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (52), 18956–18961. https://doi.org/10.1073/pnas.0508234102.

(73) Bhushan, S.; Gartmann, M.; Halic, M.; Armache, J.-P.; Jarasch, A.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. α-Helical Nascent Polypeptide Chains Visualized within Distinct Regions of the Ribosomal Exit Tunnel. *Nat. Struct. Mol. Biol.*

2010, *17* (3), 313–317. https://doi.org/10.1038/nsmb.1756.

(74) Woolhead, C. A.; McCormick, P. J.; Johnson, A. E. Nascent Membrane and Secretory Proteins Differ in FRET-Detected Folding Far inside the Ribosome and in Their Exposure to Ribosomal Proteins. *Cell* **2004**, *116* (5), 725–736. https://doi.org/10.1016/S0092-8674(04)00169-2.

(75) Holtkamp, W.; Kokic, G.; Jäger, M.; Mittelstaet, J.; Komar, A. A.; Rodnina, M. V. Cotranslational Protein Folding on the Ribosome Monitored in Real Time. *Science (80-.).* **2015**, *350* (6264), 1104–1107. https://doi.org/10.1126/science.aad0344.

(76) Tian, P.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; Von Heijne, G.; Clarke, J.; Best, R. B. Folding Pathway of an Ig Domain Is Conserved on and off the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (48), E11284–E11293. https://doi.org/10.1073/pnas.1810523115.

(77) Agirrezabala, X.; Samatova, E.; Macher, M.; Liutkute, M.; Maiti, M.; Gil-Carton, D.; Novacek, J.; Valle, M.; Rodnina, M. V. A Switch from A-helical to B-strand Conformation during Co-translational Protein Folding. *EMBO J.* **2022**, *41* (4), 1–13. https://doi.org/10.15252/embj.2021109175.

(78) Samelson, A. J.; Jensen, M. K.; Soto, R. A.; Cate, J. H. D.; Marqusee, S. Quantitative Determination of Ribosome Nascent Chain Stability. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (47), 13402–13407. https://doi.org/10.1073/pnas.1610272113.

(79) Liu, K.; Chen, X.; Kaiser, C. M. Energetic Dependencies Dictate Folding Mechanism in a Complex Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (51), 25641–25648. https://doi.org/10.1073/pnas.1914366116.

(80) Kaiser, C. M.; Goldman, D. H.; Chodera, J. D.; Tinoco, I.; Bustamante, C. The Ribosome Modulates Nascent Protein Folding. *Science (80-.).* **2011**, *334* (6063), 1723–1727. https://doi.org/10.1126/science.1209740.

(81) Liu, K.; Maciuba, K.; Kaiser, C. M. The Ribosome Cooperates with a Chaperone to Guide Multi-Domain Protein Folding. *Mol. Cell* **2019**, *74* (2), 310-319.e7. https://doi.org/10.1016/j.molcel.2019.01.043.

(82) Alexander, L. M.; Goldman, D. H.; Wee, L. M.; Bustamante, C. Non-Equilibrium Dynamics of a Nascent Polypeptide during Translation Suppress Its Misfolding. *Nat. Commun.* **2019**, *10* (1), 1–11. https://doi.org/10.1038/s41467-019-10647-6.

(83) Tanaka, T.; Hori, N.; Takada, S. How Co-Translational Folding of Multi-Domain Protein Is Affected by Elongation Schedule: Molecular Simulations. *PLoS Comput. Biol.* **2015**, *11* (7), 1–20. https://doi.org/10.1371/journal.pcbi.1004356.

(84) Dabrowski-Tumanski, P.; Piejko, M.; Niewieczerzal, S.; Stasiak, A.; Sulkowska, J. I. Protein Knotting by Active Threading of Nascent Polypeptide Chain Exiting from the Ribosome Exit Channel. *J. Phys. Chem. B* **2018**, *122* (49), 11616–11625. https://doi.org/10.1021/acs.jpcb.8b07634.

(85) Guinn, E. J.; Tian, P.; Shin, M.; Best, R. B.; Marqusee, S. A Small Single-Domain Protein Folds through the Same Pathway on and off the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (48), 12206–12211. https://doi.org/10.1073/pnas.1810517115.

(86) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(87) Weiner, P. K.; Kollman, P. A. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* **1981**, *2* (3), 287–303. https://doi.org/10.1002/jcc.540020311.

(88) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10). https://doi.org/10.1002/jcc.21287.

(89) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), 1–17. https://doi.org/10.1371/journal.pcbi.1005659.

(90) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936.

https://doi.org/10.1021/acs.chemrev.6b00163.

(91) Kar, P.; Feig, M. Hybrid All-Atom/Coarse-Grained Simulations of Proteins by Direct Coupling of CHARMM and PRIMO Force Fields. *J. Chem. Theory Comput.* **2017**, *13* (11), 5753–5765. https://doi.org/10.1021/acs.jctc.7b00840.

(92) Bayly, C. I.; Merz, K. M.; Ferguson, D. M.; Cornell, W. D.; Fox, T.; Caldwell, J. W.; Kollman, P. A.; Cieplak, P.; Gould, I. R.; Spellmeyer, D. C. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. https://doi.org/10.1021/ja00124a002.

(93) Kollman, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* **1996**, *29* (10). https://doi.org/10.1021/ar9500675.

(94) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049–1074. https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F.

(95) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function and Genetics*. 2006, pp 712–725. https://doi.org/10.1002/prot.21123.

(96) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958. https://doi.org/10.1002/prot.22711.

(97) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012. https://doi.org/10.1002/jcc.10349.

(98) Garcia, A. E.; Sanbonmatsu, K. Y. α-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (5), 2782–2787. https://doi.org/10.1073/pnas.042496899.

(99) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulation. *J. Comput. Chem.* **2004**, *25* (11), 1400–1415. https://doi.org/10.1002/jcc.20065.

(100) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. https://doi.org/10.1021/jp973084f.

(101) Feller, S. E.; MacKerell, A. D. An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. *J. Phys. Chem. B* **2000**, *104* (31), 7510–7515. https://doi.org/10.1021/jp0007843.

(102) Foloppe, N.; MacKerell, A. D. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21* (2), 86–104. https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.

(103) MacKerell, A. D.; Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21* (2), 105–120. https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P.

(104) Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water–Protein Interactions. *Biopolymers* **1984**, *23* (8), 1513–1518. https://doi.org/10.1002/bip.360230807.

(105) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; Van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103* (19), 3596–3607. https://doi.org/10.1021/jp984217f.

(106) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid

Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666. https://doi.org/10.1021/ja00214a001.

(107) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Genet.* **2006**, *65* (3), 712–725. https://doi.org/10.1002/prot.21123.

(108) Bereau, T.; Deserno, M. Generic Coarse-Grained Model for Protein Folding and Aggregation. *J. Chem. Phys.* **2009**, *130* (23). https://doi.org/10.1063/1.3152842.

(109) Bereau, T.; Bachmann, M.; Deserno, M. Interplay between Secondary and Tertiary Structure Formation in Protein Folding Cooperativity. *J. Am. Chem. Soc.* **2010**, *132* (38), 13129–13131. https://doi.org/10.1021/ja105206w.

(110) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9* (1), 687–697. https://doi.org/10.1021/ct300646g.

(111) Krainer, G.; Welsh, T. J.; Joseph, J. A.; Espinosa, J. R.; Wittmann, S.; de Csilléry, E.; Sridhar, A.; Toprakcioglu, Z.; Gudiškytė, G.; Czekalska, M. A.; Arter, W. E.; Guillén-Boixet, J.; Franzmann, T. M.; Qamar, S.; George-Hyslop, P. S.; Hyman, A. A.; Collepardo-Guevara, R.; Alberti, S.; Knowles, T. P. J. Reentrant Liquid Condensate Phase of Proteins Is Stabilized by Hydrophobic and Non-Ionic Interactions. *Nat. Commun.* **2021**, *12* (1), 1–14. https://doi.org/10.1038/s41467-021-21181-9.

(112) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved Coarse-Grained Model for Studying Sequence Dependent Phase Separation of Disordered Proteins. *Protein Sci.* **2021**, *30* (7), 1371–1379. https://doi.org/10.1002/pro.4094.

(113) Nguyen, H. T.; Hori, N.; Thirumalai, D. Condensates in RNA Repeat Sequences Are Heterogeneously Organized and Exhibit Reptation Dynamics. *Nat. Chem.* **2022**. https://doi.org/10.1038/s41557-022-00934-z.

(114) Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-Driven Coarse-Grained Model for Biomolecular Phase Separation with near-Quantitative Accuracy. *Nat. Comput. Sci.* **2021**,

1 (11), 732–743. https://doi.org/10.1038/s43588-021-00155-3.

(115) Nissley, D. A.; Vu, Q. V.; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J. Am. Chem. Soc.* **2020**, *142* (13), 6103–6110. https://doi.org/10.1021/jacs.9b12264.

(116) Jiang, Y.; Neti, S. S.; Sitarik, I.; Pradhan, P.; To, P.; Xia, Y.; Fried, S. D.; Booker, S. J.; O'Brien, E. P. How Synonymous Mutations Alter Enzyme Structure and Function over Long Timescales. *Nat. Chem.* **2023**, *15* (3), 308–318. https://doi.org/10.1038/s41557-022-01091-z.

(117) Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P. Universal Protein Misfolding Intermediates Can Bypass the Proteostasis Network and Remain Soluble and Less Functional. *Nat. Commun.* **2022**, *13* (1), 3081. https://doi.org/10.1038/s41467-022-30548-5.

(118) Best, R. B.; Chen, Y. G.; Hummer, G. Slow Protein Conformational Dynamics from Multiple Experimental Structures: The Helix/Sheet Transition of Arc Repressor. *Structure* **2005**, *13* (12), 1755–1763. https://doi.org/10.1016/j.str.2005.08.009.

(119) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Trigger Factor Slows Co-Translational Folding through Kinetic Trapping While Sterically Protecting the Nascent Chain from Aberrant Cytosolic Interactions. *J. Am. Chem. Soc.* **2012**, *134* (26), 10920–10932. https://doi.org/10.1021/ja302305u.

(120) Karanicolas, J.; Brooks, C. The Origins of Asymmetry in the Folding Transition States of Protein L and Protein G. *Protein Sci.* **2002**, *11*, 2351–2361. https://doi.org/10.2807/1560-7917.ES2014.19.46.20966.

(121) Betancourt, M. R.; Thirumalai, D. Pair Potentials for Protein Folding: Choice of Reference States and Sensitivity of Predicted Native States to Variations in the Interaction Schemes. *Protein Sci.* **2008**, *8* (2), 361–369. https://doi.org/10.1110/ps.8.2.361.

(122) Leininger, S. E.; Trovato, F.; Nissley, D. A.; O'Brien, E. P. Domain Topology, Stability, and Translation Speed Determine Mechanical Force Generation on the Ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (12), 5523–5532. https://doi.org/10.1073/pnas.1813003116.

(123) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics

Calculation of the Streptavidin-Biotin Rupture Force. *Science (80-.).* **1996**, *271* (5251), 997–999. https://doi.org/10.1126/science.271.5251.997.

(124) Binnig, G.; Quate, C. F.; Gerber, C. Atomic Force Microscope. *Phys. Rev. Lett.* **1986**, *56* (9), 930–933. https://doi.org/10.1103/PhysRevLett.56.930.

(125) Bustamante, C. J.; Chemla, Y. R.; Liu, S.; Wang, M. D. Optical Tweezers in Single-Molecule Biophysics. *Nat. Rev. Methods Prim.* **2021**, *1* (1). https://doi.org/10.1038/s43586-021-00021-6.

(126) Sarkar, R.; Rybenkov, V. V. A Guide to Magnetic Tweezers and Their Applications. *Front. Phys.* **2016**, *4* (DEC). https://doi.org/10.3389/fphy.2016.00048.

(127) Gräter, F.; Grubmüller, H. Fluctuations of Primary Ubiquitin Folding Intermediates in a Force Clamp. *J. Struct. Biol.* **2007**, *157* (3), 557–569. https://doi.org/10.1016/j.jsb.2006.11.005.

(128) Sahoo, A. K.; Bagchi, B.; Maiti, P. K. Unfolding Dynamics of Ubiquitin from Constant Force MD Simulation: Entropy-Enthalpy Interplay Shapes the Free-Energy Landscape. *J. Phys. Chem. B* **2019**, *123* (6), 1228–1236. https://doi.org/10.1021/acs.jpcb.8b09318.

(129) Vuong, Q. Van; Nguyen, T. T.; Li, M. S. A New Method for Navigating Optimal Direction for Pulling Ligand from Binding Pocket: Application to Ranking Binding Affinity by Steered Molecular Dynamics. *J. Chem. Inf. Model.* **2015**, *55* (12), 2731–2738. https://doi.org/10.1021/acs.jcim.5b00386.

(130) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. https://doi.org/https://doi.org/10.1016/0021-9991(77)90121-8.

(131) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/jcc.540130812.

(132) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91* (1–3), 275–282. https://doi.org/10.1016/0010-4655(95)00053-I.

(133) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration." *J. Chem.*

Phys. **2005**, *123* (14), 144104. https://doi.org/10.1063/1.2052648.

(134) Choudhury, N.; Montgomery Pettitt, B. Enthalpy-Entropy Contributions to the Potential of Mean Force of Nanoscopic Hydrophobic Solutes. *J. Phys. Chem. B* **2006**, *110* (16), 8459–8463. https://doi.org/10.1021/jp056909r.

(135) Chau, P.; Hardwick, A. J. A New Order Parameter for Tetrahedral Configurations. *Mol. Phys.* **1998**, *93* (3), 511–518. https://doi.org/10.1080/002689798169195.

(136) Errington, J. R.; Debenedetti, P. G. Relationship between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, *409* (6818), 318–321. https://doi.org/10.1038/35053024.

(137) Duboué-Dijon, E.; Laage, D. Characterization of the Local Structure in Liquid Water by Various Order Parameters. *J. Phys. Chem. B* **2015**, *119* (26), 8406–8418. https://doi.org/10.1021/acs.jpcb.5b02936.

(138) Frishman, D.; Argos, P. Knowledge-based Protein Secondary Structure Assignment. *Proteins Struct. Funct. Bioinforma.* **1995**, *23* (4), 566–579. https://doi.org/10.1002/prot.340230412.

(139) Mai, S. L.; Kouza, M.; Hu, C. K. Refolding upon Force Quench and Pathways of Mechanical and Thermal Unfolding of Ubiquitin. *Biophys. J.* **2007**, *92* (2), 547–561. https://doi.org/10.1529/biophysj.106.087684.

(140) Baiesi, M.; Orlandini, E.; Seno, F.; Trovato, A. Exploring the Correlation between the Folding Rates of Proteins and the Entanglement of Their Native States. *J. Phys. A Math. Theor.* **2017**, *50* (50), 504001. https://doi.org/10.1088/1751-8121/aa97e7.

(141) Kauffman, L. *Knots and Physics, XVI.*; World Scientific Pub. Co.: Singapore, 1993.

(142) Vu, Q. V; Sitarik, I.; Jiang, Y.; Yadav, D.; Sharma, P.; Fried, S. D.; Li, M. S.; O'Brien, E. P. A Newly Identified Class of Protein Misfolding in All-Atom Folding Simulations Consistent with Limited Proteolysis Mass Spectrometry. *bioRxiv* **2022**. https://doi.org/10.1101/2022.07.19.500586.

(143) Petrone, P. M.; Snow, C. D.; Lucent, D.; Pande, V. S. Side-Chain Recognition and Gating in the Ribosome Exit Tunnel. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (43), 16549–16554. https://doi.org/10.1073/pnas.0801795105.

(144) O'Brien, E. P.; Hsu, S. T. D.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Transient Tertiary Structure Formation within the Ribosome Exit Port. *J. Am. Chem. Soc.*

2010, *132* (47), 16928–16937. https://doi.org/10.1021/ja106530y.

(145) Sharma, A. K.; O'Brien, E. P. Non-Equilibrium Coupling of Protein Structure and Function to Translation–Elongation Kinetics. *Curr. Opin. Struct. Biol.* **2018**, *49*, 94–103. https://doi.org/10.1016/j.sbi.2018.01.005.

(146) Jensen, M. K.; Samelson, A. J.; Steward, A.; Clarke, J.; Marqusee, S. The Folding and Unfolding Behavior of Ribonuclease H on the Ribosome. *J. Biol. Chem.* **2020**, *295* (33), 11410–11417. https://doi.org/10.1074/jbc.RA120.013909.

(147) Cabrita, L. D.; Cassaignau, A. M. E.; Launay, H. M. M.; Waudby, C. A.; Wlodarski, T.; Camilloni, C.; Karyadi, M. E.; Robertson, A. L.; Wang, X.; Wentink, A. S.; Goodsell, L. S.; Woolhead, C. A.; Vendruscolo, M.; Dobson, C. M.; Christodoulou, J. A Structural Ensemble of a Ribosome-Nascent Chain Complex during Cotranslational Protein Folding. *Nat. Struct. Mol. Biol.* **2016**, *23* (4), 278–285. https://doi.org/10.1038/nsmb.3182.

(148) Liu, K.; Rehfus, J. E.; Mattson, E.; Kaiser, C. M. The Ribosome Destabilizes Native and Non-Native Structures in a Nascent Multidomain Protein. *Protein Sci.* **2017**, *26* (7), 1439–1451. https://doi.org/10.1002/pro.3189.

(149) Compiani, M.; Capriotti, E. Computational and Theoretical Methods for Protein Folding. *Biochemistry* **2013**, *52* (48), 8601–8624. https://doi.org/10.1021/bi4001529.

(150) Ghosh, T.; Kalra, A.; Garde, S. On the Salt-Induced Stabilization of Pair and Many-Body Hydrophobic Interactions. *J. Phys. Chem. B* **2005**, *109* (1), 642–651. https://doi.org/10.1021/jp0475638.

(151) Jo, S.; Chipot, C.; Roux, B. Efficient Determination of Relative Entropy Using Combined Temperature and Hamiltonian Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11* (5), 2234–2244. https://doi.org/10.1021/ct501034w.

(152) De Sancho, D.; Doshi, U.; Munoz, V. Protein Folding Rates and Stability: How Much Is There beyond Size? *J. Am. Chem. Soc.* **2009**, *131* (6), 2074–2075. https://doi.org/10.1021/ja808843h.

(153) Plessa, E.; Chu, L. P.; Chan, S. H. S.; Thomas, O. L.; Cassaignau, A. M. E.; Waudby, C. A.; Christodoulou, J.; Cabrita, L. D. Nascent Chains Can Form Co-Translational Folding Intermediates That Promote Post-Translational Folding Outcomes in a Disease-Causing Protein. *Nat. Commun.* **2021**, *12* (1), 6447. https://doi.org/10.1038/s41467-021-26531-1.