Critical Beginnings: Illuminating Structure, Dynamics and Interactions During the Early Stages of Protein Life

by Rachel B. Hutchinson

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the University of Wisconsin-Madison 2022

Date of final oral examination: 12/16/2022

This dissertation is approved by the following members of the Final Oral Committee:

Silvia Cavagnero, Professor, Chemistry

Randall Goldsmith, Professor/Chair, Chemistry

Judith Burstyn, Professor, Chemistry

Aaron Hoskins, Professor, Biochemistry

Acknowledgments

First, I would like to thank my doctoral advisor, Silvia Cavagnero, for her mentorship, guidance, and support throughout my time in graduate school. Her passion and enthusiasm motivated and inspired me. Her mentorship enabled me to grow as a researcher and communicator. I am thankful for all the time and energy she has put into mentoring and supporting me, so I could become the scientist I am today.

I would also like to thank the members of my thesis committee: Randall Goldsmith, Judith Burstyn, and Aaron Hoskins. Their input, questions, and guidance has helped me think critically about my work, make progress in my research, and grow as a scientist.

I am grateful to the previous and current Cavagnero Group members for all their input, encouragement, training, and contributions to this work. I have been fortunate to have such brilliant and kind people for co-workers and friends. Specifically, I would like to thank:

Angela Varela, for your encouragement and friendliness that helped me feel like a part of the group right from the start.

Rayna Addabbo, for helping me learn time-resolved fluorescence and working with me to better understand the role of chaperones in protein folding.

Matt Dalphin, for your enthusiasm, excellent teaching and training, and for making fun connections between chemistry, health, and food.

Miranda Mecha, for brightening my day with fun conversations about books, pets, and your delicious award-winning baked goods.

Hanming Yang, for helping me understand NMR theory and pulse sequences, so I was able to TA the physical chemistry lab effectively.

Valeria Guzman-Luna, for answering my countless questions, your kindness, and

encouragement.

Meranda Masse, for being an amazing friend through all the ups and downs of grad school. Thank you for always being willing to talk and listen and for all the fun adventures we've shared.

Justin Dang, for all of your thoughtful questions and invaluable recommendations and research feedback.

Siyu Li, for your thoughtful and creative problem-solving ideas.

Ummay Mahfuza Shapla, for your thoughtful questions and enthusiasm.

Heather Allaman, for your positive attitude, encouragement, enthusiasm, and thoughtful questions. It is a joy to work with you.

I especially want to thank the undergraduate students I worked with during the past years including:

Tess Carlson, for training me on the single-molecule fluorescence microscope and troubleshooting with me. You are an excellent teacher and fun to work with.

Miriam Engel, for your dedication and enthusiasm.

Ningkun Zhou, for your diligence, flexibility, and patient perseverance.

Chelsea Chen, you are diligent, a creative problem-solver, critical thinker, and excellent communicator. Your enthusiasm and energy and positive attitude are contagious, and it is a joy to work with you.

You are all excellent scientists, and I look forward to seeing what you accomplish in the future.

I also want to thank all the other friends I have made in Madison. Your friendship has made Madison a home for me. Specifically, I want to thank:

The Graduate Christian Fellowship community and Bible study group. I am grateful for all of

our conversations, meaningful discussions, and the fun times we have shared together making music, watching movies, and sharing meals. Your encouragement, support, prayers, and compassionate listening have helped me persevere through all the challenges I've experienced in grad school.

To Ojaswee Shrestha, thank you for being an amazing friend and a compassionate listener.

To Susan Hubbard, thank you for making me laugh, running with me, and including me in your fun and creative adventures.

To Josh Dietrich, thank you for being my best friend. Thank you for encouraging me daily, for giving me time and space to process my thoughts, for reminding me to take time to rest and for all the fun adventures we have shared. I am so grateful to have you in my life.

I want to thank my past mentors and teachers who inspired my love for chemistry and encouraged me to pursue graduate school especially Amy Kovach and Jason Taylor.

I also want to thank all of my family members whose love and support enabled me to make it to this point.

To my brother, Joshua Hutchinson, thank you for your encouragement and all our fun conversations and discussions.

To my sister, Sarah Johnson, thank you for being someone I can always confide in, for helping me see the bright side, and for making me laugh.

To my sister Abby Hutchinson, thank you for reminding me to take time to have fun and be creative.

To my parents, Bryan and Jennifer Hutchinson, thank you for encouraging my curiosity and love of learning. Thank you for always welcoming me home when I need time and space to rest. Thank you for supporting and encouraging me. Thank you for reminding me regularly of your

love for me.

Finally, I thank God for His love, grace, and wisdom. I thank Him for guarding my heart and mind with His peace. I am thankful He provided me with a supportive community of friends, family, co-workers, and mentors. I praise Him for His beautiful and complex creation that inspires awe and wonder. I am thankful for the opportunity to spend my time and energy learning about a small, but significant, part of His creation.

Table of Contents

Abstract ix

Chapter 1: Introduction 1

1.1 Overview 2

1.2 Protein folding and misfolding are intimately connected processes 2

1.3 Refolding of small purified proteins into buffer: experimental studies 4

1.4 Refolding of mid-size to large purified proteins into buffer: experimental studies 8

1.5 Folding of large purified proteins: experimental studies *in vitro* and in cell-like environments. 11

1.6 Kinetic trapping of the native state relative to aggregates: experimental studies in vitro and in cell-like environments 15

1.7 Major trends upon protein refolding from denaturants 19

1.8 Computational simulations of protein folding 19

1.9 Energy landscapes 21

1.10 Moving from simple model systems to more complex environments 31

1.11 Protein folding in the cell: The role of the ribosome 34

1.12 Protein folding in the cell: The role of molecular chaperones 36

1.13 Protein folding in the presence of the trigger factor chaperone 37

1.14 Trigger factor structure and function 41

1.16 Protein folding in the presence of the Hsp70 chaperone 44

1.17 Hsp70 structure and function 48

1.18 The Hsp70 chaperone cycle 49

1.19 Structure and dynamics of Hsp70-bound client proteins 51

1.20 Conclusions 57

1.21 This PhD thesis work 60

1.22 Acknowledgments 61

1.23 References 61

1.24 Appendix 100

Chapter 2: Fluorescence Anisotropy Decays and Microscale-Volume Viscometry Reveal the Compaction of Ribosome-Bound Nascent Proteins 114

2.1 Abstract 115

2.2 Introduction 116

2.3 Methods 119

2.4 Results and Discussion 124

2.5 Conclusions 156

2.6 Acknowledgments 157

2.7 References 157

2.8 Appendix 174

Chapter 3: Nascent Chain Interaction Networks and Their Effect on the Bacterial Ribosome 198

3.1 Abstract 199

3.2 Significance Statement 200

3.3 Introduction 201

3.4 Results and Discussion 203

3.5 Materials and Methods 227

3.6 Acknowledgments 228

3.7 References 228

3.8 Appendix 240

Chapter 4: Critical Beginnings: Selective Tuning of Solubility and Structural Accuracy of Newly-Synthesized Proteins by the Hsp70 Chaperone System 276

4.1 Abstract 277

4.2 Introduction 278

4.3 Results and Discussion 282

4.4 Methods 329

4.5 Conclusions 339

4.6 Acknowledgments 339

4.7 References 341

4.8 Appendix 354

Chapter 5: Distribution and Solvent Exposure of Hsp70 Chaperone Binding Sites Across the *E. coli* Proteome 391

5.1 Abstract 392

5.2 Introduction 393

5.3 Results and Discussion 400

5.4 Conclusions 416

5.5 Materials and Methods 422

5.6 Acknowledgments 425

5.7 References 425

5.8 Appendix 435

Abstract

Protein folding is essential for the survival of all organisms. Early folding stages, during translation and immediately after nascent-chain release from the ribosome, are critical because they determine a protein's ability to function for the remainder of its life. When these early folding steps go awry, proteins form harmful aggregates that cause deadly maladies and render protein production challenging and costly in the basic-research, biotechnology, and pharmaceutical settings. However, it is not well understood how proteins avoid aggregation and reach their native state in the cell.

In this thesis, I explore how single- and multi-domain proteins attain their functional state within the cellular environment. I used fluorescence lifetimes and fluorescence depolarization decays in the frequency domain to identify the structure and dynamics of ribosome-bound and released proteins. I focus on the roles of the ribosome and molecular chaperones in promoting correct folding and preventing aggregation. This thesis is divided into five chapters.

Chapter 1 summarizes what is known about protein folding and aggregation processes in buffered solutions and within the bacterial cell, focusing on the early stages of protein folding. This chapter highlights how nascent chains, the ribosome, and the molecular chaperones trigger factor and Hsp70 work together to enable proteins to overcome aggregation propensities and reach a long-lived bioactive state. The goal of this chapter is to describe the importance and complexity of folding and aggregation within the cell and describe the key open questions about the roles the ribosome, chaperones, and nascent chain in the folding processes. These unanswered questions motivate the work described in the following chapters.

Chapter 2 introduces a technology that combines fluorescence anisotropy decay and microscale-volume viscometry to explore the compaction and dynamics of ribosome-bound nascent proteins. The data in this chapter demonstrates that ribosome-bound apomyoglobin (apoMb) proteins form a compact region containing only 57-83 N-terminal amino acids (37-54% of the full-length protein). The size of the compact nascent chain demonstrates that the protein must undergo significant post-translational folding. Upon release from the ribosome, the nascent single-domain protein must incorporate more than half of its residues into the final structure. This required degree of additional folding is at risk of being accompanied by some degree of parallel aggregation processes.

Chapter 3 describes work that experimentally identifies, for the first time, the interactions between nascent proteins and the ribosome and(or) chaperones. Importantly, nascent chains bearing 140 residues interact with the ribosomal protein L23. As chain length increases, the nascent protein interaction network expands to include interactions with one additional ribosomal protein, and with the trigger factor and Hsp70 chaperones. These interactions may occur with residues within or outside the compact domain identified in Chapter 2. These interactions do not change the stability of the ribosome, peptidyl-transferase center, or ribosomal proteins to any detectable degree. Therefore, the ribosome is an unbiased nascent-protein interactor.

Chapter 4 describes work that explores how the Hsp70 molecular chaperone affects the solubility and structural accuracy of newly synthesized proteins. A significant portion of the chain must fold post-translationally (Chapter 2), so proteins are particularly aggregation-prone upon release from the ribosome. Therefore, the immediately post-translational folding period is a critical time that determines long-term protein health. We show that Hsp70 promotes protein

solubility against disruptive point mutations, upon nascent-chain release from the ribosome. Surprisingly, the resulting soluble fraction contains undesirable soluble aggregates, and higher Hsp70 chaperone concentrations are required to prevent their formation, relative to insoluble aggregates. Different protein sequences have distinct Hsp70 chaperone requirements for the prevention of soluble and insoluble aggregate formation. Yet, chaperones are not always able to prevent aggregation. Much higher than physiologically relevant concentrations of Hsp70 are sometimes required to prevent soluble-aggregate formation. In addition to highlighting the peculiar role of the Hsp70 chaperone, this chapter shows for the first time, that soluble aggregates can be generated shortly after (ca. 30-70 min) nascent protein release from the ribosome.

Chapter 5 investigates the role of Hsp70-protein interactions throughout the *E. coli* proteome. This work identifies the Hsp70 binding sites in 2,258 *E. coli* proteins and found that 99% of these proteins contained at least one Hsp70 binding site. The number of Hsp70 binding sites is directly proportional to client-protein size but only weakly correlated with protein solubility. Approximately 35% of the Hsp70 client proteins do not require chaperone assistance for solubility, but these proteins may require Hsp70 to reach the native state. Interestingly, Hsp70 binding sites are more solvent-exposed than other nonpolar sites, in protein native states, suggesting that some conformational sampling may take place within Hsp70-bound client-protein states.

The chapters in this thesis reveal the important players and processes that govern the early stages of protein folding. During translation, nascent proteins interact with the ribosome and undergo co-translational compaction that involves approximately half of the protein sequence. Upon release from the ribosome, proteins either fold into their native state or form soluble or

insoluble aggregates. Once formed, these states are kinetically trapped from each other, demonstrating the critical importance of the early folding steps for long-term protein health and function. The molecular chaperone Hsp70 acts during the immediately post-translational folding period to grant solubility to aggregation-prone protein variants and improve the structural accuracy of wild-type proteins. Overall, this thesis provides fundamental insights about cellular function and informs the future development of targeted strategies to overcome deleterious protein aggregation.

Chapter 1

Introduction

This chapter reproduces a published article (Miranda F. Mecha*, Rachel B. Hutchinson*, Jung Ho Lee, and Silvia Cavagnero (*co-first authors). *Biophysical Chemistry* **2022** 287, 106821).

1.1 Overview

For most proteins, a well-folded native three-dimensional protein structure is a prerequisite for biological activity. While intrinsically disordered proteins (IDPs) are an exception, protein folding remains a fundamental process for life on earth.[1] Yet, it is not well understood how unfolded proteins achieve their functional native state, despite the enormous number of conformations that they can potentially populate. The cell machinery ensures that predominantly native states are generated and that thermodynamically stable – and often undesirable – aggregated states are not. After being generated, many native and aggregated states of bacterial proteins remain kinetically trapped from each other, under physiologically relevant conditions. Given that the early stages of protein life are absolutely critical for the success of this process, this review will summarize current knowledge on protein folding in buffered solution and in the cell, including the early stages of protein's life. Successful folding in the complex cellular environment has clearly evolved as a team effort and is often achieved through the combined involvement of many molecular players, including the ribosome and a variety of chaperones. This review focuses on three of these major players in bacteria, namely the ribosome, and the molecular chaperones trigger factor and Hsp70.

1.2 Protein folding and misfolding are intimately connected processes

The practical consequences of aberrant protein folding are often severe and undesirable. For instance, protein overexpression in bacteria frequently leads to the formation of insoluble aggregates known as inclusion bodies. The latter species are difficult and expensive to disaggregate and to convert to the native state. This challenge often renders protein production in the basic-science, biotechnology, pharmaceutical and biomaterials settings extremely costly and

inefficient.[2, 3] In medicine, protein misfolding and aggregation in higher organisms is often associated with deadly maladies known as proteinopathies, including brain disorders like Parkinson's, Huntington's, and Alzheimer's disease.[4, 5] In summary, understanding how proteins fold is necessary to advance basic science, biotechnology and human health.

Protein folding research encompasses two major topics: the prediction of native structure from amino acid sequence and the mechanism by which proteins attain their native state. Significant advances were recently made in protein structure prediction. For instance, in the 2020 protein structure prediction challenge known as Critical Assessment of Structure Prediction (CASP), the software AlphaFold 2 (from the DeepMind artificial-intelligence company) predicted structures that matched the experimental structure of nearly two-thirds of the target proteins. This result is comparable, yet even better, to the predictions achieved with other protein-structure prediction programs (e.g., RoseTTaFold)[6] and to an earlier AlphaFold version.[7, 8] The AlphaFold family is based on a large structural-database and on sophisticated deep learning tools.[7, 9] Conveniently, AlphaFold 2 predictions have recently been integrated into online resources devoted to protein sequence and biochemical/physical properties, e.g., UniProt.[10, 11]

The second protein folding topic is the mechanism by which proteins attain their native state. In addition to enhancing basic knowledge, understanding protein folding mechanisms is a prerequisite for comprehending and controlling the relative flux through the parallel kinetic paths that lead to either folding or aggregation in Nature. Mechanistic insights into the overall folding/misfolding/aggregation process promise to yield invaluable insights to design and experimentally generate next-generation aggregation-free biomaterials, biosensors and drugs, as well as to devise better strategies to combat a variety of deadly proteinopathies.[12-14]

1.3 Refolding of small purified proteins into buffer: experimental studies

Soluble and correctly folded proteins typically bury a significant fraction (60%-80%) of their nonpolar residues inside the core to minimize exposure to the hydrophilic environment of typical intracellular media.[15-17] Thus, an essential function of protein folding is the intramolecular burial of most nonpolar amino-acid side chains, rendering them inaccessible to mostly nonpolar/nonpolar-type interactions with other proteins. The latter interactions would eventually lead to intermolecular aggregation via the hydrophobic effect.[18] Research over the past decades sought to explain how small proteins achieve their soluble functional native structure. More recently, investigations have also explored how native structure formation is coupled with the avoidance of the pervasive risk of protein aggregation.

In the 1960s, Christian Anfinsen showed that ribonuclease A and other proteins fold reversibly from a chemically denatured unfolded state.[19] These results led Anfinsen to propose the well-known "thermodynamic hypothesis," which states that the native state has the lowest free energy, out of all possible conformations. Anfinsen showed that reversible protein folding is fully determined by amino-acid sequence and environmental conditions, and that the folding process is under thermodynamic control.[20] In 1969, Cyrus Levinthal argued that because there are so many possible protein conformations, thermodynamic control is not sufficient for proteins to attain a folded state on biologically relevant timescales via a random conformational search.[21] On the other hand, it soon became clear that Levinthal's paradox[22, 23] could be resolved if proteins were to fold via specific single or multiple pathways, progressively narrowing the accessible conformational space.

More recent experimental work identified the folding pathways of a variety of single-domain globular monomeric proteins.[24-28] *In vitro* folding has typically been studied upon refolding

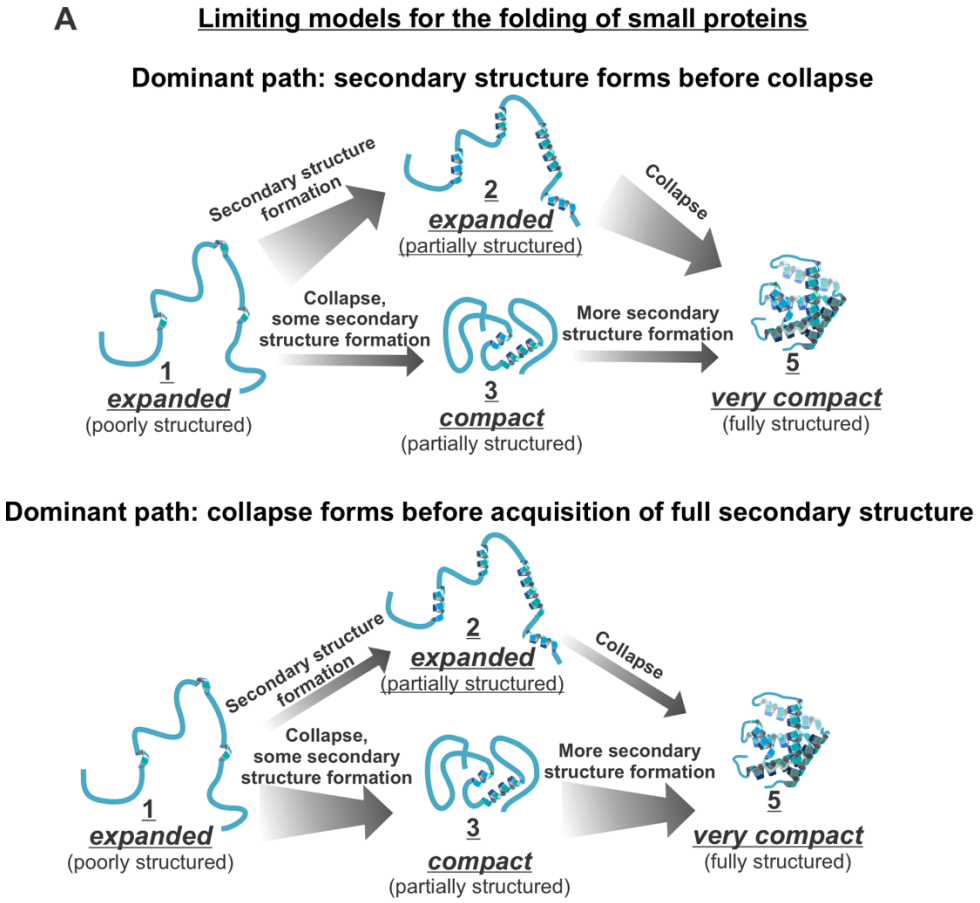
proteins from chemically or thermally denatured states. Upon refolding from denaturant for instance, proteins typically begin as a fairly expanded unfolded state bearing little or no secondary structure, and finally attain a compact, folded state bearing native secondary and tertiary structure.[29, 30] This final structure buries the nonpolar residues within the protein hydrophobic core, enabling the protein to be soluble within the hydrophilic environment of the cell.

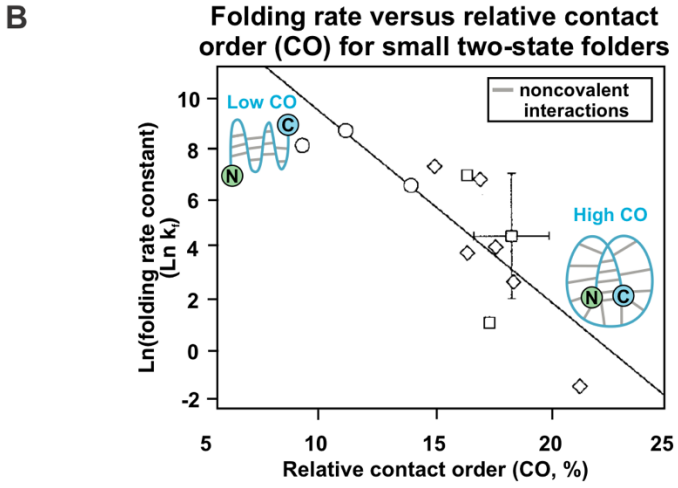
The simplest folding pathways follow two-state mechanisms and include only the unfolded and folded states and a single transition state. Small (50-60 residues) single-domain, α-helical proteins that experimentally show two-state folding can form secondary structure before chain collapse (framework mechanism) (Figure 1-1A). Alternatively, secondary structure formation and chain collapse may occur concurrently (nucleation-condensation mechanism) **(**Figure 1-1B**)**.[27, 31] For instance, engrailed homeodomain from *Drosophila melanogaster* (59 residues) folds via the framework mechanism, human TRF1 Myb domain (52 residues) folds via nucleation-condensation, and human c-Myb transforming protein (54 residues) folds with a mixed framework/nucleation-condensation mechanism[27, 32] In general, small proteins with higher local α-helical propensity are more likely to fold via the framework model (leading to considerable secondary structure formation preceding global chain collapse).[33, 34]

Remarkably, proteins were found to fold more slowly when their native state bears a greater number of long-range interactions. The latter are defined as noncovalent contacts between residues far away in sequence. This trend is described by the relative contact order (CO) parameter, which is defined as[35, 36]

$$CO = \frac{1}{L*N}\sum^{N} \Delta S_{i,j} \quad , \tag{1}$$

where L is the total number of residues, N is the total number of noncovalent contacts between


A
Limiting models for the folding of small proteins
Dominant path: secondary structure forms before collapse
Secondary structure formation
2
expanded
(partially structured)
Collapse
Collapse, some secondary structure formation
More secondary structure formation
1
expanded
(poorly structured)
3
compact
(partially structured)
5
very compact
(fully structured)
Dominant path: collapse forms before acquisition of full secondary structure
Secondary structure formation
2
expanded
(partially structured)
Collapse
Collapse, some secondary structure formation
More secondary structure formation
1
expanded
(poorly structured)
3
compact
(partially structured)
5
very compact
(fully structured)


B
Folding rate versus relative contact order (CO) for small two-state folders
noncovalent interactions
Low CO
High CO
N
C
Ln(folding rate constant) (Ln k_f)
10
8
6
4
2
0
-2
5
10
15
20
25
Relative contact order (CO, %)

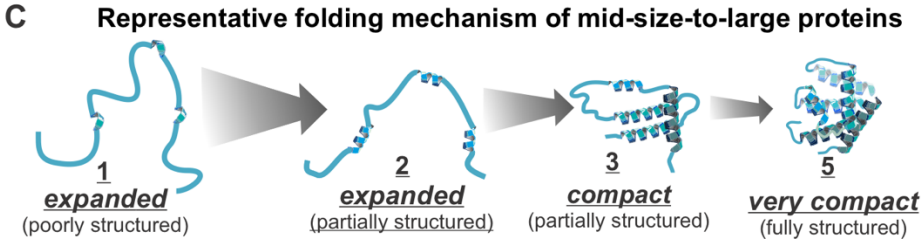

C
Representative folding mechanism of mid-size-to-large proteins
1
expanded
(poorly structured)
2
expanded
(partially structured)
3
compact
(partially structured)
5
very compact
(fully structured)

Figure 1-1. Overview of protein folding mechanisms upon dilution from denaturant or upon recovery from temperature jumps. Some small (50-60 residues) proteins fold via (A) a mechanism dominated by secondary structure formation before chain collapse, or via (B) chain collapse preceding the formation of most secondary structure. (C) Plot of folding rate as a function of relative contact order (CO) for small (60-110 residues) two-state folding proteins. The graph is reprinted with permission from Figure 1A of *J. Mol. Biol., 277*, Plaxco, K. W.; Simons, K. T.; Baker, D., 985-994, Copyright (1998).[35] (D) Larger proteins (> 60 residues) fold via more complex folding mechanisms, often including folding intermediates.

nonhydrogen atoms, and ΔSi,j is the sequence separation between interacting residues i and j.[36] As CO increases, the speed of protein folding decreases, for small two-state folding proteins (Figure 1-1C).[24, 35-37]

1.4 Refolding of mid-size to large purified proteins into buffer: experimental studies

Most proteins in the cell are larger than 100 residues. For instance, the average protein size is 360 and 530 residues in prokaryotes and eukaryotes, respectively.[38] In addition, a significant fraction of proteins have multiple domains. For instance, 40-65% of proteins in prokaryotes and 65-80% of proteins in eukaryotes have multiple domains.[39] Larger single- and multi-domain proteins have, by definition, a large number of degrees of freedom and may experience more complex folding paths.[40, 41]

Large proteins are more likely to have experimentally detectable folding intermediates.[24, 40, 42] While typical intermediates are on-path to the native state,[43-55] a few off-path intermediates have also been identified.[56] In general, unequivocally identifying folding intermediates can be challenging. For instance, it was reported that transient aggregates can sometimes be mistaken for folding intermediates.[57] Two well-studied mid-size proteins are sperm whale and horse apomyoglobin, each of which has 153 residues and a multistate folding mechanism with experimentally-detectable compact intermediates (Figure 1-1D).[58-60] Some of these intermediates are obligatory[61] and have a partially-folded structure bearing quasi-native features and lacking a few structural elements.[58, 59]The major apomyoglobin folding intermediate later resolves to a fully native conformation via slight conformational rearrangements within the early-folding A, B, G and H helices, which then enable final native-structure formation during the later stages of

folding.[62]

One model for the folding of medium-to-large proteins (100-370 residues) is the foldon model, according to which proteins fold via progressively populating small independent cooperative units known as foldons (Figure 1-2).[28, 63-66] Foldons are folding intermediates with some regions bearing native or quasi-native structure and other region unfolded or only partially folded. The presence of foldons naturally limits the dimensionality of the conformational search, providing a simple justification for how Nature avoids exhaustive sampling (Levinthal's paradox) during protein folding.[28, 67, 68] Examples of proteins that fold via foldons include: cytochrome c (104 residues),[69] RnaseH (155 residues),[70] apoflavodoxin (179 residues),[71] apomyoglobin (153 residues),[72] staphylococcal nuclease (149 residues),[73] and the following two-domain proteins: maltose binding protein (370 residues)[74] and DapA (292 residues).[75]

Similar to foldons, some multi-domain proteins fold successfully when each domain folds independently.[76] Some multi-domain proteins show independent domain folding, including titin,[77] fibronectin,[78] and the double B domain of protein A (BBdpA).[79, 80] Other proteins, including spectrin,[81, 82] phosphoglycerate kinase,[83] and the ankyrin repeat domains[80, 84] do not exhibit independently-folding domains, yet they are able to successfully refold from denaturant.[39]

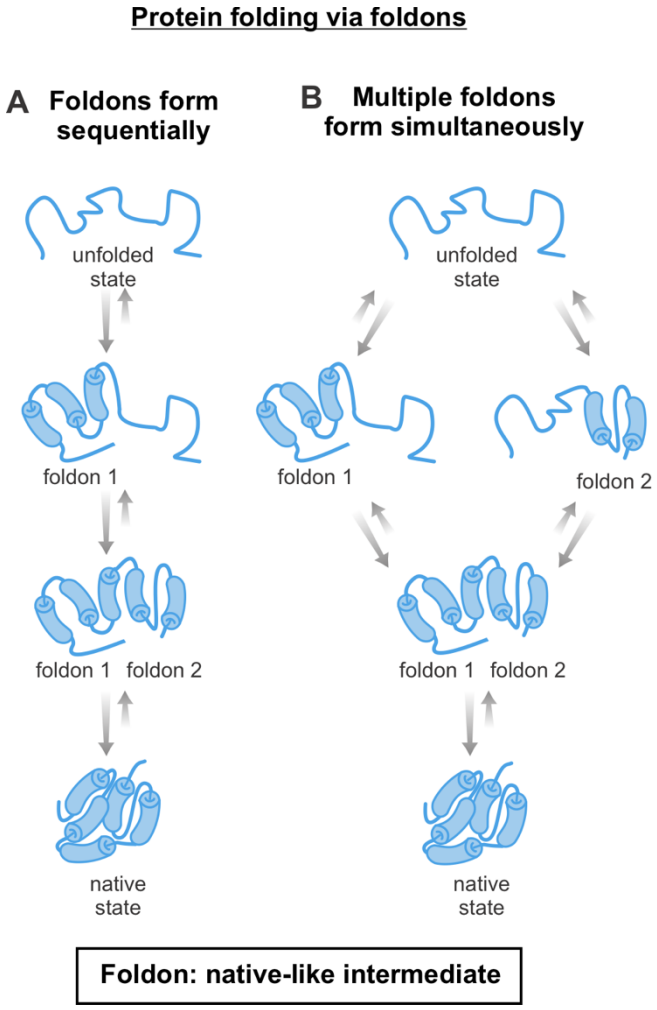


Figure 1-2. Foldons and protein folding mechanisms. Scheme illustrating how proteins may fold via native-like intermediates denoted as foldons, which are generated either (A) sequentially or (B) in parallel.

1.5 Folding of large purified proteins: experimental studies *in vitro* and in cell-like environments.

Many medium- to large-size proteins do not attain a 100% population of native state upon refolding from denaturant and give rise to some soluble or insoluble aggregates. Several examples of proteins that are known to form aggregates upon *in vitro* refolding from denaturant are shown in Table 1-1. This class of biomolecules includes proteins ranging from 153-550 residues, single and multi-domain proteins, monomeric proteins, and protein complexes. Without assistance from "folding helpers", these and likely many other proteins are unable to fully populate their native state in solution, upon refolding into buffer at physiologically relevant temperature and pH.

One key parameter that facilitates misfolding and aggregation over folding during refolding from denaturant and upon release from the ribosome in the cell is a slower folding rate than aggregation rate. Small (<50 residues) two-state folders tend to fold quickly, with folding rate constants (k_f) greater than 12,000 s^{-1} (Figure 1-3A, Supplementary Table 1-S1).[92] Yet, many larger two-state proteins fold more slowly, and some two-state folders have observed k_f values of less than 1 s^{-1}. Large proteins (>200 residues) with multi-state folding mechanisms have the slowest folding rate constants, with k_f values as low as 0.0004 s^{-1} (Figure 1-3B, Supplementary Table 1-S1).[92] Given that folding and aggregation pathways proceed in parallel, proteins that fold slowly are in general more likely to misfold and aggregate than proteins that fold rapidly.[93-95] While evolution has granted a few slow-folding proteins the ability to concurrently aggregate even more slowly (see well-behaved proteins in Figure 1-3) thereby defying aggregation, this is often not the case. For instance, many proteins undergo *insoluble* aggregate formation upon release from the ribosome in the absence of chaperones.[96] In addition, slow-folding globins

Table 1-1. List of proteins known to undergo aggregation upon *in vitro* refolding from denaturant.

Protein name	Number of residues)	Size of monomer (kDa)	Number of domains	Monomer or complex	Reference
Luciferase	550	60	2	monomer	[85]
Rhodanese	293	33	2	monomer	[86]
Rubisco	474 (large subunit), 122 (small subunit)	520 total, 50 for large subunit, 15 for small subunit	1	hexadecamer (8 small subunits, 8 large subunits)	[87]
Apomyoglobin	153	17	1	monomer	[88]
Galactitol-1-phosphate 5- dehydrogenase	346	37.4	2	tetramer	[89]
Glutamate decarboxylase alpha	466	52.7	3	hexamer	[89]
Threonyl-tRNA synthetase (ThrRS)	642	74	4	dimer	[89]
5,10-methylenetetrahydrofolate reductase	296	33.1	1	tetramer	[89]
S-adenosylmethionine synthetase	384	41.8	3	tetramer	[89]

Dihydrodipicolinate synthase (DHDPS)	292	31.3	2	tetramer	[89]
Tagatose-1,6-bisphosphate aldolase gatY (TBPA)	286	30.8	1	monomer	[89]
Tryptophanase	471	52.8	2	tetramer	[90]
α_1-antitrypsin	418	44.4	1	monomer	[91]

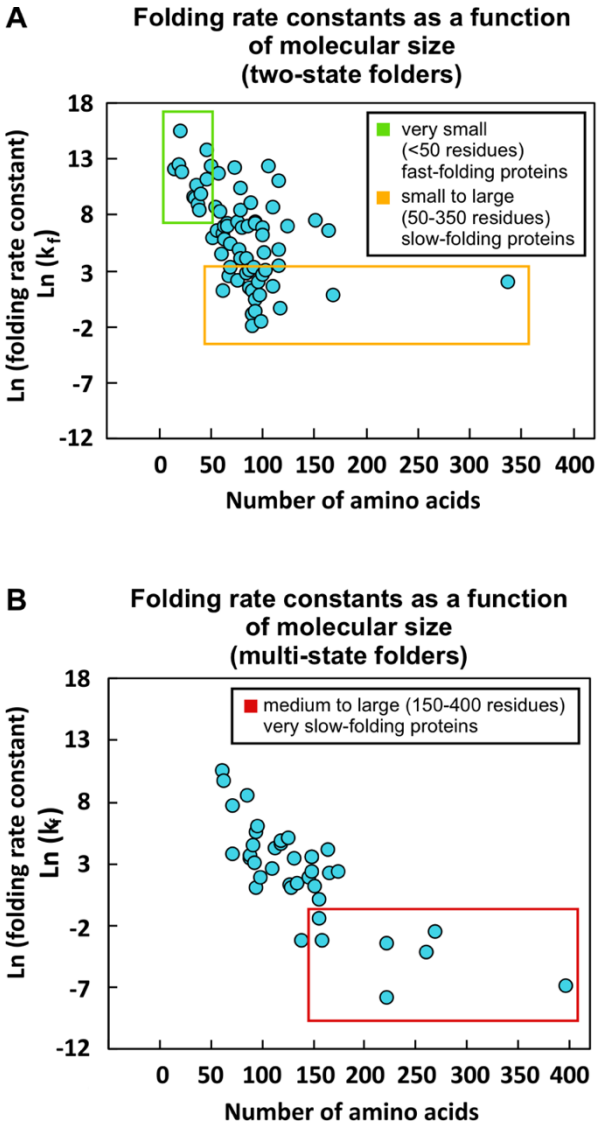


Figure 1-3. Effect of size and folding mechanism on protein folding rates. (A) Plot illustrating the dependence of protein folding rate constant (k_f) on the number of residues for two-state folding proteins. Small (<50 residues) two-state proteins fold quickly with $\ln(k_f) > 9.4$ (green box). Many larger two-state folders fold more slowly (orange box). (B) Dependence of protein folding rate constant (k_f) on the number of residues for multi-state folding proteins. Large (>200 residues) multi-state proteins have the slowest folding rates, with $\ln(k_f) < -2.5$ (red box). A list of the proteins and references for the data in this plot is available as Supplementary Information Table 1-S1.

undergo both folding and *soluble* aggregate formation upon release form the ribosome in the absence of the heme cofactor and chaperones.[93]

1.6 Kinetic trapping of the native state relative to aggregates: experimental studies in vitro and in cell-like environments

After proteins attain their native state for the first time, they continue sampling thermally accessible conformational states. Structural dynamics is often important for protein function. For instance, as pictorially described in the plots of Figure 1-4A, some non-aggregation-prone proteins routinely fold and unfold in the cell, displaying Anfinsen-like behavior.[19, 20] On the other hand, other proteins have more complex energy landscapes, which include aggregates (Figure 1-4B-C). These proteins typically experience some degree of kinetic trapping to avoid aggregation under physiologically relevant conditions. Native states can be kinetically trapped relative to aggregates or aggregation-prone intermediates, as shown in Figure 1-4B.[97-100] Alternatively, native states can also be kinetically trapped relative to their unfolded states, hence rarely unfold during the lifetime of their host organism (Figure 1-4C).[101, 102]

Kinetic trapping of the native state relative to aggregates (Figure 1-4B) has been detected under physiologically relevant conditions for a number of proteins. A few eukaryotic proteins (including bovine insulin, and human β2-microgloblin, lysozyme and αB-crystallin) are kinetically trapped and metastable relative to amyloid fibrils at pH 7.[97, 98, 103-105] In addition, Varela *et al.* showed that sperm-whale apomyoglobin and most soluble proteins of the *E. coli* bacterium (ca. 2,246 – 2,545 proteins) are kinetically trapped relative to aggregates that are not necessarily amyloid in nature, under physiologically relevant conditions.[99] This result was found to hold at concentrations much lower than physiologically relevant concentrations.[99] In this

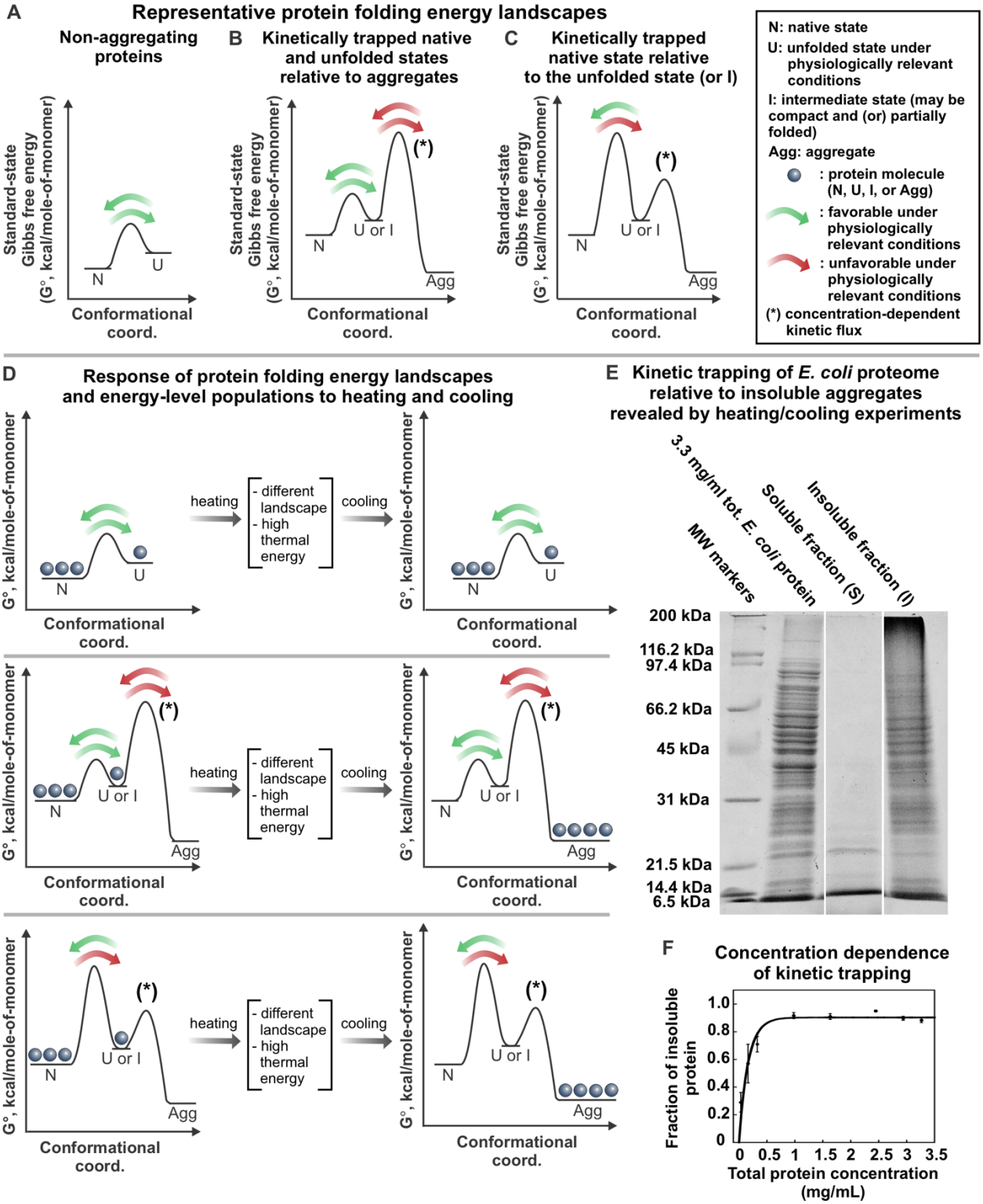


Figure 1-4. Kinetic trapping of *E. coli* proteome relative to insoluble aggregates. (A-C)

Representative standard-state Gibbs free energy landscapes for (A) non-aggregating proteins, (B) proteins that have kinetically-trapped native and unfolded states relative to aggregated states, and (C) proteins that have kinetically-trapped native states relative to unfolded states or folding intermediates. (D) Variations in the population of proteins described in panels A, B and C, respectively, after heating and cooling. (E) SDS-Page analysis of soluble *E. coli* proteome upon heating for 20 hrs at 70 °C followed by slow cooling to room temperature. Sample centrifugation generated a supernatant (S) and an insoluble pellet (I), shown separately in the gel [92]. (F) Fraction of insoluble, aggregated E. coli proteome generated by procedure described above as a function of total protein concentration. The solid line is meant to guide the eye. Error bars denote the standard error for three independent experiments [92]. Panels E and F are adapted with permission from Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S. *J. Phys. Chem. B* **2018,** *122*, 7682-7698. Copyright (2018) American Chemical Society.

study, kinetic trapping of bacterial proteins was demonstrated by heating the *E. coli* proteome, enabling it to transiently populate landscape regions inaccessible at physiological temperatures. The proteome was then slowly cooled, thus reverting back to physiologically relevant landscapes (Figure 1-4D). As shown in Figure 1-4D and 4E, after the heating-and-cooling process, most *E. coli* proteins form insoluble aggregates under both reducing and non-reducing conditions. Given the negligible extent of covalent protein modifications (assessed by mass spectrometry), this result demonstrates the presence of kinetic barriers that typically prevent conversion of several *E. coli* native proteins to the "aggregated region" of the landscape, under physiologically relevant conditions.[99] Kinetic trapping relative to insoluble aggregates occurs mostly for proteins larger than ca. 25 kDa at 0.5 – 3.5 mg/mL total protein concentration (Figure 1-4F).[99] The above findings are significant because they show that many bacterial proteins have an energy landscape that includes aggregates. Native states are kinetically trapped relative to these aggregates under physiologically relevant conditions, hence they do not convert. It is common knowledge that thermally denatured proteins are particularly aggregation-prone. Indeed, some of this aggregation is known to be due to covalent protein modifications (e.g. new disulfide bridges upon boiling an egg). However, the study by Varela *et al.* (under reducing conditions) ruled out that the observed aggregation is a consequence of covalent modifications. The work was based on the analysis of an *E. coli* S100 protein mixture. Hence, its conclusions apply only within that mixture. While pure proteins like sperm whale apomyoglobin were also found to exhibit analogous behavior, future investigations on additional isolated proteins will contribute to establish the generality of the present findings.

Importantly, aggregation rates can be modulated by protein concentration and by environmental conditions that modify protein energy landscapes. Some changes in

environmental conditions may even lead to protein covalent damage. For instance, an increase in protein concentration and kinetic-barrier curvature (for the aggregation rate-determining steps), as well as a decrease in barrier height, are sufficient to trigger pervasive aggregation. The latter phenomena are known to play a role in the case of deadly proteinopathies.

1.7 Major trends upon protein refolding from denaturants

In summary, experimental studies to date show that there are some general trends in protein folding. These include: (a) a higher flux of folding via the framework model (leading to considerable secondary structure formation preceding global chain collapse) for small proteins with high local α-helical propensity, (b) highly populated folding intermediates, some of which are native-like (i.e., foldons), in proteins larger than ca. 100 residues, (c) an inverse correlation between relative contact order CO and folding rate constants for apparent 2-state folders. Finally, small (<50 residues) two-state folders fold quickly (k_f > 12,000 s^{-1}), many larger two-state folders fold more slowly (k_f = 0.13 s^{-1} -12,000 s^{-1}), and large (>200 residues) multi-state folders fold the slowest (k_f = 0.0004-0.08 s^{-1}).[92] Yet, despite the above general trends, folding pathways upon refolding from denaturant and in the cell are overall quite diverse, for different protein folds.[25, 27, 65]

1.8 Computational simulations of protein folding

Computer simulations were extensively employed to characterize folding pathways and to define leading features of conformational energy landscapes. Computational approaches to simulate protein folding events, including molecular dynamics (MD) and Monte Carlo simulations and genetic algorithms, have been reviewed by Li *et al.*[106] The main challenge with simulating

protein folding paths is that the typical 100 μs – ms timescale for this process is significantly longer than the capabilities of the traditional MD method. The first MD protein-folding simulation, carried out in 1998, required two months. This effort focused on the folding of villin, a 36-residue protein that folds on the microsecond timescale.[107] Advances in simulation algorithms and supercomputer technologies made it possible for MD methods based on unbiased empirical force fields to simulate the folding of ~100-residue proteins on the ms-timescale.[106, 108] Yet, most proteins from eukaryotes and prokaryotes are 300 amino acids or longer,[38] so their folding cannot yet be simulated by unbiased MD techniques. One option is to simulate the folding of larger proteins with biased force fields that favor native contacts. This strategy is embraced by Gō models and structure-based models to simulate the folding of multi-domain proteins ranging from 150-400 residues [[76, 108, 109] Gō models and other native-structure-based models greatly simplify folding landscapes by allowing only native interactions.[108] While the latter methods are successful at predicting the experimental behavior of small and mid-size well-behaved proteins,[110, 111] they are likely unreliable in the case of less-well-behaved proteins. In addition, Gō models are typically inadequate to describe protein misfolding in the presence of concurrent aggregation leading to non-native-like self-associated states. Modified Gō models that incorporate misfolding have been developed.[112] These models require knowledge of high-resolution structures of both native and misfolded states. These structures are unfortunately not available in the case of most soluble misfolded states.[112] Another strategy to model the folding of large proteins is to use Markov-state models (MSM), which can model long timescale dynamics.[113, 114] MSMs partition the system into multiple states, and assume that transitions between states are memoryless. In other words, the probability of going from state x to state y only depends on states x and y, not any previously occupied states. Short MD simulations can be used to model small conformational changes within

each state and can then be combined to predict long-timescale dynamics.[113, 114] Current challenges with using MSMs to model protein folding include correctly identifying the MSM states and interpreting folding mechanisms from MSMs. Recent machine learning advances can be employed to address the latter challenges.[114] In summary, while more work is necessary to develop simulations that can successfully model the experimentally observed folding and misfolding/aggregation of midsize to large proteins, the future of this area of research holds promise.[89, 96, 115]

1.9 Energy landscapes

Experimental data are often consistent with small single-domain proteins folding through a single pathway with either no or few intermediates.[25, 116] Yet, theoretical models suggest that the unfolded states reach the native state via multiple parallel pathways.[42, 111, 117-120] It is worth noting that experimentally observed 2- or 3-state refolding kinetics is not incompatible with multiple parallel folding paths.[120] Further, single-molecule experimental studies were able to unequivocally identify the existence of multiple parallel folding pathways for a single-domain protein that shows fast, two-state folding kinetics in bulk measurements.[121] Therefore, it is likely that many proteins fold via multiple parallel pathways, even when these paths cannot be explicitly resolved in bulk refolding experiments.

One theoretical model that promotes parallel folding pathways is based on the concept of folding funnel. The existence of folding funnels was initially suggested by Dill in 1987,[117] and then thoroughly detailed by Wolynes,[110, 111, 118] Dill,[41, 119] and coworkers. Folding funnels can be portrayed as two-dimensional diagrams, as shown in Figure 1-5A.[110, 111, 118, 122, 123] The y-axis

Representative protein folding landscapes

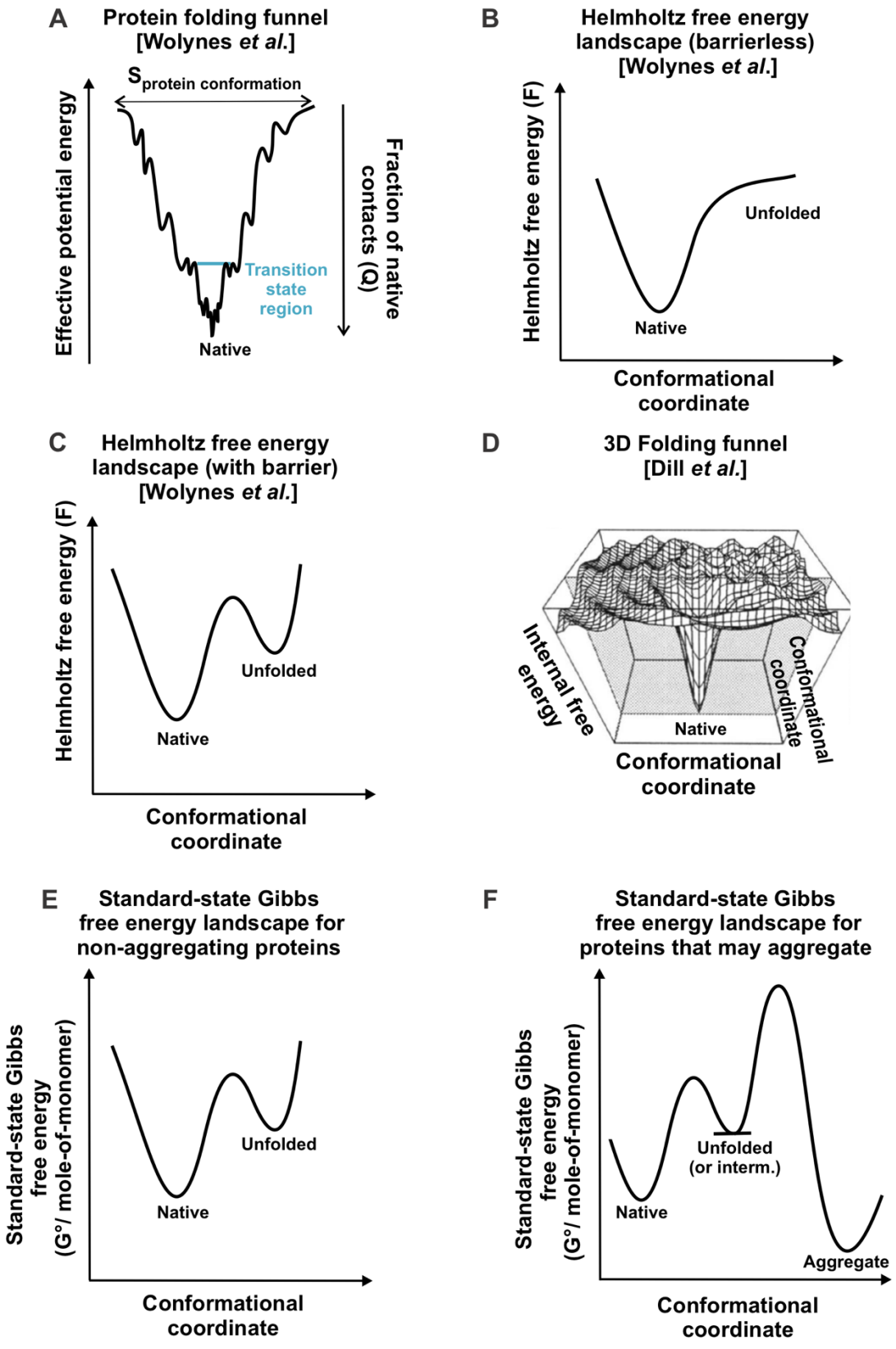


Figure 1-5. Representative protein folding energy landscapes. (A) Folding funnel proposed by Wolynes and coworkers, showing how protein conformational entropy decreases in concert with effective potential energy, as a protein folds to its native state.[111, 122] (B) Helmholtz free energy landscape for proteins that do not have a free-energy transition state for folding. (C) Helmholtz free energy landscape for proteins that have a free-energy transition state for folding. In panels A-C, the native state has 100% native contacts (Q = 1), and the unfolded state has Q = 0. (D) Multidimensional energy landscape. The vertical axis represents the potential energy of any given protein conformation plus the free energy of solvation.[119] Figure 1-1D is reprinted with permission from John Wiley and Sons[41] from Figure 1-37C in Protein Science 4, Dill, K. A.; Bromberg, S.; Yue, K.; Chan, H. S.; Ftebig, K. M.; Yee, D. P.; Thomas, P. D. Principles of Protein Folding — a Perspective from Simple Exact Models. *4*, 561-602. Copyright (1995). (E) Standard-state Gibbs free energy landscape of a protein that cannot form aggregates at a given temperature, pressure and solution conditions. (F) Gibbs free energy landscape for a protein that can form aggregates at a given temperature, pressure and solution conditions.

describes the protein's effective potential energy, which includes the potential energy of the protein chain and free-energy contributions arising from interactions with the solvent. The change in effective potential energy is proportional to the fraction of native contacts (Q). The horizontal axis represents the conformational entropy of the protein ($S_{prot,\ conf}$), so that the funnel width coincides with $S_{prot,\ conf}$. Note that the conformational-entropy term is distinct from the total entropy, and it does not include entropy contributions involving the solvent. The width of the funnel narrows as the effective potential energy decreases, showing that conformational entropy gets smaller (hence disfavoring folding), as the protein approaches the native state. In summary, the model postulates that, as each protein folds, both effective potential energy and conformational entropy decrease in concert, leading to an overall decrease in Helmholtz free energy. This process renders the overall landscape funnel-shaped.[111, 122] The diagram in Figure 1-5A qualitatively shows that there are typically very few conformations sharing the same effective potential energy and separated by significant local barriers. Hence the landscape is only very weakly "frustrated". Multiple conformations that share the same energy contribute to increasing the density of states. Thus, entropy contributions due to both density of states and to solvent-related configurations need to be considered, in addition to the conformational entropy illustrated in the diagram of Figure 1-5A, to compute the total entropy.

In the canonical ensemble, the change in total free energy of folding (ΔF_{tot}) is defined as the difference between the changes in total internal energy (ΔU_{tot}) and total entropy (ΔS_{tot}) terms according to

$$\Delta F_{tot} = \Delta U_{tot} - T\Delta S_{tot}, \qquad (2)$$

where T is temperature and ΔS_{tot} includes changes in both protein conformational entropy and entropy related to solvent molecules. The total internal energy is equal to the effective potential

energy averaged over all microstates (i.e., protein conformations).

The negative change in conformational entropy as the protein folds is energetically unfavorable and must be compensated by favorable internal energy changes and/or favorable solvent entropy changes, so that the actual folding process ends up being thermodynamically favorable ($\Delta F < 0$).[122] If the energetically favorable internal-energy and solvent-entropy changes fully compensate or override the unfavorable conformational-entropy changes, the free energy landscape is barrierless, as shown in Figure 1-5B.[118] When the energetically unfavorable changes exceed the favorable contributions, the free energy landscape bears a barrier, as shown in Figure 1-5C.[123, 124] Bryngelson *et al.* denote barrierless and barrier-containing free-energy folding scenarios as type 0 and type 1, respectively.[118]

In the case of proteins whose folding free-energy landscapes bear a thermodynamic barrier ($\Delta G^{\ddagger}_{tot} > 0$), experimental studies showed that the thermodynamic activation parameters for folding, including activation enthalpy ($\Delta H_f^{\ddagger}$) and entropy changes ($\Delta S_f^{\ddagger}$), can be energetically favorable or unfavorable, as shown in Table 1-2. The Gibbs activation free energy for folding may be entropy or enthalpy driven, depending on whether $T\Delta S_f^{\ddagger}$ or $\Delta H_f^{\ddagger}$ has a larger magnitude (Table 1-2).[125-127]

Protein-folding landscapes can also be visualized according to Dill *et al.*[41, 119] as three-dimensional curves, as shown in Figure 1-5D. In this case, the y axis is denoted as internal free energy, and is essentially equivalent to the effective potential energy of the landscapes by Wolynes *et al*. These landscapes are generated in the isothermal-isobaric ensemble, which is particularly relevant to biological systems, and assume constant temperature and pressure. It is worth noting that protein free energy landscapes are highly temperature,[124] as well as pressure-dependent.[128, 129]

Table 1-2. List of experimentally determined thermodynamic activation parameters for the folding of several proteins (unfolded, U → transition state, TS).

Protein name	$T\Delta S_f^{\ddagger}$ (U → TS) (kJ/mol)	$\Delta H_f^{\ddagger}$ (U → TS) (kJ/mol)	$\Delta G_f^{\ddagger}$ (U → TS) (kJ/mol)	Reaction is driven by:	Reference
Chymotrypsinogen	-155	-66.7	88	Entropy	[127]
Soybean trypsin inhibitor	-111	-8.0	103	Entropy	[127]
Chymotrypsin Inhibitor 2 (CI2)	27.4	59.0	31.6	Enthalpy	[125]
N-terminal domain of Ribosomal Protein L9	29.5	54.0	24.5	Enthalpy	[125]
Ig binding domain of Y43W point mutation of protein L	7.3	40.2	32.9	Enthalpy	[125]
Immunophilin protein FKBP12	15.1	53.1	38.0	Enthalpy	[125]
Transcriptional activator protein M2V GCN4-pl	4.0	23.4	19.4	Enthalpy	[125]
Cold shock protein B	19.9	44.8	24.9	Enthalpy	[125]

Cold shock protein B	-24 ± 2	31.6 ± 2.2	55.7 ± 1.0	Enthalpy	[131]
Common-type acylphosphatase	-49.3 ± -4.9	23.6 ± 2.4	72.9 ± 0.4	Entropy	[132]
Muscle acylphosphatase	-41.4 ± 4.1	40.7 ± 4.1	82.1 ± 0.4	Neither	[132]
Apocytochrome b5	-20.6 ± 5.5	42.4 ± 5.5	63 ± 8	Enthalpy	[133]
Heart cytochrome c	22 ± 9	59 ± 9	37 ± 13	Enthalpy	[134]

In the isothermal-isobaric ensemble, free-energy landscapes describe changes in the Gibbs free energy of the system (G) as a function of protein conformational coordinates. However, G varies before a chemical process reaches equilibrium (or before an irreversible reaction is complete), while standard-state free energy per mole (G°) does not.[130] Therefore, Cavagnero *et al*. proposed to plot protein folding free-energy landscapes as G° instead of G, as shown in Figure 1-5E.[99, 100] The standard-state free energy of the system G° is expressed on a per-mole-of-monomer basis, so that both monomeric and aggregated protein states can be reliably plotted within the same landscape **(**Figure 1-5, panels E and F**)**.[99, 100] As discussed in a previous section, standard-state chemical-potential landscapes were recently employed to show that most bacterial proteins are kinetically trapped relative to a variety of aggregates.[99, 100] We have also adopted this type of representation in Figure 1-4 of the present review.

Getting back to effective-potential energy landscapes, the folding-funnel concept has also been employed to explain the folding of multi-domain proteins.[76, 135] Experimental studies show that some large multi-domain proteins are able to successfully and independently refold from denaturant. These proteins include titin,[77] fibronectin,[78] and the double B domain of protein A (BBdpA).[79, 80] Computational studies proposed that these types of multi-domain proteins fold successfully via a "divide and conquer strategy," according to which each domain folds independently.[76, 135] If each domain is able to fold independently, then several smaller folding funnels can be combined into a single large funnel. Therefore, the folding process is characterized by significantly fewer degrees of freedom than if interactions between domains were to play a role during the folding process (Figure 1-6**)**.[76]

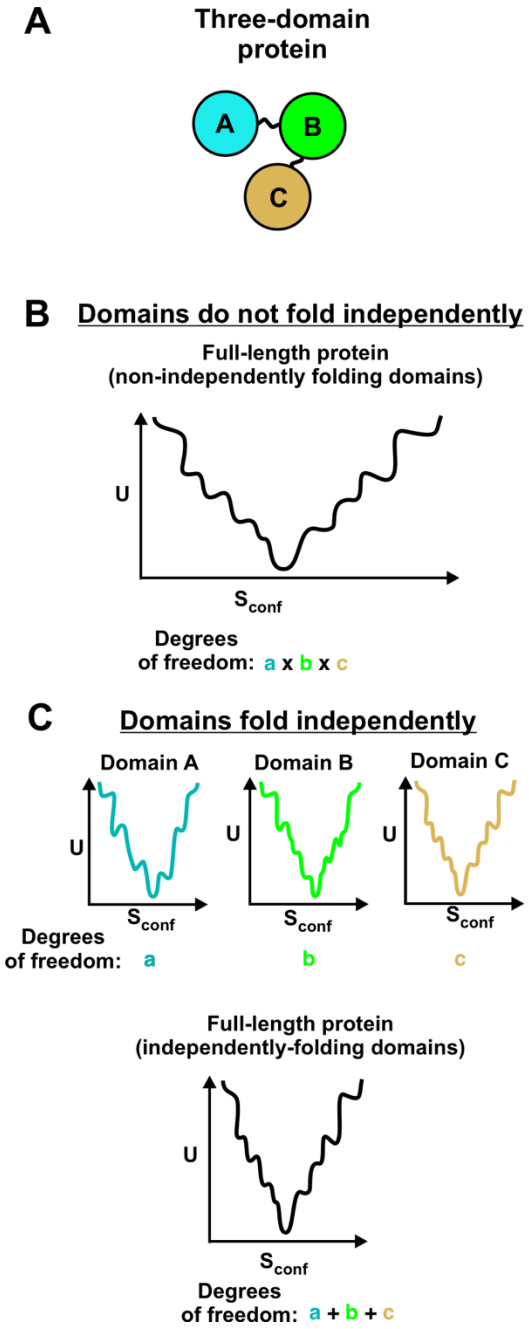


Figure 1-6. Funnel landscapes of multi-domain proteins. Some multi-domain proteins show independent folding of domains, and their folding can be described by combining the independent folding funnels for each domain.[76] (A) Cartoon of three-domain protein. (B) Folding funnel for multidomain protein with non-independently folding domains. The total

number of degrees of freedom of the full-length protein is equal to the product of the degrees of freedom of each individual domain.[76] (C) Folding funnels for individual domains and combined folding funnel for multi-domain protein whose domains fold independently. In this case, the total number of degrees of freedom for the full-length protein is equal to the sum of the degrees of freedom for each individual domain.[76]

1.10 Moving from simple model systems to more complex environments

The studies described above focused on proteins refolding from denaturant. Physiologically relevant systems, however, not only include larger proteins but also involve more complex and crowded solution environments compared to buffered solutions. These environments can affect protein aggregation propensity.

In principle, crowded environments could decrease protein aggregation propensity. Molecular crowding from large inert cosolutes tends to stabilize the native state.[123] The crowding molecules decrease the volume available to the protein, pushing the protein toward a compact state, and thus reducing the total entropy of the system.[1, 123] Yet, at high protein concentrations (> 100 g/L), interactions between the crowding molecules and proteins can decrease protein stability.[1] In some cases, crowding increases aggregation rate. For instance, the aggregation rate of α-synuclein is 10-fold greater in the presence of crowding agents than in plain buffer.[136]

In general, crowding alone cannot fully explain the effect of many types of cosolutes on protein folding because it does not account for electrostatic and hydrophobic interactions. Some cosolutes, including osmolytes, stabilize the native state via repulsive interactions with the protein.[123] Site-specific ligand binding can also increase the stability of the native state. Other cosolutes, including denaturants, destabilize the native state via non-specific binding to the protein surface. Salt ions also affect protein stability, and their effect differs depending on the ion concentration as well as the location of charge in the folded and unfolded protein.[123, 137-139] Interestingly, the presence of other proteins with different sequences does not significantly change protein aggregation propensity, within complex mixtures. For instance, when bovine serum albumin and consensus tetratricopeptide repeat are mixed in solution, the solubility of each protein depends on its individual concentration and is minimally affected by the concentration of the other protein.[140]

The refoldability of proteins after chemical denaturation within a complex protein collection from *E. coli* lysate was recently analyzed.[115] This study employed limited proteolysis via proteinase K to determine whether proteins refold to the native state. The results showed that 33% of *E coli* proteins do not refold to the native state after denaturation. Even more proteins may exhibit this characteristic behavior, upon taking soluble aggregates into account. Proteins with many domains are more likely to misfold than proteins with a single domain.[115]

In a different study, Niwa *et al*. measured the solubility of all *E. coli* proteins within an *E. coli* cell-free system upon release from the ribosome in the absence of molecular chaperones.[96] Their results showed that only 28% of non-membrane proteins are ≥80% soluble (see eSol database: http://www.tanpaku.org/tp-esol/index.php?lang=en).[96, 141] Larger proteins are more likely to form insoluble aggregates.[96] Namely, while 42% of small proteins (<30 kDa) are soluble, only 14% of large proteins are soluble.[96] Although the above experiments were carried out in the presence of the strong T7 promoter, the expressed-protein concentration range (2-100 μg/mL, average = 33 μg/mL) is comparable to the endogenous concentration range of most proteins in *E. coli* (4.7 - 153 μg/mL [0.11-4.30 μM], excluding outliers, median = 29.6 μg/mL [0.87 μM]).[142] *E. coli* cellular concentrations were estimated from experimental copy numbers, assuming an *E. coli* cell volume of 10^{-12} mL. Proteins with concentrations larger than the third quartile value plus the quartile range (> 153 μg/mL, [>4.30 μM]) were considered outliers. This category included 17% of the 1,103 proteins quantified in this study.[142] In principle, higher concentration could decrease solubility relative to the results by Niwa *et al*. Apart from these selected high-abundance proteins, we expect the concentration of individual proteins generated via the above two methods to be similar, given the similar concentration ranges. Codon usage is not an issue, and any differences in translation rates in cell-free systems *vs in vivo* expression is

would likely affect all proteins to a similar degree. The cell-free experiments performed by Niwa employ the strong T7 promoter, which leads to overexpression of bacterial proteins relative to conventional cellular production levels. However, cell-free systems have a lower percent of active ribosomes compared to live cells. Specifically, in the PURE system employed by Niwa *et al.,* approximately 40% of ribosomes are active in protein synthesis at any given time, compared to 80% in *E. coli* cells.[143, 144] The lower ribosome activity of cell-free systems explains why the protein concentration in these experiments is comparable to cellular protein concentrations, despite the use of the T7 promoter. Note that the cell-free system employed in the study by Niwa *et al.* does not perfectly represent the cellular environment. This system lacks heme and other cofactors that may be required for correct folding of some proteins.[145] In addition, each protein was expressed individually. Therefore, proteins that give rise to hetero-complexes in live cells might show higher than regular aggregation levels in the work by Niwa *et al.* This outcome is likely in cases when the concentration of these proteins is higher than that of their complexation counterparts. Importantly, while the all aggregates detected by Niwa *et al*. were insoluble, it is known that some proteins can also form soluble aggregates upon release form the ribosome, if chaperones are not present.[93] Hence, the results by Niwa *et al.* may underestimate the actual extent of protein aggregation upon release from the ribosome, in the absence of molecular chaperones.

Once formed, soluble and insoluble aggregates are often kinetically trapped in *E. coli*, relative to the native state.[99] This phenomenon is responsible for the persistence of long-lived aggregation-free bioactive conformations. Further, amyloid aggregates are typically highly thermodynamically stable.[42, 95] While cellular quality-control systems can disaggregate and degrade misfolded proteins later in life, these processes are energetically costly.[146-149] Thus,

correct folding in the early stages of protein life, including cotranslational and immediately post-translational folding, is critical for long-term protein solubility and function. The above experimental studies, performed in the absence of molecular chaperones, show significant levels of aggregation. Yet, in living cells molecular chaperones are present and enable the correct refolding of the numerous proteins that would otherwise aggregate, upon release from the ribosome.

1.11 Protein folding in the cell: The role of the ribosome

The ribosome alters the folding energy landscape because many proteins begin folding cotranslationally, before they are fully synthesized and while they are still bound to the ribosome.[24, 150-158] Translation is vectorial and enables some proteins to fold sequentially, with N-terminal regions folding before C-terminal regions,[155, 159-162] and sometimes enabling separate domains to fold independently. Independent folding of domains decreases the protein's number of degrees of freedom (Figure 1-6)[76] and could prevent inter-domain misfolding interactions.[163] Rare codons clusters that slow down translation may provide more time for cotranslational folding.[164, 165] Rare codons often appear within protein domains and separate small structural units.[164] Synonymous codon substitution that alters translation rate can cause proteins to misfold, suggesting that Nature has optimized codon usage for correct folding.[166]

The ribosome also reduces the number of accessible conformations by interacting with the nascent chain and spatially confining nascent chains motions.[151, 158, 167, 168] Nascent chains can interact with the ribosomal tunnel[169, 170] or surface.[171-175] The ribosomal tunnel is approximately 100 Å long and 10-20 Å wide.[176, 177] The tunnel can hold approximately 30-40 amino acids, depending on the nascent protein structure,[168, 178-180] and it can fit more residues if the protein

forms tertiary structure within the tunnel.[181, 182] Nascent proteins can form alpha-helical secondary structure,[151, 168, 183, 184] tertiary interactions,[152, 185] and even fully folded structures[181, 182, 186] within the tunnel. Nascent-chain compaction is a prerequisite for the folding of globular proteins and typically occurs after 54-59 nascent-chain residues have been synthesized.[155, 181] Larger tertiary structures can form within the vestibule[154, 187, 188] and outside the ribosomal tunnel.[155, 189] Ribosome-bound conformations may be dynamic and flexible[155, 185] The ribosome can destabilize full-length ribosome-bound protein structures outside the ribosome tunnel compared to released folded proteins.[190]

Most single-domain proteins cannot fully fold into the native state until they are released from the ribosome and their C-terminal residues are available for folding. The C-terminal residues are usually important for folding because they bear key interaction counterparts, including sometimes residues expected to establish contacts with N-terminal regions of the protein chain.[37, 191, 192] Indeed, protein fragments lacking C-terminal residues are often insoluble.[192] Fortunately, the ribosome grants solubility to partially synthesized nascent chains.[93] Immediately post-translational folding sometimes involves structure formation by significant portions of the protein. For instance, apomyoglobin must incorporate at least 60 residues (40% of the total number of residues) into the native structure post-translationally.[156] Immediately after release from the ribosome, the nascent-protein region that becomes solvent-exposed may include a significant fraction of nonpolar residues. These nonpolar residues can either be buried intra- or inter-molecularly, giving rise to folding or aggregation, respectively. Therefore, the immediately post-translational steps are critical for the kinetic channeling of the nascent chain towards intramolecular folding, as opposed to intermolecular aggregation.[93, 155, 156] Once formed, most native and aggregated states in bacteria are kinetically trapped from each other, rendering later

interconversion between these states highly unlikely.[99]

Translation through the ribosome is sometimes sufficient to grant solubility to released proteins.[93, 96] Many proteins, however, require additional assistance from molecular chaperones to reach their soluble native structure.[93, 96, 141]

1.12 Protein folding in the cell: The role of molecular chaperones

Molecular chaperones act both co- and post-translationally and are able to prevent, and in some cases reverse, protein aggregation. Importantly, only 28% of the proteins synthesized with an *E. coli* cell-free system lacking molecular chaperones is soluble (excluding membrane proteins).[96] Remarkably, molecular chaperones increase the solubility of 97% of these aggregation-prone proteins.[141] Correct folding and solubility are promoted by chaperones via a variety of mechanisms. Chaperones can catalyze conformational changes, including folding of unfolded states and unfolding of misfolded states, utilizing energy from ATP hydrolysis.[193, 194] Chaperones can also simply bind proteins and, in so doing, bury solvent exposed nonpolar regions and transiently decreases free-protein concentration. This chaperone action does not typically require ATP hydrolysis.[193, 194] Chaperones may be especially important to promote the folding of large and multi-domain proteins which tend to fold more slowly,[36, 195-198] and are more likely to aggregate than two-state folding proteins.[96, 115]

Bacterial cells contain a wide variety of chaperones that effectively mitigate the detrimental effect of misfolding and aggregation, including trigger factor (TF), the Hsp70 system, and GroEL/GroES.[148, 199-201] TF associates with nascent chains as they emerge from ribosomes, thus contributing to the prevention of aggregation and to the protein's folding efficiency.[202-206] The affinity of TF for unfolded proteins is lower than the affinity of other chaperones. This

thermodynamic property is accompanied by the rapid binding and release of client proteins,[207] compatible with efficient translation.[208] DnaK, which is a prominent Hsp70 protein in prokaryotes, interacts with its substrates co- and post-translationally.[209] GroEL/GroES, a prokaryotic Hsp60 chaperone, acts downstream of DnaK upon *de novo* protein folding and facilitates correct folding through several functions including isolating proteins within its chamber (a.k.a. Anfinsen cage), catalyzing protein folding, and unfolding misfolded states.[210-212] ClpB, a prokaryotic heat-shock protein belonging to the Hsp100 class, solubilizes protein aggregates by threading protein chains through its central hexameric channel, thus facilitating disaggregation either alone[213, 214] or in combination with the Hsp70 chaperone network.[215] Figure 1-7 shows a graphical representation of the molecular chaperones that have currently been identified in *E. coli*. Given that the early stages of protein life are vital for long-term solubility and function, we focus on chaperones that act cotranslationally and immediately post-translationally. These chaperones include trigger factor and the Hsp70 system.

1.13 Protein folding in the presence of the trigger factor chaperone

Trigger factor (TF) is the only known ribosome-associated chaperone in bacteria (Figure 1-8). It was discovered by Crooke & Wickner who demonstrated that TF promotes the folding of the pro-OmpA protein to its membrane-assembly-competent form.[216] TF is both a chaperone and a *cis/trans* prolyl- isomerase,[217] and it binds the ribosome with a 1:1 stoichiometry (Figure 1-8A). The cellular TF concentration is ~50 μM.[218, 219] This value is comparable to the ribosome concentration, though the latter varies as a function of cell growth rate. Ribosome-unbound TF undergoes a monomer-dimer equilibrium.[219] TF does not bind ATP and interacts with nascent

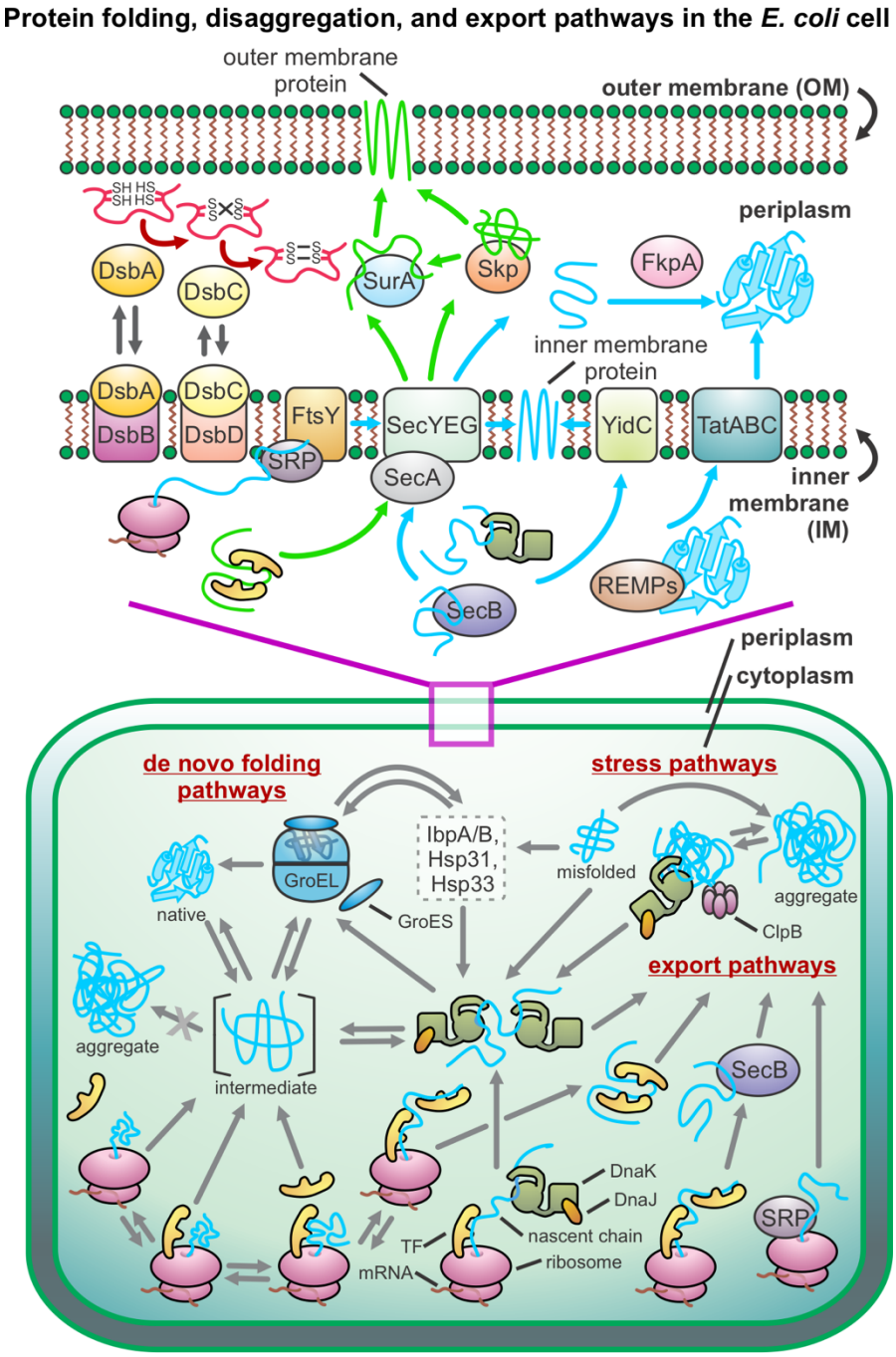


Figure 1-7. Pictorial representation of major pathways leading to protein native-structure formation in prokaryotes. This scheme applies to gram-negative bacteria, e.g., *E. coli*. Abbreviations: SRP = signal recognition particle, REMPs = redox enzyme maturation protein, TatABC = twin-arginine translocation, TF = trigger factor.

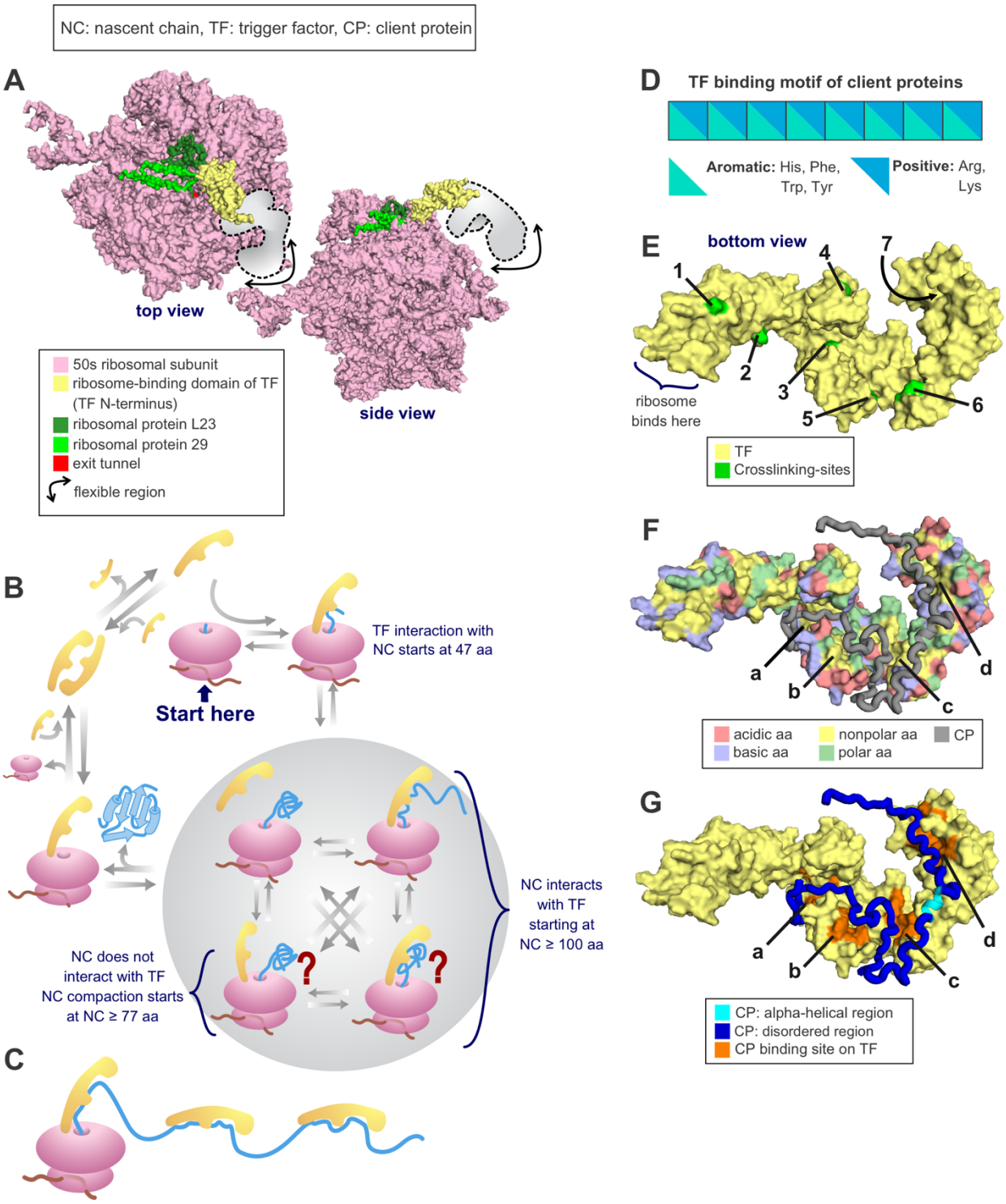


Figure 1-8. Protein folding in the presence of the trigger factor (TF) chaperone. (A) Crystal structure of the RBD of TF bound to the 50S unit of the ribosome from eubacterium

Deinococcus radiodurans. PDB: 2AAR.[231] (B) TF cycle. Note that TF is in rapid equilibrium with the ribosome. (C) Multiple TFs can be associated with the same nascent chain during translation or with the client protein in solution. (D) Nascent chain binding site for TF. (E) Crosslinking sites used to track the progression of the nascent chain as it travels throughout the TF.[230] PDB: 2MLX. (F) *E. coli* TF amino acids (aa) highlighted according to type. Note that the TF binding sites for PhoA, shown in the next panel, are all either nonpolar or polar. PDB: 2MLX. (G) *E. coli* TF associated with the 220-310 fragment of the PhoA client protein. PDB: 2MLX (Abbreviations: NC = nascent chain, TF = trigger factor, residues = amino acids).

chains cotranslationally (Figure 1-8B).[194, 202] Deletion of TF in *E. coli* under regular growth conditions is not lethal, but the combined deletion of TF and DnaK causes protein aggregation and cell death.[201, 209, 220] TF binds to ribosome-bound nascent chains of most cytosolic proteins, outer membrane proteins, and periplasmic proteins.[202, 221] TF was also found to assist the refolding of some denatured proteins *in vitro.*[207, 222-224] Upon binding nascent proteins, TF delays acquisition of the fully native state and increases the ultimate yield of bioactive protein.[147, 225] Off the ribosome, TF binds client proteins in a predominantly unfolded conformation.[226] On the ribosome, TF reduces the force exerted by a cotranslationally folding chain, suggesting that it increases the population of unfolded nascent protein.[227] TF was also proposed to generate a "protected" space where nascent chains may be shielded from degradation and aggregation and may potentially fold cotranslationally (Figure 1-8B).[202, 228-230] See additional comments in the section titled "Structure and dynamics of trigger factor client proteins".

1.14 Trigger factor structure and function

TF is a 48 kDa (432 residues) protein comprising a ribosome-binding N-terminal domain, a peptidyl-prolyl isomerase (PPIase) domain, and a C-terminal domain.[229, 232] TF was described as having a dragon-shaped structure, with the N-terminal domain as the tail, the PPIase domain as the head, and the C-terminal domain, located in the central portion of the structure, forming the two arms.[229] The N-terminal domain binds ribosomal protein L23 and can also interact with L29.[229, 231, 233, 234] TF's PPIase activity has been demonstrated *in vitro*.[217, 235] Interestingly, this domain is not necessary for TF's *in vivo* chaperone function.[220, 223, 236] The C-terminal domain performs the main chaperone function and TF fragments containing only the C-terminal domain prevent aggregation and promote folding *in vitro*[236] while fragments lacking the C-terminal

domain show decreased chaperone activity.[237]

1.15 Structure and dynamics of trigger factor client proteins

Nascent proteins can interact with all three domains of TF.[226] TF typically binds nonpolar regions of nascent proteins,[208, 238] though it can also interact with hydrophilic regions.[202, 203] The PPIase domain of TF binds eight-residue sequences enriched in aromatic and basic amino acids (Figure 1-8D).[239] *In vitro* experiments featuring purified TF-client protein complexes revealed that this chaperone binds proteins with 40 or more residues.[170, 230, 240] On the other hand, *in vivo* investigations showed that TF binds ribosome-bound nascent proteins of 100 or more residues.[221] It has a higher affinity for ribosomes carrying nascent chains than for empty ribosomes, supporting its cotranslational role.[208, 241] TF's affinity for the ribosome[155, 241] and nascent chains[242] increases with chain length, likely due to increased interactions between nascent chains and TF. Nascent proteins appear to move along the TF structure as they emerge from the ribosome. For instance, ribosome-bound isocitrate dehydrogenase interacts with the N-terminal domain first and, as the nascent chain elongates, it proceeds through the TF's arms and then reaches the PPIase domain.[230] Some client proteins bind concurrently bind multiple TF proteins. For instance, PhoA can bind to up to three TF proteins (Figure 1-8C, E-G).[226]

Crystal structures[231, 234] and a cryoEM structure[242] show that, when TF binds the ribosome, its N-terminal ribosome binding domain undergoes a conformational change that exposes a nonpolar region to the ribosomal tunnel. This conformational transition enables TF to interact with nonpolar regions of unfolded nascent proteins.[226, 227, 238, 243, 244] TF was also proposed to create a shielded environment supporting aggregation-free cotranslational folding.[228] A crystal structure of an *E. coli* TF bound to *Haloarcula marismortui* ribosomes[229] show there is sufficient

space between TF and the ribosome for a small to medium single-domain nascent protein to fold. This potential folding cavity was also shown in cryoEM structure of an *E. coli* TF and ribosome complex.[230] Yet, another crystal structure of the *D. radiodurans* TF and ribosome shows a much smaller space underneath TF that may not accommodate cotranslational folding.[234]

Fluorescence anisotropy-decay showed that ribosome-bound nascent proteins form a compact structure both in the absence and presence of TF.[155] Hence TF is not necessary for nascent-chain compaction, though it could affect its population.[155] The average residence time for TF binding to ribosomes is at least 10 s.[208, 219, 238, 241, 245] This time is sufficient for the translation 100-200 amino acids in *E. coli*[202] and for the concurrent binding and unbinding of TF to nascent chains as they elongate, which occurs on the ms timescale.[207, 246]

The two TF modes of action outline above, namely enhancing the population of unfolded clients and providing a protected environment supporting some nascent-chain compaction, are not mutually exclusive.

The known conformational flexibility of TF[229-231, 234, 242] is consistent with its ability to interact with a variety of nascent chains.[247] Finally, TF can also act post-translationally, by binding ribosome-released proteins[207] or by remaining bound to nascent chains after they are released from the ribosome.[203, 238] Post-translational interaction with TF may help stabilize protein monomers until they are assembled into complexes. This proposed role of TF in complex assembly is supported by the observation that cells lacking TF show a ribosome assembly defect under heat-stress conditions.[203]

1.16 Protein folding in the presence of the Hsp70 chaperone

In 1962, Ferruccio Ritossa observed that *Drosophila* larvae under heat stress show a "puffing pattern" around chromosomes that was later shown to result from an upregulation of the heat shock protein now known as Hsp70[248-255] Later, the presence of Hsp70 chaperones was identified within wide a variety of organisms.[254]

Hsp70 chaperones are ATP-dependent proteins that are routinely produced within the cell cytosol under non-stress conditions and that are also upregulated upon heat stress. Hsp70s are highly conserved and very important for maintaining cellular life.[256, 257] Hsp70s and Hsp70-like proteins are found across a wide variety of organisms, including prokaryotes, eukaryotes and even most archaea,[159, 256] which are missing several other classes of chaperones (e.g., Hsp100 and Hsp90/83.[258] The only chaperone more universally represented than Hsp70 is the Hsp60 chaperonin (known as GroEL in bacteria), which evolved first and is present in all living organisms.[259] While Hsp70 is widespread and generally ubiquitous, it is not universally represented and is missing from the genome of most hyperthermophiles[258, 260, 261] and two specific classes of bacteria.[262]

The Hsp70 chaperone system includes Hsp70 and its cochaperones. In *E. coli*, this system includes DnaK (*E. coli* Hsp70) and cochaperones DnaJ (Hsp40) and GrpE. The latter is a nucleotide exchange factor (NEF). The Hsp70 chaperone system is considered a "central hub" in *E. coli* cells (Figure 1-7) due to its ability to interact with a wide variety of client proteins and due to its capability to influence a variety of cellular processes, spanning from *de novo* protein folding to protein transport and disaggregation.[209, 263] The concentration of Hsp70 within an *E. coli* cell is approximately 30-50 μM,[264] and the total *E. coli* protein concentration is 5-8 mM.[265] Therefore, not all proteins in a cell can associate with the Hsp70 chaperone at once.[266, 267] DnaK displays a

preference for 30-75 kDa client proteins and binds ~20% of newly synthesized proteins in the *E. coli* proteome (Figure 1-9A).[209, 268] The Hsp70 chaperone system maintains cell homeostasis by holding unfolded proteins to prevent aggregation and by unfolding misfolded client proteins, so they can fold correctly.[85, 267]

The Hsp70 chaperone interacts with client proteins both co- and post-translationally to promote correct *de novo* folding of nascent chains.[155, 201, 263] Hsp70 assists the disassembly of protein complexes during bacteriophage replication,[269], protein transport across membranes,[270] and promotes the assembly of tail-anchored proteins within the cell membrane.[271] The Hsp70 chaperone system was reported to help disaggregating small aggregates and, in conjunction with other chaperones (e.g., ClpB), it was shown to assist the disaggregation of large aggregates.[272-275] The Hsp70 chaperone system can either take over client proteins from other chaperone systems and/or transfer them to other chaperone networks. Relevant chaperone networks include GroEL/ES,[199] heat shock protein 90 (Hsp90)[278] and other small heat shock proteins like IbpA, IbpB, or inclusion-body binding proteins.[279, 280]

Interestingly, Hsp70 is capable of preventing harm arising from deleterious mutations, thus granting key benefits to the parent organism in terms of both health and evolutionary rates.[281, 282] Therefore, Hsp70 and other chaperones allow organisms to experience greater genetic variation without harmful effects on fitness and could increase the species' ability to evolve.[281, 283] In proteo-bacteria, client proteins with a high binding affinity for Hsp70 evolve faster than client proteins with low binding affinities for this chaperone.[281, 282] Given that Hsp70 increases protein evolution rate, overexpression of this chaperones may promote the efficiency of directed evolution.[282] Yet, chaperones do not always promote evolution, and other studies showed that Hsp70 and other chaperones sometimes decrease the client-protein evolution rate.[284]

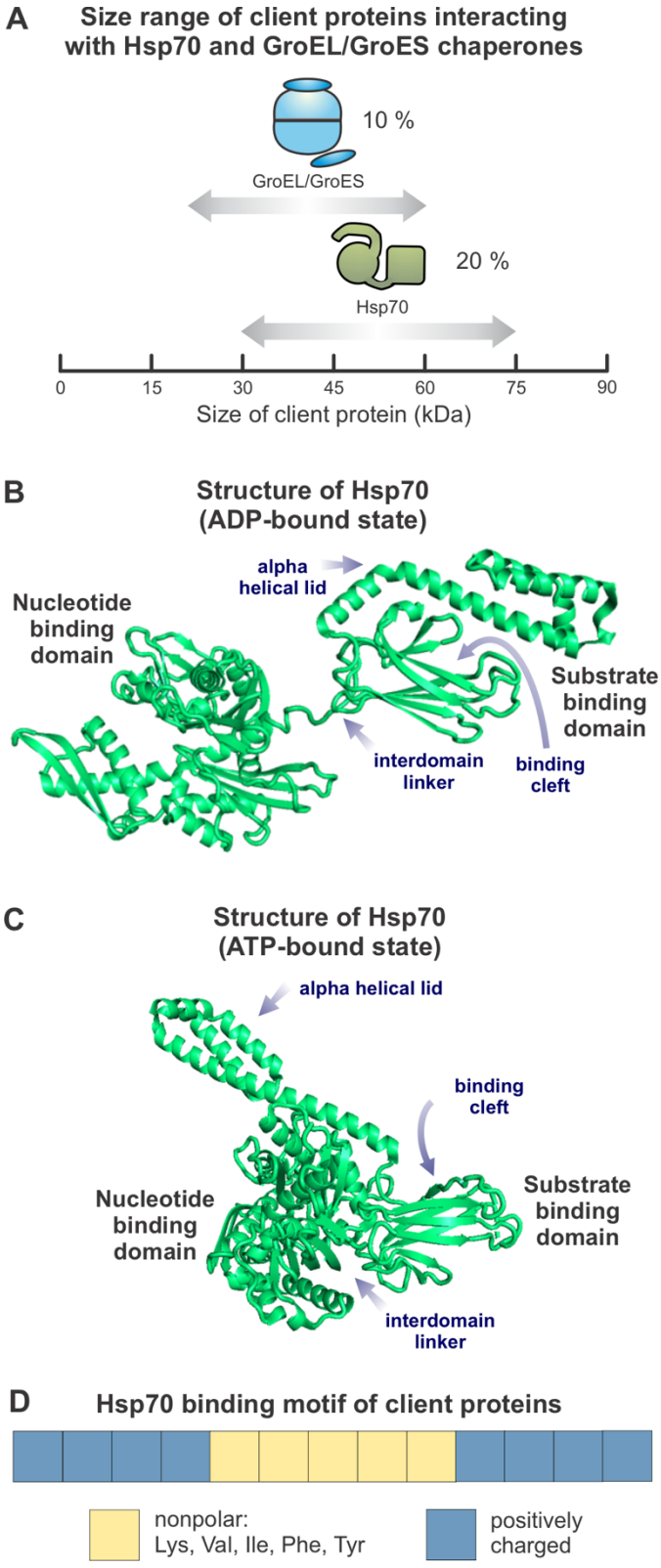

A
Size range of client proteins interacting with Hsp70 and GroEL/GroES chaperones
10 %
GroEL/GroES
20 %
Hsp70
0
15
30
45
60
75
90
Size of client protein (kDa)
B
Structure of Hsp70 (ADP-bound state)
alpha helical lid
Nucleotide binding domain
Substrate binding domain
interdomain linker
binding cleft
C
Structure of Hsp70 (ATP-bound state)
alpha helical lid
binding cleft
Nucleotide binding domain
Substrate binding domain
interdomain linker
D
Hsp70 binding motif of client proteins
nonpolar: Lys, Val, Ile, Phe, Tyr
positively charged

Figure 1-9. Key structural features required for the interactions of a client protein with the Hsp70 chaperone system. (A) Structure of ADP-bound (or nucleotide-free) Hsp70 chaperone (DnaK from *E. coli*). PDB ID: 2KHO. (B) Structure of ATP-bound DnaK chaperone. PDB ID: 4B9Q. (C) Client-protein binding motif for interaction with the *E. coli* Hsp70 chaperone DnaK, defined according to.[276, 277] Note that the positively charged residues flanking the central nonpolar core are progressively less important, as the sequence separation from the core increases.

Additional future studies are necessary to fully understand the link between Hsp70 and evolutionary rates.

Not surprisingly, suboptimal Hsp70 function is linked to disease. If genes encoding the trigger factor (TF) and Hsp70 chaperones are concurrently knocked out, *E. coli* cells are no longer viable. This finding implies that the combination of TF and DnaK is essential for *E. coli* life. If only one of the two chaperone systems is knocked out, cells can survive but are more susceptible to stress.[201, 220, 256] Eukaryotic Hsp70 knockout or downregulation leads to increased levels of amyloid plaques in neurodegenerative diseases including Alzheimer's and Huntington's. Interestingly, while Hsp70 upregulation reduces the aggregation of plaque-forming proteins (a favorable effect), it also disfavors the apoptosis of cancerous cells (a deleterious effect).[285, 286] Therefore, a carefully balanced chaperone concentration is required to support optimal health.

1.17 Hsp70 structure and function

DnaK, the *E. coli* Hsp70 chaperone, consists of two domains comprising a 45 kDa nucleotide-binding domain (NBD) and a 25 kDa substrate-binding domain (SBD).[287] The NBD contains two lobes that form a cleft that contains a binding site for a nucleotide (ATP or ADP) and specific cations (Mg^{2+} and 2 K^{+}).[288-290] The nucleotide state determines the conformation of Hsp70. If the chaperone is nucleotide-free or bound to ADP, then the two domains behave independently (Figure 1-9B). If Hsp70 is bound to ATP, then the chaperone lobes within the NBD rotate which subsequently cause the NBD and SBD domains dock to each other (Figure 1-9C).[290, 291]

The SBD contains two subdomains: the alpha-helical lid (SBDα), and the β-sheet pocket

(SBDβ).[292] SBDβ contains two beta sheets and two loops that form the pocket where the client protein binds.[292, 293] The conformation of the SBD varies between an open state when ATP is bound to the NBD and a closed state when the NBD is ADP-bound or nucleotide-free.[289] Crystal structures show that the binding pocket of ATP-bound DnaK exists in multiple open conformations and is likely dynamic and flexible.[290, 293, 294] The binding pocket preferentially interacts with a 4-5 residue long client-protein motif comprising aromatic (Phe, Tyr) or aliphatic (Val, Leu, Ile) nonpolar residues flanked by ca. four positively charged amino acids (Figure 1-9D**)**. The characteristics of the amino acids towards the center of this motif are more important than the outer amino acids, for predicting binding to Hsp70.[277] Interestingly, this binding motif occurs on average every 36 residues in most client proteins.[277, 295]

1.18 The Hsp70 chaperone cycle

Hsp70 chaperone activity proceeds via a functional cycle, which includes the Hsp40 (a J-domain protein) and nucleotide exchange factor (NEF) cochaperones..[296-298] This cycle can be split into four main steps, as shown in Figure 1-10. The different stages of the Hsp70 chaperone cycle are briefly outlined below.

In the first step, the ATP-bound Hsp70 binds the client protein via hydrogen bonds and van der Waals interactions. However, that binding is very transient unless accompanied by ATP hydrolysis shown in step two.[292] ATP hydrolysis can occur in the absence and presence of client protein. However, the reaction is significantly slower in the absence of a bound client protein. To achieve maximal rate of ATP hydrolysis, a J protein (Hsp40s, DnaJ in *E. coli*) is needed. The J protein binds the client and transfers it to Hsp70.[257, 299-301] J proteins have slightly different binding motifs to client proteins when compared to Hsp70. The J-protein binding motif enables

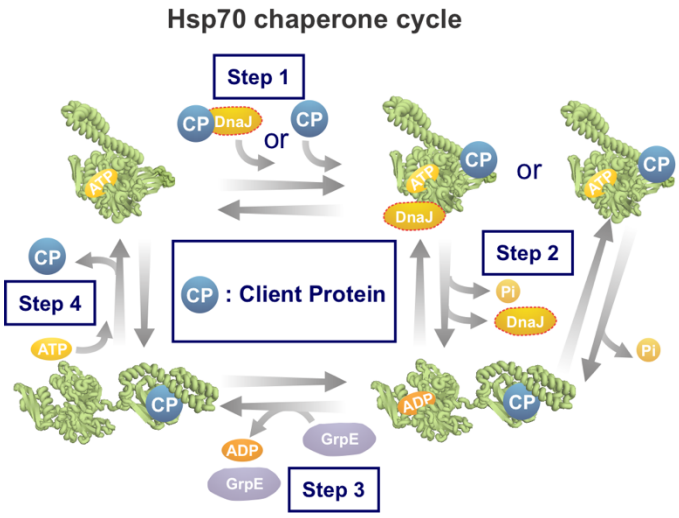


Figure 1-10. Scheme illustrating the major steps of the *E. coli* Hsp70 (a.k.a. DnaK) chaperone cycle. Hsp70 cooperates with co-chaperones DnaJ and GrpE through an ATP-dependent cycle to promote the folding of client proteins.

the ultimate binding of Hsp70 to a wider range of misfolded or aggregated proteins, and likely targets this class of client proteins to Hsp70 because they may not be able to directly bind the Hsp70 binding motif described above (Figure 1-9D).[257, 300, 302, 303]

In the second step, ATP hydrolysis causes the SBD and NBD of ADP-Hsp70 to undock and behave independently, while staying covalently connected to the inter-domain linking region.[304, 305] The alpha-helical lid lowers towards the client protein, leading to increased client-protein affinity and slower client dissociation rate.[257]

The third step involves departure of the ADP nucleotide from Hsp70. In bacteria, this step is rate-limiting and requires a nucleotide-exchange factor (NEF), e.g., GrpE (in *E. coli*) to promote the release of ADP, leaving Hsp70 in a nucleotide-free state.[306, 307] Upon nucleotide removal, NEF remains associated with Hsp70 and is thought to prevent ADP from rebinding.[289, 308, 309] While the α-helical lid is considered "closed" when DnaK is either ADP-bound or nucleotide-free, several studies showed that the lid is subject to slow dynamics and it occasionally reopens, thus allowing client proteins to bind/unbind the chaperone.[290, 298, 310]

In the fourth step, ATP binds Hsp70 within the NBD cleft. The highly conserved nonpolar linker between the NBD and SBD upon ATP-binding, pulls the two domains together until they are firmly docked (Figure 1-9C).[291] In addition, nucleotide-binding causes the α-helical lid to lift, enabling the client-protein to be released and the cycle to start anew.[257, 289, 296, 305, 308, 309]

1.19 Structure and dynamics of Hsp70-bound client proteins

The client protein can also change conformation during the Hsp70 cycle. The first NMR study of Hsp70-bound peptides showed that client proteins bound to nucleotide-free bacterial Hsp70 have a more extended conformation than in chaperone-free solution.[311] Later studies

using electron paramagnetic resonance spectroscopy (EPR) confirmed these results for peptides bound to nucleotide-free, ADP-bound, or ATP-bound Hsp70.[312] NMR studies on N-terminal fragments of apoMb alone showed that this protein has some helical structure.[313] Binding to Hsp70 unwinds the local helix structure of residues in the Hsp70 binding site.[313] Yet, regions distant from the binding site form non-native α-helical structure.[314] Single-molecule Főrster resonance energy transfer (FRET) experiments showed that the protein rhodanese (296 residues) lacks stable tertiary structure when bound to bacterial ADP-Hsp70.[315]

NMR studies by Lee and coworkers[316] showed that the drkN SH3 client protein, an N-terminal SH3 domain from *Drosophila*, populates multiple globally unfolded interconverting states while bound to ADP-Hsp70. The bound protein also populates additional spectroscopically undetectable states that account for 43% of the entire chaperone-bound population. This result is important because it shows that conformational sampling takes place while the client protein is bound to the ADP-Hsp70 chaperone.[316] Therefore, Hsp70-bound client proteins are dynamic chains that are able to sample distinct conformational states (and potentially fold, partially fold or unfold) while chaperone bound. Clearly, additional research needs to be performed to define the nature of the Hsp70-bound client more accurately-protein states and how they depend on client-protein amino-acid sequence.

Hsp70 promotes folding and prevents aggregation via two mechanisms. First, this chaperone "holds" (i.e., binds) client proteins in a predominantly unfolded or partially folded state, thus preventing aggregation by effectively lowering the concentration of client-protein conformations in solution. Second, Hsp70 promotes conformational changes within the bound client proteins that enable the conversion of misfolded client proteins to the folded state. In this review, we denote the first mechanism as "hold-only" and the second as "fold-promoting" behavior (Figure

1-11). Note that we use these terms instead of the more common “holdase” and “foldase” descriptors. A brief justification follows. In the biochemical literature, the “ase” suffix is typically employed to denote enzymes that catalyze reactions that lead to bond breaking or covalent scission of substrates into smaller components. While the Hsp70 chaperone system can lead to faster generation of the client-protein native state,[85] in some cases, Hsp70 slows down native-state formation.[317] In both cases no covalent cuts are introduced. Therefore, we opted not to use the “ase” terminology.

According to the “hold-only” mechanism, Hsp70 transiently binds client proteins whether or not they are aggregation-prone, thus decreasing the concentration of free proteins in solution (Figure 1-11A).[95, 256, 286] This mode of action prevents aggregation because the nucleation and elongation rates of individual molecules undergoing nucleated-polymerization-like aggregation are concentration dependent. On the other hand, the rate of folding of monomeric proteins is not concentration dependent.[319-321]

Several chaperone systems can adopt a hold-only-type mechanism, including the Hsp40, Hsp70 and GroEL/ES machineries.[322-325] Note that the Hsp70 and GroEL/ES chaperones may also facilitate folding and prevent aggregation via a fold-promoting-type mechanism.[210] The hold-only mode of Hsp70 action is supported by size-exclusion chromatography experiments that determined the degree of chaperone association of three distinct non-aggregation-prone model client proteins.[318] This study showed that, during folding away from equilibrium, Hsp70 interacts mostly with slow-folding proteins. In contrast, at equilibrium Hsp70 interacts preferentially with thermodynamically unstable proteins.[318] Similar conclusions were reached in separate experimental studies focusing on the apparent folding rate of RNAse H in the absence

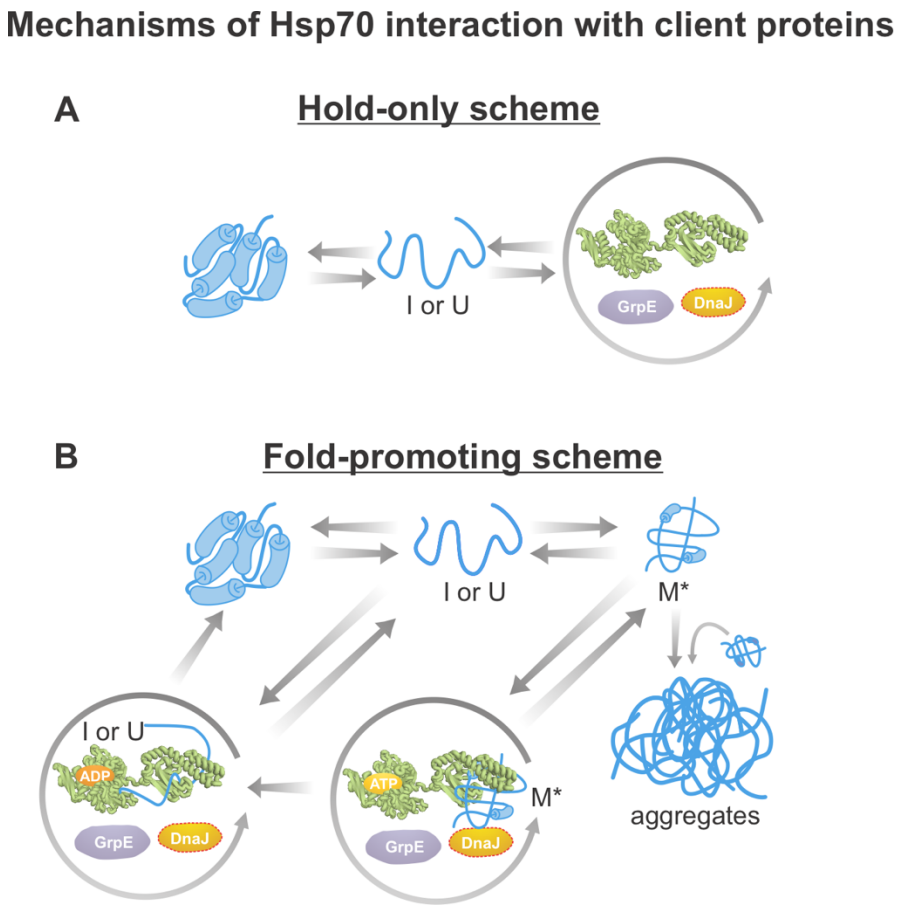


Figure 1-11. Simplified schemes illustrating chaperone-assisted protein folding. The diagrams in this figure are consistent with experimental results achieved with distinct classes of client proteins. (A) Hold-only model consistent with both computational and experimental results on non-aggregation-prone proteins bearing one Hsp70 binding site.[317, 318] (B) Fold-promoting models consistent with experimental results obtained with aggregation-prone client proteins bearing multiple chaperone binding sites per molecule. For instance, firefly luciferase (fluc) populate their native states more quickly and avoid generating aggregates in the presence of the Hsp70 chaperone system.[85] According to this fold-promoting model, the Hsp70 chaperone system catalyzes the conversion of misfolded monomers (M*) to the native state and, in so doing, increases the yields and observed rates of native-structure formation.

and presence of the Hsp70 chaperone system.[317]

Optical tweezer experiments showed that Hsp70 binds and stabilizes unfolded maltose binding protein.[326] Similar results were obtained for Hsp70-bound drkN SH3 by NMR, except that additional conformations were found.[316] Earlier NMR studies with client peptides showed an effectively unfolded (conformationally expanded, β-sheet-like) population of Hsp70-bound peptides.[311] Similar results were later obtained by electron paramagnetic resonance (EPR).[312] Single molecule FRET studies showed that the large client protein rhodanese concurrently binds several Hsp70 molecules.[315] Hsp70 binding to unfolded client proteins (e.g., RNAse H) slows down the observed rate of native-state acquisition.[317] Chaperone binding can occur directly to the ADP-bound state of Hsp70, suggesting that the hold-only mechanism does not require ATP hydrolysis for client binding.[312, 316] Therefore, interestingly, Hsp70 can both accelerate or slow-down acquisition of native structure via the fold-promoting and hold-only mechanisms, respectively. The experimental evidence available so far is consistent with the fact that Hsp70 uses the fold-promoting mechanism when interacting with aggregation-prone proteins that proceed via one or more misfolded intermediates.[85] Conversely, Hsp70 employs the hold-only mechanism upon interacting with non-aggregation-prone proteins,[318, 327] which are characterized by free-energy landscapes similar to those of Figure 1-4A. The latter scenario is facilitated in the case of thermodynamic unstable and(or) slow-folding client proteins.[318] Additional work in this area is necessary, to more comprehensively characterize all viable scenarios. For instance, aggregation-prone client proteins that do not significantly populate misfolded intermediates may interact with Hsp70 via the hold-only mode.

The second mechanism adopted by Hsp70 to prevent client protein aggregation is denoted here as “fold-promoting” behavior. According to this mechanism, Hsp70 binds and unfolds

misfolded proteins. In this way, Hsp70 enables misfolded states to bypass kinetic trapping relative to the native state give rise to the native conformation.[328] Fluorescence studies in the bulk and at the single-molecule level showed that ATP- and client-protein-bound Hsp70 undergoes ATP hydrolysis concurrently with unfolding of misfolded client proteins. Further, the unfolding of the misfolded state does not take place in the absence of ATP.[85, 328] This combined evidence strongly suggests that Hsp70 uses energy from ATP hydrolysis to unfold misfolded luciferase.[85, 328] Upon release from the chaperone, luciferase can then fold to the native state. Hydrolysis of five ATPs is required to enable the correct folding of one single luciferase protein, suggesting that one out of five unfolded proteins folds correctly to the native state, while the others misfold. Therefore, multiple cycles of chaperone binding and release are required for the client protein to fold correctly.[85, 328]

Client proteins bearing more than one Hsp70 binding site may interact with multiple Hsp70 chaperones at once.[315, 329] Binding multiple chaperones causes steric repulsion that causes client proteins to adopt an expanded state.[85, 315] When multiple chaperones are bound to a single client protein, release from the chaperone is likely asynchronous. This causes the client protein to spend more time in a chaperone-associated state (bearing at least one bound chaperone) than if it were bound to one single chaperone.[315, 330, 331] Asynchronous release may allow different regions of the client protein to fold independently, similar to the foldon mechanism, and prevent misfolding interactions.

It was reported that Hsp70 can also assist protein disaggregation. This process has low efficiency in the presence of Hsp70 alone[274] and is much more effective when Hsp70 cooperates with other chaperones including Hsp100-type disaggregases including bacterial ClpB.[148, 332, 333]

1.20 Conclusions

A critical event in protein folding is the burial of most nonpolar residues away from the hydrophilic solvent. This process is accompanied by the formation of a spatially organized 3D structure, which is in most cases kinetically trapped from a variety of aggregated states, under physiologically relevant conditions. Some proteins successfully fold fast and independently. This category includes small, mid-size and large proteins. On the other hand, many proteins, especially some mid-size and most large multi-state-folding proteins, fold independently but slowly. Many mid-size and large proteins, bearing molecular-weight ranges highly represented in bacteria, are very aggregation-prone upon refolding from denaturant or as they emerge from the ribosome. These proteins require the cellular machinery in the early stages of their life, so that they populate bioactive states that remain kinetically trapped from misfolded aggregates under physiologically relevant conditions. In bacteria, the relevant machinery responsible for the formation of the native state at birth includes the ribosome and the molecular chaperones trigger factor and Hsp70. It is therefore clear that the early steps in protein folding in the cell, including co- and immediately post-translational folding, are essential for long-term protein solubility and function. They key aspects of this process are schematically illustrated in Figure 1-12.

In conclusion, while some proteins are capable of folding independently when diluted from denaturant or upon release from the ribosome, many proteins need assistance from the cellular machinery in the early stages of their life. In this way, they remain bioactive and kinetically trapped from harmful aggregates over extremely long time spans.

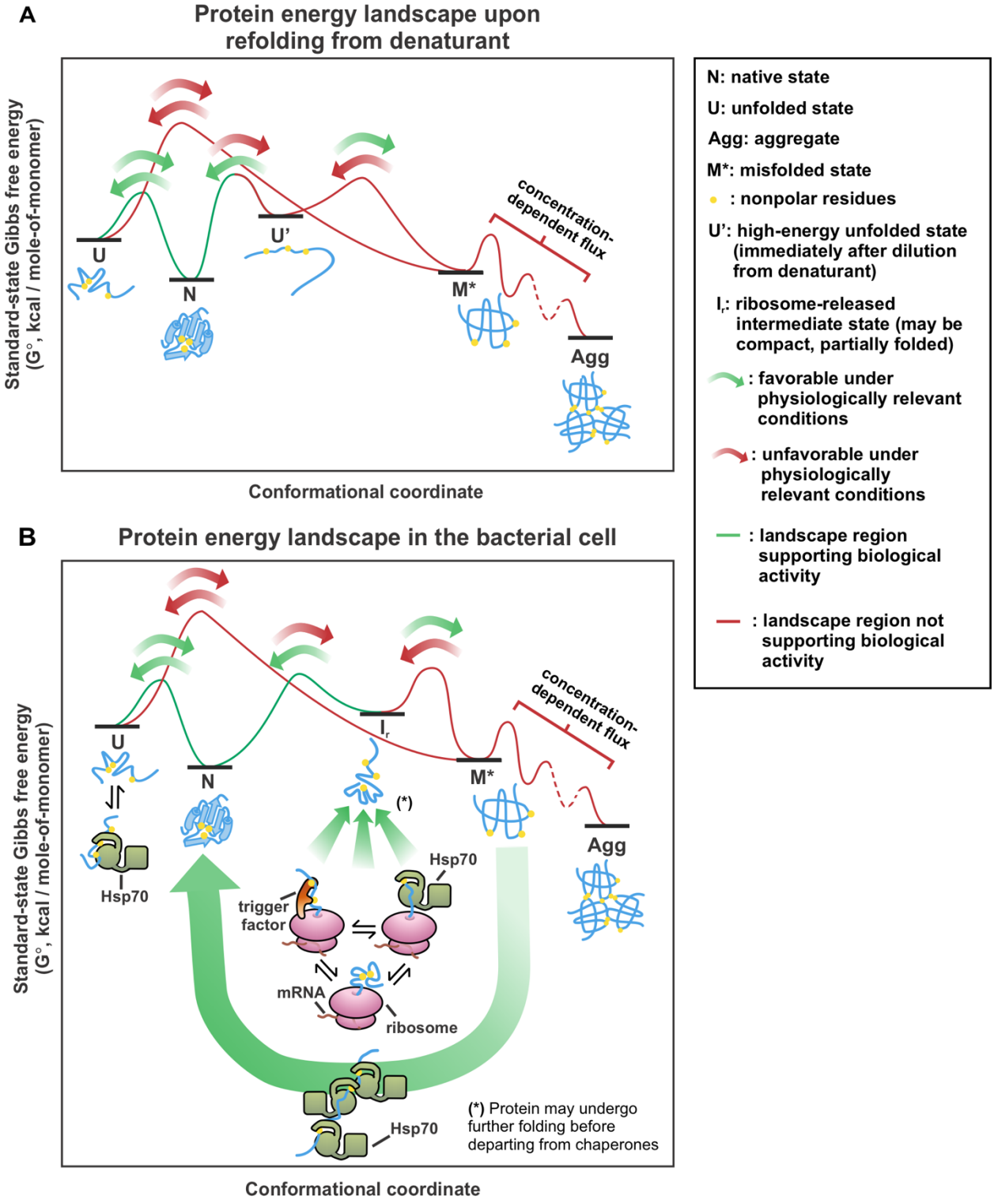

A
Protein energy landscape upon refolding from denaturant
Standard-state Gibbs free energy (G°, kcal / mole-of-monomer)
U
N
U'
M*
Agg
concentration-dependent flux
Conformational coordinate
N: native state
U: unfolded state
Agg: aggregate
M*: misfolded state
: nonpolar residues
U': high-energy unfolded state (immediately after dilution from denaturant)
I_r: ribosome-released intermediate state (may be compact, partially folded)
: favorable under physiologically relevant conditions
: unfavorable under physiologically relevant conditions
: landscape region supporting biological activity
: landscape region not supporting biological activity
B
Protein energy landscape in the bacterial cell
Standard-state Gibbs free energy (G°, kcal / mole-of-monomer)
U
N
I_r
M*
Agg
concentration-dependent flux
(*)
Hsp70
trigger factor
mRNA
ribosome
Hsp70
(*) Protein may undergo further folding before departing from chaperones
Conformational coordinate

Figure 1-12. Summary of protein folding standard-state Gibbs free energy landscapes. (A) Protein folding *in vitro* upon dilution from denaturant, and (B) ribosome and chaperone-assisted protein folding within the bacterial cell.

1.21 This PhD thesis work

This thesis highlights the essential early stages in single- and multi-domain protein folding and describes how the ribosome and molecular chaperones enable newly synthesized proteins to form their functional native structure in the cell. As proteins are synthesized, they undergo co-translational compaction, involving only 40-50% of nascent-protein residues (Chapter 2). Interactions between globin chains (140 residues) and the ribosomal protein L23 were experimentally demonstrated, for the first time. As the protein chain length increases, their interaction network expands to include one additional ribosomal protein, and the trigger factor and Hsp70 chaperones (Chapter 3). These nascent-chain interactions may involve residues within or outside the N-terminal compact protein domain. During release from the ribosome, newly synthesized proteins must incorporate at least 50% of the nascent-protein residues into final structure. These early post-translational folding steps are crucial, and proteins are particularly prone to aggregation while completing their folding. Molecular chaperones, including Hsp70, are often required to enable correct folding. Hsp70 grants solubility to aggregation-prone globin variants. Yet, soluble proteins are not always structurally accurate. Hsp70 not only prevents insoluble aggregation but also reduces soluble aggregate formation, enabling a higher fraction of protein to reach the native state (Chapter 4). The Hsp70 chaperone concentration required to prevent soluble and insoluble aggregate formation depends on the protein sequence. However, chaperones are not a perfect solution to aggregation. Physiologically relevant concentrations of Hsp70 are sometimes insufficient to prevent soluble-aggregate formation. Nearly all *E. coli* proteins contain at least one Hsp70 binding site, including proteins that do not require chaperone assistance for solubility under normal growth conditions (Chapter 5). These proteins may require Hsp70 to reach the native state. In summary, this work provides insights into the critical early co-

and immediately post-translational folding events that are a prerequisite for long-term folding and function in the cell.

1.22 Acknowledgments

We are thankful to Xuhui Huang for helpful discussions, and we are grateful to Dmitry Ivankov for sharing the compiled data for protein folding rates that we used to generate Figure 1-3, and Supplementary Tables S1 and S2. We thank the National Science Foundation (NSF) for funding (grants MCB-2124672 and CBET 1912259 to S.C.). In addition, R.H. thanks the National Institute of General Medical Sciences of the National Institutes of Health for a TEAM-Science Fellowship (award number R25GM083252).

1.23 References

(1) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J., et al. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014,** *114*, 6661-6714.

(2) Yamaguchi, H.; Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **2014,** *4*, 235-251.

(3) Betts, S. D.; King, J. Cold rescue of the thermolabile tailspike intermediate at the junction between productive folding and off-pathway aggregation. *Protein Sci.* **1998,** *7*, 1516-23.

(4) Speed, M. A.; Wang, D. I.; King, J. Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotechnol.* **1996,** *14*, 1283-1287.

(5) Tycko, R. Amyloid polymorphism: Structural basis and neurobiological relevance. *Neuron* **2015,** *86*, 632-45.

(6) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D., et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021,** *373*, 871-876.

(7) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596*, 583-589.

(8) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C. L.; Zidek, A.; Nelson, A. W. R.; Bridgland, A., et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020,** *577*, 706-710.

(9) Miller, J. L. Deep learning opens up protein science's next frontiers. *Phys. Today* **2021,** *74*.

(10) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Ar-Ganiska, J.; Bely, B.; Bingley, M., et al. Uniprot: A hub for protein information. *Nucleic Acids Res.* **2015,** *43*, D204-D212.

(11) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B., et al. Uniprot: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021,** *49*, D480-D489.

(12) Leader, B.; Baca, Q. J.; Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **2008,** *7*, 21-39.

(13) Greenwald, E. C.; Mehta, S.; Zhang, J. Genetically encoded fluorescent biosensors illuminate the spatiotemporal regulation of signaling networks. *Chem. Rev.* **2018,** *118*, 11707-11794.

(14) Gebauer, M.; Skerra, A. Engineered protein scaffolds as next-generation therapeutics. *Annual Review of Pharmacology and Toxicology* **2020,** *60*, 391-415.

(15) Richards, F. M. Structure of proteins. *Annu. Rev. Biochem.* **1963,** *32*, 269-300.

(16) Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **1974,** *248*, 338-339.

(17) Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **1976,** *105*, 1-12.

(18) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A view of the hydrophobic effect. *J. Phys. Chem. B* **2002,** *106*, 521-533.

(19) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **1961,** *47*, 1309.

(20) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **1973,** *181*, 223-230.

(21) Levinthal, C. How to fold graciously. *Mossbauer spectroscopy in biological systems* **1969,** *67*, 22-24.

(22) Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985,** *24*, 1501-1509.

(23) Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*, 20-22.

(24) Fedyukina, D. V.; Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **2011,** *40*, 337-359.

(25) Sosnick, T. R.; Barrick, D. The folding of single domain proteins—have we reached a consensus? *Curr. Opin. Struct. Biol.* **2011,** *21*, 12-24.

(26) Englander, S. W.; Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111*, 15873-80.

(27) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W. N.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 13286-13291.

(28) Englander, S. W. Protein folding intermediates and pathways studied by hydrogen exchange. *Annual Review of Biophysics and Biomolecular Structure* **2000,** *29*, 213-238.

(29) Radford, S. E. Protein folding: Progress made and promises ahead. *Trends Biochem. Sci.* **2000,** *25*, 611-618.

(30) Nelson, D. L.; Lehninger, A. L.; Cox, M. M. *Lehninger principles of biochemistry*. W. H. Freeman: New York, 2008.

(31) Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **2003,** *28*, 18-25.

(32) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **2003,** *421*, 863-867.

(33) White, G. W. N.; Gianni, S.; Grossmann, J. G.; Jemth, P.; Fersht, A. R.; Daggett, V. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* **2005,** *350*, 757-775.

(34) Arai, M. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophys. Rev.* **2018,** *10*, 163-181.

(35) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998,** *277*, 985-994.

(36) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* **2003,** *12*, 2057-2062.

(37) Kurt, N.; Mounce, B. C.; Ellison, P. A.; Cavagnero, S. Residue-specific contact order and contact breadth in single-domain proteins: Implications for folding as a function of chain elongation. *Biotechnol. Prog.* **2008,** *24*, 570-575.

(38) Wang, M.; Kurland, C. G.; Caetano-Anollés, G. Reductive evolution of proteomes and protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, 11954.

(39) Han, J.-H.; Batey, S.; Nickson, A. A.; Teichmann, S. A.; Clarke, J. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **2007,** *8*, 319-330.

(40) Gianni, S.; Ivarsson, Y.; Jemth, P.; Brunori, M.; Travaglini-Allocatelli, C. Identification and characterization of protein folding intermediates. *Biophys. Chem.* **2007,** *128*, 105-113.

(41) Dill, K. A.; Bromberg, S.; Yue, K.; Chan, H. S.; Ftebig, K. M.; Yee, D. P.; Thomas, P. D. Principles of protein folding — a perspective from simple exact models. *Protein Sci.* **1995,** *4*, 561-602.

(42) Baldwin, R. L. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR* **1995,** *5*, 103-109.

(43) Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S. Direct observation of the three-state folding of a single protein molecule. *Science* **2005,** *309*, 2057-60.

(44) Chamberlain, A. K.; Marqusee, S. Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* **1998,** *37*, 1736-1742.

(45) Griko, Y. V.; Freire, E.; Privalov, P. L. Energetics of the alpha-lactalbumin states - a calorimetric and statistical thermodynamic study. *Biochemistry* **1994,** *33*, 1889-1899.

(46) Jha, S. K.; Udgaonkar, J. B. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 12289-12294.

(47) Rami, B. R.; Udgaonkar, J. B. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* **2002,** *41*, 1710-1716.

(48) Schulman, B. A.; Kim, P. S.; Dobson, C. M.; Redfield, C. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol.* **1997,** *4*, 630-634.

(49) Shimizu, A.; Ikeguchi, M.; Sugai, S. Unfolding of the molten globule state of alpha-lactalbumin studied by H-1-NMR. *Biochemistry* **1993,** *32*, 13198-13203.

(50) Feng, H.; Zhou, Z.; Bai, Y. A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 5026.

(51) Laurents, D. V.; Bruix, M.; Jamin, M.; Baldwin, R. L. A pulse-chase-competition experiment to determine if a folding intermediate is on or off-pathway: Application to ribonuclease A. *J. Mol. Biol.* **1998,** *283*, 669-678.

(52) Bai, Y. W. Kinetic evidence of an on-pathway intermediate in the folding of lysozyme. *Protein Sci.* **2000,** *9*, 194-196.

(53) Bai, Y. W. Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 477-480.

(54) Heidary, D. K.; O'Neill, J. C.; Roy, M.; Jennings, P. A. An essential intermediate in the folding of dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97*, 5866-5870.

(55) Spudich, G. M.; Miller, E. J.; Marqusee, S. Destabilization of the Escherichia coli RNase H kinetic intermediate: Switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* **2004,** *335*, 609-618.

(56) Lindhoud, S.; Pirchi, M.; Westphal, A. H.; Haran, G.; van Mierlo, C. P. M. Gradual folding of an off-pathway molten globule detected at the single-molecule level. *J. Mol. Biol.* **2015,** *427*, 3148-3157.

(57) Silow, M.; Oliveberg, M. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 6084.

(58) Dyson, H. J.; Wright, P. E. How does your protein fold? Elucidating the apomyoglobin folding pathway. *Acc. Chem. Res.* **2017,** *50*, 105-111.

(59) Uzawa, T.; Akiyama, S.; Kimura, T.; Takahashi, S.; Ishimori, K.; Morishima, I.; Fujisawa, T. Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 1171.

(60) Eliezer, D.; Jennings, P. A.; Wright, P. E.; Doniach, S.; Hodgson, K. O.; Tsuruta, H. The radius of gyration of an apomyoglobin folding intermediate. *Science* **1995,** *270*, 487-487.

(61) Tsui, V.; Garcia, C.; Cavagnero, S.; Siuzdak, G.; Dyson, H. J.; Wright, P. E. Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* **1999,** *8*, 45-49.

(62) Aoto, P. C.; Nishimura, C.; Dyson, H. J.; Wright, P. E. Probing the non-native H helix translocation in apomyoglobin folding intermediates. *Biochemistry* **2014,** *53*, 3767-3780.

(63) Bédard, S.; Krishna, M. M.; Mayne, L.; Englander, S. W. Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 7182-7187.

(64) Englander, S. W.; Mayne, L. The case for defined protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114*, 8253.

(65) Englander, S. W.; Mayne, L.; Krishna, M. M. Protein folding and misfolding: Mechanism and principles. *Q. Rev. Biophys.* **2007,** *40*, 1-41.

(66) Englander, S. W.; Mayne, L.; Kan, Z.-Y.; Hu, W. Protein folding—how and why: By hydrogen exchange, fragment separation, and mass spectrometry. *Annu. Rev. Biophys.* **2016,** *45*, 135-152.

(67) Chamberlain, A. K.; Marqusee, S. Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Adv. Protein Chem.* **2000,** *53*, 283-328.

(68) Krishna, M. M.; Englander, S. W. A unified mechanism for protein folding: Predetermined pathways with optional errors. *Protein Sci.* **2007,** *16*, 449-464.

(69) Bai, Y. W.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein-folding intermediates - native-state hydrogen-exchange. *Science* **1995,** *269*, 192-197.

(70) Hu, W. B.; Walters, B. T.; Kan, Z. Y.; Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110*, 7684-7689.

(71) Bollen, Y. J. M.; Kamphuis, M. B.; van Mierlo, C. P. M. The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 4095-4100.

(72) Uzawa, T.; Nishimura, C.; Akiyama, S.; Ishimori, K.; Takahashi, S.; Dyson, H. J.; Wright, P. E. Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 13859-13864.

(73) Bedard, S.; Mayne, L. C.; Peterson, R. W.; Wand, A. J.; Englander, S. W. The foldon substructure of staphylococcal nuclease. *J. Mol. Biol.* **2008,** *376*, 1142-1154.

(74) Walters, B. T.; Mayne, L.; Hinshaw, J. R.; Sosnick, T. R.; Englander, S. W. Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 18898.

(75) Georgescauld, F.; Popova, K.; Gupta, A. J.; Bracher, A.; Engen, J. R.; Hayer-Hartl, M.; Hartl, F. U. GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **2014,** *157*, 922-934.

(76) Wang, Y.; Chu, X.; Suo, Z.; Wang, E.; Wang, J. Multidomain protein solves the folding problem by multifunnel combined landscape: Theoretical investigation of a y-family DNA polymerase. *J. Am. Chem. Soc.* **2012,** *134*, 13755-13764.

(77) Scott, K. A.; Steward, A.; Fowler, S. B.; Clarke, J. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **2002,** *315*, 819-829.

(78) Steward, A.; Adhya, S.; Clarke, J. Sequence conservation in lg-like domains: The role of highly conserved proline residues in the fibronectin type III superfamily. *J. Mol. Biol.* **2002,** *318*, 935-940.

(79) Arora, P.; Hammes, G. G.; Oas, T. G. Folding mechanism of a multiple independently-folding domain protein: Double B domain of protein A. *Biochemistry* **2006,** *45*, 12312-12324.

(80) Petersen, M.; Barrick, D. Analysis of tandem repeat protein folding using nearest-neighbor models. *Annu. Rev. Biophys.* **2021,** *50*, 245-265.

(81) Batey, S.; Clarke, J. Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 18113-18118.

(82) Batey, S.; Randles, L. G.; Steward, A.; Clarke, J. Cooperative folding in a multi-domain protein. *J. Mol. Biol.* **2005,** *349*, 1045-1059.

(83) Osvath, S.; Kohler, G.; Zavodszky, P.; Fidy, J. Asymmetric effect of domain interactions on the kinetics of folding in yeast phosphoglycerate kinase. *Protein Sci.* **2005,** *14*, 1609-1616.

(84) Mello Cecilia, C.; Barrick, D. An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 14102-14107.

(85) Imamoglu, R.; Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein. *Nat. Commun.* **2020,** *11*, 365.

(86) Mendoza, J. A.; Rogers, E.; Lorimer, G. H.; Horowitz, P. M. Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. *J. Biol. Chem.* **1991,** *266*, 13044-13049.

(87) Liu, C.; Young, A. L.; Starling-Windhof, A.; Bracher, A.; Saschenbrecker, S.; Rao, B. V.; Rao, K. V.; Berninghausen, O.; Mielke, T.; Hartl, F. U., et al. Coupled chaperone action in folding and assembly of hexadecameric rubisco. *Nature* **2010,** *463*, 197-202.

(88) Chow, C.; Kurt, N.; Murphy, R. M.; Cavagnero, S. Structural characterization of apomyoglobin self-associated species in aqueous buffer and urea solution. *Biophys. J.* **2006,** *90*, 298-309.

(89) Kerner, M. J.; Naylor, D. J.; Ishihama, Y.; Maier, T.; Chang, H.-C.; Stines, A. P.; Georgopoulos, C.; Frishman, D.; Hayer-Hartl, M.; Mann, M., et al. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **2005,** *122*, 209-220.

(90) London, J.; Skrzynia, C.; Goldberg, M. E. Renaturation of Escherichia coli tryptophanase after exposure to 8 m urea. *Eur. J. Biochem.* **1974,** *47*, 409-415.

(91) Kim, D.; Yu, M.-H. Folding pathway of human α1-antitrypsin: Characterization of an intermediate that is active but prone to aggregation. *Biochem. Biophys. Res. Commun.* **1996,** *226*, 378-384.

(92) Corrales, M.; Cuscó, P.; Usmanova, D. R.; Chen, H.-C.; Bogatyreva, N. S.; Filion, G. J.; Ivankov, D. N. Machine learning: How much does it tell about protein folding rates? *PLoS One* **2015,** *10*, e0143166.

(93) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(94) Cox, B. G. *Modern liquid phase kinetics*. University Press: Oxford; New York, 1994.

(95) Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G.; Dobson, C. M. Kinetic partitioning of protein folding and aggregation. *Nat. Struct. Biol.* **2002,** *9*, 137-143.

(96) Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 4201.

(97) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L., et al.

Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **2011,** *133*, 14160-14163.

(98) Yoshimura, Y.; Lin, Y. X.; Yagi, H.; Lee, Y. H.; Kitayama, H.; Sakurai, K.; So, M.; Ogi, H.; Naiki, H.; Goto, Y. Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109*, 14446-14451.

(99) Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S. Kinetic trapping of folded proteins relative to aggregates under physiologically relevant conditions. *J. Phys. Chem. B* **2018,** *122*, 7682-7698.

(100) Varela, A. E.; England, K. A.; Cavagnero, S. Kinetic trapping in protein folding. *Protein Eng. Des. Sel.* **2019,** *32*, 103-108.

(101) Manning, M.; Colón, W. Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. *Biochemistry* **2004,** *43*, 11248-11254.

(102) Xia, K.; Zhang, S.; Solina, B. A.; Barquera, B.; Colón, W. Do prokaryotes have more kinetically stable proteins than eukaryotic organisms? *Biochemistry* **2010,** *49*, 7239-7241.

(103) Ciryam, P.; Antalek, M.; Cid, F.; Tartaglia, G. G.; Dobson, C. M.; Guettsches, A. K.; Eggers, B.; Vorgerd, M.; Marcus, K.; Kley, R. A., et al. A metastable subproteome underlies inclusion formation in muscle proteinopathies. *Acta Neuropathol. Commun.* **2019,** *7*.

(104) Kundra, R.; Ciryam, P.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M. Protein homeostasis of a metastable subproteome associated with Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114*, E5703-E5711.

(105) Noji, M.; Samejima, T.; Yamaguchi, K.; So, M.; Yuzu, K.; Chatani, E.; Akazawa-Ogawa, Y.; Hagihara, Y.; Kawata, Y.; Ikenaka, K., et al. Breakdown of supersaturation barrier links protein folding to amyloid formation. *Commun. Biol.* **2021,** *4*.

(106) Li, B.; Fooksa, M.; Heinze, S.; Meiler, J. Finding the needle in the haystack: Towards solving the protein-folding problem computationally. *Critical Reviews in Biochemistry and Molecular Biology* **2018,** *53*, 1-28.

(107) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998,** *282*, 740-744.

(108) Gershenson, A.; Gosavi, S.; Faccioli, P.; Wintrode, P. L. Successes and challenges in simulating the folding of large proteins. *J. Biol. Chem.* **2020,** *295*, 15-33.

(109) Li, W.; Terakawa, T.; Wang, W.; Takada, S. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 17789.

(110) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92*, 3626.

(111) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995,** *267*, 1619-1620.

(112) Trovato, F.; O'Brien, E. P. Fast protein translation can promote co- and posttranslational folding of misfolding-prone proteins. *Biophys. J.* **2017,** *112*, 1807-1819.

(113) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018,** *140*, 2386-2396.

(114) Konovalov, K. A.; Unarta, I. C.; Cao, S.; Goonetilleke, E. C.; Huang, X. Markov state models to study the functional dynamics of proteins in the wake of machine learning. *JACS Au* **2021,** *1*, 1330-1341.

(115) 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.

(116) Jackson, S. E. How do small single-domain proteins fold? *Folding and Design* **1998,** *3*, R81-R91.

(117) Dill, K. A. The stabilities of globular proteins. In *Protein eng.*, Oxender, D. L.; Fox, C. F., Eds.; Alan R. Liss, Inc: New York, 1987; pp 187-192.

(118) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **1995,** *21*, 167-95.

(119) Dill, K. A.; Chan, H. S. From levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997,** *4*, 10-19.

(120) Ellison, P. A.; Cavagnero, S. Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics. *Protein Sci.* **2006,** *15*, 564-582.

(121) Schönfelder, J.; Perez-Jimenez, R.; Muñoz, V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat. Commun.* **2016,** *7*, 11777.

(122) Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **2011,** *7*, 401-404.

(123) Gruebele, M.; Dave, K.; Sukenik, S. Globular protein folding in vitro and in vivo. *Annu. Rev. Biophys.* **2016,** *45*, 233-251.

(124) Becktel, W. J.; Schellman, J. A. Protein stability curves. *Biopolymers* **1987,** *26*, 1859-1877.

(125) Akmal, A.; Muñoz, V. The nature of the free energy barriers to two-state folding. *Proteins* **2004,** *57*, 142-152.

(126) Liu, Z.; Chan, H. S. Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol.* **2005,** *349*, 872-889.

(127) Lumry, R.; Eyring, H. Conformation changes of proteins. *J. Phys. Chem.* **1954,** *58*, 110-120.

(128) Zipp, A.; Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry* **1973,** *12*, 4217-4228.

(129) Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; García-Moreno E., B.; Royer, C. A. Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 6945-6950.

(130) Levine, I. N. *Physical chemistry*. 6th edition ed.; McGraw-Hill: Boston, 2009.

(131) Schindler, T.; Schmid, F. X. Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry* **1996,** *35*, 16833-42.

(132) Taddei, N.; Chiti, F.; Paoli, P.; Fiaschi, T.; Bucciantini, M.; Stefani, M.; Dobson, C. M.; Ramponi, G. Thermodynamics and kinetics of folding of common-type acylphosphatase: Comparison to the highly homologous muscle isoenzyme. *Biochemistry* **1999,** *38*, 2135-2142.

(133) Manyusa, S.; Whitford, D. Defining folding and unfolding reactions of apocytochrome b5 using equilibrium and kinetic fluorescence measurements. *Biochemistry* **1999,** *38*, 9533-9540.

(134) Pascher, T. Temperature and driving force dependence of the folding rate of reduced horse heart cytochrome c. *Biochemistry* **2001,** *40*, 5812-5820.

(135) Wolynes, P. G. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 6170-6175.

(136) Horvath, I.; Kumar, R.; Wittung-Stafshede, P. Macromolecular crowding modulates α-synuclein amyloid fiber growth. *Biophys. J.* **2021,** *120*, 3374-3381.

(137) Record, M. T.; Anderson, C. F.; Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* **1978,** *11*, 103-178.

(138) Record, M. T., Jr.; Zhang, W.; Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem.* **1998,** *51*, 281-353.

(139) Sukenik, S.; Sapir, L.; Gilman-Politi, R.; Harries, D. Diversity in the mechanisms of cosolute action on biomolecular processes. *Faraday Discuss.* **2013,** *160*, 225-237.

(140) Bianco, V.; Alonso-Navarro, M.; Di Silvio, D.; Moya, S.; Cortajarena, A. L.; Coluzza, I. Proteins are solitary! Pathways of protein folding and aggregation in protein mixtures. *J. Phys. Chem. Lett.* **2019,** *10*, 4800-4804.

(141) Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H. Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 8937.

(142) Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F. U.; Kerner, M. J.; Frishman, D. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **2008,** *9*, 102.

(143) Forchhammer, J.; Lindahl, L. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15. *J. Mol. Biol.* **1971,** *55*, 563-568.

(144) Kempf, N.; Remes, C.; Ledesch, R.; Züchner, T.; Höfig, H.; Ritter, I.; Katranidis, A.; Fitter, J. A novel method to evaluate ribosomal performance in cell-free protein synthesis systems. *Sci. Rep.* **2017,** *7*, 46753.

(145) Kuruma, Y.; Ueda, T. The pure system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* **2015,** *10*, 1328-1344.

(146) Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **2013,** *82*, 323-355.

(147) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **2016,** *353*, aac4354.

(148) Mogk, A.; Bukau, B.; Kampinga, H. H. Cellular handling of protein aggregates by disaggregation machines. *Mol. Cell* **2018,** *69*, 214-226.

(149) Doyle, S. M.; Genest, O.; Wickner, S. Protein rescue from aggregates by powerful molecular chaperone machines. *Nat. Rev. Mol. Cell Biol.* **2013,** *14*, 617.

(150) Waudby, C. A.; Dobson, C. M.; Christodoulou, J. Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* **2019,** *44*, 914-926.

(151) Lu, J.; Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2005,** *12*, 1123-1129.

(152) Kosolapov, A.; Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2009,** *16*, 405-411.

(153) Tu, L. W.; Deutsch, C. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J. Mol. Biol.* **2010,** *396*, 1346-1360.

(154) Tu, L.; Khanna, P.; Deutsch, C. Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *J. Mol. Biol.* **2014,** *426*, 185-198.

(155) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(156) Hutchinson, R. B.; Chen, X.; Zhou, N.; Cavagnero, S. Fluorescence anisotropy decays and microscale-volume viscometry reveal the compaction of ribosome-bound nascent proteins. *J. Phys. Chem. B* **2021,** *125*, 6543-6558.

(157) Hardesty, B.; Kramer, G. Folding of a nascent peptide on the ribosome. *Prog. Nucleic Acid Res. Mol. Biol.* **2001,** *66*, 41-66.

(158) Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci.* **2009,** *18*, 2003-2015.

(159) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Hartl, F. U. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Biol.* **1999,** *6*, 697-705.

(160) Evans, M. S.; Sander, I. M.; Clark, P. L. Cotranslational folding promotes β-helix formation and avoids aggregation in vivo. *J. Mol. Biol.* **2008,** *383*, 683-692.

(161) Kim, S. J.; Yoon, J. S.; Shishido, H.; Yang, Z.; Rooney, L. A.; Barral, J. M.; Skach, W. R. Translational tuning optimizes nascent protein folding in cells. *Science* **2015,** *348*, 444.

(162) 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.

(163) Bitran, A.; Jacobs, W. M.; Zhai, X.; Shakhnovich, E. Cotranslational folding allows misfolding-prone proteins to circumvent deep kinetic traps. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117*, 1485.

(164) Chaney, J. L.; Steele, A.; Carmichael, R.; Rodriguez, A.; Specht, A. T.; Ngo, K.; Li, J.; Emrich, S.; Clark, P. L. Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput. Biol.* **2017,** *13*, e1005531.

(165) Zhao, V.; Jacobs, W. M.; Shakhnovich, E. Effect of protein structure on evolution of cotranslational folding. *Biophys. J.* **2020,** *119*, 1123-1134.

(166) 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.

(167) Ziv, G.; Haran, G.; Thirumalai, D. Ribosome exit tunnel can entropically stabilize α-helices. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 18956-18961.

(168) 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*, 725-736.

(169) Cruz-Vera, L. R.; Rajagopal, S.; Squires, C.; Yanofsky, C. Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell* **2005,** *19*, 333-343.

(170) Houben , E. N. G.; Zarivach , R.; Oudega , B.; Luirink , J. Early encounters of a nascent membrane protein : Specificity and timing of contacts inside and outside the ribosome. *J. Cell Biol.* **2005,** *170*, 27-35.

(171) Ullers, R. S.; Houben, E. N.; Raine, A.; ten Hagen-Jongman, C. M.; Ehrenberg, M.; Brunner, J.; Oudega, B.; Harms, N.; Luirink, J. Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J. Cell Biol.* **2003,** *161*, 679-684.

(172) Peterson, J. H.; Woolhead, C. A.; Bernstein, H. D. The conformation of a nascent polypeptide inside the ribosome tunnel affects protein targeting and protein folding. *Mol. Microbiol.* **2010,** *78*, 203-217.

(173) Wang, S.; Jomaa, A.; Jaskolowski, M.; Yang, C.; Ban, N.; Shan, S. The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat. Struct. Mol. Biol.* **2019,** *26*, 919-929.

(174) Guzman-Luna, V.; Fuchs, A. M.; Allen, A. J.; Staikos, A.; Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Commun. Biol.* **2021,** *4*, 1236.

(175) Cassaignau, A. M. E.; Włodarski, T.; Chan, S. H. S.; Woodburn, L. F.; Bukvin, I. V.; Streit, J. O.; Cabrita, L. D.; Waudby, C. A.; Christodoulou, J. Interactions between nascent proteins and the ribosome surface inhibit co-translational folding. *Nat. Chem.* **2021**.

(176) Dao Duc, K.; Batra, S. S.; Bhattacharya, N.; Cate, Jamie H. D.; Song, Yun S. Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res.* **2019,** *47*, 4198-4210.

(177) Noeske, J.; Wasserman, M. R.; Terry, D. S.; Altman, R. B.; Blanchard, S. C.; Cate, J. H. D. High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.* **2015,** *22*, 336-341.

(178) Tsalkova, T.; Odom, O. W.; Kramer, G.; Hardesty, B. Different conformations of nascent peptides on ribosomes. *J. Mol. Biol.* **1998,** *278*, 713-723.

(179) Kramer, G.; Ramachandiran, V.; Hardesty, B. Cotranslational folding — omnia mea mecum porto? *Int. J. Biochem. Cell B.* **2001,** *33*, 541-553.

(180) Malkin, L. I.; Rich, A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* **1967,** *26*, 329-346.

(181) Nilsson, Ola B.; Hedman, R.; Marino, J.; Wickles, S.; Bischoff, L.; Johansson, M.; Müller-Lucks, A.; Trovato, F.; Puglisi, Joseph D.; O'Brien, Edward P., et al. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* **2015,** *12*, 1533-1540.

(182) Marino, J.; von Heijne, G.; Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **2016,** *590*, 655-660.

(183) Bhushan, S.; Gartmann, M.; Halic, M.; Armache, J.-P.; Jarasch, A.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. A-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2010,** *17*, 313-317.

(184) Lu, J.; Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **2005,** *44*, 8230-8243.

(185) Liutkute, M.; Maiti, M.; Samatova, E.; Enderlein, J.; Rodnina, M. V. Gradual compaction of the nascent peptide during cotranslational folding on the ribosome. *eLife* **2020,** *9*, e60895.

(186) Farías-Rico, J. A.; Ruud Selin, F.; 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.

(187) Tian, P.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; von Heijne, G.; Clarke, J., et al. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115*, E11284-E11293.

(188) Nilsson, O. B.; Nickson, A. A.; Hollins, J. J.; Wickles, S.; Steward, A.; Beckmann, R.; von Heijne, G.; Clarke, J. Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **2017,** *24*, 221-225.

(189) Mercier, E.; Rodnina, M. V. Co-translational folding trajectory of the HemK helical domain. *Biochemistry* **2018**.

(190) 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.

(191) Mounce, B. C.; Kurt, N.; Ellison, P. A.; Cavagnero, S. Nonrandom distribution of intramolecular contacts in native single-domain proteins. *Proteins* **2009,** *75*, 404-12.

(192) Kurt, N.; Cavagnero, S. The burial of solvent-accessible surface area is a predictor of polypeptide folding and misfolding as a function of chain elongation. *J. Am. Chem. Soc.* **2005,** *127*, 15690-15691.

(193) Frydman, J. Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu. Rev. Biochem.* **2001,** *70*, 603-647.

(194) Koubek, J.; Schmitt, J.; Galmozzi, C. V.; Kramer, G. Mechanisms of cotranslational protein maturation in bacteria. *Front. Mol. Biosci.* **2021,** *8*.

(195) Finkelstein, A. V.; Badretdinov, A. Y. Physical reason for fast folding of the stable spatial structure of proteins: A solution of the levinthal paradox. *Mol. Biol.* **1997,** *31*, 391-398.

(196) Finkelstein, A. V.; Badretdinov, A. Y. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold Des.* **1997,** *2*, 115-121.

(197) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. Chain length scaling of protein folding time. *Phys. Rev. Lett.* **1996,** *77*, 5433-5436.

(198) Thirumalai, D. From minimal models to real proteins - time scales for protein-folding kinetics. *J. Phys. I* **1995,** *5*, 1457-1467.

(199) Langer, T.; Lu, C.; Echols, H.; Flanagan, J.; Hayer, M. K.; Hartl, F. U. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **1992,** *356*, 683-9.

(200) Martin, J.; Langer, T.; Boteva, R.; Schramel, A.; Horwich, A. L.; Hartl, F. U. Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature* **1991,** *352*, 36.

(201) Deuerling, E.; Schulze-Specking, A.; Tomoyasu, T.; Mogk, A.; Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **1999,** *400*, 693-696.

(202) Hoffmann, A.; Bukau, B.; Kramer, G. Structure and function of the molecular chaperone trigger factor. *Biochim. Biophys. Acta Mol. Cell Res.* **2010,** *1803*, 650-661.

(203) Martinez-Hackert, E.; Hendrickson, W. A. Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* **2009,** *138*, 923-934.

(204) Valent, Q. A.; Kendall, D. A.; High, S.; Kusters, R.; Oudega, B.; Lurink, J. Early events in preprotein recognition in *e.Coli*: Interaction of SRP and trigger factor with nascent polypeptides. *EMBO J.* **1995,** *14*, 12.

(205) Schaffitzel, E.; Rudiger, S.; Bukau, B.; Deuerling, E. Functional dissection of trigger factor and DnaK: Interactions with nascent polypeptides and thermally denatured proteins. *Biol. Chem.* **2001,** *382*, 9.

(206) Bukau, B.; Deuerling, E.; Pfund, C.; Craig, E. A. Getting newly synthesized proteins into shape. *Cell* **2000,** *101*, 119-122.

(207) Maier, R.; Scholz, C.; Schmid, F. X. Dynamic association of trigger factor with protein substrates1. *J. Mol. Biol.* **2001,** *314*, 1181-1190.

(208) Rutkowska, A.; Mayer, M. P.; Hoffmann, A.; Merz, F.; Zachmann-Brand, B.; Schaffitzel, C.; Ban, N.; Deuerling, E.; Bukau, B. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **2008,** *283*, 4124-4132.

(209) Teter, S. A.; Houry, W. A.; Ang, D.; Tradler, T.; Rockabrand, D.; Fischer, G.; Blum, P.; Georgopoulos, C.; Hartl, F. U. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **1999,** *97*, 755-765.

(210) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. Recent advances in understanding catalysis of protein folding by molecular chaperones. *FEBS Lett.* **2020,** *594*, 2770-2781.

(211) Ewalt, K. L.; Hendrick, J. P.; Houry, W. A.; Hartl, F. U. In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* **1997,** *90*, 491-500.

(212) Houry, W. A.; Frishman, D.; Eckerskorn, C.; Lottspeich, F.; Hartl, F. U. Identification of in vivo substrates of the chaperonin GroEL. *Nature* **1999,** *402*, 147-154.

(213) Haslberger, T.; Zdanowicz, A.; Brand, I.; Kirstein, J.; Turgay, K.; Mogk, A.; Bukau, B. Protein disaggregation by the AAA+ chaperone ClpB involves partial threading of looped polypeptide segments. *Nat. Struct. Mol. Biol.* **2008,** *15*, 641-650.

(214) Baneyx, F.; Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **2004,** *22*, 1399-408.

(215) Rosenzweig, R.; Moradi, S.; Zarrine-Afsar, A.; Glover, J. R.; Kay, L. E. Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. *Science* **2013,** *339*, 1080-1083.

(216) Crooke, E.; Wickner, W. Trigger factor: A soluble protein that folds pro-OmpA into a membrane-assembly-competent form. *Proc. Natl. Acad. Sci. U.S.A.* **1987,** *84*, 5.

(217) Stoller, G.; Rucknagel, K. P.; Nierhaus, K. H.; Schmid, F. X.; Fischer, G.; Rahfeld, J.-U. A ribosome-associated peptidyl-prolyl *cis/trans* isomerse identified as the trigger factor. *EMBO J.* **1995,** *14*, 10.

(218) Lill, R.; Crooke, E.; Guthrie, B.; Wickner, W. The “trigger factor cycle” includes ribosomes, presecretory proteins, and the plasma membrane. *Cell* **1988,** *54*, 1013-1018.

(219) Patzelt, H.; Kramer, G.; Rauch, T.; Schönfeld, H.-J.; Bukau, B.; Deuerling, E. Three-state equilibrium of Escherichia coli trigger factor. **2002,** *383*, 1611-1619.

(220) Genevaux, P.; Keppel, F.; Schwager, F.; Langendijk-Genevaux, P. S.; Hartl, F. U.; Georgopoulos, C. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* **2004,** *5*, 195-200.

(221) Oh, E.; Becker, Annemarie H.; Sandikci, A.; Huber, D.; Chaba, R.; Gloge, F.; Nichols, Robert J.; Typas, A.; Gross, Carol A.; Kramer, G., et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **2011,** *147*, 1295-1308.

(222) Kramer, G.; Rutkowska, A.; Wegrzyn, R. D.; Patzelt, H.; Kurz, T. A.; Merz, F.; Rauch, T.; Vorderwulbecke, S.; Deuerling, E.; Bukau, B. Functional dissection of Escherichia coli trigger factor: Unraveling the function of individual domains. *J. Bacteriol.* **2004,** *186*, 3777-3784.

(223) Kramer, G.; Patzelt, H.; Rauch, T.; Kurz, T. A.; Vorderwulbecke, S.; Bukau, B.; Deuerling, E. Trigger factor peptidyl-prolyl cis/trans isomerase activity is not essential for the folding of cytosolic proteins in Escherichia coli. *J. Biol. Chem.* **2004,** *279*, 14165-14170.

(224) Huang, G.-C.; Li, Z.-Y.; Zhou, J.-M.; Fischer, G. Assisted folding of d-glyceraldehyde-3-phosphate dehydrogenase by trigger factor. *Protein Sci.* **2000,** *9*, 1254-1261.

(225) Agashe, V. R.; Guha, S.; Chang, H. C.; Genevaux, P.; Hayer-Hartl, M.; Stemp, M.; Georgopoulos, C.; Hartl, F. U.; Barral, J. M. Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. *Cell* **2004,** *117*, 199-209.

(226) Saio, T.; Guan, X.; Rossi, P.; Economou, A.; Kalodimos, C. G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **2014,** *344*, 1250494.

(227) Nilsson, O. B.; Muller-Lucks, A.; Kramer, G.; Bukau, B.; von Heijne, G. Trigger factor reduces the force exerted on the nascent chain by a cotranslationally folding protein. *J. Mol. Biol.* **2016,** *428*, 1356-1364.

(228) Hoffmann, A.; Merz, F.; Rutkowska, A.; Zachmann-Brand, B.; Deuerling, E.; Bukau, B. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.* **2006,** *281*, 6539-6545.

(229) Ferbitz, L.; Maier, T.; Patzelt, H.; Bukau, B.; Deuerling, E.; Ban, N. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **2004,** *431*, 590-596.

(230) Merz, F.; Boehringer, D.; Schaffitzel, C.; Preissler, S.; Hoffmann, A.; Maier, T.; Rutkowska, A.; Lozza, J.; Ban, N.; Bukau, B., et al. Molecular mechanism and structure of trigger factor bound to the translating ribosome. *EMBO J.* **2008,** *27*, 1622.

(231) Baram, D.; Pyetan, E.; Sittner, A.; Auerbach-Nevo, T.; Bashan, A.; Yonath, A. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 12017-12022.

(232) Zarnt, T.; Tradler, T.; Stoller, G.; Scholz, C.; Schmid, F. X.; Fischer, G. Modular structure of the trigger factor required for high activity in protein folding. *J. Mol. Biol.* **1997,** *271*, 827-837.

(233) Kramer, G.; Rauch, T.; Rist, W.; Vorderwülbecke, S.; Patzelt, H.; Schulze-Specking, A.; Ban, N.; Deuerling, E.; Bukau, B. L23 protein functions as a chaperone docking site on the ribosome. *Nature* **2002,** *419*, 171-174.

(234) Schlünzen, F.; Wilson, D. N.; Tian, P.; Harms, J. M.; McInnes, S. J.; Hansen, H. A. S.; Albrecht, R.; Buerger, J.; Wilbanks, S. M.; Fucini, P. The binding mode of the trigger factor on the ribosome: Implications for protein folding and SRP interaction. *Structure* **2005,** *13*, 1685-1694.

(235) Hesterkamp, T.; Hauser, S.; Lutcke, H.; Bukau, B. *Escherichia coli* trigger factoris a prolyl isomerase that associates with nascent polypeptide chains. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93*, 5.

(236) Merz, F.; Hoffmann, A.; Rutkowska, A.; Zachmann-Brand, B.; Bukau, B.; Deuerling, E. The C-terminal domain of Escherichia coli trigger factor represents the central module of its chaperone activity. *J. Biol. Chem.* **2006,** *281*, 31963-31971.

(237) Zeng, L.-L.; Yu, L.; Li, Z.-Y.; Perrett, S.; Zhou, J.-M. Effect of C-terminal truncation on the molecular chaperone function and dimerization of Escherichia coli trigger factor. *Biochimie* **2006,** *88*, 613-619.

(238) Kaiser, C. M.; Chang, H. C.; Agashe, V. R.; Lakshmipathy, S. K.; Etchells, S. A.; Hayer-Hartl, M.; Hartl, F. U.; Barral, J. M. Real-time observation of trigger factor function on translating ribosomes. *Nature* **2006,** *444*, 455-460.

(239) Patzelt, H.; Rudiger, S.; Brehmer, D.; Kramer, G.; Vorderwulbecke, S.; Schaffitzel, E.; Waitz, A.; Hesterkamp, T.; Dong, L.; Schneider-Mergener, J., et al. Binding specificity of Escherichia coli trigger factor. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 14244-9.

(240) Lakshmipathy, S. K.; Tomic, S.; Kaiser, C. M.; Chang, H. C.; Genevaux, P.; Georgopoulos, C.; Barral, J. M.; Johnson, A. E.; Hartl, F. U.; Etchells, S. A. Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* **2007,** *282*, 12186-12193.

(241) Raine, A.; Lovmar, M.; Wikberg, J.; Ehrenberg, M. Trigger factor binding to ribosomes with nascent peptide chains of varying lengths and sequences. *J. Biol. Chem.* **2006,** *281*, 28033-8.

(242) Deeng, J.; Chan, K. Y.; van der Sluis, E. O.; Berninghausen, O.; Han, W.; Gumbart, J.; Schulten, K.; Beatrix, B.; Beckmann, R. Dynamic behavior of trigger factor on the ribosome. *J. Mol. Biol.* **2016,** *428*, 3588-3602.

(243) Hoffmann, A.; Becker, A. H.; Zachmann-Brand, B.; Deuerling, E.; Bukau, B.; Kramer, G. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* **2012,** *48*, 63-74.

(244) Mashaghi, A.; Kramer, G.; Bechtluft, P.; Zachmann-Brand, B.; Driessen, A. J. M.; Bukau, B.; Tans, S. J. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **2013,** *500*, 98.

(245) Maier, R.; Eckert, B.; Scholz, C.; Lilie, H.; Schmid, F.-X. Interaction of trigger factor with the ribosome. *J. Mol. Biol.* **2003,** *326*, 585-592.

(246) Scholz, C.; Stoller, G.; Zarnt, T.; Fischer, G.; Schmid, F. X. Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO J.* **1997,** *16*, 54-58.

(247) Lakshmipathy, S. K.; Gupta, R.; Pinkert, S.; Etchells, S. A.; Hartl, F. U. Versatility of trigger factor interactions with ribosome-nascent chain complexes. *J. Biol. Chem.* **2010,** *285*, 27911-27923.

(248) Ritossa, F. A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia* **1962,** *18*, 571-573.

(249) Ritossa, F. New puffs induced by temperature shock, DNP and salicylate in salivary chromosomes of drosophila melanogaster. *Drosoph. Inf. Serv.* **1963,** *37*, 122-123.

(250) Ritossa, F. Experimental activation of specific loci in polytene chromosomes of drosophila. *Exp. Cell Res.* **1964,** *35*, 601-607.

(251) Ritossa, F. Discovery of the heat shock response. *Cell Stress Chaperones* **1996,** *1*, 97.

(252) Craig, E. A.; Gross, C. A. Is Hsp70 the cellular thermometer? *Trends Biochem. Sci.* **1991,** *16*, 135-140.

(253) Tissiéres, A.; Mitchell, H. K.; Tracy, U. M. Protein synthesis in salivary glands of drosophila melanogaster: Relation to chromosome puffs. *J. Mol. Biol.* **1974,** *84*, 389-398.

(254) Lindquist, S.; Craig, E. A. The heat-shock proteins. *Annu. Rev. Genet.* **1988,** *22*, 631-677.

(255) Kampinga, H. H.; Hageman, J.; Vos, M. J.; Kubota, H.; Tanguay, R. M.; Bruford, E. A.; Cheetham, M. E.; Chen, B.; Hightower, L. E. Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress and Chaperones* **2009,** *14*, 105-111.

(256) Goloubinoff, P.; De Los Rios, P. The mechanism of Hsp70 chaperones:(entropic) pulling the models together. *Trends Biochem. Sci.* **2007,** *32*, 372-380.

(257) Mayer, M. P.; Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* **2005,** *62*, 670.

(258) Laksanalamai, P.; Whitehead, T. A.; Robb, F. T. Minimal protein-folding systems in hyperthermophilic archaea. *Nat. Rev. Microbiol.* **2004,** *2*, 315-324.

(259) Rebeaud, M. E.; Mallik, S.; Goloubinoff, P.; Tawfik, D. S. On the evolution of chaperones and cochaperones and the expansion of proteomes across the tree of life. *Proc. Natl. Acad. Sci. U.S.A.* **2021,** *118*, e2020885118.

(260) Macario, A. J.; Lange, M.; Ahring, B. K.; De Macario, E. C. Stress genes and proteins in the archaea. *Microbiology and Molecular Biology Reviews* **1999,** *63*, 923-967.

(261) Gribaldo, S.; Lumia, V.; Creti, R.; Conway de Macario, E.; Sanangelantoni, A.; Cammarano, P. Discontinuous occurrence of the Hsp70 (DnaK) gene among archaea and sequence features of Hsp70 suggest a novel outlook on phylogenies inferred from this protein. *J. Bacteriol.* **1999,** *181*, 434-443.

(262) Warnecke, T. Loss of the DnaK-DnaJ-GrpE chaperone system among the aquificales. *Molecular biology and evolution* **2012,** *29*, 3485-3495.

(263) Calloni, G.; Chen, T.; Schermann, S. M.; Chang, H. C.; Genevaux, P.; Agostini, F.; Tartaglia, G. G.; Hayer-Hartl, M.; Hartl, F. U. DnaK functions as a central hub in the E. coli chaperone network. *Cell Reports* **2012,** *1*, 251-64.

(264) Hesterkamp, T.; Bukau, B. Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in e.Coli. *EMBO J.* **1998,** *17*, 4818-4828.

(265) Zimmerman, S. B.; Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **1991,** *222*, 599-620.

(266) Hartl, F. U.; Hayer-Hartl, M. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **2002,** *295*, 8.

(267) Mogk, A.; Tomoyasu, T.; Goloubinoff, P.; Rüdiger, S.; Röder, D.; Langen, H.; Bukau, B. Identification of thermolabile Escherichia coli proteins: Prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* **1999,** *18*, 6934-6949.

(268) Hartl, F. U.; Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **2009,** *16*, 574-81.

(269) Zylicz, M.; Ang, D.; Liberek, K.; Georgopoulos, C. Initiation of lambda DNA replication with purified host-and bacteriophage-encoded proteins: The role of the DnaK, DnaJ and GrpE heat shock proteins. *EMBO J.* **1989,** *8*, 1601-1608.

(270) Meimaridou, E.; Gooljar, S. B.; Chapple, J. P. From hatching to dispatching: The multiple cellular roles of the Hsp70 molecular chaperone machinery. *Journal of molecular endocrinology* **2009,** *42*, 1-9.

(271) Peschke, M.; Le Goff, M.; Koningstein, G. M.; Karyolaimos, A.; de Gier, J.-W.; van Ulsen, P.; Luirink, J. Srp, FtsY, DnaK and YidC are required for the biogenesis of the E. coli tail-anchored membrane proteins DjlC and Flk. *J. Mol. Biol.* **2018,** *430*, 389-403.

(272) Mogk, A.; Bukau, B. Molecular chaperones: Structure of a protein disaggregase. *Curr. Biol.* **2004,** *14*, R78-80.

(273) Aguado, A.; Fernandez-Higuero, J. A.; Moro, F.; Muga, A. Chaperone-assisted protein aggregate reactivation: Different solutions for the same problem. *Arch Biochem Biophys* **2015,** *580*, 121-34.

(274) Diamant, S.; Ben-Zvi, A. P.; Bukau, B.; Goloubinoff, P. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem.* **2000,** *275*, 21107-21113.

(275) Zolkiewski, M. ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation: A novel multi-chaperone system from Escherichia coli. *J. Biol. Chem.* **1999,** *274*, 28083-28086.

(276) Rudiger, S.; Schneider-Mergener, J.; Bukau, B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J.* **2001,** *20*, 1042-1050.

(277) Rüdiger, S.; Germeroth, L.; Schneider-Mergener, J.; Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **1997,** *16*, 1501-1507.

(278) Goloubinoff, P.; Mogk, A.; Zvi, A. P. B.; Tomoyasu, T.; Bukau, B. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 13732-13737.

(279) Veinger, L.; Diamant, S.; Buchner, J.; Goloubinoff, P. The small heat-shock protein ibpb from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J. Biol. Chem.* **1998,** *273*, 11032-11037.

(280) Żwirowski, S.; Kłosowska, A.; Obuchowski, I.; Nillegoda, N. B.; Piróg, A.; Ziętkiewicz, S.; Bukau, B.; Mogk, A.; Liberek, K. Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *EMBO J.* **2017,** *36*, 783-796.

(281) Aguilar-Rodríguez, J.; Sabater-Muñoz, B.; Montagud-Martínez, R.; Berlanga, V.; Alvarez-Ponce, D.; Wagner, A.; Fares, M. A. The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol. Evol.* **2016,** *8*, 2979-2991.

(282) Kadibalban, A. S.; Bogumil, D.; Landan, G.; Dagan, T. DnaK-dependent accelerated evolutionary rate in prokaryotes. *Genome Biol. Evol.* **2016,** *8*, 1590-1599.

(283) Rutherford, S. L. Between genotype and phenotype: Protein chaperones and evolvability. *Nat. Rev. Genet.* **2003,** *4*, 263-274.

(284) Victor, M. P.; Acharya, D.; Chakraborty, S.; Ghosh, T. C. Chaperone client proteins evolve slower than non-client proteins. *Funct. Integr. Genomics* **2020,** *20*, 621-631.

(285) Brodsky, J. L.; Chiosis, G. Hsp70 molecular chaperones: Emerging roles in human disease and identification of small molecule modulators. *Curr. Top. Med. Chem.* **2006,** *6*, 1215-1225.

(286) Monsellier, E.; Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **2007,** *8*, 737-742.

(287) Bertelsen, E. B.; Chang, L.; Gestwicki, J. E.; Zuiderweg, E. R. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 8471-6.

(288) Sondermann, H.; Scheufler, C.; Schneider, C.; Höhfeld, J.; Hartl, F.-U.; Moarefi, I. Structure of a Bag/Hsc70 complex: Convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* **2001,** *291*, 1553-1557.

(289) Harrison, C. J.; Hayer-Hartl, M.; Di Liberto, M.; Hartl, F.-U.; Kuriyan, J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **1997,** *276*, 431-435.

(290) Kityk, R.; Kopp, J.; Sinning, I.; Mayer, M. P. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell* **2012,** *48*, 863-74.

(291) Buchberger, A.; Theyssen, H.; Schröder, H.; McCarty, J. S.; Virgallita, G.; Milkereit, P.; Reinstein, J.; Bukau, B. Nucleotide-induced conformational changes in the ATPase and substrate binding domains of the DnaK chaperone provide evidence for interdomain communication. *J. Biol. Chem.* **1995,** *270*, 16903-16910.

(292) Zhu, X.; Zhao, X.; Burkholder, W. F.; Gragerov, A.; Ogata, C. M.; Gottesman, M. E.; Hendrickson, W. A. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **1996,** *272*, 1606-1614.

(293) Liu, Q. L.; Liang, C.; Zhou, L. Structural and functional analysis of the Hsp70/Hsp40 chaperone system. *Protein Sci.* **2020,** *29*, 378-390.

(294) Qi, R.; Sarbeng, E. B.; Liu, Q.; Le, K. Q.; Xu, X.; Xu, H.; Yang, J.; Wong, J. L.; Vorvis, C.; Hendrickson, W. A., et al. Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. *Nat. Struct. Mol. Biol.* **2013,** *20*, 900-907.

(295) Vega, C. A.; Kurt, N.; Chen, Z.; Rüdiger, S.; Cavagnero, S. Binding specificity of an α-helical protein sequence to a full-length Hsp70 chaperone and its minimal substrate-binding domain. *Biochemistry* **2006,** *45*, 13835-13846.

(296) Mayer, M. P.; Rudiger, S.; Bukau, B. Molecular basis for interactions of the DnaK chaperone with substrates. *Biol. Chem.* **2000,** *381*, 9.

(297) Schmid, D.; Baici, A.; Gehring, H.; Christen, P. Kinetics of molecular chaperone action. *Science* **1994,** *263*, 971-973.

(298) Mayer, M. P.; Schröder, H.; Rüdiger, S.; Paal, K.; Laufen, T.; Bukau, B. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **2000,** *7*, 586-593.

(299) McCarty, J. S.; Buchberger, A.; Reinstein, J.; Bukau, B. The role of ATP in the functional cycle of the DnaK chaperone system. *J. Mol. Biol.* **1995,** *249*, 126-37.

(300) Clerico, E. M.; Tilitsky, J. M.; Meng, W.; Gierasch, L. M. How Hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.* **2015,** *427*, 1575-88.

(301) Laufen, T.; Mayer, M. P.; Beisel, C.; Klostermeier, D.; Mogk, A.; Reinstein, J.; Bukau, B. Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 5452-5457.

(302) Mayer, M. P. The Hsp70-chaperone machines in bacteria. *Front. Mol. Biosci.* **2021,** *8*.

(303) Barends, T. R.; Brosi, R. W.; Steinmetz, A.; Scherer, A.; Hartmann, E.; Eschenbach, J.; Lorenz, T.; Seidel, R.; Shoeman, R. L.; Zimmermann, S., et al. Combining crystallography and epr: Crystal and solution structures of the multidomain cochaperone DnaJ. *Acta Crystallogr. D Biol. Crystallogr.* **2013,** *69*, 1540-52.

(304) Vogel, M.; Mayer, M. P.; Bukau, B. Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. *J. Biol. Chem.* **2006,** *281*, 38705-11.

(305) Swain, J. F.; Dinler, G.; Sivendran, R.; Montgomery, D. L.; Stotz, M.; Gierasch, L. M. Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol. Cell* **2007,** *26*, 27-39.

(306) Gething, M.-J. Molecular chaperones: Clasping the prize. *Curr. Biol.* **1996,** *6*, 1573-1576.

(307) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 å resolution. *J. Mol. Biol.* **1986,** *192*, 133-154.

(308) Buchberger, A.; Schröder, H.; Büttner, M.; Valencia, A.; Bukau, B. A conserved loop in the ATPase domain of the DnaK chaperone is essential for stable binding of GrpE. *Nat. Struct. Biol.* **1994,** *1*, 95-101.

(309) Schönfeld, H.-J.; Schmidt, D.; Schröder, H.; Bukau, B. The DnaK chaperone system of Escherichia coli: Quaternary structures and interactions of the DnaK and GrpE components. *J. Biol. Chem.* **1995,** *270*, 2183-2189.

(310) Lai, A. L.; Clerico, E. M.; Blackburn, M. E.; Patel, N. A.; Robinson, C. V.; Borbat, P. P.; Freed, J. H.; Gierasch, L. M. Key features of an Hsp70 chaperone allosteric landscape revealed by ion-mobility native mass spectrometry and double electron-electron resonance. *J. Biol. Chem.* **2017,** *292*, 8773-8785.

(311) Landry, S. J.; Jordan, R.; McMacken, R.; Gierasch, L. M. Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature* **1992,** *355*, 455.

(312) Popp, S.; Packschies, L.; Radzwill, N.; Vogel, K. P.; Steinhoff, H.-J.; Reinstein, J. Structural dynamics of the DnaK–peptide complex. *J. Mol. Biol.* **2005,** *347*, 1039-1052.

(313) Chen, Z.; Kurt, N.; Rajagopalan, S.; Cavagnero, S. Secondary structure mapping of DnaK-bound protein fragments: Chain helicity and local helix unwinding at the binding site. *Biochemistry* **2006,** *45*, 12325-12333.

(314) Kurt, N.; Cavagnero, S. Nonnative helical motif in a chaperone-bound protein fragment. *Biophys. J.* **2008,** *94*, L48-L50.

(315) Kellner, R.; Hofmann, H.; Barducci, A.; Wunderlich, B.; Nettels, D.; Schuler, B. Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, 13355-13360.

(316) Lee, J. H.; Zhang, D.; Hughes, C.; Okuno, Y.; Sekhar, A.; Cavagnero, S. Heterogeneous binding of the SH3 client protein to the DnaK molecular chaperone. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, E4206-E4215.

(317) Sekhar, A.; Santiago, M.; Lam, H. N.; Lee, J. H.; Cavagnero, S. Transient interactions of a slow-folding protein with the Hsp70 chaperone machinery. *Protein Sci.* **2012,** *21*, 1042-1055.

(318) Sekhar, A.; Lam, H. N.; Cavagnero, S. Protein folding rates and thermodynamic stability are key determinants for interaction with the Hsp70 chaperone system. *Protein Sci.* **2012,** *21*, 1489-1502.

(319) Cohen, S. I.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. From macroscopic measurements to microscopic mechanisms of protein aggregation. *J. Mol. Biol.* **2012,** *421*, 160-171.

(320) Ciryam, P.; Kundra, R.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M. Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. *Trends Pharmacol. Sci.* **2015,** *36*, 72-77.

(321) Hayashi, H.; Kimura, N.; Yamaguchi, H.; Hasegawa, K.; Yokoseki, T.; Shibata, M.; Yamamoto, N.; Michikawa, M.; Yoshikawa, Y.; Terao, K. A seed for Alzheimer amyloid in the brain. *J. Neurosci.* **2004,** *24*, 4894-4902.

(322) Jewett, A.; Shea, J.-E. Folding on the chaperone: Yield enhancement through loose binding. *J. Mol. Biol.* **2006,** *363*, 945-957.

(323) Jakob, U.; Gaestel, M.; Engel, K.; Buchner, J. Small heat shock proteins are molecular chaperones. *J. Biol. Chem.* **1993,** *268*, 1517-1520.

(324) Buchner, J.; Schmidt, M.; Fuchs, M.; Jaenicke, R.; Rudolph, R.; Schmid, F. X.; Kiefhaber, T. Groe facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **1991,** *30*, 6.

(325) Cry, D. M. Cooperation of the molecular chaperone Ydj1 with specific Hsp70 homologs to suppress protein aggregation. *FEBS Lett.* **1995,** *359*, 129-132.

(326) Mashaghi, A.; Bezrukavnikov, S.; Minde, D. P.; Wentink, A. S.; Kityk, R.; Zachmann-Brand, B.; Mayer, M. P.; Kramer, G.; Bukau, B.; Tans, S. J. Alternative modes of client binding enable functional plasticity of Hsp70. *Nature* **2016,** *539*, 448-451.

(327) Mecha, M. F.; Macchi, J. K.; Winkler, G. L. W.; Yang, H.; Chen, X.; Hutchinson, R. B.; Cavagnero, S. In praise of imperfection: The Hsp70 chaperone prevents aggregation of a predictable class of proteins over a limited time span. *In preparation*.

(328) Sharma, S. K.; De Los Rios, P.; Christen, P.; Lustig, A.; Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **2010,** *6*, 914-920.

(329) Noguchi, A.; Ikeda, A.; Mezaki, M.; Fukumori, Y.; Kanemori, M. DnaJ-promoted binding of DnaK to multiple sites on σ32 in the presence of ATP. *J. Bacteriol.* **2014,** *196*, 1694-1703.

(330) Sekhar, A.; Nagesh, J.; Rosenzweig, R.; Kay, L. E. Conformational heterogeneity in the Hsp70 chaperone-substrate ensemble identified from analysis of NMR-detected titration data. *Protein Sci.* **2017,** *26*, 2207-2220.

(331) Rosenzweig, R.; Sekhar, A.; Nagesh, J.; Kay, L. E. Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *eLife* **2017,** *6*, e28030.

(332) Clerico, E. M.; Meng, W. L.; Pozhidaeva, A.; Bhasne, K.; Petridis, C.; Gierasch, L. M. Hsp70 molecular chaperones: Multifunctional allosteric holding and unfolding machines. *Biochem. J.* **2019,** *476*, 1653-1677.

(333) Rosenzweig, R.; Nillegoda, N. B.; Mayer, M. P.; Bukau, B. The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* **2019,** *20*, 665-680.

1.24 Appendix

Supporting Tables

Table 1-S1. Protein length (number of amino acids) and folding rate constant (k_f) values of two-state folders plotted in Figure 1-3A.

Name	PDB	Length	ln k_f	Reference
C-terminal β-hairpin of protein GB1	1PGB	16	12.0	[1, 2]
Trp-cage protein	1L2Y	20	12.5	[1, 3]
Alanine-based peptide	NA	21	15.5	[1, 4]
BBA5 mini-protein	1T8J	23	11.8	[1, 5]
Pin WW domain	1PIN	34	9.5	[1, 6]
Villin headpeace subdomain	1VII	36	9.4	[1, 7]
Formin-binding protein	1E0L	37	10.6	[1, 8]
Prototype WW domain	1E0M	38	8.9	[1, 8]
Yes kinase-associated protein	1JMQ	40	8.4	[1, 8]
Peripheral subunit-binding domain	2PDD	41	9.8	[1, 9]
E3-binding domain of BBL	2WXC	47	11.2	[1]
GA module of albumin binding domain	1PRB	47	13.8	[1, 10]
POB	1W4J	51	12.3	[1]
TRF1 Myb domain	1BA5	53	5.9	[1, 11]
c-Myb-transforming protein	1GV2	55	8.7	[1, 12]
N-terminal domain from ribosomal	1DIV	56	6.6	[1, 13]

protein L9				
B-domain of staphylococcal proteinA	1BDD	58	11.7	[1, 14]
RAP1 Myb domain	1FEX	59	8.2	[1, 11]
B1 domain of streptococcal protein G	3GB1	62	6.3	[1]
Src SH3 domain	1RLQ	62	4.4	[1, 13]
α-spectrin SH3 domain	1SHG	62	1.1	[1, 13]
B1 domain of streptococcal protein G	3GB1	62	6.3	[13]
Sso7d protein, Y34W	1BNZ	64	7.0	[1]
Sso7d protein, Y34W	1C8C	64	6.95	[15]
Chymotrypsin inhibitor 2	2CI2	65	5.8	[1, 13]
Cold shock protein B	1C9O	66	7.2	[1, 16]
Cold shock protein B	1G6P	66	6.3	[1, 16]
Cold shock protein B	1CSP	67	6.5	[1, 16]
LysM domain	1E0G	66	7.0	[1]
Actin binding protein ABP1 SH3 domain	1JO8	68	2.5	[1, 13]
Photosystem I accessory protein E	1PSF	69	3.2	[1]
Cold shock protein A	1MJC	70	5.3	[1, 17]
Immunoglobulin light chain-binding domain of protein L	1HZ5	72	4.1	[1]
α3D	2A3D	73	12.2	[1, 18]
Sho1 SH3 domain	2VKN	76	2.1	[1, 13]
Ubiquitin	1UBQ	76	7.3	[1, 13]

Fyn SH3 domain	1AVZ	78	4.9	[1, 13]
Immunoglobulin light chain-binding domain of protein L	2PTL	79	4.1	[13]
Ras-binding domain of C-raf-1	1RFA	80	8.4	[1, 13]
l-repressor	1LMB	80	10.4	[13]
Activation domain of procarboxypeptidase A2	1O6X	81	6.8	[1, 13]
l-repressor	1LMB	81	10.4	[1]
Histidine-containing phosphocarrier protein	1POH	85	2.7	[1, 19]
SH3-like domain of virulence protein internalin B	1M9S	85	4.0	[1, 13]
Acyl-coenzyme A binding protein	1NTI	86	7.0	[1, 13]
CAfn2 (B. circulans)	1K85	88	1.4	[1]
N-domain of spore coat protein S	1PRS	88	3.0	[1]
Hypothetical protein encoded by the Yjbj gene from E.coli	1JYG	89	9.1	[1, 13]
Ninth fibronectin type III module of fibronectin	1FNF	90	-0.9	[1, 20]
SH3 domain of the p85α subunit of phosphatidylinositol 3'-kinase	1PNJ	90	-1.0	[1, 21]
Third fibronectin type III repeat of tenascin	1TEN	90	1.1	[22]

C-domain of spore coat protein S	1PRS	91	-2.0	[1]
C-terminal domain from ribosomal protein L9	1DIV	92	3.3	[1, 13]
Third fibronectin type III repeat of tenascin	1TEN	92	1.1	[1]
18th module of muscle protein twitchin	1WIT	93	0.4	[1]
Colicin E9 immunity protein	1IMQ	93	7.3	[1, 13]
hbLBD	1K8M	93	-0.7	[1]
18th module of muscle protein twitchin	1WIT	93	0.4	[23]
Colicin E7 immunity protein	1AYI	94	7.2	[1, 13]
Ribosomal protein L23	1N88	96	2.0	[1, 13]
Common-type acylphosphatase	2VH7	98	0.8	[1]
Common-type acylphosphatase	2ACY	98	0.8	[24]
Muscle acylphosphatase	1APS	99	-1.6	[1, 13]
Death domain	1E41	100	6.9	[1]
Ribosomal protein S6	1RIS	101	6.1	[1, 13]
Ubiquitin related modifyer 1	2PKO	101	2.6	[1, 13]
Spliceosomal protein U1A	1AUD	102	4.6	[1, 13]
apocytochrome b5	1HKO	104	3.0	[1]
Cytochrome b562	256B	106	12.3	[1, 25]
FK506 binding protein	1FKF	109	1.6	[1, 13]
Src SH2 domain	1IS0	110	8.7	[1, 13]
P13	1QTU	117	-0.4	[1]

15th domain of brain α-spectrin	1U5P	118	11.0	[1, 26]
16th domain of brain α-spectrin	1CUN	118	4.8	[1, 26]
17th domain of brain α-spectrin	1CUN	118	3.4	[1, 26]
Hypothetical protein Tm1083	1J5U	124	6.9	[1, 13]
Chemotaxis protein CheW	1K0S	151	7.4	[1, 13]
Cyclophilin A	1LOP	164	6.6	[1, 27]
Apoflavodoxin (Anabaena sp.)	1RCF	169	0.8	[1]
Lyme desease variable surface antigen	1L8W	338	2.0	[1, 13]

Table 1-S2. Protein chain length and folding rate constant (k_f) values of multi-state folders of Figure 1-3B.

Name	PDB	Length	ln k_f	Reference
Engrailed homeodomain	1ENH	61	10.5	[1, 11]
Phage 434 cro protein	2CRO	71	3.7	[1, 28]
FF domain	1UZC	71	7.7	[1]
ACBP (Yeast)	1ST7	86	8.5	[1]
Immunoglobulin domain of cardiac titin	1TIT	89	3.6	[1, 29]
Barstar	1BRS	89	3.4	[1, 30]
N-terminal domain of HypF	1GXT	91	4.4	[1, 31]
Third PDZ domain from PSD-95	1TP3	93	3.0	[1]
Tenth fibronectin type III domain of fibronectin	1FNF	94	5.5	[1, 32]
PDZ2 domain from PTP-BL	1GM1	94	1.0	[1]
FRB	1AUE	95	6.0	[1]
C-terminal domain of the cell-surface receptor protein CD2	1HNG	98	1.8	[1, 33]
Barnase	1BNI	110	2.6	[1, 34]
Cell-cycle regulatory protein p13suc1	SUC1	113	4.2	[1, 35]
Hisactophilin	1HCD	118	4.6	[1]
Myotrophin	2MYO	118	4.8	[1]
Villin 14T, N-terminal domain of villin	2VIK	126	5	[1, 36]

Ileal lipid binding protein	1EAL	127	1.3	[1, 37]
Chemotactic protein	3CHY	129	1	[1, 38]
Intestinal fatty acid binding protein	1IFC	131	3.4	[1, 39]
Cellular retinol binding protein II	1OPA	134	1.4	[1, 39]
Cellular retinoic acid binding protein I	1CBI	138	-3.2	[1, 39]
RNase-H (C. tepidum)	3H08	146	1.9	[1]
Tumour suppressor protein p16	2A5E	148	3.5	[1, 40]
Apoflavodoxin (D. desulfuricans)	3F6R	148	3.5	[1]
Staphylococcal nuclease	1SNQ	149	2.3	[1]
Apomyoglobin	1A6N	151	1.1	[1, 41]
C-terminally truncated fragment of GroEL apical domain	1AON	155	-1.5	[1, 42]
Ribonuclease HI	2RN2	155	0.1	[1, 43]
Dihydrofolate reductase	1RA9	159	-3.2	[1, 44]
T4 lysozyme	2LZM	164	4.1	[1, 43]
p19INK4d	1BI8	166	2.2	[1]
N-terminal domain of phosphoglycerate kinase	1PHP	175	2.3	[1, 43]
C-terminal domain of phosphoglycerate kinase	1PHP	221	-3.5	[1, 45]
Sigps	1IGS	222	-7.8	[1]
Carbonic anhydrase	5A25	260	-4.2	[1]
Tryptophan synthase α subunit	1QOP	268	-2.5	[1, 46]

Tryptophan synthase β2 subunit	1QOP	396	-6.9	[1, 47]

Supporting References

(1) Corrales, M.; Cuscó, P.; Usmanova, D. R.; Chen, H.-C.; Bogatyreva, N. S.; Filion, G. J.; Ivankov, D. N. Machine learning: How much does it tell about protein folding rates? *PLoS ONE* **2015,** *10*, e0143166.

(2) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **1997,** *390*, 196-199.

(3) Qiu, L. L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.* **2002,** *124*, 12952-12953.

(4) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. Laser temperature jump study of the helix⇌coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **1997,** *36*, 9200-9210.

(5) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002,** *420*, 102-106.

(6) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The folding mechanism of a β-sheet: The WW domain. *J. Mol. Biol.* **2001,** *311*, 373-393.

(7) Wang, M. H.; Tang, Y. F.; Sato, S. S.; Vugmeyster, L.; McKnight, C. J.; Raleigh, D. P. Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. *J. Am. Chem. Soc.* **2003,** *125*, 6032-6033.

(8) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci. USA* **2001,** *98*, 13002-13007.

(9) Spector, S.; Raleigh, D. P. Submillisecond folding of the peripheral subunit-binding domain. *J. Mol. Biol.* **1999,** *293*, 763-768.

(10) Wang, T.; Zhu, Y. J.; Gai, F. Folding of a three-helix bundle at the folding speed limit. *J. Phys. Chem. B* **2004,** *108*, 3694-3697.

(11) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W. N.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 13286-13291.

(12) Gianni, S.; Ivarsson, Y.; Jemth, P.; Brunori, M.; Travaglini-Allocatelli, C. Identification and characterization of protein folding intermediates. *Biophys. Chem.* **2007,** *128*, 105-113.

(13) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; de los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J. C.; Bona, D.; Miller, E. J., et al. Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **2005,** *14*, 602-616.

(14) Myers, J. K.; Oas, T. G. Preorganized secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.* **2001,** *8*, 552-558.

(15) Guerois, R.; Serrano, L. The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **2000,** *304*, 967-982.

(16) Perl, D.; Welker, C.; Schindler, T.; Schroder, K.; Marahiel, M. A.; Jaenicke, R.; Schmid, F. X. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol.* **1998,** *5*, 229-235.

(17) Reid, K. L.; Rodriguez, H. M.; Hillier, B. J.; Gregoret, L. M. Stability and folding properties of a model beta-sheet protein, Escherichia coli CspA. *Protein Sci.* **1998,** *7*, 470-479.

(18) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C. Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. Ultrafast folding of alpha(3): A de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 15486-15491.

(19) Van Nuland, N. A. J.; Meijberg, W.; Warner, J.; Forge, V.; Scheek, R. M.; Robillard, G. T.; Dobson, C. M. Slow cooperative folding of a small globular protein HPr. *Biochemistry* **1998,** *37*, 622-637.

(20) Plaxco, K. W.; Spitzfaden, C.; Campbell, I. D.; Dobson, C. M. A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J. Mol. Biol.* **1997,** *270*, 763-770.

(21) Guijarro, J. I.; Morton, C. J.; Plaxco, K. W.; Campbell, I. D.; Dobson, C. M. Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* **1998,** *276*, 657-667.

(22) Clarke, J.; Hamill, S. J.; Johnson, C. M. Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **1997,** *270*, 771-778.

(23) Clarke, J.; Cota, E.; Fowler, S. B.; Hamill, S. J. Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* **1999,** *7*, 1145-1153.

(24) Taddei, N.; Chiti, F.; Paoli, P.; Fiaschi, T.; Bucciantini, M.; Stefani, M.; Dobson, C. M.; Ramponi, G. Thermodynamics and kinetics of folding of common-type acylphosphatase: Comparison to the highly homologous muscle isoenzyme. *Biochemistry* **1999,** *38*, 2135-2142.

(25) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. Cytochrome folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein. *Proc. Natl. Acad. Sci. USA* **1999,** *96*, 6587-6590.

(26) Scott, K. A.; Batey, S.; Hooton, K. A.; Clarke, J. The folding of spectrin domains i: Wild-type domains have the same stability but very different kinetic properties. *J. Mol. Biol.* **2004,** *344*, 195-205.

(27) Ikura, T.; Hayano, T.; Takahashi, N.; Kuwajima, K. Fast folding of Escherichia coli cyclophilin a: A hypothesis of a unique hydrophobic core with a phenylalanine cluster11edited by c. R. Matthews. *J. Mol. Biol.* **2000,** *297*, 791-802.

(28) Laurents, D. V.; Corrales, S.; Elías-Arnanz, M.; Sevilla, P.; Rico, M.; Padmanabhan, S. Folding kinetics of phage 434 cro protein. *Biochemistry* **2000,** *39*, 13963-13973.

(29) Fowler, S. B.; Clarke, J. Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state. *Structure* **2001,** *9*, 355-366.

(30) Schreiber, G.; Fersht, A. R. The refolding of cis-peptidylprolyl and trans-peptidylprolyl isomers of barstar. *Biochemistry* **1993,** *32*, 11195-11203.

(31) Calloni, G.; Taddei, N.; Plaxco, K. W.; Ramponi, G.; Stefani, M.; Chiti, F. Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding. *J. Mol. Biol.* **2003,** *330*, 577-591.

(32) Cota, E.; Clarke, J. Folding of beta-sandwich proteins: Three-state transition of a fibronectin type III module. *Protein Sci.* **2000,** *9*, 112-120.

(33) Parker, M. J.; Dempsey, C. E.; Lorch, M.; Clarke, A. R. Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. *Biochemistry* **1997,** *36*, 13396-13405.

(34) Matouschek, A.; Kellis, J. T.; Serrano, L.; Bycroft, M.; Fersht, A. R. Transient folding intermediates characterized by protein engineering. *Nature* **1990,** *346*, 440-445.

(35) Schymkowitz, J. W. H.; Rousseau, F.; Irvine, L. R.; Itzhaki, L. S. The folding pathway of the cell-cycle regulatory protein p13suc1: Clues for the mechanism of domain swapping. *Structure* **2000,** *8*, 89-100.

(36) Choe, S. E.; Matsudaira, P. T.; Osterhout, J.; Wagner, G.; Shakhnovich, E. I. Folding kinetics of villin 14t, a protein domain with a central beta-sheet and two hydrophobic cores. *Biochemistry* **1998,** *37*, 14508-14518.

(37) Dalessio, P. M.; Ropson, I. J. Beta-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry* **2000,** *39*, 860-871.

(38) Munoz, V.; Lopez, E. M.; Jager, M.; Serrano, L. Kinetic characterization of the chemotactic protein from Escherichia-coli, chey - kinetic-analysis of the inverse hydrophobic effect. *Biochemistry* **1994,** *33*, 5858-5866.

(39) Burns, L. L.; Dalessio, P. M.; Ropson, I. J. Folding mechanism of three structurally similar beta-sheet proteins. *Proteins: Struct. Funct. Genet.* **1998,** *33*, 107-118.

(40) Tang, K. S.; Guralnick, B. J.; Wang, W. K.; Fersht, A. R.; Itzhaki, L. S. Stability and folding of the tumour suppressor protein p1611. *J. Mol. Biol.* **1999,** *285*, 1869-1886.

(41) Cavagnero, S.; Dyson, H. J.; Wright, P. E. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **1999,** *285*, 269-282.

(42) Golbik, R.; Zahn, R.; Harding, S. E.; Fersht, A. R. Thermodynamic stability and folding of GroEL minichaperones. *J. Mol. Biol.* **1998,** *276*, 505-515.

(43) Parker, M. J.; Marqusee, S. The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **1999,** *293*, 1195-1210.

(44) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262*, 892.

(45) Parker, M. J.; Sessions, R. B.; Badcoe, I. G.; Clarke, A. R. The development of tertiary interactions during the folding of a large protein. *Fold Des.* **1996,** *1*, 145-156.

(46) Ogasahara, K.; Yutani, K. Unfolding-refolding kinetics of the tryptophan synthase alpha-subunit by cd and fluorescence measurements. *J. Mol. Biol.* **1994,** *236*, 1227-1240.

(47) Goldberg, M. E.; Semisotnov, G. V.; Friguet, B.; Kuwajima, K.; Ptitsyn, O. B.; Sugai, S. An early immunoreactive folding intermediate of the tryptophan synthase β2 subunit is a 'molten globule'. *FEBS Lett.* **1990,** *263*, 51-56.

Chapter 2

Fluorescence Anisotropy Decays and Microscale-Volume Viscometry Reveal the Compaction of Ribosome-Bound Nascent Proteins

This chapter reproduces a published article (Rachel B. Hutchinson, Xi Chen, Ningkun Zhou, and Silvia Cavagnero. *Journal of Physical Chemistry B* **2021**, 125, 6543-6558). R.B.H. and S.C. designed the project. R.B.H participated in sample preparation and data collection for fluorescence lifetime, fluorescence anisotropy decay, and SDS-PAGE gel experiments. R.B.H. performed FoldX free energy calculations. R.B.H. and X. C. participated in viscosity measurements. X.C. generated elongation NECNOP plots. R.B.H., X.C., N.Z., and S.C. participated in data analysis, interpretation, discussions, and manuscript writing.

2.1 Abstract

This work introduces a technology that combines fluorescence anisotropy decay with microscale-volume viscometry to investigate the compaction and dynamics of ribosome-bound nascent proteins. Protein folding in the cell, especially when nascent chains emerge from the ribosomal tunnel, is poorly understood. Previous investigations based on fluorescence anisotropy decay determined that a portion of the ribosome-bound nascent protein apomyoglobin (apoMb) forms a compact structure. This work, however, could not assess the size of the compact region. The combination of fluorescence anisotropy with microscale-volume viscometry, presented here, enables identifying the size of compact nascent-chain subdomains using a single fluorophore label. Our results demonstrate that the compact region of nascent apoMb contains 57-83 amino acids, and lacks residues corresponding to the two native C-terminal helices. These amino acids are necessary for fully burying the nonpolar residues in the native structure, yet they are not available for folding before ribosome release. Therefore, apoMb requires a significant degree of post-translational folding for the generation of its native structure. In summary, the combination of fluorescence anisotropy decay and microscale-volume viscometry is a powerful approach to determine the size of independently tumbling compact regions of biomolecules. This technology is of general applicability to compact macromolecules linked to larger frameworks.

2.2 Introduction

Correct protein folding is essential for the proper function of living organisms and for the efficient large-scale production of biomedically relevant proteins in the context of biotechnology and pharmaceutical applications. While *in vitro* refolding of pure proteins has helped explain general aspects of the process and underlying trends,[1-3] protein folding within the cellular environment, including the role of the ribosome and molecular chaperones, is still poorly understood. [4-7]

Many proteins begin folding cotranslationally before they are released from the ribosome.[4, 8-15] The ribosome plays an important role in protein folding. On one hand, it may act as a chaperone.[16-21] On the other hand, the ribosome is known to influence cotranslational folding by confining the motion of nascent chains,[9, 15, 22, 23] by enabling interactions between the nascent chain and the ribosomal tunnel[24, 25] and(or) ribosomal surface[26-29] and by promoting nascent-chain solubility.[30] Immediately after release from the ribosome, single-domain nascent proteins must be properly kinetically channeled to their native state to prevent the formation of aggregated states.[30] The ribosomal exit tunnel, which is approximately 100 Å long and has a width of 10-20 Å, can be divided into two regions: the tunnel core, and the vestibule, which encompasses the last 20 Å of the tunnel and is wider than the rest of the ribosomal tunnel.[31, 32] The exit tunnel generally hosts approximately 30 – 40 amino acids, depending on the sequence and structure of the nascent protein.[23, 33-35] More residues can be hosted if the nascent protein adopts tertiary structure within the tunnel.[36, 37] Within the tunnel core, nascent chains can form α-helical secondary structure, [9, 23, 38, 39] tertiary interactions,[10, 40] and even fully folded states.[36, 37, 41] In addition, larger tertiary structures can form within the tunnel vestibule.[12, 42, 43]

In order to better understand the extent of protein folding during protein biogenesis, Ellis *et*

al. addressed the development of nascent-protein compaction within stalled ribosome-bound nascent chains (RNCs).[13] The protein apomyoglobin from sperm whale (apoMb) was selected as a model system for this work, and transcription-translation was carried out in cell-free systems.[4, 44] These studies addressed RNC compaction as a function of chain elongation by fluorescence anisotropy-decay analysis.[13] The investigations by Ellis *et al.* revealed two rotational motions of short nascent chains (less than 57 amino acids) that correspond solely to the dynamics of the ribosome and the fluorescently-labeled N-terminal region of the nascent chain.[13] However, apoMb RNCs bearing nascent chains longer than 57 residues display a third component corresponding to a rotational correlation time of several ns.[13] This component is absent from the anisotropy decay of RNCs carrying an intrinsically-disordered nascent protein, and it is characterized by a smaller rotational correlation time (ca. 4-7 ns) than the ribosome-released folded protein (ca. 40-50 ns). This anisotropy-decay component established the presence of a compact subdomain within the nascent chain.[13]

The abovementioned investigations, however, were unable to unequivocally determine the number of amino acids belonging to the compact nascent-chain domain, given that they relied on indirect information on solution viscosity inferred from fluorescence data. Specifically, the above studies estimated the viscosity of RNC solutions from the rotational correlation time of either a small fluorescent dye or ribosome-released apoMb.[13] Given that the solution viscosity deduced from these two approaches were significantly different, it was impossible to draw clear conclusions on the size of the compact RNC subdomain. Therefore, while previous work demonstrated that RNCs longer than 57 residues bear a compact nascent chain, the number of amino acids corresponding to this compact region could not be assessed. This concept is pictorially illustrated in Figure 2-1, which shows that conformations a-c can be excluded due to

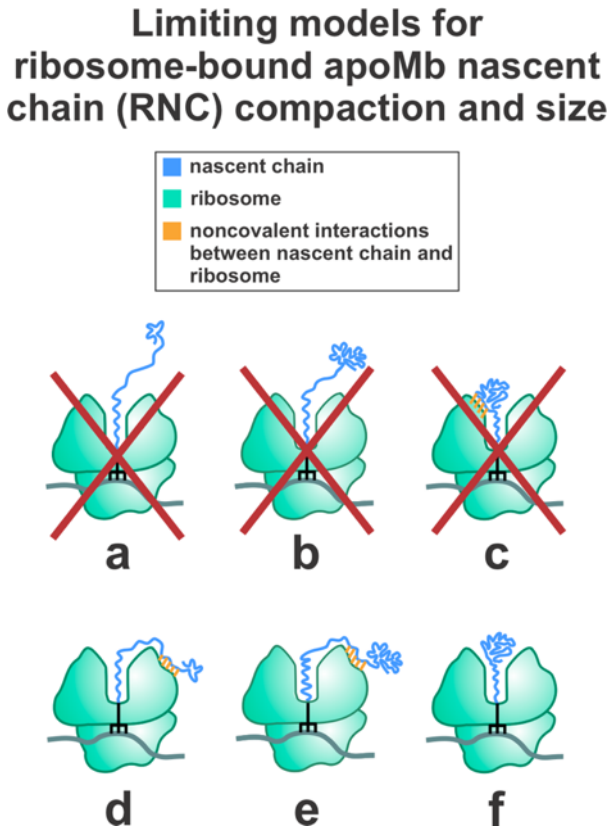


Figure 2-1. Hypothetical limiting models describing the ribosome-bound apoMb nascent chain (RNC) compaction. Models a and b can be eliminated due to the known highly spatially confined environment of the nascent chain (cone semiangle = 20 ± 1°).[15] Model c can be eliminated because the compact region explored in this work independently tumbles on the low-ns timescale (see fluorescence anisotropy decay analysis). This timescale is incompatible with the much slower tumbling of the ribosome.[13] Models d, e, and f are consistent with previously published work but could not be discriminated from each other in the past.

previous studies reporting small-amplitude local RNC motions on the ns timescale.[13] Until now, however, it has been impossible to discriminate between small (species d) and large (species e and f) nascent-chain conformations.

In this work, we address the above gap of knowledge by combining fluorescence anisotropy decays with direct viscosity measurements to identify the size of compact nascent-chain domains of apoMb RNCs. Our results show that the compact nascent chain comprises 37 - 54% of the amino-acid sequence, hence it lacks C-terminal residues corresponding to 46 - 63% of the chain. Therefore, a significant portion of the RNC must fold post-translationally. In all, this work demonstrates that fluorescence anisotropy decays can be synergistically combined with microscale-volume viscometry to unveil the compaction of nascent chains. The technology introduced here is of wide applicability to any compact macromolecule linked to larger frameworks. In addition, this approach is particularly straightforward as it only requires a single fluorophore (extrinsic or intrinsic) per biomolecule of interest.

2.3 Methods

Generation of RNC complexes. An *E. coli* cell-free transcription-translation system was used to produce ribosome-bound apoMb$_{153}$ nascent-chain complexes as described.[13, 45] This cell-free system included an *E. coli* S30 extract prepared from the K12 A19 strain[45, 46] and a pET-Blue1 plasmid encoding the sperm-whale apomyoglobin gene with an *E. coli*-optimized nucleotide sequence.[47] In addition, the cell-free system included oligodeoxynucleotides designed to carry a complementary sequence to that of the mRNA region immediately before the stop codon (see Supporting Information for more details). Consistent with the oligodeoxynucleotide-directed mRNA cleavage approach,[4] the oligodeoxynucleotides bind the mRNA, followed by site-specific

DNA-RNA-hybrid cleavage by endogenous RNaseH.[48] The resulting stalled ribosome-bound nascent-chain complex carried the full-length apoMb protein. An anti-SsrA oligodeoxynucleotide was added to the cell-free system to prevent tmRNA-mediated release of nascent proteins from the ribosome.[49] The N-terminal methionine was labeled with BODIPY-FL using BODIPY-FL Met-tRNA^{fMet}, which was prepared and added to the cell-free system as described.[13] Cell-free components were assembled at room temperature (0.75 – 1.05 ml total volume) followed by splitting into 75 μL equivolume aliquots and parallel incubation of all aliquots at 37 °C for 30 min.

Purification of RNC complexes. After cell-free transcription-translation, ribosome-bound nascent-chain complexes were purified upon centrifugation (160,000 rcf for 60 min at 4 °C) of each of the aliquots over a 150 μL of sucrose cushion (1.1 M sucrose, 20 mM tris-HCl, 10 mM Mg^{2+} acetate, 500 mM NH_4Cl, 0.5 mM EDTA and 1 mM DL-dithiothreitol, pH adjusted to 7.0).[50] Pellets containing ribosome-bound nascent-chain complexes were then solubilized in 15 μL of resuspension buffer (10 mM Tris HCl, 10 mM $Mg(OAc)_2$, 60 mM NH_4Cl, 0.5 mM EDTA, 1 mM dithiothreitol, pH 7.0). All 15 μL aliquots were then combined, mixed, and then further split for anisotropy decay and viscosity data collection, which were run in parallel each day. The ribosome-bound status of the nascent chains was verified via a puromycin assay (see Supporting Information for details and representative SDS-PAGE data) performed on the same samples previously subject to fluorescence measurements.[51, 52]

Concentration of purified RNCs. The total ribosome concentration of resuspended RNCs was assessed from sample absorbances at 260 nm, considering the extinction coefficient of the 70S *E. coli.* ribosome (3.91×10^7 M^{-1} s^{-1}).[53] The concentration of BODIPY-FL labeled RNCs was determined from the fluorescence intensity of RNC gel bands, upon comparison with the

fluorescence intensity of reference gel bands of BODIPY-FL Met-tRNA^{fMet} of known concentration. ApoMb gel band intensities were determined with the ImageJ software.[54, 55]

Fluorescence anisotropy decay: Data collection and analysis. Fluorescence anisotropy-decay data were collected with a Chronos frequency-domain fluorometer (ISS Inc.). Samples were excited at 477 nm with a laser diode. A 480 ± 5 nm band-pass filter was used for the excitation channel and a 520 ± 20 nm band-pass filter (Chroma Technology) was used for the emission channel. The excitation polarizer was set to vertical for lifetime and anisotropy measurements. The emission polarizer was set to 54.7° for lifetime measurements. Samples were incubated at 25 °C for at least 30 min prior to fluorescence measurements. Sample temperature was maintained at 25 ± 0.1 °C during fluorescence measurements with a circulating water bath.

Lifetime and anisotropy decay data were analyzed with the Globals software package (Laboratory for Fluorescence Dynamics, LFD).[56] The standard deviation for phase and modulation were set to 0.2° and 0.004, respectively, to estimate reduced χ^2 values.[57] Lifetime data were fit to three discrete exponential decay components, one of which was fixed to 1 ps to account for light scattering, and the other two, corresponding to the actual fluorophore lifetimes, were allowed to float. Anisotropy data were corrected taking the experimental frequency-dependent G-factor into account. The latter parameter was determined on the same day as the anisotropy decay measurements. Then anisotropy data were then fit to multiexponential decays. The fundamental anisotropy r_0 was fixed at 0.37. In order to assess best fits, all data were independently fit to both two and three-component decays and χ^2 values were compared. Three-component fits were chosen if their χ^2 values were $\leq$ 2.5x smaller than the χ^2 values for two-component fit. The number of amino acids in the compact domain of the nascent protein was determined upon evaluating both spherical, prolate (representative axial ratio = 3.5) and oblate

(representative axial ratio = 0.5) ellipsoid models. Order parameters and cone semiangles were determined as described.[15] See Supporting Information for further details on χ^2 and compact-domain calculations. Anisotropy-decay simulations were performed with the Vinci software (ISS, Inc.).

Derivation of equations used for ellipsoidal models. Spherical, prolate ellipsoidal, and oblate ellipsoidal models were used to calculate the number of amino acids contained in the compact subdomain of the nascent protein. We used ellipsoids of revolution for the ellipsoid models, which have two axes with equivalent length (Figure 2-8). These ellipsoids can have three different rotational correlation times, but these three correlation times are nearly identical for oblate ellipsoids and prolate ellipsoids with an axial ratio less than five and are too close to be resolved via our method.[58, 59] Therefore, for the ellipsoidal models, the harmonic-mean rotational correlation time τ_H was used, and determined according to[60]

$$\tau_H = \left(\frac{0.4}{\tau_1} + \frac{0.4}{\tau_2} + \frac{0.2}{\tau_3}\right)^{-1} \quad (1)$$

where τ_1, τ_2 and τ_3 were determined from the parallel and perpendicular components of the rotational diffusion coefficient according to

$$\tau_1 = (D_\parallel + 5D_\perp)^{-1} \quad (2)$$

$$\tau_2 = (4D_\parallel + 2D_\perp)^{-1} \quad (3)$$

$$\tau_3 = (6D_\perp)^{-1} \quad (4)$$

Substituting eqs 2 - 4 into eq 1 gives

$$\tau_H = \left(0.4(D_\parallel + 5D_\perp) + 0.4(4D_\parallel + 2D_\perp) + 0.2(6D_\perp)\right)^{-1} \quad (5)$$

which simplifies to

$$\tau_H = (2D_{\parallel} + 4D_{\perp})^{-1} \quad (6)$$

$D_{\parallel}$ and $D_{\perp}$ are defined as

$$D_{\parallel} = \left(\frac{3\rho(\rho-b)}{2(\rho^2-1)}\right) D_{sphere} \quad (7)$$

$$D_{\perp} = \left(\frac{3\rho[(2\rho^2-1)b-\rho)]}{2(\rho^4-1)}\right) D_{sphere} \quad (8)$$

where ρ is the aspect ratio and b is defined as

$$b = (\rho^2-1)^{-1/2}\ln\left[\rho + (\rho^2+1)^{1/2}\right] \text{ for prolate ellipsoids} \quad (9)$$

$$b = (1-\rho^2)^{-1/2}\arctan\left[\frac{(1-\rho^2)^{1/2})}{\rho}\right] \text{ for oblate ellipsoids} \quad (10)$$

Viscosity measurements. The viscosity of apoMb RNC solutions was measured with a microVISC™ viscometer (Rheosense, Inc.), an apparatus tailored to small-volume samples (15-60 µL depending on shear rate). Samples were allowed to flow though the viscometer channel at constant shear rate, and pressure differences as the fluid flow within the channel were measured, to assess sample viscosity. The viscosity of apoMb RNC solutions was determined at shear rates 1500, 3000, 5000, 6500, and 8000 s^{-1}, with a temperature control module set to 25 °C. Viscosity measurements with an R^2 value lower than 0.995 were discarded. Plots of viscosity *versus* shear rate and student's *t*-test calculations were used to compare viscosity values. This analysis let us to conclude that all samples analyzed in this work exhibit Newtonian behavior.

Fluorescence-Detected Gel Electrophoresis. SDS-PAGE via three-layer tris-tricine low-pH gels[61] was employed to analyze RNCs in the absence and presence of puromycin.[45, 62] Fluorophore-labeled peptidyl tRNAs (carrying nascent proteins) were visualized directly on the gels with a FLA 9500 Typhoon Gel Imager (GE Healthcare Life Sciences, 473 nm excitation laser, BPB1 emission filter (525/50 nm band pass filter).

2.4 Results and Discussion

Experimental design. This study employs apoMb as a model system. The holo form of this protein, containing the heme cofactor, occupies a special place in chemistry and biology as it was the very first protein whose structure was solved by X-ray crystallography. In addition, apoMb carries the ubiquitous all-α-helical globin fold.[63, 64] and its *in vitro* refolding mechanism was extensively characterized via classical refolding experiments from denaturant [65-72] as well as within cell-relevant environments.[13, 15, 30, 45, 73] The apoMb plasmid is codon-usage optimized for expression in *E. coli*.[47]

To explore the compaction and dynamics of apoMb in a cell-relevant environment, we generated stalled full-length RNCs of this protein in an *E. coli* cell-free system. The N-terminal methionine was labeled with the fluorophore BODIPY-FL. Ribosomes harboring nascent protein chains were purified and resuspended in a buffer solution to remove free fluorophore, tRNA, and released protein from the solution before performing fluorescence anisotropy decay and viscosity measurements. We verified that the nascent chain was bound to the ribosome with a puromycin assay, after completing the fluorescence anisotropy decay and viscosity measurements (Figure 2-S1).

The overall experimental design is schematically illustrated in Figure 2-2. Briefly, we assessed fluorescence anisotropy decays in the frequency domain[58, 74] and combined this technique with direct microscale-volume viscosity measurements on the same samples. In this way, we could determine the number of amino acids contained in the compact portion of the nascent chain. The rotational correlation time (τ_c) of a molecular system of interest depends on its size (e.g., its number of amino acids, #aa), shape, and solution viscosity (η). Therefore, the τ_c of the nascent chain was determined via fluorescence anisotropy decay and the η was assessed

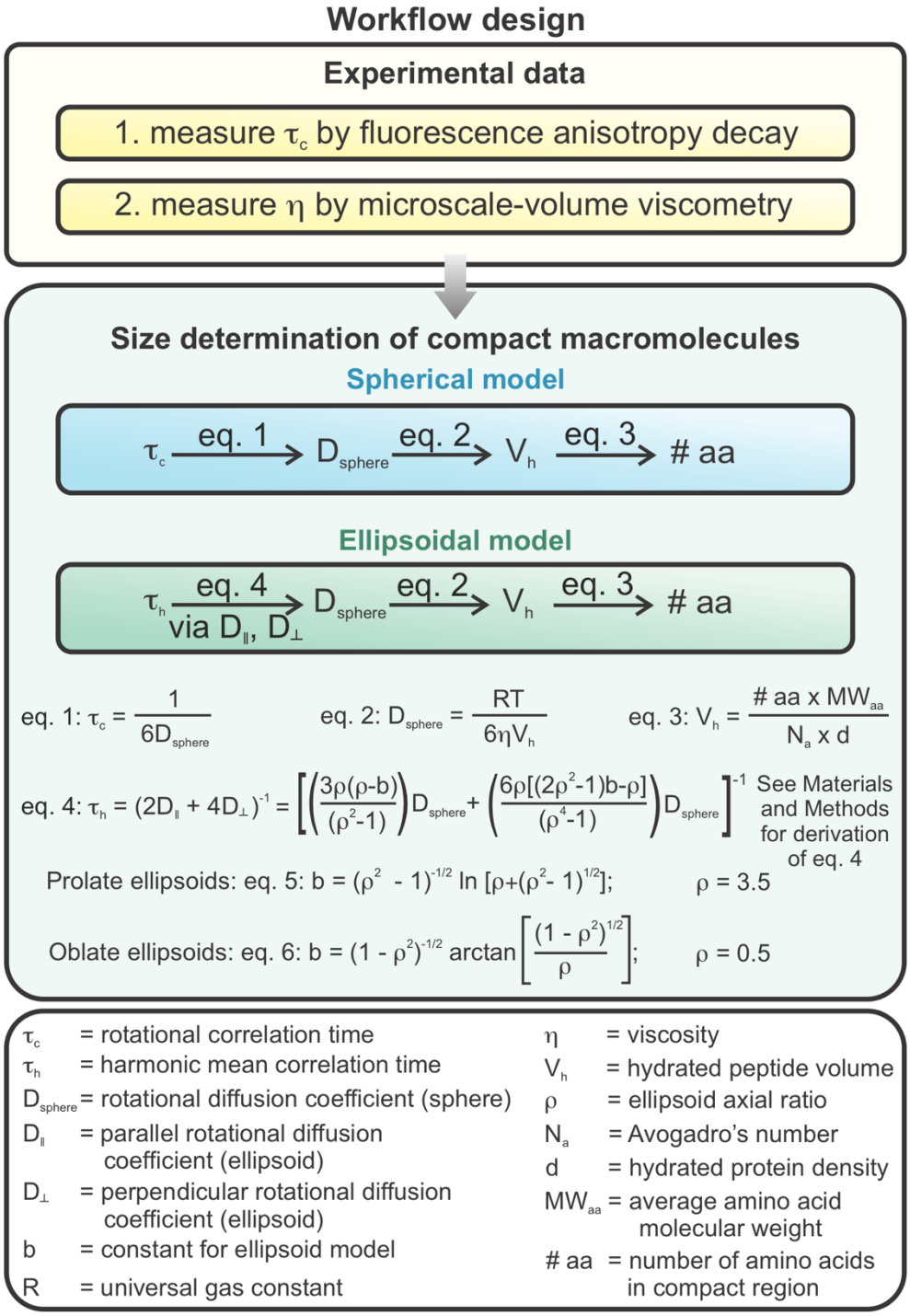


Figure 2-2. Summary of workflow design to determine degree of compaction of RNCs. The rotational correlation time (τ_c) describing nascent-chain compact-subdomain motions and RNC macroscopic viscosity (η_{macro}) were determined via fluorescence anisotropy decay in the

frequency domain and microscale-volume viscometry, respectively. Experimentally determined values were employed to determine the size of the compact subdomain. Spherical, prolate ellipsoidal, and oblate ellipsoidal models were used to model the compact region of the nascent chain.

with a microscale-volume viscometer. Then, we employed the Stokes-Einstein Debye equation to determine the hydrated volume (V_h) of the compact domain of the nascent chain from experimental rotational correlation time and solution viscosity values, assuming spherical, prolate or oblate ellipsoidal shapes of representative aspect ratios ρ=3.5 or ρ=0.5, respectively. We then calculated the number of amino acids (# aa) from the volume (V_h) considering that the standard protein dry volume of 0.75 mL/g includes 0.2 mL of water per g of protein[60] and the average molecular weight of amino acids is 110 g/mol.

Combining anisotropy decay with microscale-volume viscometry is a convenient approach to determine the size of independently tumbling compact regions of biomolecules, including RNCs. Fluorescence anisotropy decay and microscale-volume viscometry can be synergistically combined to determine the size and dynamics of compact independently tumbling regions of biomolecules. This method is non-perturbative and can be used to study both compaction and dynamics of nascent proteins in solution. Our approach does not need extrinsic SecM sequences to stall translation, and is particularly convenient because it only requires a single fluorophore, making it easily applied to any protein of interest. We attached the fluorophore to the protein N-terminus, but the fluorescent label can be located anywhere on the independently tumbling region of the macromolecule where it does not interact with immobilized or slowly tumbling species.

Fluorescence anisotropy decay alone has been previously used to study conformation and dynamics of biomolecules such as foldable and intrinsically-disordered proteins,[75-77] nucleic acids,[78, 79] the actin-myosin system,[75] and antibodies.[75] Fluorescence anisotropy decay can be used to resolve local and global dynamics not only in nascent-protein-ribosome complexes, but also in ribosome-released foldable and intrinsically-disordered proteins[13, 80] and in protein

fibrils.[81]

The method introduced here is versatile because it can be straightforwardly applied to any complex biomolecule containing an independently tumbling compact region.

Viscosity measurements and Newtonian *versus* non-Newtonian fluids. The rotational correlation time of a molecule depends on its size and solution viscosity, as shown in Figure 2-2. Therefore, viscosity measurements are required to determine the size of compact domains of RNCs. Here, we solve this challenge by performing direct solution viscosity measurements *in situ*. As shown in the sections below, we also took into account how shear rate, molecular crowding, and the highly negatively charged surface of the ribosome may affect the local viscosity in proximity of the nascent protein surface.

Solution viscosity may depend on the shear rate, (i.e., the flow rate) of the fluid of interest. Newtonian fluids have a constant viscosity that is not affected by shear rate. Non-Newtonian fluids include shear-thinning fluids, which undergo a decrease in viscosity as shear rate increases, and shear-thickening solutions, which experience an increase in viscosity as shear rate increases.[82] Many fluids show Newtonian behavior at low shear rates and then Non-Newtonian behavior at higher shear rates.[83-85] However, several globular proteins including globulins, especially bovine globulin serum albumin, ovalbumin, and β-lactoglobulin, display shear-thinning behavior at low shear rates and Newtonian behavior at higher shear rates.[86-91] For these proteins, shear-thinning behavior occurs at low shear rates (less than 100 s^{-1}) because the proteins form a film at the liquid-gas interface. Within this film, the solution is viscoelastic and experiences a higher viscosity compared to the experimentally measured macroscopic viscosity η_{macro}. As the shear rate increases, the increased flow breaks up the film, decreasing the solution viscosity.[90] Non-globulin proteins,[92] *E. coli* ribosomes,[93] mixtures of RNA and proteins,[94] and

DNA solutions,[95, 96] on the other hand, show Newtonian behavior at low shear rates. In order to assess whether a given solution displays Newtonian or non-Newtonian behavior, it is necessary to perform experimental measurements at different shear rates.

Effect of molecular crowding and excluded volume on viscosity. The Stokes-Einstein (SE) and Stokes-Einstein-Debye (SED) relations (see eqs 11 and 12 below, respectively) show how a particle's translational and rotational diffusion depend on solution viscosity

$$D_t = \frac{k_B T}{6\pi\eta r} \tag{11}$$

$$D_{r,sphere} = \frac{k_B T}{8\pi\eta r^3} = \frac{RT}{6\eta V_h} \tag{12}$$

where k_B is the Boltzmann constant, T is temperature in Kelvin, η is the solution viscosity, r is the radius of the particle, V_h is the hydrated volume of the molecule of interest, and R is the universal gas constant. The above equations assume homogeneous solution viscosity and non-interacting spherical particles. Within physiologically relevant systems, however, molecular crowding may cause experimental diffusion coefficients to differ from the values predicted from experimental macroscopic viscosity (η_{macro}) and the SE and SED equations.[97-101] Both negative and positive deviations, leading to faster and slower experimental diffusion than predicted from eqs 11 and 12, respectively, were observed to date. Negative deviations were observed for protein translational and rotational diffusion in the presence of polymer[102-105] and selected protein[106] crowders. This solution behavior was predicted by a Brownian dynamics model.[107] This negative deviation from the SE and SED equations can be explained by the excluded volume effect. The space unavailable to a molecule, known as the excluded volume, includes the space occupied by other molecules and a depletion layer, which is the region surrounding a particle that is depleted in other particles compared to the bulk solution (Figure 2-3).[108-110] Given that the depletion layer is inaccessible to crowder molecules, the viscosity within this space,

denoted as η_{dl}, is lower than η_{macro}, and approaches the pure solvent viscosity η_s.[101, 108] In summary, $\eta_s \leq \eta_{dl} \leq \eta_{macro}$ (Figure 2-3c).When η_{dl} is less than η_{macro}, both translational and rotational diffusion can be faster than the diffusion behavior predicted from eqs 11 and 12. The deviation for rotational diffusion, however, is greater because rotational tumbling occurs entirely within the depletion layer.[100, 101]

In the case of most protein crowders, however, experimental results[97] and molecular dynamics simulations[111] showed positive deviations. These results are explained by transient cluster formation between proteins, which becomes significant at high protein concentration (>100 mg/mL) and increases the apparent radius of the rotating species, thus slowing down translational and rotational diffusion.[97, 111] In these concentrated protein solutions, the effect of transient associative interactions counteracts the effect of the depletion layer.

In our experiments, we employed a BODIPY-FL labeled RNC concentrations of ca. 0.1 mg/mL (40 nM) and total ribosome concentrations of ca. 2 mg/mL (790 nM). See the Methods section for procedures employed to determine RNC concentrations. Given that these values are significantly lower than 100 mg/mL, we do not expect clustering to play a significant role.

Effect of charge on viscosity. The local environment of the nascent chain is affected by the presence of the ribosome. The ribosomal RNA and charge segregation of ribosomal proteins[112] render the ribosomal surface highly negatively charged, with a formal charge of nearly -4000.[32, 113] In general, the charge of polyelectrolytes alters solution characteristics relative to solutions of uncharged polyelectrolytes. This phenomenon is known as the "polyelectrolyte effect."[114, 115] Specifically, the negative surface charge of the ribosome[32, 112, 113] is likely surrounded by a high-density mostly positively-charged counterion layer, based on model studies on RNA and proteins.[116] The Coulombic interactions between counterion layer and negatively charged

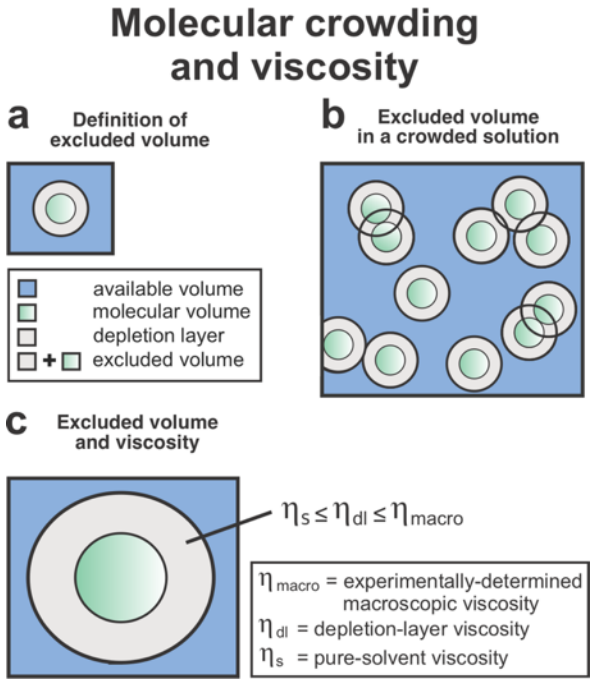


Figure 2-3. Effect of molecular crowding and excluded volume on microscale-volume viscosity. (a) The excluded volume comprises the space occupied by molecules (green) and the depletion layer surrounding them (gray). (b) Within a crowded solution, two molecules cannot occupy the same space, but their depletion layers can overlap. (c) The viscosity within the depletion layer (η_{dl}) surrounding a particle is between the pure-buffer (solvent system) viscosity (η_s) and the experimentally measured macroscopic viscosity (η_{macro}).[108]

ribosome surface shields repulsive interactions.[112] Therefore, given that the environment surrounding ribosome-bound nascent proteins has a higher ion concentration than the surrounding solution, we asked whether this environment could affect the local viscosity experienced by the nascent chain.

Under high-salt solution conditions relevant to our work (0.1 – 1 M), the experimental viscosity of highly charged negatively or positively electrolytes approaches the viscosity of the pure buffer.[117, 118] Therefore, we measured the viscosity of the pure buffer (η_s) and the viscosity of the RNC solution in buffer (η_{macro}) to assess the effect of charge on the local viscosity near the surface of the ribosome.

Experimental viscosity of apoMb RNC solutions. The viscosities of apoMb RNC solutions at shear rates ranging from 1500 s^{-1} to 8000 s^{-1} were not statistically different, according to the two-tailed Student t test (Figure 2-4). This result demonstrates that the apoMb RNC solution behaves as a Newtonian fluid over the tested shear-rate range. We could not directly measure the viscosity at shear rates lower than 1500 s^{-1} due to the limitations of the microscale-volume viscometer. However, we expect RNC solutions to also show Newtonian behavior at lower shear rates because most fluids including DNA, non-globulin proteins, and ribosomes in aqueous solutions show Newtonian behavior at low shear rates.[83-85, 92-96] It is worth mentioning that some globulin and albumin protein solutions display shear-thinning behavior at low shear rates due to their tendency to bind other molecules at the air-water surface interface.[86-91] We expect the our RNC samples, however, to show Newtonian behavior at low shear rates given that surface effects are negligible and that fluorescence anisotropy is a solution property, not pertinent to the air-water interface of RNC samples. In addition, previous work showed that ribosome solutions show Newtonian behavior at low shear rates.[93]

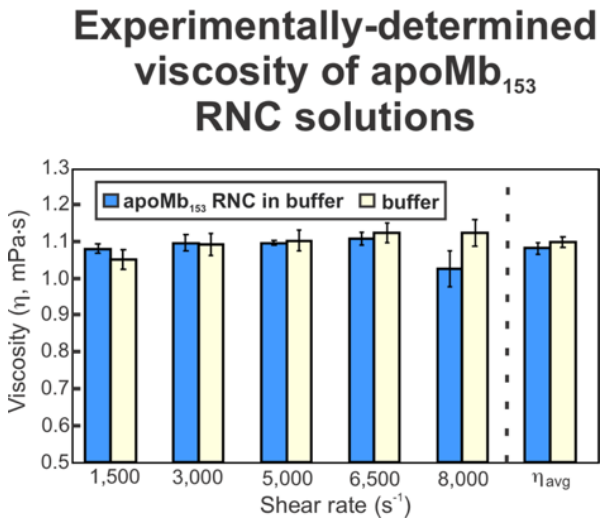


Figure 2-4. Experimentally measured viscosity of RNC solutions (blue) and buffer (yellow).

Regarding the potential effect of transient clustering (i.e., non-specific macromolecular interactions), we expect it to be negligible because the total ribosome concentration in our samples (ca. 2 mg/mL) is significantly lower than the concentration required for significant clustering (>100 mg/mL). We considered the effects of the depletion layer and charge on local viscosity by comparing η_{macro} of the RNC solution to the pure buffer viscosity (η_s). For our sample, η_s (1.10 ± 0.06 mPa-s) is the same within error as η_{macro} for the RNC solution (1.08 ± 0.02 mPa-s) (Figure 2-4). These viscosity values are close to the previously reported value of 0.94 ± 0.02 mPa-s for ribosome solutions.[93] Given that η_{macro} and η_s are identical within error, we determined that the effective rotational viscosity is the same as the η_{macro} measured via microscale-volume viscometry.

In conclusion, the viscosity of our RNC solutions is not significantly affected by depletion layer, clustering, or charge effects. Therefore, the experimentally measured RNC solution viscosity (η_{macro}) over all tested shear rates was used to assess the number of amino acids corresponding to the nascent-chain compact subdomain.

Fluorescence anisotropy decays are an effective probe of rotational dynamics. During anisotropy measurements, a sample labeled with a fluorescent molecule is irradiated with polarized light, which preferentially excites fluorophores with transition dipoles aligned with the electric field of the incoming light.[58] Rotation of the fluorophores while in the excited state results in a depolarization of the emitted light compared to the excitation light. The degree of depolarization can be described quantitatively by the fluorescence anisotropy, which is defined as

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{13}$$

$I_{\parallel}$ and $I_{\perp}$ are the parallel and perpendicular intensities of the emitted light, respectively, and r is

the fluorescence anisotropy.[119]

The anisotropy of a frozen sample that does not rotate is the fundamental anisotropy r_0, which depends on the angle ξ between the absorption and emission transition dipole moments as described by:[120]

$$r_0 = \frac{3\cos^2 \xi - 1}{5} \quad (14)$$

For a sample in solution, rotation of excited-state fluorophores leads to emission-dipole spatial displacement relative to a corresponding frozen sample, resulting in depolarization of the emitted light and a fluorescence anisotropy value smaller than r_0. The degree of depolarization depends on the extent of rotational diffusion of the fluorophores during their fluorescent lifetime. Therefore, anisotropy measurements can be used to measure the rotational diffusion of the sample, which can provide information about the shape, size, and dynamics of the molecule. Importantly, fluorescence anisotropy can only sense dynamics that occur on timescales similar or faster than the lifetime of the fluorophore, which typically range from 0.1-20 ns.[121] While steady-state anisotropy reports on the average anisotropy for all rotational motions, anisotropy decay measurements can resolve the timescales and amplitudes of different motions experienced by the fluorophore (Figure 2-5). In this work, we measure fluorescence anisotropy decay in the frequency domain to resolve the rotational motion of the compact portion of the nascent chain from the motions of the ribosome and fluorescent probe (see Supporting Information and Figure 2-S2 for further discussion on steady-state *versus* time-decay and time-domain *versus* frequency-domain approaches).

Order parameters and cone semiangles define the spatial confinement of rotational motions. In addition to sensing the compaction of nascent chains, fluorescence anisotropy decay measurements also provide information on the spatial confinement of nascent-chain dynamics.

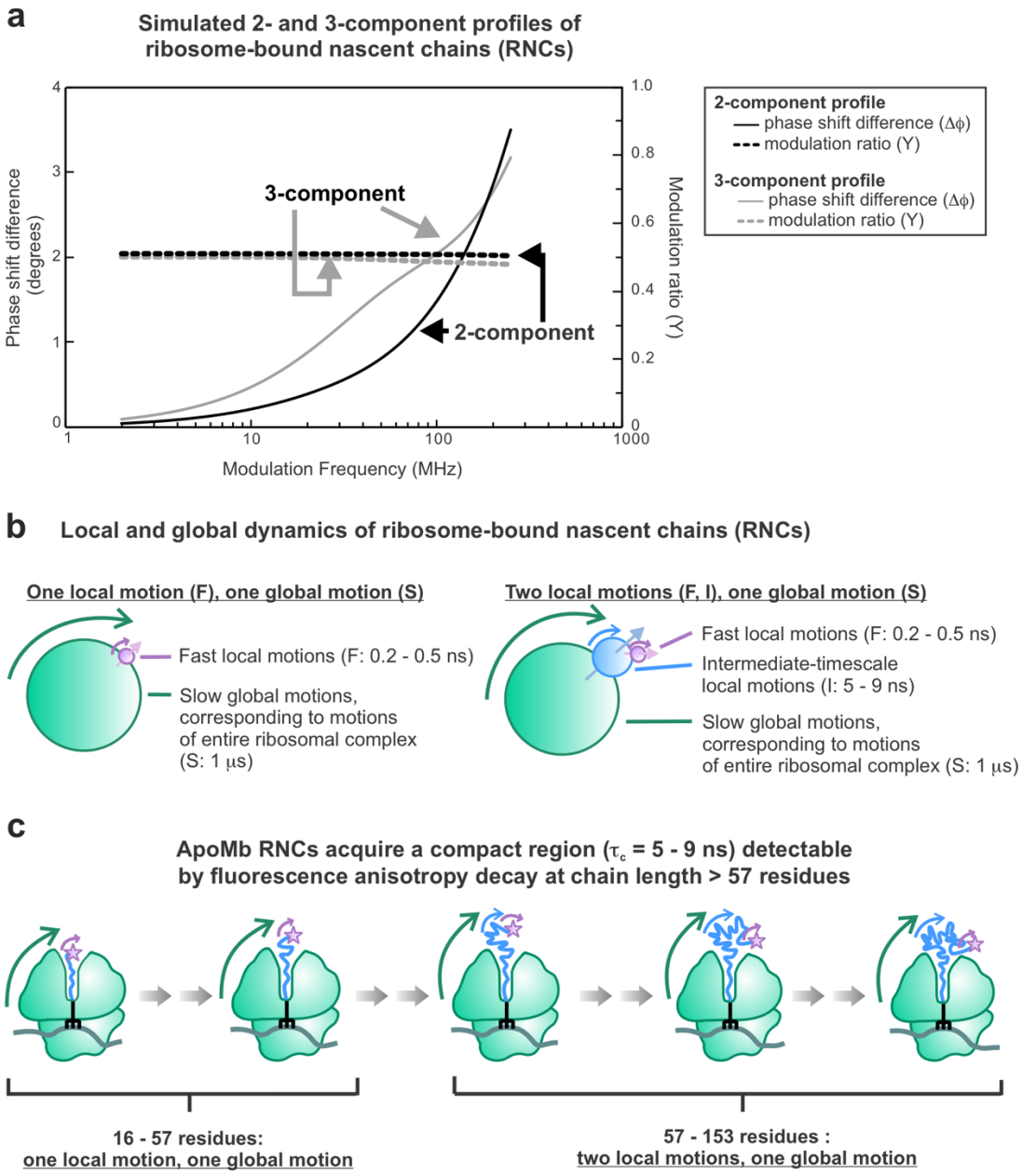


Figure 2-5. Fluorescence anisotropy decay in the frequency domain enables resolving multiple rotational modes, including one corresponding to an apoMb RNC compact subdomain (rotational correlation time τ_c = 5-9 ns). (a) Simulations of 2- and 3-component frequency-domain anisotropy decay data. (b) Representative physical models for samples with one global tumbling

motion one or two local tumbling motions. (c) Prior anisotropy decay data revealed that ApoMb nascent chains become compact and tumble independently starting from chain lengths of ca. 57 residues.[13]

The Lipari-Szabo approach provides a model-free method to determine an order parameter that describes the spatial confinement of a rotational motion based on that motion's fractional contribution to the anisotropy decay.[58, 122] Order parameters range from 0 to 1, where higher order parameters describe more highly constrained motions. The Lipari-Szabo method does not require a specific model to describe the motions, but it does depend on the following five assumptions: (i) The fluorophore has axial symmetry. (ii) Either the fluorescence absorption or emission dipole is colinear with the fluorophore's symmetry axis. (iii) The local fast motions are random and do not depend on the azimuthal angle ϕ (Figure 2-6). (iv) The local and global motions are independent of each other. (v) The rotational correlation times of the motions differ by an order of magnitude or more.[58] For a system that meets these requirements, the anisotropy decay is described by eqs 15 and 16 in the case of two and three rotational motions, respectively[15]

$$\frac{r(t)}{r(0)} = (1 - S_F^2)e^{-t/\tau_{c,F}} + (S_F^2)e^{-t/\tau_{c,S}} \quad (15)$$

$$\frac{r(t)}{r(0)} \cong (1 - S_F^2)e^{-t/\tau_{c,F}} + (S_F^2)(1 - S_I^2)e^{-t/\tau_{c,I}} + S_F^2 S_I^2 e^{-t/\tau_{c,S}} \quad (16)$$

Specific models can be used to gain a more intuitive appreciation for the nature of the spatial confinement encoded by the order parameter. We employed a model by Kinosita *et al.* that assumes local motions to be confined across a cone according to a square-well potential. [123] In this way, we were able to determine limiting cone semiangles from order parameters, thus gaining a more physically meaningful description, as shown in Figure 2-6.[123]

The position of the symmetry axis of the rotating species ($\vec{\mu}$) can be described by[123]

$$p_{eq}(\theta)=[2\pi(1-\cos\theta_0)]^{-1} \text{ for } 0 \leq \theta \leq \theta_0 \quad (17)$$

$$p_{eq}(\theta) = 0 \; for \; \theta > \theta_0 \quad (18)$$

where $p_{eq}(\theta)$ is the normalized equilibrium distribution of the angle θ, and θ_0 is the limiting cone

semiangle. Within this model, the order parameter can be converted to the cone semiangle according to

$$S_i^2 = \left[\frac{1}{2}\cos\theta_0(1 - cos\,\theta_0)\right]^2 \quad (19)$$

The abovementioned equations are powerful because they enable assessing the spatial confinement of rotational motions.

Biologically significant application: Identification of size and dynamics of a ribosome-bound nascent-chain. Full-length apoMb ribosome-nascent protein complexes displayed three rotational motions, consistent with previous results.[13] Figure 2-7 shows representative fluorescence anisotropy decays in the frequency domain. Lifetime and anisotropy decay data for all samples are shown in Supporting Figure 2-S3. The rotational correlation times corresponding to these motions are summarized in Table 2-1. The rotational correlation time for the global tumbling of the ribosomal complex was fixed to 1000 ns. This global rotational correlation time cannot be detected directly because it occurs on a significantly longer timescale than the lifetime of BODIPY-FL (5.9 ns). On the other hand, its presence (in the context of spatially biased faster local motions) is evident from the fact that the anisotropy does not completely decay to zero.[13] The fast motions (τ_c = 0.14 ± 0.01 ns) have a rotational correlation time corresponding to free BODIPY-FL linked to 1-2 residues, hence these motions describe the dynamics of the RNC N-terminus.[15] The intermediate-timescale motions (τ_c = 3.7 ± 0.5 ns) correspond to the motions of a compact region of the nascent chain that tumbles independently from the ribosome. These results agree with previous anisotropy-decay investigations.[13]

In order to determine the size of the compact domain, apoMb nascent chains were modeled as either spheres, prolate ellipsoids or oblate ellipsoids (Figure 2-8). Full-length apoMb nascent chains are unlikely to exist as an extended chain because the persistence length of proteins is

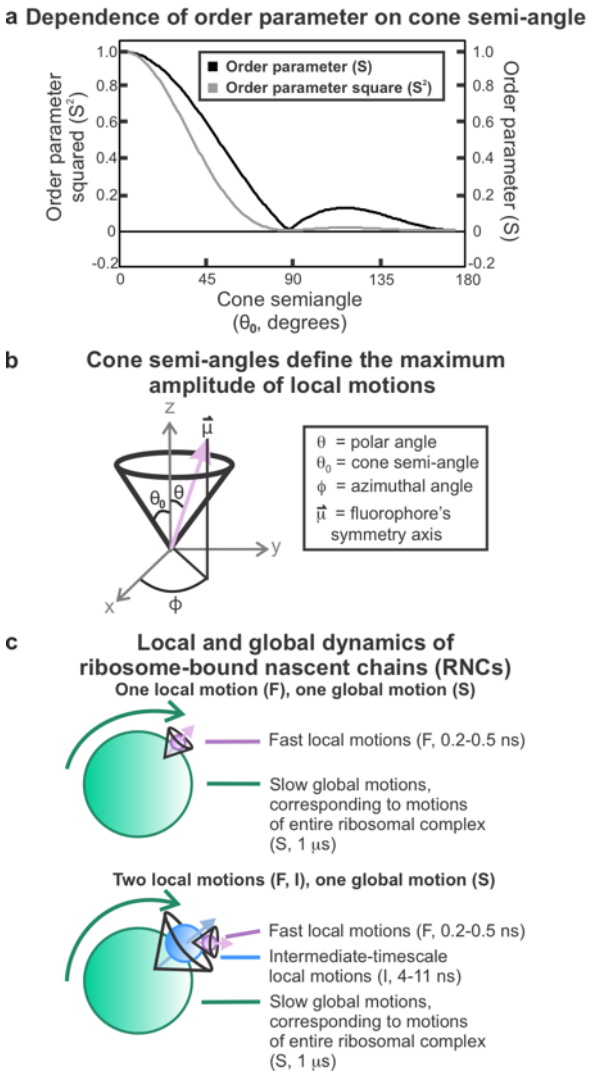


Figure 2-6. Order parameters and cone semiangles define the spatial confinement of nascent chain dynamics. (a) Dependence of order parameter on cone semiangle, assuming a square-well potential. (b) General features of cone-semi angle models. The vector $\vec{\mu}$ defines the fluorophore's symmetry axis, whose orientation can fluctuate so that its angle with the z-axis (θ) is less than or equal to the cone semiangle θ_0. The motions are randomly distributed within the XY plane and can assume any value of the azimuthal angle ϕ) The z-axis is defined to be perpendicular to the macromolecular surface the tumbling species is attached to. (c) Cartoon description of macromolecules bearing either one or two local motions that are spatially confined within a cone.

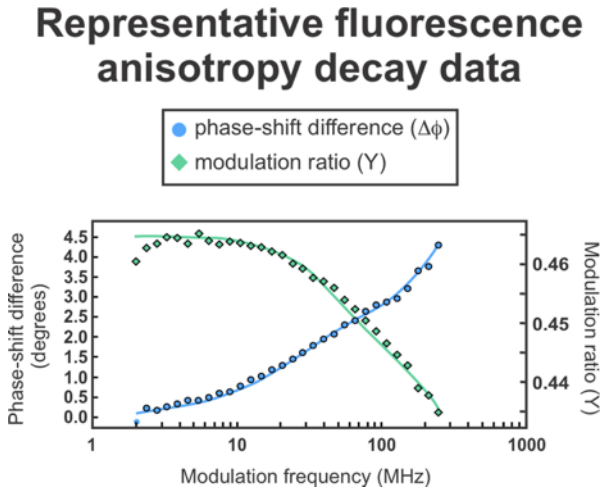


Figure 2-7. Representative frequency-domain fluorescence anisotropy-decay data for full-length apoMb RNCs.

Table 2-1. Summary of rotational correlation times and number of amino acids in compact nascent-chain domains.

Regimes of Motion[a]			Number of amino acids in compact domain			Percent of chain in compact domain		
Slow (~microsecond or slower)	Intermediate (~ns)	Fast (~ps)						
$\tau_{c,S}$ (ns)	$\tau_{c,I}$ (ns)	$\tau_{c,F}$ (ns)	Sphere	Prolate ellipsoid (ρ = 3.5)	Oblate ellipsoid (ρ = 0.5)	Sphere	Prolate ellipsoid	Oblate ellipsoid
1000	3.7 ± 0.5	0.14 ± 0.01	83 ± 10	57 ± 7	68 ± 9	54% ± 7%	37% ± 5%	45% ± 6%

[a] Data are reported as average ± SE for n = 9.

typically only 4-6 amino acids.[13] The experimentally-measured solution viscosity and rotational correlation times were used to determine the number of amino acids contained in the compact region of the nascent chain. The SED relation (eq 12) was used to determine the size of the compact domain assuming spherical shape. The nascent chain was also modeled as a prolate ellipsoid (representative axial ratio of 3.5) and an oblate ellipsoid (representative axial ratio 0.5). The equations used for these models are included in Figure 2-2 and in the Supporting Information.

Our results show that the RNC compact region comprises 57 to 83 amino acids (37% - 54% of the nascent chain, see Table 2-1). Therefore, we are able to eliminate Figure 2-1's model d, which entails a small-size compact subdomain (Figure 2-9). It is worth noting that our data are insensitive to other potential extended or compact RNC conformations interacting with the ribosome or molecular chaperones.

Spatial confinement of nascent-chain motions. In addition to determining the size of the compact domain, we also determined the spatial confinement of the nascent chain dynamics (see Table 2-2). The cone semiangles for the fast and intermediate motions are 24.3° ± 0.8° and 15.3° ± 0.7° respectively.[15] These cone semiangles are quite small as the maximum size for a cone semiangle is 180°, indicating that the amplitude of the compact RNC-domain motions are spatially constrained.[15] The small cone semiangles suggest that the compact subdomain may be confined within the ribosomal vestibule or near the ribosome surface (Figure 2-10). Previous work showed that the ribosomal vestibule is sufficiently large to accommodate nascent protein tertiary structure[10, 12] and even proteins as large as 108 amino acids.[43] Therefore, it is possible the vestibule accommodates the apoMb compact structure which contains 57-83 amino acids. Alternatively, a portion of the chain prior to compact region may interact with the outer

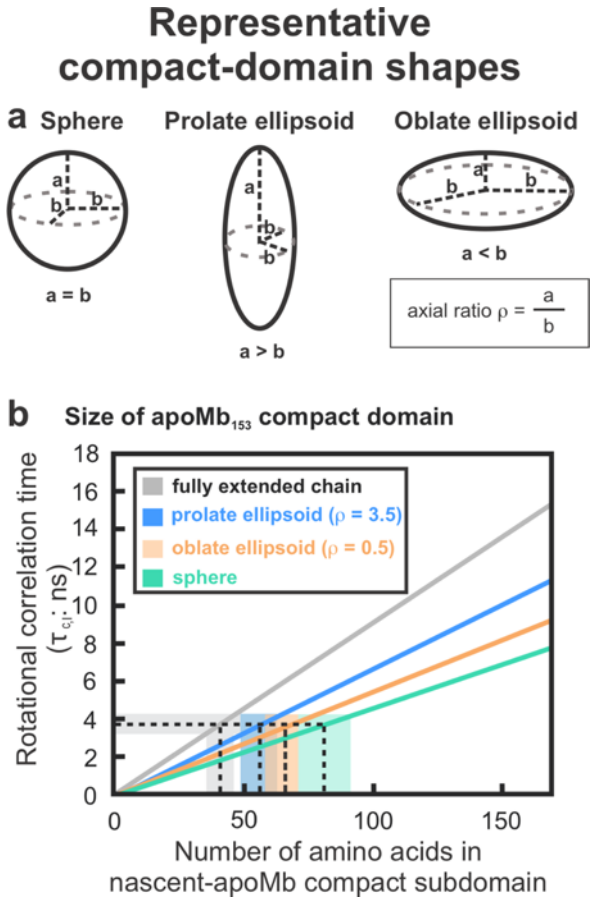


Figure 2-8. The size of the RNC compact subdomain can be deduced from the experimental rotational correlation time (for the low-ns motion) and from the experimentally determined viscosity. Three limiting molecular shapes were considered: a. spherical, prolate ellipsoid (axial ratio = 3.5) and oblate ellipsoid (axial ratio = 0.5). b. Plot illustrating the experimentally determined rotational correlation time for the intermediate-timescale motions ($\tau_{c,I}$) as a function of number of amino acids in the compact subdomain. Domain sizes compatible with experimentally determined rotational correlation times are shaded in color (blue, orange, green).

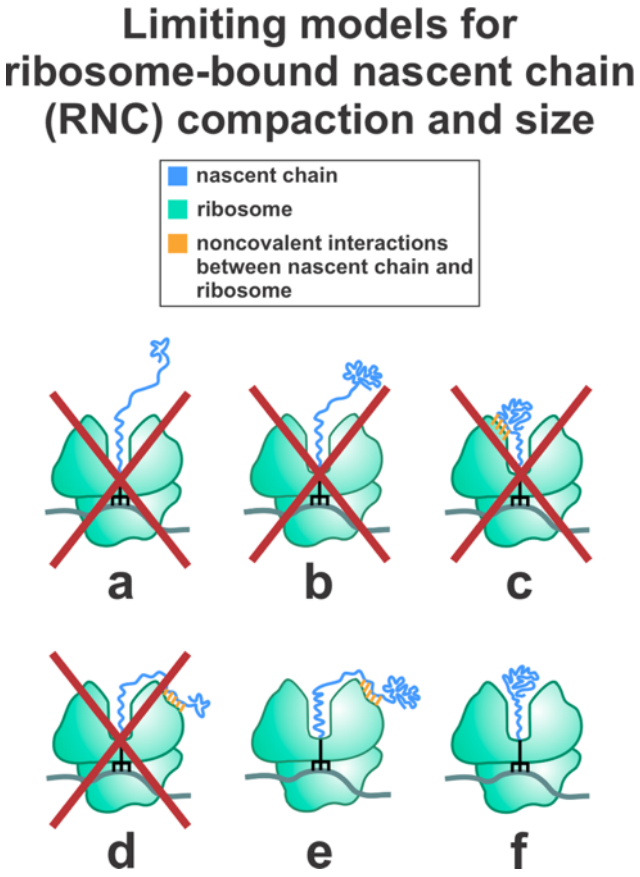


Figure 2-9. RNC models compatible with experimentally determined rotational correlation times and viscosity values.

Table 2-2. Summary of order parameters and cone semiangles for apoMb rotational motions.

Regimes of Motion[a]			
Intermediate (~microsecond or slower)		Fast (~ps)	
S_I	$\theta_0(°)$	S_F	$\theta_0(°)$
0.947 +/- 0.005	15.3 ± 0.7	S_F = 0.87 +/- 0.01	24.3 ± 0.8

[a] Data are reported as average ± SE for n = 9.

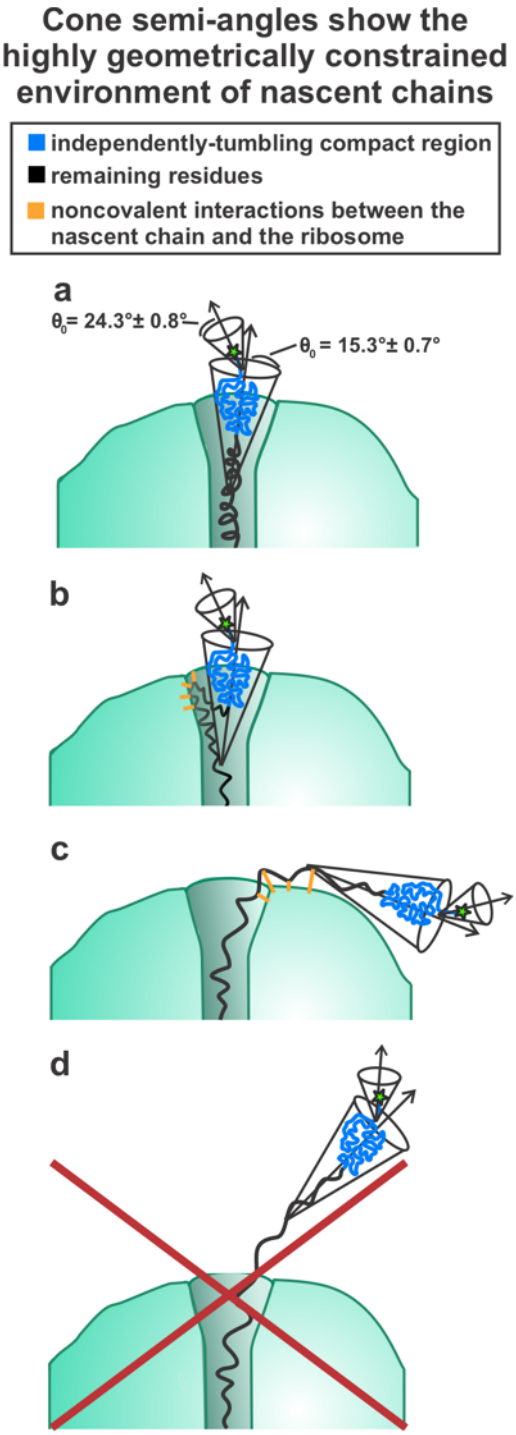


Figure 2-10. RNC models compatible with spatially-confined dynamics described by experimental cone-semi angles. The spatial confinement of the nascent chain dynamics suggests that the compact domain lies within the ribosomal vestibule or is somehow spatially confined within the outer surface of the ribosome.

ribosomal surface, causing the compact subdomain to be outside of the ribosomal tunnel altogether.

Potential interactions of nascent chains and ribosome (or chaperones) do not affect the estimated size of the compact region of the nascent chain. Until now, this work disregarded any potential interactions between nascent chains and the ribosomal surface or molecular chaperones. In principle, these interactions may affect the size of the compact region. We show here that the presence of any interactions does not affect the estimated size of the compact region.

Recent studies show that some nascent chains include populations experiencing noncovalent contacts with the ribosome[26, 124, 125] and/or chaperones.[13, 126-129]A study by Ellis *et al.*[13] showed that the observed rotational correlation time of apoMb RNCs is the same in the absence and presence of trigger factor and DnaK. This work showed that, although chaperone-bound RNC populations are likely present, their effect on fluorescence anisotropy decays is spectroscopically undetectable.[13] The arguments below show that a similar conclusion applies to nascent-chain interactions with the ribosomal surface.

First, even the most rapid (i.e., diffusion-controlled) conformational exchange of RNCs with the ribosomal surface is too slow to be fully averaged during the fluorescence anisotropy decay timescale (see derivation in Supporting Information). Thus, any interacting populations are in slow exchange on the fluorescence timescale of our measurements, and therefore correspond to distinct non-averaged contributions. Second, the effect of ribosome-interacting and non-interacting populations on the observed anisotropy-decay parameters was explicitly evaluated and shown not to affect observed rotational correction times corresponding to low-ns intermediate-timescale RNC motions.

Briefly, let us denote independently tumbling compact RNC populations as contributing (C) and populations interacting with the ribosomal surface as noncontributing (NC). Noncontributing RNCs show two rotational correlation times (fast and slow) and contributing RNCs show a third intermediate-timescale (ns) rotational correlation time corresponding to the independently tumbling nascent protein. A mixture including both populations has an anisotropy decay described by the relation below[15]

$$\frac{r_{obs}(t)}{r_0} = \left[x_C\left(1 - S_{F,C}^2\right) + x_{NC}\left(1 - S_{F,C}^2\right)\right] e^{\left(\frac{-t}{\tau_{c,F}}\right)} + x_C S_{F,C}^2\left(1 - S_{I,C}^2\right) e^{\left(\frac{-t}{\tau_{c,I}}\right)} +$$

$$\left(x_C S_{I,C}^2 S_{F,C}^2 + x_{NC} S_{F,NC}^2\right) e^{\left(\frac{-t}{\tau_{c,S}}\right)} \quad , \qquad (20)$$

where x_C and x_{NC} are the mole fractions of the contributing and noncontributing species, respectively. Equation 20 is explicitly derived in the Supporting Information. This relation follows a three-component exponential decay.

In summary, eq 20 shows that the presence of RNCs interacting with the ribosome does not affect the experimental rotational correlation times and the estimated size of the compact domain experiencing the low-ns intermediate-timescale motions.

Next, we asked whether the presence of ribosome (or chaperone) nascent-chain interactions may affect the observed amplitude of RNC ns motions. On the other hand, eq 20 also shows that the presence of any ribosome-interacting population is expected to affect the observed order parameter and cone semiangle corresponding to the intermediate-timescale motion. Hence these parameters are numerically different in the absence and presence of interactions with the ribosomal surface, as shown in Supporting Table 2-S2. Interestingly, Table 2-S2 shows that even if the contributing (i.e., non-interacting) species has a mole fraction of only 0.2, the cone-semi angle for the intermediate timescale motions is at most 39°, which is significantly less than the

maximum value of 180°. This small cone semiangle value confirms that, regardless of any ribosome (or chaperone) interactions with the nascent protein, the nascent-chain motions are highly spatially confined.

Compact region of full-length apoMb RNCs lacks residues corresponding to the two native C-terminal helices. The combined anisotropy/microscale-volume viscometry approach reveals that the compact subdomain of full-length apoMb RNCs lacks residues corresponding to the two native C-terminal helices, denoted as the G and H helices (Figure 2-11). While previous work demonstrated that the C-terminal H helix plays an important role in enabling apoMb to fold into its native state, this study provides the first experimental evidence that the major RNC compact subdomain of apoMb does not contain the residues corresponding to the G helix.

Most single-domain proteins require the C-terminal region in order to fold into their native state. The C-terminal region is important for folding because it buries nonpolar resides.[130] Fragments of these proteins without the C-terminal region have more nonpolar solvent accessible surface area than the full-length native structure, making them more likely to misfold and aggregate.[130] The C-terminal region is also important for folding because the native structures of many single domain proteins have interactions between the protein N and C-terminal regions.[131, 132] These long range interactions with the C-terminal region make post-translational folding important for the protein to reach its final native structure.[30, 133]

The experimental finding that the residues corresponding to the G helix are not contained in the compact region can be rationalized with the help of NECNOP plots. These plots predict whether proteins are capable of being folded or disordered at ambient temperature and pressure based solely on net-charge and nonpolar content.[134] NECNOP plots contain a discriminant line that separates folded proteins (right side of the line) from unfolded proteins or

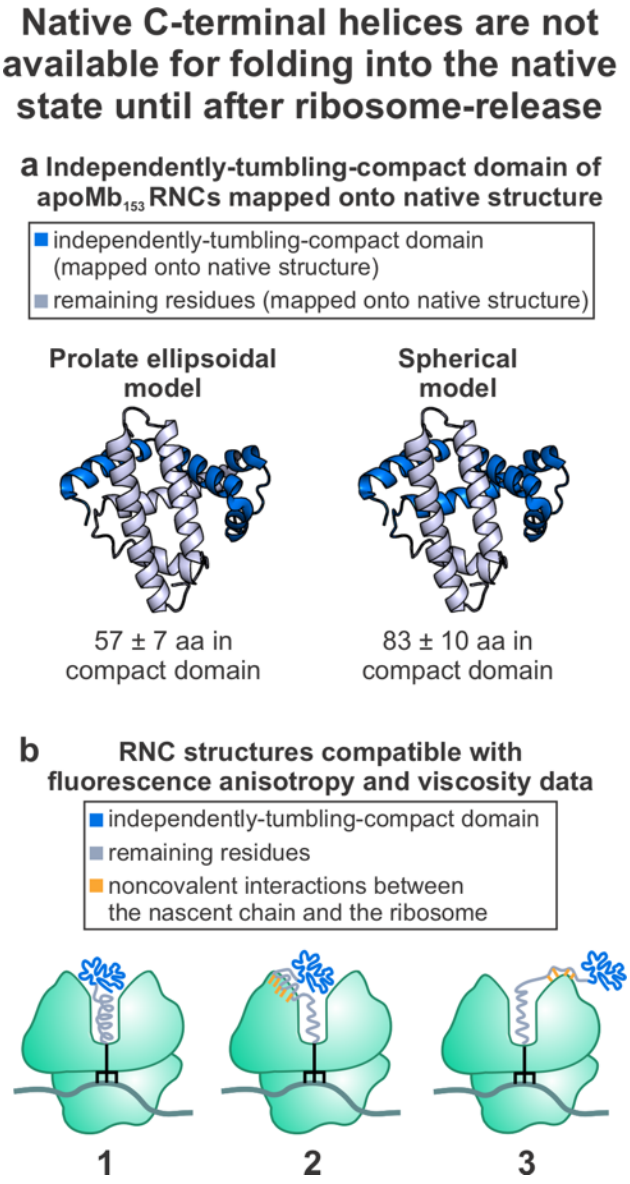


Figure 2-11. Compact nascent-apoMb subdomain does not include residues corresponding to the native C-terminal helices. a. Size of independently tumbling compact subdomain mapped onto the structure of native apoMb. b. RNC structures compatible with our experimental data.

intrinsically-disordered proteins (left side of line). The discriminant line is defined as[134]

$$|MNC| = 12.0698 \text{ x MNPC-}8.4815 \quad (21)$$

where |MNC| is the absolute value of mean net charge per residue and MNPC is the mean nonpolar content per residue.

Importantly, the NECNOP plot demonstrates that the G helix adds nonpolar residues to the chain, increasing the overall hydrophobicity of the protein. However, our work demonstrates that the full-length ribosome-bound apoMb does not bury the G-helix residues in a compact structure. Additionally, the predicted folding free energy of apoMb fragments containing the residues corresponding to the G helix but excluding the residues corresponding to the H helix is unfavorable (Figure 2-12d, predicted folding free energy calculated with FoldX version 5[135]). Folding into the native state without the H helix present is likely unfavorable because the G-helix residues do not play a significant role in burying nonpolar residues, as shown by Kurt et al. and Figure 2-12c.[130] Given that residues corresponding to the native G helix are not included in the compact subdomain of ribosome-bound apoMb, they must bury their nonpolar content in another way, perhaps by interacting with the ribosome or by forming some structure (potentially helical)[136] a compatible with the ribosomal tunnel core (Figure 2-11b).

The NECNOP plot also shows that the H-helix residues decrease the overall hydrophobicity of the protein. Therefore, the residues belonging to the H helix are not particularly nonpolar, but they are very effective at burying nonpolar residues (Figure 2-12c).[130] Additionally, the H helix has a higher residue-specific contact order (RCO) and residue contact breadth (RCB) compared to the middle region of the protein.[131] RCO is a measure of the number of long-range interactions, such as interactions between the C-terminal residues and the N-terminal residues, within a given region of a protein. RCB is a measure of the spread in sequence of the

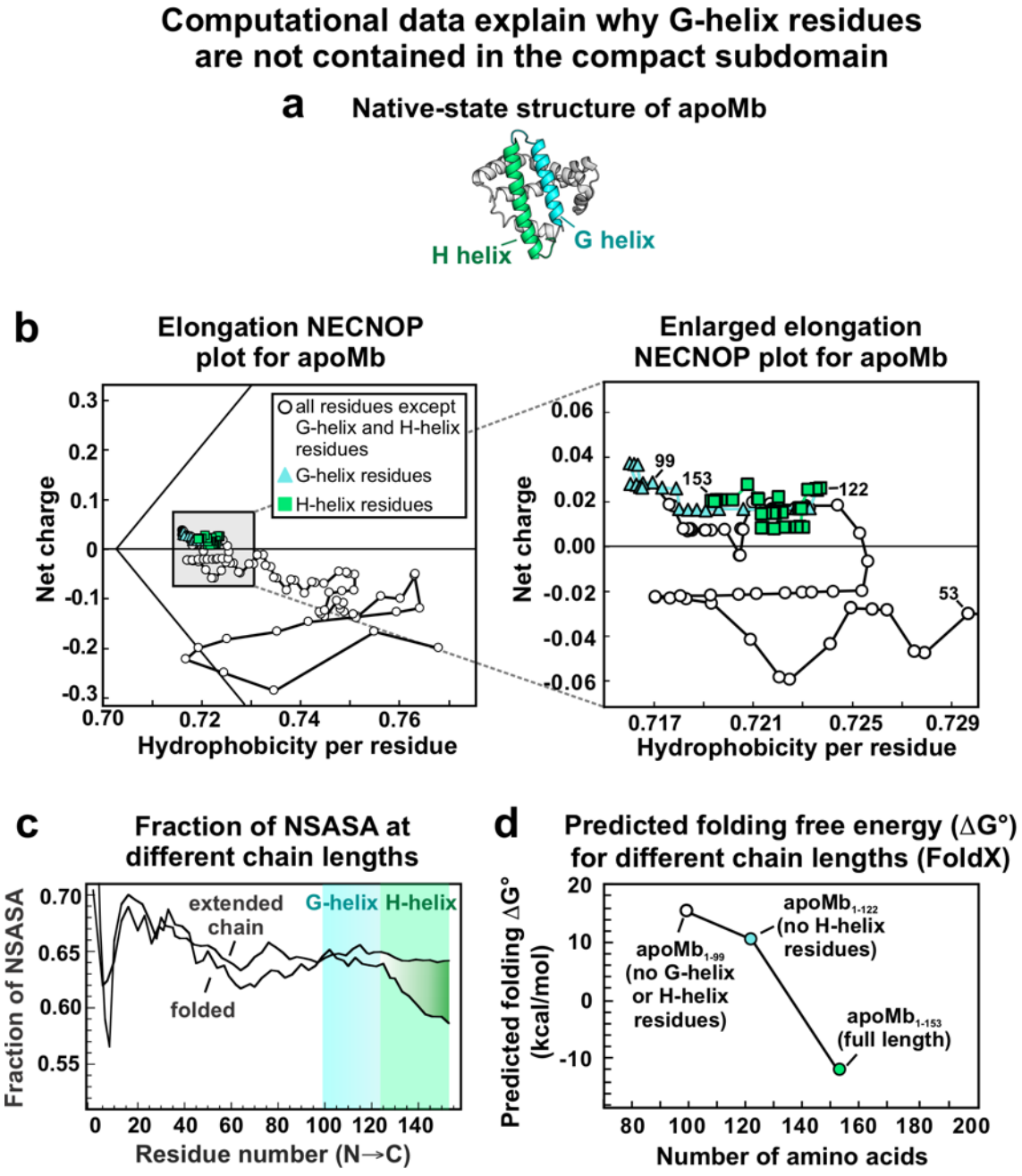


Figure 2-12. Computational data rationalizing why the nascent-apoMb compact subdomain lacks residues corresponding to the native G helix. (a) Native apoMb structure of with G and H helices colored in blue and green, respectively. (b) Elongation net-charge nonpolar (NECNOP) plot for apoMb generated according to Yaeger-Weiss *et al.*[134] The black discriminant line

separates regions corresponding to folded (right of line) and disordered proteins (left of line). (c) Fraction of nonpolar solvent-accessible surface area (NSASA) at different apoMb chain lengths. Adapted with permission from Kurt, N.; Cavagnero, S. J. Am. Chem. Soc. 2005, 127 (45), 15690-15691.[130] Copyright (2005) American Chemical Society. (d) Predicted free energy of folding for different apoMb chain lengths. See Figure 2-S6 for predicted folding free energies of additional chain lengths. All free energy calculations were carried out with FoldX version 5.[135] PDB files for apoMb (RCSB PDB: 1mbc) were generated with Pymol version 2.0.0.

interactions.[131] Therefore, the relatively high RCO and RCB of the H-helix residues demonstrate that this region interacts with regions spanning a large sequence range of the protein, including the N-terminal residues. Therefore, the H-helix residues form an integral part of the native structure. The breadth of interactions that the H-helix residues have with the rest of the protein and their ability to bury nonpolar residues contribute to the favorable folding energy when the H helix is present (Figure 2-12d).

Implications for cotranslational conformation of full-length apoMb RNCs. Our results demonstrate that the residues corresponding to the G helix of full-length ribosome-bound apoMb do not belong to the compact subdomain. Therefore, this region must bury its nonpolar residues through other interactions, which suggests that these residues either interact with the ribosomal vestibule or surface, or are involved in tight secondary structure within spatially constrained core regions of the ribosomal tunnel (Figure 2-11b) Additionally, because the compact subdomain does not contain the H helix that plays an important role in burying nonpolar residues, this compact region may be a non-native conformation that buries the N-terminal nonpolar residues until the H helix is available for folding. The compaction of nascent chain and the interaction of the residues corresponding to the G helix with the ribosome may help to keep the nascent chain soluble and prevent it from misfolding until the H-helix residues are available for folding into the native structure.

Implications for post-translational folding. Previous work demonstrated that cotranslational folding is not sufficient for apoMb to fold into its native structure, but rather that immediately post-translational events are vital for folding.[30] Immediately post-translational folding is a critical and potentially dangerous time during which nascent chains must be kinetically channeled to their native state, or they risk forming an aggregated state.[30] Our findings explain why this post-

translational period is so critical because they quantify the amount of required post-translational folding, revealing that apoMb must incorporate more than 60 amino acids into the final structure post-translationally, not just the final 30-40 residues contained within the ribosomal exit tunnel. These results demonstrate that while apoMb undergoes cotranslational compaction, this protein still requires significant post-translational folding to obtain its final native structure.

2.5 Conclusions

In summary, our results demonstrate that only a portion (57-83 amino acids, 37% - 54% of the chain) of the full-length apoMb ribosome-bound nascent chain forms a compact core. Specifically, residues belonging to the native G and H helices do not belong to this compact domain. In all, our results demonstrate that the model single-domain protein apoMb must undergo considerable additional folding after ribosome release. Therefore, our data suggest that the immediately post-translational folding of single-domain proteins is a prominent theme whose importance has been underestimated to date. Further, our findings underscore the importance of a proper supporting machinery (e.g., molecular chaperones) to prevent protein aggregation upon nascent-protein release from the ribosome. The spatial confinement of the nascent-protein local dynamics suggest that the compact subdomain is located within the ribosomal vestibule or within spatially biased regions of the outer surface of the ribosome.

Furthermore, our results demonstrate that microscale-volume viscometry combined with fluorescence anisotropy decay is a powerful approach to investigate the compaction and dynamics of ribosome-bound nascent chains. The technology presented in this work is simple, versatile, and of general applicability. Conveniently, this method only requires a single fluorophore and no C-terminal linkers.

2.6 Acknowledgments

We thank to the National Science Foundation (NSF) for funding (grants MCB-1616459 and CBET 1912259 to S.C.). In addition, R.H. thanks the National Institute of General Medical Sciences of the National Institutes of Health for a TEAM-Science Fellowship (award number R25GM083252). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are thankful for a grant from the Graduate School and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison (to S.C.). We are grateful to Valeria Guzman-Luna for help with the generation of Figure 2-S5. Finally, we thank Grace Baek and Gordon Stack (Rheosense), Javier Delgado Blanco (FoldX expert), Shih-Chu Liao (ISS) and all members of the Cavagnero group for technical assistance and helpful discussions.

2.7 References

(1) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012,** *338*, 1042-1046.

(2) Sosnick, T. R.; Barrick, D. The folding of single domain proteins—have we reached a consensus? *Curr. Opin. Struct. Biol.* **2011,** *21*, 12-24.

(3) Udgaonkar, J. B. Polypeptide chain collapse and protein folding. *Arch. Biochem. Biophys.* **2013,** *531*, 24-33.

(4) Fedyukina, D. V.; Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **2011,** *40*, 337-359.

(5) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **2016,** *353*, aac4354.

(6) Braselmann, E.; Chaney, J. L.; Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* **2013,** *38*, 337-344.

(7) Gruebele, M.; Dave, K.; Sukenik, S. Globular protein folding in vitro and in vivo. *Annu. Rev. Biophys.* **2016,** *45*, 233-251.

(8) Waudby, C. A.; Dobson, C. M.; Christodoulou, J. Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* **2019,** *44*, 914-926.

(9) Lu, J.; Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2005,** *12*, 1123-1129.

(10) Kosolapov, A.; Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2009,** *16*, 405-411.

(11) Tu, L. W.; Deutsch, C. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J. Mol. Biol.* **2010,** *396*, 1346-1360.

(12) Tu, L.; Khanna, P.; Deutsch, C. Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *J. Mol. Biol.* **2014,** *426*, 185-198.

(13) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(14) Hardesty, B.; Kramer, G. Folding of a nascent peptide on the ribosome. *Prog. Nucleic Acid Res. Mol. Biol.* **2001,** *66*, 41-66.

(15) Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci.* **2009,** *18*, 2003-2015.

(16) Kudlicki, W.; Coffman, A.; Kramer, G.; Hardesty, B. Ribosomes and ribosomal rna as chaperones for folding of proteins. *Fold. Des.* **1997,** *2*, 101-108.

(17) Chattopadhyay, S.; Das, B.; Dasgupta, C. Reactivation of denatured proteins by 23s ribosomal rna: Role of domain v. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93*, 8284-8287.

(18) Das, B.; Chattopadhyay, S.; Dasgupta, C. Reactivation of denatured fungal glucose 6-phosphate dehydrogenase and E. coli alkaline phosphatase with E. coli ribosome. *Biochem. Biophys. Res. Commun.* **1992,** *183*, 774-780.

(19) Samanta, D.; Das, A.; Bhattacharya, A.; Basu, A.; Das, D.; DasGupta, C. Mechanism of ribosome assisted protein folding: A new insight into rRNA functions. *Biochem. Biophys. Res. Commun.* **2009,** *384*, 137-140.

(20) Das, D.; Samanta, D.; Hasan, S.; Das, A.; Bhattacharya, A.; Dasgupta, S.; Chakrabarti, A.; Ghorai, P.; Das Gupta, C. Identical rna-protein interactions in vivo and in vitro and a scheme of folding the newly synthesized proteins by ribosomes. *J. Biol. Chem.* **2012,** *287*, 37508-37521.

(21) Pang, Y. H.; Kovachev, P.; Sanyal, S. Ribosomal rna modulates aggregation of the podospora prion protein HET-s. *Int. J. Mol. Sci.* **2020,** *21*, ijms21176340.

(22) Ziv, G.; Haran, G.; Thirumalai, D. Ribosome exit tunnel can entropically stabilize α-helices. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 18956-18961.

(23) 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*, 725-736.

(24) Cruz-Vera, L. R.; Rajagopal, S.; Squires, C.; Yanofsky, C. Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell* **2005,** *19*, 333-343.

(25) Houben , E. N. G.; Zarivach , R.; Oudega , B.; Luirink , J. Early encounters of a nascent membrane protein : Specificity and timing of contacts inside and outside the ribosome. *J. Cell Biol.* **2005,** *170*, 27-35.

(26) Guzman-Luna, V.; Fuchs, A. M.; Cavagnero, S. Interaction of an intrinsically disordered nascent protein with specific regions of the ribosomal surface. **2021**, Under revision.

(27) Ullers, R. S.; Houben, E. N.; Raine, A.; ten Hagen-Jongman, C. M.; Ehrenberg, M.; Brunner, J.; Oudega, B.; Harms, N.; Luirink, J. Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J. Cell Biol.* **2003,** *161*, 679-684.

(28) Peterson, J. H.; Woolhead, C. A.; Bernstein, H. D. The conformation of a nascent polypeptide inside the ribosome tunnel affects protein targeting and protein folding. *Mol. Microbiol.* **2010,** *78*, 203-217.

(29) Wang, S.; Jomaa, A.; Jaskolowski, M.; Yang, C.; Ban, N.; Shan, S. The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat. Struct. Mol. Biol.* **2019,** *26*, 919-929.

(30) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(31) Dao Duc, K.; Batra, S. S.; Bhattacharya, N.; Cate, Jamie H. D.; Song, Yun S. Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res.* **2019,** *47*, 4198-4210.

(32) Noeske, J.; Wasserman, M. R.; Terry, D. S.; Altman, R. B.; Blanchard, S. C.; Cate, J. H. D. High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.* **2015,** *22*, 336-341.

(33) Tsalkova, T.; Odom, O. W.; Kramer, G.; Hardesty, B. Different conformations of nascent peptides on ribosomes. *J. Mol. Biol.* **1998,** *278*, 713-723.

(34) Kramer, G.; Ramachandiran, V.; Hardesty, B. Cotranslational folding — omnia mea mecum porto? *Int. J. Biochem. Cell B.* **2001,** *33*, 541-553.

(35) Malkin, L. I.; Rich, A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* **1967,** *26*, 329-346.

(36) Nilsson, Ola B.; Hedman, R.; Marino, J.; Wickles, S.; Bischoff, L.; Johansson, M.; Müller-Lucks, A.; Trovato, F.; Puglisi, Joseph D.; O'Brien, Edward P., et al. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* **2015,** *12*, 1533-1540.

(37) Marino, J.; von Heijne, G.; Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **2016,** *590*, 655-660.

(38) Bhushan, S.; Gartmann, M.; Halic, M.; Armache, J.-P.; Jarasch, A.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. A-helical nascent polypeptide chains

visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2010,** *17*, 313-317.

(39) Lu, J.; Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **2005,** *44*, 8230-8243.

(40) Liutkute, M.; Maiti, M.; Samatova, E.; Enderlein, J.; Rodnina, M. V. Gradual compaction of the nascent peptide during cotranslational folding on the ribosome. *eLife* **2020,** *9*, e60895.

(41) Farías-Rico, J. A.; Ruud Selin, F.; 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. USA* **2018,** *115*, E9280-E9287.

(42) Tian, P.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; von Heijne, G.; Clarke, J., et al. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci. USA* **2018,** *115*, E11284-E11293.

(43) Nilsson, O. B.; Nickson, A. A.; Hollins, J. J.; Wickles, S.; Steward, A.; Beckmann, R.; von Heijne, G.; Clarke, J. Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **2017,** *24*, 221-225.

(44) Liutkute, M.; Samatova, E.; Rodnina, M. V. Cotranslational folding of proteins on the ribosome. *Biomolecules* **2020,** *10*, biom10010097.

(45) Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S. In vitro expression and characterization of native apomyoglobin under low molecular crowding conditions. *Protein Expr. Purif.* **2006,** *45*, 381-392.

(46) Gesteland, R. F. Isolation and characterization of ribonuclease I mutants of Escherichia coli. *J. Mol. Biol.* **1966,** *16*, 67-84.

(47) Springer, B. A.; Sligar, S. G. High-level expression of sperm whale myoglobin in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84*, 8961-8965.

(48) Behrmann, M.; Koch, H. G.; Hengelage, T.; Wieseler, B.; Hoffschulte, H. K.; Müller, M. Requirements for the translocation of elongation-arrested, ribosome-associated OmpA across the plasma membrane of Escherichia coli. *J. Biol. Chem.* **1998,** *273*, 13898-13904.

(49) Hanes, J.; Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94*, 4937-4942.

(50) Isolation and analysis of ribosomes from prokaryotes, eukaryotes, and organelles. In *Ribosomes and protein synthesis: A practical approach*, Spedding, G., Ed.; Oxford University Press: Oxford, U.K., 1990; pp 1-27.

(51) Odom, O. W.; Hardesty, B. Use of 50 S-binding antibiotics to characterize the ribosomal site to which peptidyl-tRNA is bound. *J. Biol. Chem.* **1992,** *267*, 19117-19122.

(52) Ziehr, D. R.; Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Production of ribosome-released nascent proteins with optimal physical properties. *Anal. Chem.* **2010,** *82*, 4637-4643.

(53) Moore, S. D.; Baker, T. A.; Sauer, R. T. Forced extraction of targeted components from complex macromolecular assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 11685-11690.

(54) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9*, 671-675.

(55) Abràmoff, M. D.; Magalhães, P. J.; Ram, S. J. Image processing with ImageJ. *Biophotonics international* **2004,** *11*, 36-42.

(56) Beechem, J. M.; Gratton, E. Fluorescence spectroscopy data analysis environment: A second generation global analysis program. In *Time-resolved laser spectroscopy in biochemistry*, Lakowicz, J. R., Ed.; SPIE, Bellingham, 1988; pp 70-81.

(57) Ross, J. A.; Jameson, D. M. Time-resolved methods in biophysics. 8. Frequency domain fluorometry: Applications to intrinsic protein fluorescence. *Photochem. Photobiol. Sci.* **2008,** *7*, 1301-1312.

(58) Weinreis, S. A.; Ellis, J. P.; Cavagnero, S. Dynamic fluorescence depolarization: A powerful tool to explore protein folding on the ribosome. *Methods* **2010,** *52*, 57-73.

(59) Small, E. W.; Isenberg, I. Hydrodynamic properties of a rigid molecule: Rotational and linear diffusion and fluorescence anisotropy. *Biopolymers* **1977,** *16*, 1907-1928.

(60) Lakowicz, J. R. *Principles of fluorescence spectroscopy*. 3rd ed.; New York : Plenum Press: 2006.

(61) Kirchdoerfer, R. N.; Huang, J. J.; Isola, M. K.; Cavagnero, S. Fluorescence-based analysis of aminoacyl- and peptidyl-tRNA by low-pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* **2007,** *364*, 92-4.

(62) Schägger, H.; von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **1987,** *166*, 368-79.

(63) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** *181*, 662-666.

(64) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. Structure of myoglobin: A three-dimensional fourier synthesis at 2 å. Resolution. *Nature* **1960,** *185*, 422-427.

(65) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262*, 892.

(66) Nishimura, C.; Dyson, H. J.; Wright, P. E. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J. Mol. Biol.* **2006,** *355*, 139-156.

(67) Eliezer, D.; Wright, P. E. Is apomyoglobin a molten globule? Structural characterization by NMR. *J. Mol. Biol.* **1996,** *263*, 531-8.

(68) Dyson, H. J.; Wright, P. E. How does your protein fold? Elucidating the apomyoglobin folding pathway. *Acc. Chem. Res.* **2017,** *50*, 105-111.

(69) Armstrong, B. D.; Choi, J.; Lopez, C.; Wesener, D. A.; Hubbell, W.; Cavagnero, S.; Han, S. Site-specific hydration dynamics in the nonpolar core of a molten globule by dynamic nuclear polarization of water. *J. Am. Chem. Soc.* **2011,** *133*, 5987-5995.

(70) Samatova, E. N.; Melnik, B. S.; Balobanov, V. A.; Katina, N. S.; Dolgikh, D. A.; Semisotnov, G. V.; Finkelstein, A. V.; Bychkova, V. E. Folding intermediate and folding nucleus for i→n and u→i→n transitions in apomyoglobin: Contributions by conserved and nonconserved residues. *Biophys. J.* **2010,** *98*, 1694-1702.

(71) Balobanov, V. A.; Katina, N. S.; Finkelstein, A. V.; Bychkova, V. E. Intermediate states of apomyoglobin: Are they parts of the same area of conformations diagram? *Biochem.-Moscow* **2017,** *82*, 625-631.

(72) Ptitsyn, O. B. Molten globule and protein folding. *Adv. Protein Chem.* **1995,** *47*, 83-229.

(73) Jungbauer, L. M.; Bakke, C. K.; Cavagnero, S. Experimental and computational analysis of translation products in apomyoglobin expression. *J. Mol. Biol.* **2006,** *357*, 1121-1143.

(74) Jameson, D. M.; Gratton, E.; Hall, R. D. The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry. *Appl. Spectrosc. Rev.* **1984,** *20*, 55-106.

(75) Bucci, E.; Steiner, R. F. Anisotropy decay of fluorescence as an experimental approach to protein dynamics. *Biophys. Chem.* **1988,** *30*, 199-224.

(76) Feinstein, E.; Deikus, G.; Rusinova, E.; Rachofsky, E. L.; Ross, J. B. A.; Laws, W. R. Constrained analysis of fluorescence anisotropy decay:Application to experimental protein dynamics. *Biophys. J.* **2003,** *84*, 599-611.

(77) Majumdar, A.; Mukhopadhyay, S. Fluorescence depolarization kinetics to study the conformational preference, structural plasticity, binding, and assembly of intrinsically disordered proteins. *Methods Enzymol.* **2018,** *611*, 347-381.

(78) Shi, X.; Herschlag, D. Fluorescence polarization anisotropy to measure rna dynamics. *Methods Enzymol.* **2009,** *469*, 287-302.

(79) Liu, C.; Liang, G.; Liu, Z.; Zu, L. Time-resolved fluorescence anisotropy study of the interaction between DNA and a peptide truncated from the p53 protein core domain. *J. Fluoresc.* **2014,** *24*, 533-539.

(80) Jain, N.; Narang, D.; Bhasne, K.; Dalal, V.; Arya, S.; Bhattacharya, M.; Mukhopadhyay, S. Direct observation of the intrinsic backbone torsional mobility of disordered proteins. *Biophys. J.* **2016,** *111*, 768-774.

(81) Sahay, S.; Anoop, A.; Krishnamoorthy, G.; Maji, S. K. Site-specific fluorescence dynamics of α-synuclein fibrils using time-resolved fluorescence studies: Effect of familial Parkinson's disease-associated mutations. *Biochemistry* **2014,** *53*, 807-809.

(82) Motyka, A. L. An introduction to rheology with an emphasis on application to dispersions. *J. Chem. Educ.* **1996,** *73*, 374-380.

(83) Jary, D.; Sikorav, J. L.; Lairez, D. Nonlinear viscoelasticity of entangled DNA molecules. *EPL* **1999,** *46*, 251-255.

(84) Picchi, D.; Poesio, P.; Ullmann, A.; Brauner, N. Characteristics of stratified flows of newtonian/non-newtonian shear-thinning fluids. *Int. J. Multiph. Flow* **2017,** *97*, 109-133.

(85) Chhabra, R. P. Non-newtonian fluids: An introduction. In *Rheology of complex fluids* Krishnan, J.; Deshpande, A.; Kumar, P., Eds.; Springer: New York, 2010; pp 3-34.

(86) Ikeda, S.; Nishinari, K. Solid-like mechanical behaviors of ovalbumin aqueous solutions. *Int. J. Biol. Macromol.* **2001,** *28*, 315-320.

(87) Ikeda, S.; Nishinari, K. On solid-like rheological behaviors of globular protein solutions. *Food Hydrocoll.* **2001,** *15*, 401-406.

(88) Matsumoto, T.; Inoue, H. Colloidal structure and properties of bovine serum globulin aqueous systems using SAXS and rheological measurements. *Chem. Phys.* **1996,** *207*, 167-172.

(89) Inoue, H.; Matsumoto, T. Viscoelastic characterization of solid-like structure in aqueous colloids of globular proteins. *Colloids Surf. A Physicochem. Eng. Asp.* **1996,** *109*, 89-96.

(90) Sharma, V.; Jaishankar, A.; Wang, Y.-C.; McKinley, G. H. Rheology of globular proteins: Apparent yield stress, high shear rate viscosity and interfacial viscoelasticity of bovine serum albumin solutions. *Soft Matter* **2011,** *7*, 5150-5160.

(91) Lu, J. R.; Su, T. J.; Thomas, R. K. Structural conformation of bovine serum albumin layers at the air–water interface studied by neutron reflection. *J. Colloid Interface Sci.* **1999,** *213*, 426-437.

(92) Gouveia, S. M.; Tiffany, J. M. Human tear viscosity: An interactive role for proteins and lipids. *Biochim. Biophys. Acta. Proteins Proteom.* **2005,** *1753*, 155-163.

(93) Allen, S. H.; Wong, K.-P. A comparative study on the hydrodynamic shape, conformation, and stability of E. coli ribosomal subunits in reconstitution buffer. *Arch. Biochem. Biophys.* **1979,** *195*, 112-120.

(94) Taylor, N.; Elbaum-Garfinkle, S.; Vaidya, N.; Zhang, H.; Stone, H. A.; Brangwynne, C. P. Biophysical characterization of organelle-based rna/protein liquid phases using microfluidics. *Soft Matter* **2016,** *12*, 9142-9150.

(95) Juarez, G.; Arratia, P. E. Extensional rheology of DNA suspensions in microfluidic devices. *Soft Matter* **2011,** *7*, 9444-9452.

(96) Ciccolini, L. A. S.; Ayazi Shamlou, P.; Titchener-Hooker, N. J.; Ward, J. M.; Dunnill, P. Rheological properties of chromosomal and plasmid DNA during alkaline lysis reaction. *Bioprocess Eng.* **1999,** *21*, 231-237.

(97) Wang, Y.; Li, C.; Pielak, G. J. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* **2010,** *132*, 9392-9397.

(98) Mika, J. T.; Poolman, B. Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotechnol.* **2011,** *22*, 117-126.

(99) Dix, J. A.; Verkman, A. S. Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.* **2008,** *37*, 247-263.

(100) Tabaka, M.; Kalwarczyk, T.; Szymanski, J.; Hou, S.; Holyst, R. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front. Phys.* **2014,** *2*, fphy.2014.00054.

(101) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J., et al. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014,** *114*, 6661-6714.

(102) Li, C.; Wang, Y.; Pielak, G. J. Translational and rotational diffusion of a small globular protein under crowded conditions. *J. Phys. Chem. B* **2009,** *113*, 13390-13392.

(103) Kuttner, Y. Y.; Kozer, N.; Segal, E.; Schreiber, G.; Haran, G. Separating the contribution of translational and rotational diffusion to protein association. *J. Am. Chem. Soc.* **2005,** *127*, 15138-15144.

(104) Wiśniewska, A.; Sozański, K.; Kalwarczyk, T.; Kędra-Królik, K.; Pieper, C.; Wieczorek, S. A.; Jakieła, S.; Enderlein, J.; Hołyst, R. Scaling of activation energy for macroscopic flow in poly(ethylene glycol) solutions: Entangled – non-entangled crossover. *Polymer* **2014,** *55*, 4651-4657.

(105) Zosel, F.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Schuler, B. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117*, 13480-13489.

(106) Zorrilla, S.; Hink, M. A.; Visser, A. J. W. G.; Lillo, M. P. Translational and rotational motions of proteins in a protein crowded environment. *Biophys. Chem.* **2007,** *125*, 298-305.

(107) McGuffee, S. R.; Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput. Biol.* **2010,** *6*, e1000694.

(108) Tuinier, R.; Dhont, J. K. G.; Fan, T. H. How depletion affects sphere motion through solutions containing macromolecules. *EPL* **2006,** *75*, 929-935.

(109) Asakura, S.; Oosawa, F. Interaction between particles suspended in solutions of macromolecules. *J. Polym. Sci.* **1958,** *33*, 183-192.

(110) Vrij, A. Polymers at interfaces and the interactions in colloidal dispersions. *Pure Appl. Chem.* **1976,** *48*, 471-483.

(111) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116*, 9843-9852.

(112) Fedyukina, D. V.; Jennaro, T. S.; Cavagnero, S. Charge segregation and low hydrophobicity are key features of ribosomal proteins from different organisms. *J. Biol. Chem.* **2014,** *289*, 6740-6750.

(113) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 10037-10041.

(114) Record, M. T., Jr.; Zhang, W.; Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, hofmeister effects, and osmotic effects of salts. *Adv. Protein Chem.* **1998,** *51*, 281-353.

(115) Anderson, C. F.; Record, M. T. Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.* **1995,** *46*, 657-700.

(116) García-García, C.; Draper, D. E. Electrostatic interactions in a peptide--rna complex. *J. Mol. Biol.* **2003,** *331*, 75-88.

(117) Suresha, P. R.; Badiger, M. V.; Wolf, B. A. Polyelectrolytes in dilute solution: Viscometric access to coil dimensions and salt effects. *RSC Adv.* **2015,** *5*, 27674-27681.

(118) Izzo, D.; Cloitre, M.; Leibler, L. The viscosity of short polyelectrolyte solutions. *Soft Matter* **2014,** *10*, 1714-1722.

(119) Jameson, D. M.; Hazlett, T. L. In *Biophysical and biochemical aspects of fluorescence spectroscopy*, Dewey, T. G., Ed.; Plenum Press: New York, 1991; pp 105-133.

(120) Cantor, C. R.; Schimmel, P. R. *Biophysical chemistry*. W.H. Freeman and Company: New York, 1980.

(121) Berezin, M. Y.; Achilefu, S. Fluorescence lifetime measurements and biological imaging. *Chem. Rev.* **2010,** *110*, 2641-2684.

(122) Lipari, G.; Szabo, A. Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophys. J.* **1980,** *30*, 489-506.

(123) Kinosita, K.; Ikegami, A.; Kawato, S. On the wobbling-in-cone analysis of fluorescence anisotropy decay. *Biophys. J.* **1982,** *37*, 461-464.

(124) Knight, A. M.; Culviner, P. H.; Kurt-Yilmaz, N.; Zou, T.; Ozkan, S. B.; Cavagnero, S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **2013,** *8*, 1195-1204.

(125) Deckert, A.; Waudby, C. A.; Wlodarski, T.; Wentink, A. S.; Wang, X.; Kirkpatrick, J. P.; Paton, J. F. S.; Camilloni, C.; Kukic, P.; Dobson, C. M., et al. Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113*, 5012-5017.

(126) Hoffmann, A.; Merz, F.; Rutkowska, A.; Zachmann-Brand, B.; Deuerling, E.; Bukau, B. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.* **2006,** *281*, 6539-6545.

(127) Hoffmann, A.; Bukau, B.; Kramer, G. Structure and function of the molecular chaperone trigger factor. *Biochim. Biophys. Acta Mol. Cell Res.* **2010,** *1803*, 650-661.

(128) Kramer, G.; Shiber, A.; Bukau, B. Mechanisms of cotranslational maturation of newly synthesized proteins. *Annu. Rev. Biochem.* **2019,** *88*, 337-364.

(129) Mayer, M. P.; Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **2005,** *62*, 670-684.

(130) Kurt, N.; Cavagnero, S. The burial of solvent-accessible surface area is a predictor of polypeptide folding and misfolding as a function of chain elongation. *J. Am. Chem. Soc.* **2005,** *127*, 15690-15691.

(131) Kurt, N.; Mounce, B. C.; Ellison, P. A.; Cavagnero, S. Residue-specific contact order and contact breadth in single-domain proteins: Implications for folding as a function of chain elongation. *Biotechnol. Prog.* **2008,** *24*, 570-575.

(132) Mounce, B. C.; Kurt, N.; Ellison, P. A.; Cavagnero, S. Nonrandom distribution of intramolecular contacts in native single-domain proteins. *Proteins* **2009,** *75*, 404-412.

(133) Zhao, V.; Jacobs, W. M.; Shakhnovich, E. Effect of protein structure on evolution of cotranslational folding. *Biophys. J.* **2020,** *119*, 1123-1134.

(134) Yaeger-Weiss, S. K.; Jennaro, T. S.; Mecha, M.; Becker, J. H.; Yang, H.; Winkler, G. L. W.; Cavagnero, S. Net charge and nonpolar content guide the identification of folded and prion proteins. *Biochemistry* **2020,** *59*, 1881-1895.

(135) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The foldx web server: An online force field. *Nucleic Acids Res.* **2005,** *33*, W382-W388.

(136) 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*, 274-282.

2.8 Appendix

Supporting Information Description

Puromycin-assay description and representative SDS-PAGE gels, comparisons between steady-state anisotropy and time-domain and frequency-domain anisotropy decays, determination of reduced χ^2 values, potential effect of ribosome−nascent-chain interactions on anisotropy parameters, oligonucleotide sequences used in cell-free reactions, anisotropy-decay graphs for all the data collected in this work, and predicted standard-state folding free energies for several nascent-chain lengths.

Supporting Text

Puromycin assay. A puromycin assay was used to verify that nascent proteins are ribosome-bound.[1, 2] For this assay, the RNC sample was split into two aliquots. Puromycin was added to one aliquot to reach a final puromycin concentration of 1 mM. This aliquot was incubated for 30 min at 37 °C. The two aliquots were analyzed with an SDS-PAGE gel as shown in Figure 2-S1. The sample treated with puromycin shows a smaller molecular weight than the sample lacking puromycin, demonstrating that this nascent chain had been released from the ribosome, and confirming that the sample was ribosome-bound prior to the addition of puromycin. The puromycin treatment was performed on the samples employed for fluorescence measurements after completion of fluorescence and microscale-volume viscometry data collection.

Steady-state versus anisotropy-decay measurements. In steady-state fluorescence anisotropy, the excitation light source is continuous, and it has a constant intensity. The excitation time is considerably longer than the sample fluorescent lifetime, so the measured fluorescence anisotropy is averaged over time. Time-averaged anisotropy and rotational correlation time

provide information about the average size and shape for the whole sample. However, steady-state methods cannot resolve independent rotational motions within the same molecule, each associated with a different rotational correlation time.

In contrast, fluorescence anisotropy decay measurements are based on a more complex excitation scheme. Pulsed excitation is used for time-domain anisotropy measurements and sinusoidally-modulated excitation amplitude is used for frequency-domain measurements. In both cases, changing the intensity of the excitation source enables determining fluorescence-anisotropy decays, as long as they take place within the fluorophore lifetime. Multiple fluorescence anisotropy decays occurring within the same sample lead to the detection of multiple rotational correlation times provided that these times are (a) comparable to or shorter in duration than the fluorophore's fluorescent lifetime and (b) they significantly differ from each other, typically by at least a factor of five.

Time-dependent fluorescence anisotropy as a function of time is defined as

$$r(t) = \frac{(I_{\parallel}(t) - I_{\perp}(t))}{(I_{\parallel}(t) - 2I_{\perp}(t))} \quad . \tag{1}$$

For a spherical species with a single fluorescent lifetime τ_F and a global rotational correlation time τ_c, $I_{\parallel}(t)$ and $I_{\perp}(t)$ are

$$I_{\parallel}(t) = \left[\frac{1}{3} + \left(\frac{4}{15}\right) e^{-\frac{t}{\tau_c}\frac{(3\cos^2\xi - 1)}{2}}\right] e^{-\frac{t}{\tau_F}} \tag{2}$$

and

$$I_{\perp}(t) = \left[\frac{1}{3} - \left(\frac{2}{15}\right) e^{-\frac{t}{\tau_c}\frac{(3\cos^2\xi - 1)}{2}}\right] e^{-\frac{t}{\tau_F}} \quad . \tag{3}$$

Substitution of eqs 2 and 3 into eq 1 gives eq 4, which defines the time-dependent fluorescence anisotropy of a sample with a single rotational correlation time

$$r(t) = r_0 e^{-t/\tau_c} \quad , \tag{4}$$

where r_0 is the fundamental anisotropy at time = 0. The steady-state fluorescence anisotropy is the normalized average of the time-dependent anisotropy decay *r(t)* weighted by the fluorescence intensity as shown in eq 5

$$r = \frac{\int_0^\infty r(t)I(t)dt}{\int_0^\infty I(t)dt} \quad . \tag{5}$$

Fluorescence anisotropy decay measurements can resolve the timescales and amplitudes of multiple motions experienced by complex fluorophore-labeled macromolecules, including global and local motions. The fluorescence anisotropy decay of a sample characterized by multiple rotational motions is defined as

$$r(t) = r_0\left(\sum_i F_i e^{-t/\tau_{c,i}}\right) \ . \tag{6}$$

Time-domain versus frequency-domain anisotropy-decay measurements. As mentioned in the previous section, frequency-domain (FD) measurements involve sample excitation via a continuous monochromatic light source with sinusoidally-modulated intensity.[3] Time-domain (TD) and FD data can be interconverted via Laplace transforms (Figure 2-S2). The direct conversion of raw FD data into raw TD data, however, is unnecessary. Therefore the FD and TD methodologies are of comparable practical convenience.[4] While TD instruments are often more sensitive than FD instruments, FD measurements are typically able to more effectively resolve multi-component anisotropy decays (especially if they encompass a wide temporal regime, e.g. from ps to 50-100 ns) than TD measurements.[3] In this work, we exploited the FD methodology to monitor fluorescence anisotropy decays.

In FD measurements, the sinusoidally-modulated intensity of the excitation light results in sinusoidal variation in the intensity of the emitted light that is offset in phase and amplitude relative to the excitation light. Before performing fluorescence anisotropy measurements,

fluorophore lifetimes are determined from the phase shift (ϕ) and modulation (M) of the emitted light, upon exciting the sample with unpolarized light. The modulation M is defined as

$$M = \frac{AC_{EM}/DC_{EM}}{AC_{EX}/DC_{EX}} \quad , \tag{7}$$

where AC is the amplitude of the alternating current, DC is the magnitude of the direct current, and the subscripts EX and EM denote the excitation and emission light respectively.[3]

After determining the fluorescence lifetime(s) of the fluorophore, polarized excitation is used to measure the anisotropy decay. For these measurements, the phase-shift difference ($\Delta\phi$) and modulation ratio (Y) are measured for a range of excitation light modulation frequencies. $\Delta\phi$ and Y are defined as

$$\Delta\phi = \phi_{\parallel} - \phi_{\perp} \tag{8}$$

$$\mathrm{Y} = \frac{AC_{EM,\perp}}{AC_{EM,\parallel}} \tag{9}$$

$\phi_{\parallel}$ and $\phi_{\perp}$ are the phase shifts of the parallel and perpendicular components of the polarized emission, respectively.[5] Nonlinear least squares fitting of the graphs of $\Delta\phi$ and Y versus the modulation frequency (ω) is used to determine the rotational correlation time of the fluorophore.[5] As mentioned above, FD measurements often provide higher resolution of multi-component anisotropy decays than TD measurements. This feature may be contributed by the fact that FD anisotropy measurements entail collecting two independent quantities, $\Delta\phi(\omega)$ and $Y(\omega)$, to determine rotational correlation times, while TD anisotropy measurements rely on the assessment of a single parameter, $r(t)$.[6] FD fluorescence anisotropy decay measurements can differentiate the local dynamics of a fluorophore from the rotational motions of the macromolecule to which it is connected. These measurements can resolve up to three rotational motions as long as the timescale for the motions is faster or comparable to the lifetime of the fluorophore, and the different rotational correlation times differ by at least a factor of ca. five.[3]

Multi-component anisotropy decays depend on two parameters, the rotational correlation times and the fractional contributions of the different motions, as described by

$$r(t) = r_0\left(\sum_i F_i e^{-t/\tau_{c,i}}\right) \quad (10)$$

where t is time, F_i is the fractional contribution of each anisotropy decay term, and $\tau_{c,i}$ is the rotational correlation time for each anisotropy decay term.[3] RNCs can show either two or three rotational motions, so their anisotropy decays are described by either eq 11 (two motions) or eq 12 (three motions).

$$\frac{r(t)}{r(0)} = F_s e^{-t/\tau_{c,S}} + F_F e^{-t/\tau_{c,F}} \quad (11)$$

$$\frac{r(t)}{r(0)} = F_s e^{-t/\tau_{c,S}} + F_I e^{-t/\tau_{c,I}} + F_F e^{-t/\tau_{c,F}} \quad , \quad (12)$$

where F_s, F_I, and F_F are the fractional contribution to the anisotropy decay from the slow, intermediate, and fast rotational motions, respectively.[3] $\tau_{c,S}$, $\tau_{c,I}$, and $\tau_{c,F}$ denote the rotational correlation times for the slow, intermediate, and fast motions. Figure 2-5a shows simulated anisotropy decay data for RNCs with 2 and 3 rotational motions. RNCs with three rotational motions show an increase in the phase shift difference at intermediate modulation frequencies, resulting in a "bump" in the phase change data. The "bump" spectral feature indicates the presence of motions on the ns timescale.[3] In order to determine whether a sample shows two or three rotational motions, the anisotropy decay data is fit with both a two-component and a three-component exponential decay and the χ^2 values for both fits are compared. The three-component fit must have a χ^2 value at least 2.5-fold smaller than the χ^2 value for the two-component fit for the presence of three rotational motions to be confirmed.[3]

Determination of reduced χ^2 values. In order to determine whether the fluorescence anisotropy decay data fit best to a two- or three-component model, reduced χ^2 values resulting from fits to

both models (evaluated independently) were computed. Reduced χ^2 values show how well curve fits match experimental results. The following equation was used to determine reduced χ^2 values

$$\chi^2 = \frac{1}{2n-f-1}\sum_{\omega}\left(\frac{\Delta\phi_{\omega}-\Delta\phi_{c\omega}}{\delta\Delta\phi}\right)^2 + \frac{1}{2n-f-1}\sum_{\omega}\left(\frac{Y_{\omega}-Y_{c\omega}}{\delta Y}\right)^2 \quad , \tag{13}$$

where n is the number of experimental light intensity modulation frequencies, f is the number of free-floating parameters used in the least-squares analysis, $\Delta\phi_{\omega}$ and $\Delta\phi_{c\omega}$ denote the experimental and calculated phase shift difference, Y_{ω} and $Y_{c\omega}$ are the experimental and calculated modulation ratios, and $\delta\Delta\phi$ and δY are the uncertainties in the phase shift difference and modulation ratio.[3, 7] The last two parameters were set to 0.2° and 0.004, respectively.[3, 5] Lowest possible reduced χ^2 values (< 5) are desirable. If χ^2 decreased by a factor of >2.5 for the three-component fit relative to the two-component fit, the three-component fit was determined to be the best fit for the data, consistent analogous criteria employed in the literature.[3, 8]

RNC populations that may be interacting with ribosomal surface (or chaperones) are in slow exchange on the fluorescence-lifetime timescale. As schematically illustrated in Figure 2-S4, it is in principle possible that any nascent-chain interactions with the ribosomal surface may affect observed fluorescence-anisotropy decay parameters. In order to test whether this is true, we started by estimating the diffusion-limited association rate constant and matching timescales for interactions between the nascent protein and the ribosome. The goal of this section is to use kinetic arguments to show that RNC populations that interact or do not interact with the ribosome (and molecular chaperones) are in slow exchange on the fluorescence-lifetime timescale. The next section goes on to prove that, under slow-exchange conditions, experimental rotational correlation times for intermediate-timescale ns motions are not affected by the presence of nascent-chain interactions with the ribosomal surface.

We modeled the fastest possible interactions of nascent chains with the ribosomal surface as diffusion-limited intramolecular collisions of small particles (i.e., compact nascent chains) towards larger tethered particles (i.e., the ribosome). In general, for the diffusion of particles toward an "absorbing sphere", the number of collisions per second, or current (*I(r)*, in units s^{-1}), is equal to the number of particles colliding per unit area per second (flux, *J(r)*) multiplied by the sphere's surface area, as shown below[9]

$$I(r) = J(r)4\pi r^2 = -4\pi D c_\infty r \ , \tag{14}$$

where *D* is the translational diffusion coefficient of the nascent chain and *r* is the radius of the ribosome (114.1 Å, determined from Pymol measurements of PDB: 4ybb). The association rate is described by the rate constant $k_{assoc.}$, which denotes the frequency of collisions through the solution volume per molecule, in units of cm^3s^{-1}. Now, $k_{assoc.}$ for a diffusion-limited process can be estimated via the relation

$$k_{assoc.} = 4\pi D r \ , \tag{15}$$

and the association lifetime ($t_{assoc.}$), corresponding to the average time between collisions, can be determined by dividing the volume sampled by the tethered particle (*V*) by the rate constant

$$t_{assoc.} = \frac{V}{k_{assoc.}} \ . \tag{16}$$

Eqs 15-16 were used to estimate the association lifetime ($t_{assoc.}$) for the interaction between nascent chain and ribosome. For this system, $t_{assoc.}$ is equal to the average time between collisions between the ribosome and the nascent protein. *D* was estimated via the Stokes-Einstein equation (eq 11 in main text) with radius = 16.1 Å (determined from Pymol measurements of PDB: 1mbc), viscosity = 1.08 mPa-s, and temperature = 298 K. The range of values for the volume (*V*) available for nascent-chain local diffusion was estimated from two limiting cases. Namely, *model a*, which assumes that the compact RNC region is located within the ribosomal vestibule

(Figure 2-S5a), and *model b*, which assumes that the nascent chain lies on the outer surface of the ribosome (Figure 2-S5b). For *model a,* the volume of the vestibule was estimated as 1/5th of the total ribosomal tunnel volume giving an approximate volume of 7.7 nm^3 (total tunnel volume = 38.5 nm^3, Figure 2-S5a).[10] In *model b*, the smallest diffusion volume was estimated from a cone with a height of 56 Å and a semi-angle of 15° (experimental value of cone semi-angle) yielding a volume of ~13.2 nm^3 (Figure 2-S5b). The cone height in *model b* was chosen to accommodate the expected size of the RNC globular domain (assuming spherical shape) within a cone bearing an apex angle of 30° (i.e., twice the experimental cone semi-angle). Using the above volume estimates, the estimated $t_{assoc.}$ was estimated to be 0.4 – 0.7 ns. The diffusion volumes and $t_{assoc.}$ values for these models are summarized in Table 2-S2. The limiting estimates in Table 2-S2 refer to the smallest possible diffusion volumes, yielding the fastest possible association rate constants. Therefore, the actual association rate is likely slower than these estimates. Even considering this worst-case scenario, the shortest possible association lifetime (0.4 – 0.7 ns) is of similar order of magnitude to the fluorescence lifetime of the fluorophore BODIPY-FL.[8] Therefore, if interactions between the nascent chain and ribosome occur, they are expected not to be fully averaged relative to the timescale of the fluorescence lifetime of the BODIPY-FL label. Note that the fluorescence lifetime establishes the relevant timescale of anisotropy decay measurements, given the definition of fluorescence anisotropy.

In summary, compact independently tumbling nascent chains are expected to be in slow exchange with any ribosome-bound (or large-chaperone-bound) conformations, on the fluorescence anisotropy-decay timescale.

Derivation of anisotropy-decay equation for mixture of multiple RNC populations. Given that any potential interactions between nascent chain and the ribosome are in slow exchange

relative to the fluorescent lifetime of the fluorophore, we considered how a mixture of multiple RNC populations may affect experimentally observed anisotropy decay parameters. Here, we show the derivation of the anisotropy decay equation for a mixture of two species: RNCs that tumble independently from the ribosome and RNCs that interact with the ribosome (Figure 2-S4). The non-interacting species has three rotational correlation times (slow, intermediate, and fast timescales) and is labeled contributing (C) because it contributes to the intermediate-timescale rotational correlation time. The interacting species has two rotational correlation times (slow and fast timescales) and is labeled noncontributing (NC). If both species are labeled with the same fluorophore and the fluorophores experience the same environment, then the observed anisotropy decay for the mixture is a sum of their anisotropy decays multiplied by the mole fraction of each species, as shown in eq 17 (see eq 11 in Knight *et al.* supporting information[11])

$$r_{obs}(t) = \sum_i r_{0,i} x_i r_i(t) \quad . \tag{17}$$

Both species have the same r_0 value because they have the same fluorophore, so eq 17 becomes

$$\frac{r_{obs(t)}}{r_0} = x_C r_C(t) + x_{NC} r_{NC}(t) \tag{18}$$

where x_C and x_{NC} are the mole fractions for the contributing and noncontributing species. The anisotropy decay values $r_C(t)$ and $r_{NC}(t)$ are defined as[12]

$$\frac{r_C(t)}{r_0} = \left(1 - S_{F,C}^2\right) e^{\left(\frac{-t}{\tau_{c,F}}\right)} + S_{F,C}^2 \left(1 - S_{I,C}^2\right) e^{\left(\frac{-t}{\tau_{c,I}}\right)} + S_{I,C}^2 S_{F,C}^2 e^{\left(\frac{-t}{\tau_{c,S}}\right)} \tag{19}$$

and

$$\frac{r_{NC}(t)}{r_0} = \left(1 - S_{F,NC}^2\right) e^{\left(\frac{-t}{\tau_{c,F}}\right)} + S_{F,NC}^2 e^{\left(\frac{-t}{\tau_{c,S}}\right)} \quad . \tag{20}$$

In addition, the following constraints are applied. First, both free and bound nascent-chains have rotational correlation times for the slow motions (corresponding to the global tumbling of ribosomal complexes) set to the same nominal limiting value of 1 □s. Second, the fast local

motions of the nascent-chain N terminus have very similar experimental rotational correlation times (0.1-0.3 ns) under conditions where a different extent of interactions are expected (i.e., in the absence and presence of molecular chaperones). Therefore, rotational correlation times for the fast motions are assumed to be the same for free and bound species. Under the above conditions, eqs 19 and 20 can be substituted into eq 18 to yield

$$\frac{r_{obs}(t)}{r_0} = x_C\left[\left(1-S_{F,C}^2\right)e^{\left(\frac{-t}{\tau_{c,F}}\right)} + S_{F,C}^2\left(1-S_{I,C}^2\right)e^{\left(\frac{-t}{\tau_{c,I}}\right)} + S_{I,C}^2S_{F,C}^2e^{\left(\frac{-t}{\tau_{c,S}}\right)}\right] +$$

$$x_{NC}\left[\left(1-S_{F,NC}^2\right)e^{\left(\frac{-t}{\tau_{c,F}}\right)} + S_{F,NC}^2e^{\left(\frac{-t}{\tau_{c,S}}\right)}\right] . \quad (21)$$

The latter expression can be simplified to yield eq 22

$$\frac{r_{obs}(t)}{r_0} = \left[x_C\left(1-S_{F,C}^2\right) + x_{NC}\left(1-S_{F,C}^2\right)\right]e^{\left(\frac{-t}{\tau_{c,F}}\right)} + x_CS_{F,C}^2\left(1-S_{I,C}^2\right)e^{\left(\frac{-t}{\tau_{c,I}}\right)} +$$

$$\left(x_CS_{I,C}^2S_{F,C}^2 + x_{NC}S_{F,NC}^2\right)e^{\left(\frac{-t}{\tau_{c,S}}\right)} . \quad (22)$$

Importantly, eq 22 demonstrates that the presence of ribosome-associated populations (interacting species) does not lead to any changes in rotational correlation time.

On the other hand, the presence of an interacting species does lead to changes in observed order parameters. The order parameters for a mixture can be determined from the fractional anisotropy decay intensities via eqs 23-26 as follows[12]

$$f_{S,obs} = \left(x_CS_{I,C}^2S_{F,C}^2 + x_{NC}S_{F,NC}^2\right) , \quad (23)$$

$$f_{I,obs} = x_CS_{F,C}^2\left(1-S_{I,C}^2\right) , \quad (24)$$

$$f_{F,obs} = x_C\left(1-S_{F,C}^2\right) + x_{NC}\left(1-S_{F,C}^2\right) , \quad (25)$$

$$x_C + x_{NC} = 1 . \quad (26)$$

Eqs 23-26 can be solved upon assuming that the S^2 values for the fast motions are the same for free and bound species. This assumption is reasonable because it was experimentally found that

the fluorescent dye has no affinity for the ribosome.[12] The order parameters and cone semi-angles for a range of mole fraction values for the independently tumbling (contributing) species are shown in Table 2-S3. This table shows that even if the independently tumbling population is very small (i.e., 0.2), the cone-semi angle for the intermediate timescale motions is at most 39º. This value is significantly smaller than the maximum value of 180º. This small cone semi-angle illustrates the fact that, regardless of any ribosome (or chaperone) interactions with the nascent protein, nascent chain motions are highly spatially confined.

Oligonucleotide sequences used in cell-free reaction. In order to create stalled ribosome-nascent chain complexes, cell-free expression of full-length apoMb ($apoMb_{153}$) was performed in the presence of oligodeoxynucleotides bearing a sequence complementary to the sequence of the desired mRNA region to be subsequently cleaved by the endogenous *E. coli* RNAse H enzyme.[13] The target mRNA region was designed to "hide" the apoMb gene stop codon, to generate RNCs of full-length apoMb. In addition, an anti-SSRA oligonucleotide was added to the cell-free reaction to prevent tmRNA-mediated release of the nascent chain from the stalled ribosome.[14] The final cell-free concentration of the above oligonucleotides was 0.15 μg/μL. The sequences of the oligonucleotides are

5'- ACCCTGGTAACCCAGTTCTTTGTACTTAGCAGCGATAT-3', full-length ApoMb RNC sequence, and

5'- TTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCGACTA-3', anti-ssrA oligonucleotide.

Supporting Figures

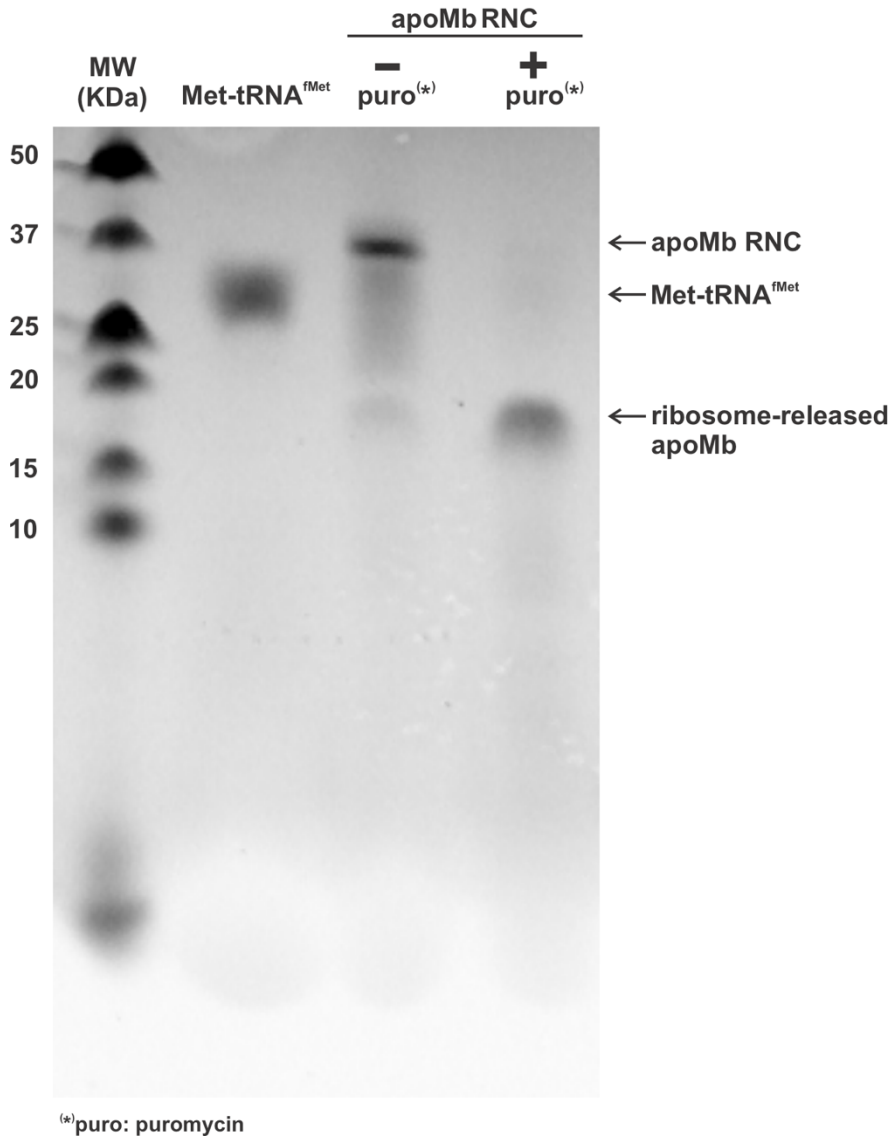


Figure 2-S1. Representative gel illustrating a typical puromycin assay performed on RNCs of full-length apoMb (apoMb$_{153}$). Bands are visualized via in-gel fluorescence. The shift to lower molecular weight upon addition of the puromycin (puro) antibiotic proves that the nascent chain was ribosome-bound (therefore an actual RNC) prior to treatment with puro. Met-tRNA^{fMet} and

apoMb RNCs are site-specifically labeled with BODIPY-FL at the N terminus.

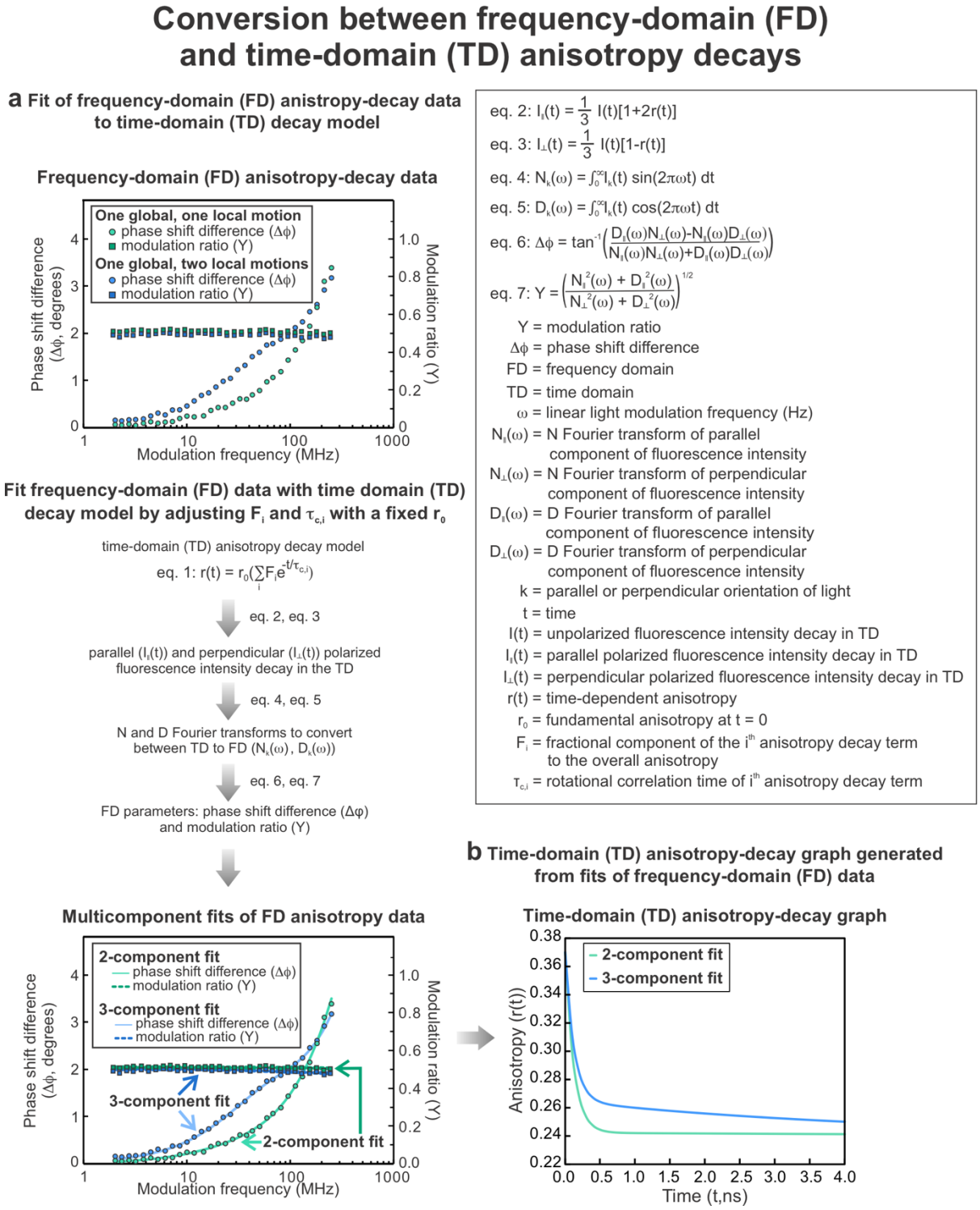


Figure 2-S2. (a) Procedure employed to fit FD anisotropy-decay data to time-domain anisotropy decay models. (b) TD anisotropy-decay graph generated from fits of FD data.

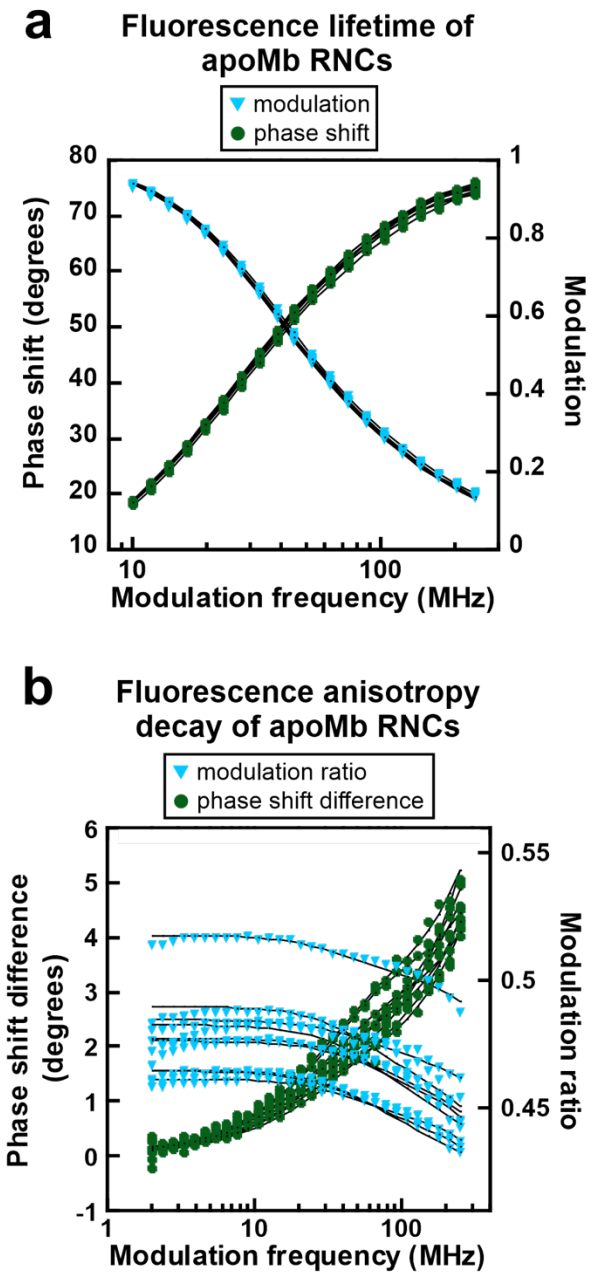


Figure 2-S3. (a) Fluorescence lifetime and (b) anisotropy-decay graphs for all the data collected in this work.

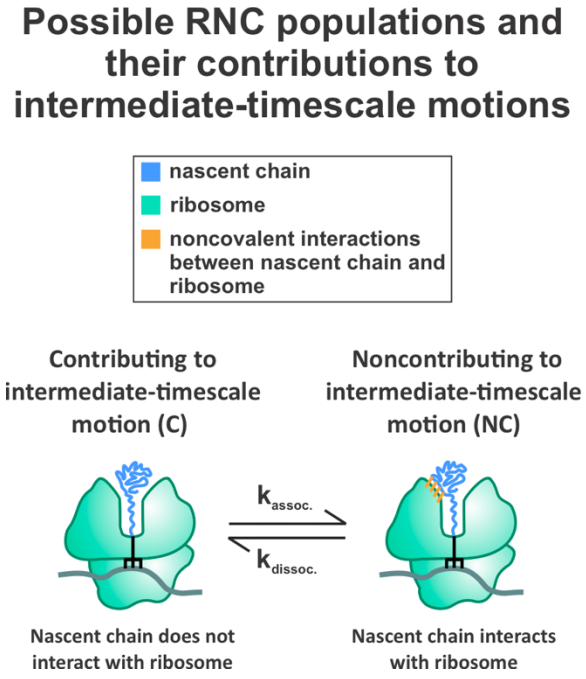


Figure 2-S4. Simplified models illustrating RNC conformations consistent with the data collected in this work. RNCs that do not interact with the ribosome show a three-component anisotropy decay and contribute to the intermediate timescale motion (C). RNCs that interact with the ribosome show a two-component anisotropy decay and do not contribute to the intermediate-timescale motion (NC).

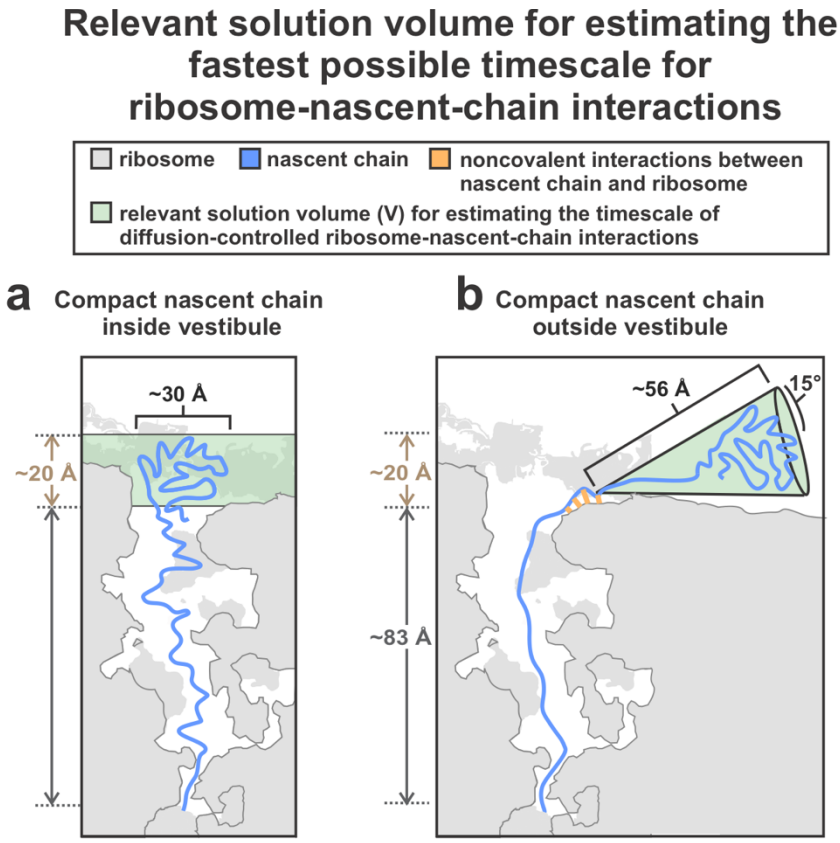


Figure 2-S5. Limiting models employed to estimate relevant solution volumes (V) for diffusion-controlled association of nascent chains with ribosomal surface. (a) Model assuming that the compact nascent chain-region is contained within the ribosomal vestibule. (c) Model assuming that the compact nascent chain-region is outside of the ribosomal vestibule. Diffusion volumes estimated from the experimental cone semi-angle of 15°. Green regions indicate relevant solution volumes for the hypothetical diffusion-controlled association between the nascent chain and the ribosome. Ribosome dimensions were estimated from Dao Duc *et al.*[10]

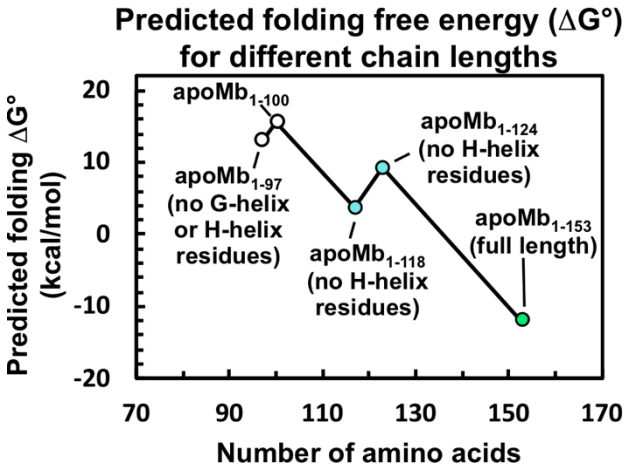


Figure 2-S6. Predicted folding free energy of apoMb N-terminal fragments of increasing length, assuming a native-like conformation. This plot is a complement to the main-article Figure 2-12. Additional chain lengths ($apoMb_{1-100}$ and $apoMb_{1-118}$) are included here, to show the predicted thermodynamic contribution of loop regions between helices. $ApoMb_{1-100}$ includes the unstructured region between helices F and G. $ApoMb_{1-124}$ includes the unstructured region between helices G and H. All free energy calculations were carried out with FoldX version 5.[15] PDB files for apoMb (RCSB PDB: 1mbc) were generated with Pymol version 2.0.0.

Supporting Tables

Table 2-S1. Statistical p values for the viscosities of the apoMb$_{153}$ RNC solution measured at shear rates ranging from 1500 s^{-1} to 8000 s^{-1} shown in Figure 2-4. P values were determined from a two-tailed Student's test assuming unequal variances. P values > 0.05 indicate data that are not statistically different (orange).

		Shear rate (s^{-1})				
		1,500	3,000	5,000	6,500	8,000
Shear rate (s^{-1})	1,500	–	0.58	0.29	0.27	0.37
	3,000		–	0.98	0.70	0.28
	5,000			–	0.56	0.28
	6,500				–	0.23
	8,000					–

Table 2-S2. Statistical p values comparing the viscosities of the apoMb$_{153}$ RNC solution and buffer system lacking RNCs measured at shear rates ranging from 1500 s^{-1} to 8000 s^{-1} shown in Figure 2-4. P values were determined from a two-tailed Student's test assuming unequal variances. P values > 0.05 indicate data that are not statistically different (orange).

Shear rate (s^{-1})	apoMb$_{153}$ RNC solution vs buffer system without RNCs
1,500	0.40
3,000	0.93
5,000	0.62
6,500	0.35
8,000	0.14

Table 2-S3. Estimates of relevant solution volume for estimating fastest possible timescale for ribosome-nascent-chain interactions as shown in Figure 2-S5

Model	Diffusion volume (nm^3)[a]	$t_{assoc.}$ (ns)
Compact subdomain contained in vestibule (Figure 2-S5b)	7.7	0.43
Compact subdomain outside vestibule. Volume is a cone with semi-angle of 15° (Figure 2-S5c)	13.2	0.73

[a]Diffusion volume is the relevant solution volume for association between the nascent chain and the ribosome.

Table 2-S4. Intermediate-timescale motion cone semi-angles and order parameter values for different mole fractions of contributing species

Mole fraction of contributing species (x_C)	S_I^2	S_I	Cone semi-angle (θ_I, deg)
0.1	-0.03 ± 0.09	(Cannot calculate from negative S_I^2)	(Cannot calculate from negative S_I^2)
0.2	0.48 ± 0.04	0.70 ± 0.06	39 ± 2
0.3	0.66 ± 0.03	0.81 ± 0.04	30 ± 2
0.4	0.74 ± 0.02	0.86 ± 0.03	25 ± 1
0.5	0.79 ± 0.02	0.89 ± 0.02	22 ± 1
0.6	0.83 ± 0.01	0.91 ± 0.02	20.0 ± 0.9
0.7	0.85 ± 0.01	0.92 ± 0.01	18.4 ± 0.8
0.8	0.87 ± 0.01	0.93 ± 0.01	17.2 ± 0.8
0.9	0.89 ± 0.01	0.94 ± 0.01	16.1 ± 0.7
1	0.897 ± 0.005	0.947 ± 0.005	15.3 ± 0.7

Supporting References

(1) Odom, O. W.; Hardesty, B. Use of 50 S-binding antibiotics to characterize the ribosomal site to which peptidyl-tRNA is bound. *J. Biol. Chem.* **1992,** *267*, 19117-19122.

(2) Ziehr, D. R.; Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Production of ribosome-released nascent proteins with optimal physical properties. *Anal. Chem.* **2010,** *82*, 4637-4643.

(3) Weinreis, S. A.; Ellis, J. P.; Cavagnero, S. Dynamic fluorescence depolarization: A powerful tool to explore protein folding on the ribosome. *Methods* **2010,** *52*, 57-73.

(4) Jameson, D. M.; Gratton, E.; Hall, R. D. The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry. *Appl. Spectrosc. Rev.* **1984,** *20*, 55-106.

(5) Ross, J. A.; Jameson, D. M. Time-resolved methods in biophysics. 8. Frequency domain fluorometry: Applications to intrinsic protein fluorescence. *Photochem. Photobiol. Sci.* **2008,** *7*, 1301-1312.

(6) Lakowicz, J. R. *Principles of fluorescence spectroscopy*. 3rd ed.; New York : Plenum Press: 2006.

(7) Lakowicz, J. R.; Laczko, G.; Cherek, H.; Gratton, E.; Limkeman, M. Analysis of fluorescence decay kinetics from variable-frequency phase shift and modulation data. *Biophys. J* **1984,** *46*, 463-477.

(8) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(9) Dill, K. A. *Molecular driving forces: Statistical thermodynamics in chemistry and biology*. 2nd ed.; New York : Garland Science: 2011.

(10) Dao Duc, K.; Batra, S. S.; Bhattacharya, N.; Cate, Jamie H. D.; Song, Yun S. Differences in the path to exit the ribosome across the three domains of life. *Nucleic Acids Res.* **2019,** *47*, 4198-4210.

(11) Knight, A. M.; Culviner, P. H.; Kurt-Yilmaz, N.; Zou, T.; Ozkan, S. B.; Cavagnero, S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **2013,** *8*, 1195-1204.

(12) Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci.* **2009,** *18*, 2003-2015.

(13) Behrmann, M.; Koch, H. G.; Hengelage, T.; Wieseler, B.; Hoffschulte, H. K.; Müller, M. Requirements for the translocation of elongation-arrested, ribosome-associated OmpA across the plasma membrane of Escherichia coli. *J. Biol. Chem.* **1998,** *273*, 13898-13904.

(14) Hanes, J.; Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94*, 4937-4942.

(15) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The foldx web server: An online force field. *Nucleic Acids Res.* **2005,** *33*, W382-W388.

Chapter 3

Nascent-Chain Interaction Networks and Their Effect on the Bacterial Ribosome

This chapter reproduces a submitted article by Meranda M. Masse, Valeria Guzman-Luna, Angela E. Varela, Rachel B. Hutchinson, Aniruddha Srivastava, Wanting Wei, Andrew M. Fuchs, and Silvia Cavagnero. M.M.M., A.E.V. and S.C. designed the project. All coauthors except for S.C. collected and interpreted the data, with R.B.H. focusing on acquiring and processing the frequency-domain anisotropy data, M.M.M. taking care of cross-linking urea titrations, sucrose gradients and some Western blotting, V.G.-L. collecting and processing most Western-Blotting data, and A.E.V. performing some urea titrations experiments. M.M.M., V.G.-L., R.B.H., A.E.V. and S.C. participated in most of the data organization, discussions, and manuscript writing.

3.1 Abstract

In order to become bioactive, proteins need to be biosynthesized and protected from aggregation during translation. The ribosome and molecular chaperones contribute to both tasks. While it is known that some ribosomal proteins (r-proteins) interact with ribosome-bound nascent chains (RNCs), specific interaction networks and their role within the ribosomal machinery remain poorly characterized and understood. Here, we find that RNCs of variable sequence and length (beyond the 1st C-terminal reside) do not modify the apparent stability of the peptidyl-transferase center (PTC) and r-proteins. Thus, RNC/r-protein interaction networks close to the PTC have no effect on the apparent stability of ribosome-RNC complexes. Further, fluorescence anisotropy decay, chemical-crosslinking and Western blots show that RNCs of the foldable protein apoHmp$_{1\text{-}140}$ have an N-terminal compact region (63– 94 residues) and interact specifically with r-protein L23 but not with L24 or L29, at the ribosomal-tunnel exit. Longer RNCs bear a similar compact region and interact either with L23 alone or with L23 and another unidentified r-protein, or with molecular chaperones. The apparent strength of RNC/r-protein interactions does not depend on RNC sequence. Taken together, our findings show that RNCs encoding foldable protein sequences establish an expanding specific interaction network as they get longer, including L23, another r-protein and chaperones. Interestingly, the ribosome alone (i.e., in the absence of chaperones) provides indiscriminate support to RNCs bearing up to ca. 190 residues, regardless of nascent-chain sequence and foldability. In all, this study highlights the unbiased features of the ribosome as a powerful nascent-protein interactor.

3.2 Significance Statement

The presence of interactions between nascent chains bearing a foldable amino-acid sequence (with no signal or arrest tags) and specific ribosomal proteins has been suggested but never experimentally demonstrated, up to now. Here, we identify the ribosomal protein L23 as a specific nascent-chain-interacting partner. L23 establishes noncovalent contacts with nascent chains of the multi-domain foldable model protein apoHmp, which lacks signal/arrest sequences. As nascent chains elongate, their interaction network expands to another ribosomal protein and to the trigger-factor and Hsp70 chaperones. Interestingly, ribosomal-protein/nascent-chain complexes have similar apparent stability regardless of nascent-chain sequence and degree of foldability. These findings are significant because they advance our knowledge on ribosome-mediated nascent-protein interaction networks and suggest avenues to prevent undesirable aggregation.

3.3 Introduction

Recent evidence suggests that the ribosome plays an active role in cotranslational protein folding.[1-6] During translation, the nascent chain traverses the ribosomal exit tunnel, which is ca. 80-100 Å long, 10-35 Å wide[7-10] and can fit 30 to 40 nascent residues.[11-16] Within the ribosomal exit tunnel and its nearby regions across the highly negatively charged outer surface of the ribosome,[3] nascent chains encoding single-domain proteins become compact[17, 18] and acquire some secondary[19-23] and tertiary structure.[5, 24-28] This set of observations proves the importance of the ribosome in nascent-protein structure formation.

During translation, the ribosome influences nascent protein chains at different levels. For instance, the inner geometry of the ribosomal exit tunnel favors formation of secondary nascent-chain structure, especially of α-helical[22, 29-31] or β-sheet nature.[23] In addition, recent work showed that the ribosomal exit tunnel and vestibule enable acceleration of folding – but not unfolding – of a small single-domain protein, thereby stabilizing nascent protein chains relative to their free state in solution.[32] On the other hand, the ribosome may also destabilize single-protein domains, in case the domain is far removed from the peptidyl transferase center.[33] Further, the ribosome renders nascent chains soluble relative to the corresponding ribosome-released protein chains, thereby supporting cotranslational folding.[34] Collectively, these results highlight the influence of the ribosome on nascent protein folding.

The ribosome is known to establish physical noncovalent interactions with nascent chains. As summarized in Table 3-S1, these interactions have been identified in a variety of experimental studies and can be divided into three categories. Namely, *(i)* interactions between the ribosome and nascent chains carrying an N-terminal signal sequence,[22, 35-39] *(ii)* interactions between the ribosome and nascent chains bearing C-terminal ribosome-stalling or arrest

sequences,[40-47] and *(iii)* interactions between nascent chains that lack N- or C-terminal ribosome-stalling or arrest sequences.[48] Additional studies are consistent with the presence of ribosome-nascent-chain interactions, though they do not directly prove their existence.[49-52]

While a body of research has explicitly addressed how the ribosome influences nascent polypeptides and proteins, little is known about how the nascent chain affects the properties of the ribosome. For instance, empty 70S ribosomes are known to be more prone to dissociation than ribosomes bearing both mRNA and peptidyl tRNA. This conclusion was reached upon addition of either ribosome-dissociation factors[53] or Hofmeister cosolutes.[54-56] Other researchers established a similar finding upon depletion of magnesium ions.[57-59] In a different study, addition of Hofmeister salts were employed to show that translation initiation complexes (including 70S in complex with initiator tRNA) disassemble more easily than peptidyl tRNAs bearing nascent chains.[60] In this work, ribosomes carrying longer nascent chains were found to be less prone to dissociation.[60] Other work has examined the effect of magnesium ions and other Hofmeister ions on empty 70S ribosome disassembly and how it changes sedimentation coefficients.[61]

Yet, there is only a limited number of studies targeting the effect of non-Hofmeister denaturing agents on the ribosome. For instance, it is known that the 30S subunit disassembles in the presence of 6 molar urea.[62] In addition, urea lowers the melting temperature and sedimentation coefficient of the 50S ribosomal subunit.[63] The 30S subunit is more sensitive to thermal denaturation than the 50S subunit, and the 70S ribosome is most thermally stable.[64] Most notably, 70S ribosomes bearing a nascent chain are less prone to chemical denaturation than empty ribosomes.[33]

However, the effect of urea on the disassembly of various ribosome components (e.g. peptidyl transferase center -- a.k.a. PTC -- and ribosomal proteins) as a function of specific RNC

characteristics (e.g., length and amino-acid sequence, interactions with ribosomal proteins) has not been characterized yet, to date. The present work moves initial steps towards filling this gap of knowledge.

We find that short peptidyl-tRNAs (snc-tRNAs) stabilize the 70S ribosome against denaturation by the non-Hofmeister cosolute urea, and we propose a multi-step model for the disassembly of ribosome-RNC complexes consistent with our findings. In addition, RNCs up to chain length 140 interact only with one ribosomal protein (r-protein), i.e., L23, in the vicinity of the tunnel exit. A wider interaction network, including one additional ribosomal protein and the trigger factor (TF) and Hsp70 chaperones, gets established as the nascent chain elongates further. Finally, our data also suggest that the interaction strength of individual RNC/r-protein populations does not vary significantly with RNC sequence and length. In all, our results show that (a) the interaction of foldable RNCs with the L23 ribosomal protein has been explicitly identified experimentally for the first time, and (b) the ribosome provides even, indiscriminate assistance to newly synthesized nascent protein chains, whether they are foldable or intrinsically disordered, via a powerful and highly specific interaction network.

3.4 Results and Discussion

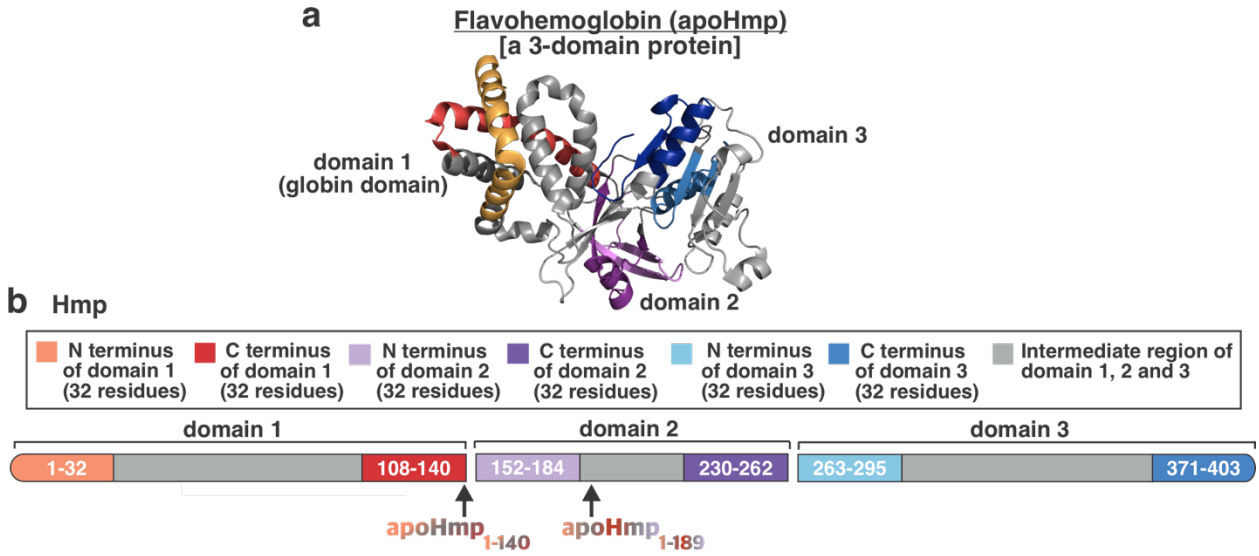
The presence of very short nascent chains stabilizes the 70S ribosomal complex. We started our investigations by performing a series of sucrose-gradient studies on *E. coli* empty ribosomes and nascent-chain-loaded ribosomes. Our results, detailed in SI-Appendix Figure 3-S1and S2, showed that empty-70S ribosomes are more sensitive to urea denaturation than ribosomes bearing tRNAs linked to longer nascent chains. It appears that the aminoacyl-tRNA is responsible for most of the stabilizing effect (SI-Appendix, Figs. S1 and S2). Interestingly, the

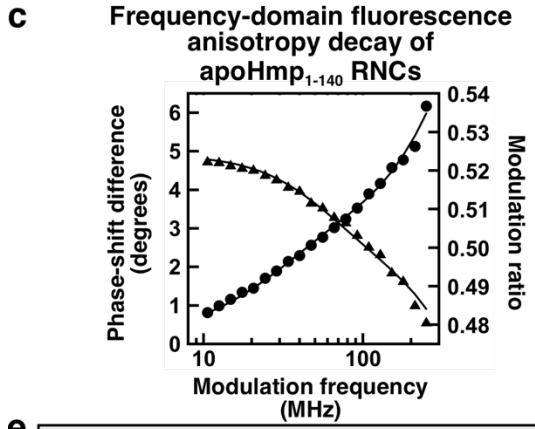
length and amino-acid sequence of the nascent protein does not influence the urea sensitivity of ribosome-RNC complexes.

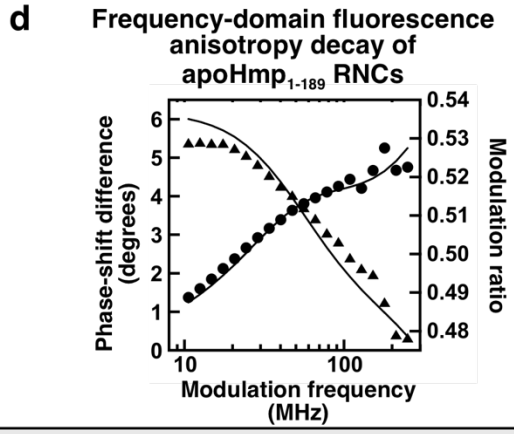
The sequence dependence of ribosome-nascent-chain interactions has been further explored in other parts of this work.

Extending the nascent-protein-chain portfolio. The experiments described in the next sections employed a larger set of RNCs than the sucrose-gradient studies. Our specific purpose was to test the effect of RNC chain length, sequence and foldability on the apparent stability of the ribosome and to experimentally characterize the RNC-ribosome interaction network. First, we examined several constructs derived from the *Escherichia coli* protein flavohemoglobin (apoHmp). The structure and key building blocks of this protein analyzed in this work shown in Figure 3-1a,b. The crystal structure of apoHmp comprises three domains, an N-terminal heme-binding (domain 1), a flavin adenine dinucleotide-binding (domain 2) and a C-terminal nicotinamide adenine dinucleotide-binding domain (domain 3), as schematically illustrated in Figure 3-1b.[65] Hmp plays a key role in O_2, NO and CO transport in *E. coli* and is involved in a variety of signaling pathways.[66, 67] We also studied the behavior of the phosphorylated insulin receptor interacting region of the growth factor receptor-bound protein 14 (Grb14) from *Rattus norvegicus.* This protein, which is denoted here as PIR, is intrinsically disordered,[68] i.e., an IDP. The specific nascent chain constructs of both proteins analyzed in this work are illustrated in SI-Appendix Figure 3-S4.

Ribosome-bound apoHmp nascent chains of variable length are compact. Next, we performed fluorescence depolarization decay experiments in the frequency domain[69-71] to probe the rotational dynamics of nascent chains encoding foldable sequences. This technique has been previously employed to assess the rotational correlation time (τ_c) and amplitude of rotational





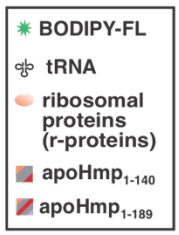


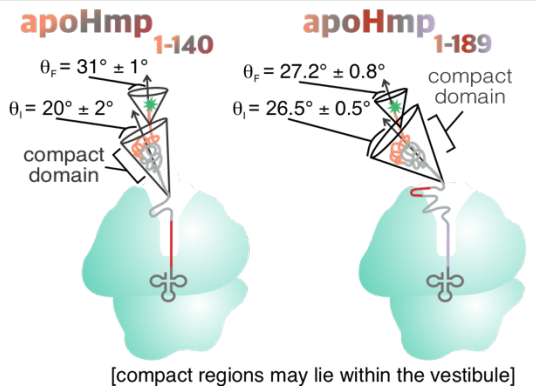
e

RNC sample	regimes of motion							reduced χ^2	
	slow (≥ μs)	intermediate (ns)			fast (~ps)				
	$\tau_{c,S}$ (ns)	$\tau_{c,I}$ (ns)	S_I	θ_I (deg)	$\tau_{c,F}$ (ns)	S_F	θ_F (deg)	2-component	3-component
apoHmp$_{1-140}$	1,000	4.3 ± 0.2	0.91 ± 0.02	20 ± 2	0.17 ± 0.03	0.79 ± 0.01	31 ± 1	11 ± 4	0.4 ± 0.1
apoHmp$_{1-189}$	1,000	4.1 ± 0.4	0.85 ± 0.01	26.5 ± 0.5	0.13 ± 0.02	0.84 ± 0.01	27.2 ± 0.8	36 ± 5	1.5 ± 0.2

f

RNC sample	number of amino acids in compact domain		
	sphere	oblate ellipsoid (ρ = 0.5)	prolate ellipsoid (ρ = 3.5)
apoHmp$_{1-140}$	94 ± 5	78 ± 4	65 ± 4
apoHmp$_{1-189}$	91 ± 10	75 ± 8	63 ± 7





[compact regions may lie within the vestibule]

Figure 3-1. ApoHmp structure and and fluorescence-anisotropy decayed reveals compaction of apoHmp RNCs. a) *E. coli* flavohemoglobin (apoHmp) has three domains (shown in red, purple and blue). PDB code: 1GVH. **b)** Scheme illustrating the three domains of apoHmp and their corresponding N-and C-termini. Representative frequency-domain fluorescence anisotropy decay data for **c)** apoHmp1-140 and **d)** apoHmp1-189 RNCs. **e)** Table 3-Summarizing anisotropy decay parameters including rotational correlation times (r_c), order parameters (S) and cone semi-angles (8). The S,I, and F subscripts denote slow-, intermediate-, and fast- timescale motions, respectively. Uncertainties are reported as ± SE for n=3-5. Three-component anisotropy fits were selected as best fits if they led to a 2.5-fold decrease in reduced χ^2, relative to two-component fits. The χ^2 of the chosen model is shown in bold. **f)** Table 3-Summarizing the number of amino acids comprising the RNC compact region, deduced from the $r_{c,I}$ rotational correlation time and assuming spherical, oblate ellipsoid, or prolate ellipsoid nascent-chain shapes. The parameter p denotes the axial ratio. **e)** Cartoon representation of apoHmp1-140 and apoHmp1-189 RNCs based on fluorescence anisotropy-decay data.

motions of RNCs.[17, 18, 34, 49, 72] The goal of this experiment was to probe whether RNCs harboring long nascent chains display any degree of compaction. We focused on RNCs of apoHmp$_{1\text{-}140}$, corresponding to the N-terminal domain 1 of Hmp, and RNCs of apoHmp$_{1\text{-}189}$, comprising Hmp domain 1 plus 49 C-terminal residues belonging to domain 2. Nascent proteins were site-specifically labeled at their N terminus with the BODIPY-FL fluorophore as described.[17] Once information on nascent-chain compaction is in hand, the interplay between ribosome and nascent-chain sensitivity to urea denaturation can be more rationally explored and understood, as apparent in the sections below.

Representative data for apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ are shown in panels c and d of Figure 3-1, respectively. Both RNCs display informative frequency-domain anisotropy decay profiles. As shown in Figure 3-1e and consistent with the very low reduced χ^2 values, the fits that include 3 rotational-tumbling components give the best results. Importantly, panels e and f of Figure 3-1 show that both apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ RNCs are characterized by an N-terminal compact domain. In both cases, this domain spans ca. 63 to 94 residues, depending on the exact shape. Note that RNC shape assessment is beyond the scope of this work. Regardless of the actual overall morphology of the compact domains, the fact that a similar-size compact domain is observed for both apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ suggests that both constructs undergo some partial folding on the ribosome. Surprisingly, the observed size of the compact domain of apoHmp$_{1\text{-}189}$ RNCs is significantly smaller than domain 2, which comprises 140 residues (Figure 3-1b). Therefore, biosynthesis of the additional 49 C-terminal amino acids is not sufficient to lead to complete folding of the N-terminal domain, for this protein.

In addition, the cone semi-angle analysis of the fluorescence anisotropy decay data (Figure 3-1f) shows that the compact domain of Hmp$_{1\text{-}189}$ RNCs spans a slightly wider cone semi-angle

($26.5^{o} \pm 0.5^{o}$ *vs* $20^{o} \pm 0.2^{o}$) than $Hmp_{1\text{-}140}$ RNCs, consistent with the fact that the latter construct is likely projecting further out from the ribosomal surface than the shorter $Hmp_{1\text{-}140}$ construct.

In all, the fluorescence anisotropy data show that the $Hmp_{1\text{-}140}$ and $Hmp_{1\text{-}189}$ nascent chains are both comparably compact and no more than partially folded, while on the ribosome, with $Hmp_{1\text{-}189}$ spanning a slightly wider cone semi-angle.

The peptidyl transferase center site is largely unaffected by nascent-chain sequence and length, beyond 32 residues. Next, we probed whether nascent chains of different length, amino-acid sequence and foldability affect the apparent stability of specific regions of the ribosome. We directed our focus on the peptidyl transferase center (PTC) of the *E. coli* ribosome, and we explored its urea sensitivity via a nascent-chain ribosome-release assay mediated by puromycin. The results of puromycin-release-detected urea titrations are shown in the SI Appendix Figure 3-S5 and further described in the SI Appendix. Overall, the urea titration data show that the nature of the nascent chain has a weak effect on the urea sensitivity of the PTC, in *E. coli*. Next, we explored the effect of nascent-chain properties on the apparent stability of ribosomal proteins.

The global urea sensitivity of ribosomal proteins is largely unaffected by nascent-chain sequence and length. The global urea sensitivity of ribosomal proteins (r-proteins) was probed via urea titrations based on Trp fluorescence emission. Trp is a well-known fluorescent reporter, and its emission properties are highly environmentally sensitive. Upon inspection of the *E. coli* ribosomal-protein sequences via the 2WWL and 2WWQ Protein Data Bank (PDB) files, we ascertained that the r-proteins of the ribosome comprise a total of 32 Trps (SI-Appendix Figure 3-S6a) roughly uniformly dispersed throughout the ribosomal structure. To gain insights at the highest possible resolution, we focused on all the apoHmp nascent chains listed in SI-Appendix

Figure 3-S4e, and on the PIR nascent chain, also listed in the same figure. Note that apoHmp nascent chains only contribute two additional Trps at positions 120 and 149, and PIR only contributes on Trp at position 44. We regard these residues, present in some of the constructs, as contributing negligibly to the overall Trp fluorescence emission. In essence, the readout of this assay is dominated by the much larger contributions arising from the 32 Trps interspersed across the r-proteins. Urea titrations were carried out and Trp fluorescence emission was monitored (SI-Appendix Figure 3-S6b,c). Spectral shifts were regarded as reporters of r-protein folding, and centers of mass of emission spectra were assessed to generate titration curves reporting on the urea sensitivity of r-proteins. Urea titration data were processed according to Santoro and Bolen.[73, 74] Individual representative titration curves are shown in SI-Appendix Figure 3-S6d.

The $\Delta G°_{app,\ unfold}$ for each construct are plotted in SI-Appendix Figure 3-S6e and corresponding t-test values are tabulated in SI-Appendix Figure 3-S6f. Nearly all the constructs show statistically similar results, with $\Delta G°_{app,\ unfold}$ values ranging from 2 to 5 kcal/mol. Hence, the presence of peptidyl tRNA, regardless of nascent-chain characteristics, does not affect the urea sensitivity of r-proteins. As shown in sections below, some nascent chains interact with specific ribosomal proteins. On the other hand, these interactions are not sufficiently strong to be detected via this assay, which monitors the overall sensitivity to urea of all r-proteins.

Nascent chains of apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ interact with ribosomal protein L23. The data of SI-Appendix Figure 3-S6 showed that the apparent stability of r-proteins is not influenced by the presence of either intrinsically disordered or foldable nascent chains of variable length and physical properties. On the other hand, nascent chains encoding the intrinsically disordered protein PIR are known to interact with the r-protein L23 and L29.[48] It is reasonable to imagine that these interactions are not sufficiently strong to be detected by the Trp-fluorescence assay of

SI-Appendix Figure 3-S6, which monitors Trp environment across the entire ribosome. Hence, the data in SI-Appendix Figure 3-S6 do not preclude the presence of RNC/ribosome interactions. Hence, it is compelling to explore whether these interaction exist. As shown in Figure 3-2a, the *E. coli* ribosome comprises several proteins within the region near the exit-tunnel.

In this work, we focused on the L23, L24 and L29 r-proteins, which are closest to the vestibule and outer region of the ribosomal exit tunnel. We explored the interaction patterns of the apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ nascent chains, which populate compact states while ribosome-bound (Figure 3-1 c-g). The well-characterized zero-length chemical crosslinker carbodiimide 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC) was employed.[48, 75, 76] To carry out chemical crosslinking of nascent chains, we followed known procedures[48] involving a combination of low-pH SDS-PAGE[77] and Western blotting in the absence and presence of the trigger factor (TF) chaperone. Notably, EDC enables detecting existing noncovalent interactions, though it does not provide an accurate quantitation of interacting populations, as discussed at length by Guzman-Luna *et al.*.[48] Yet, EDC is an extremely valuable tool to detect protein-protein interactions within the ribosome-nascent-chain complex. Site-specific fluorescence labeling of nascent proteins at their N terminus enables focusing exclusively on interactions involving the nascent protein. Low-pH-gel and Western-blot were collected to explore interactions between nascent chains and r-proteins. These interactions have not been directly detected before, in the case nascent chains lacking arrest or signal sequences. It is worth noting that EDC does not have high accessibility within the exit-tunnel core.[48] Therefore, detecting interactions within the tunnel core, including corresponding Western-blotting work, would not have been informative

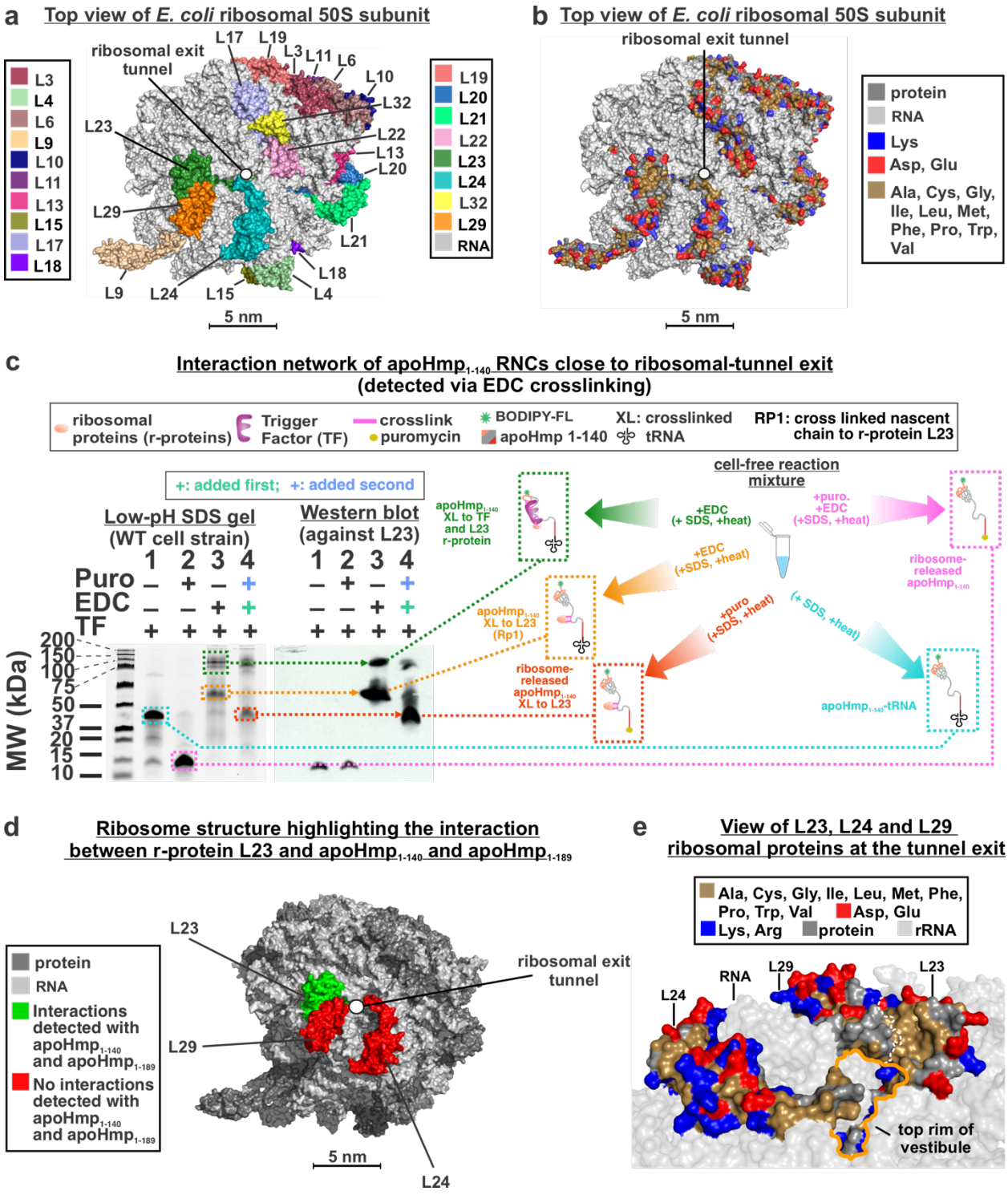


Figure 3-2. Crosslinking patterns of apoHmp$_{1-140}$ RNCs and identification of interacting r-proteins. **a)** Top view of 50S subunit of the *E. coli* ribosome highlighting the ribosomal proteins (r-proteins). Figure has been modified from [48] under a Creative Commons Attribution 4.0

International license. **b)** Top view of 50S *E. coli* ribosome highlighting r-protein charged and nonpolar residues. Figure has been modified from *Guzman, et. al* under a Creative Commons Attribution 4.0 International license. **c)** SDS-PAGE and Western blot data identifying r-proteins interacting with apoHmp$_{1\text{-}140}$ RNCs in the absence and presence of the EDC crosslinker and the ribosome-release agent puromycin. Data show that the L23 r-protein interacts with apoHmp$_{1\text{-}140}$ RNCs. Corresponding data employing antibodies against L24 and L29 r-proteins, showing no interactions, are available in the SI Appendix. **d)** Top view of *E. coli* 50S ribosomal subunit highlighting the r-proteins that either interact (green) or do not interact (red) with apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ RNCs. **e)** Side view of r-proteins near the vestibule of the ribosomal exit tunnel. Figure has been modified from *Guzman, et. al*. under a Creative Commons Attribution 4.0 International license.

within our experimental setup.

It is also important to mention that, under our experimental conditions, EDC does not report on interactions involving nascent protein chains and ribosomal RNA (rRNA). In the presence of imidazole, crosslinks between RNA 5' phosphate and aliphatic amines of proteins are known to take place.[75] However, our samples did not contain imidazole, and this chemical would anyways be unable to detect interactions not involving the 5’ end of RNA. Therefore, even in the presence of imidazole, EDC would likely underestimate all potential interactions with RNA. Thus, interactions between nascent proteins and r-RNA are beyond the scope of this study.

The data for apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$, shown in Figures 3-2c and 3-3a,b, respectively, indicate that both nascent proteins interact with ribosomal protein L23. Western blotting was also carried out with antibodies against ribosomal proteins L24, L29 (SI Appendix, Figs. S9 and S11). However, no interactions between these two r-proteins and apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ were detected.

These results show for the first time that two RNCs lacking signal or arrest sequences and encoding nascent proteins that partially fold on the ribosome experience interaction with a specific ribosomal protein. The intrinsically disordered PIR RNC, which was studied before,[48] was found to have a ca. 50% population of RNC interacting with L23 and L29. In contrast, here we use two nascent chains corresponding to the foldable apoHmpHa$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ sequences, whose N-terminal regions are compact (Figure 3-1 c-g). These apoHmpH RNCs interact only with L23 and not with L29.

The interactions experienced by apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ involve the dominant fraction of the apoHmpH RNC population, which is found to nearly-quantitatively crosslink with the L23 ribosomal protein (Figure 3-2 and 3). This scenario is different from the case of intrinsically

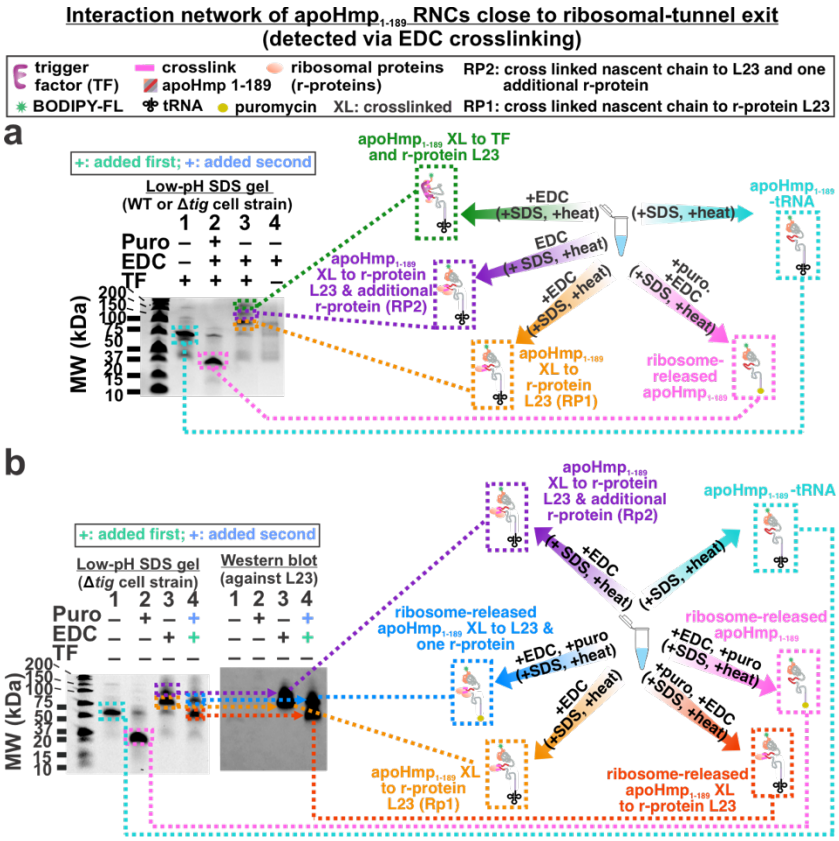


Figure 3-3. Crosslinking patterns of apoHmp$_{1-189}$ RNCs and identification of interacting proteins. **a)** Low-pH SDS-PAGE analysis of apoHmp$_{1-189}$ RNCs in the absence and presence of the EDC crosslinker, TF chaperone and the RNC ribosome-release agent puromycin. **b)** Side-by-side SDS-PAGE and Western blot data identifying the interaction network of apoHmp$_{1-189}$ RNCs in the absence and presence of EDC, puromycin and TF chaperone. The L23 r-protein is found to interact with apoHmp$_{1-189}$ RNCs. Corresponding data employing antibodies against L24 and L29 r-proteins, showing no interactions, are available in the SI Appendix.

disordered PIR RNCs,[48] which crosslink only in part, under the same experimental conditions. The larger extent of crosslinking of the foldable apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ nascent chains, however, may simply result from the greater number of EDC-reactive residues of apoHmp$_{1\text{-}140}$ (25 EDC-reactive residues, 20 beyond the tunnel core) and apoHmp$_{1\text{-}189}$ (36 EDC-reactive residues, 30 beyond the tunnel core) relative to PIR (14 EDC-reactive residues, 12 beyond the tunnel core). Interestingly, as shown in the sections below, the urea sensitivity of the L23-RNC complexes is similar for all RNCs, suggesting comparable interaction strengths.

Importantly, given that the fluorescence anisotropy-decay data of Figure 3-1 show that RNCs of both apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ have "freely tumbling" compact regions, our data suggest that the compact regions are not engaged in direct interactions with the ribosomal surface.

In the case of apoHmp$_{1\text{-}189}$, there is an additional interacting population, which we denoted as RP2 (Figure 3-3). This population includes r-protein L23 (Figure 3-3b) and one additional unidentified protein of c.a. 6-10 kDa. Our Western Blots indicate that L29 (7 kDa) is not present in the RP2 band (SI Appendix, Figure 3-S11b), based on molecular-weight arguments. Other cytoplasmic *E. coli* chaperones and ribosome interactors (GroEL, GroES, SecB, K/J/E, SRP and ClpB; MW range: 48-80 kDa) are ruled out as they would appear well above the RP2-containing band in our gels (Figure 3-3). Because of its close spatial proximity to L23 (Figure 3-2a) and based on molecular weight arguments, we believe that it is possible that RP2 comprises L23 and L29. However, our monoclonal antibodies were not able to capture an L29 epitope, suggesting the utility of experiments employing polyclonal L29 antibodies. These are experiments are planned in future investigations.

Finally, the data for Figures 3-4 and 3-5 also show that a fraction of the RNCs interacts with the trigger factor (TF) chaperone (presumably coupled to the ribosome via L23), rather than with

the L23 r-protein. The role of the interactions between apoHmpH RNCs and TF are beyond the scope of this work and have already been explored elsewhere in the case of other client proteins.[39, 78-81] These studies showed that TF interacts with clients that bear an expanded conformation in their bound state.

In all, our data show that apoHmpH$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ RNCs interact with either the L23 r-protein or with the TF chaperone. We propose that these two classes of interactions may play a similar role, and that therefore L23 may be a chaperone-like ribosomal protein that contributes to maintaining RNC solubility during translation. Future studies will focus on genetic r-protein modifications aimed at disrupting the detected interactions. The large surface-exposed nonpolar patch of the L23 ribosomal protein (Figure 3-2b) suggests that the interactions involving apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$ RNCs may be contributed by the hydrophobic effect. This nonpolar patch appears an ideal candidate for genetic modifications aimed at significantly perturbing RNC/r-protein interactions.

Nascent chain-L23 complexes have the same apparent stability regardless of RNC sequence. To further explore the nature of the interactions between the L23 r-protein and nascent chains of increasing length and variable sequence, we performed urea titrations with chemical crosslinking detection (Figure 3-4a,b). EDC readily reacts with amines and carboxylic acid functional groups, and there is no loss of EDC reactivity even in the presence of high urea concentrations.[82] It is worth noting that the interactions identified in this work are not induced by the covalently N-terminal-linked BODIPY-502 fluorophore, as previous work has shown that this fluorophore does not interact with resuspended ribosomes under conditions like those of the present study.[17] Therefore, by unfolding the complex in the presence of urea and subsequently adding EDC, we expect to gain insights into the urea sensitivity of nascent chain-L23 complexes.

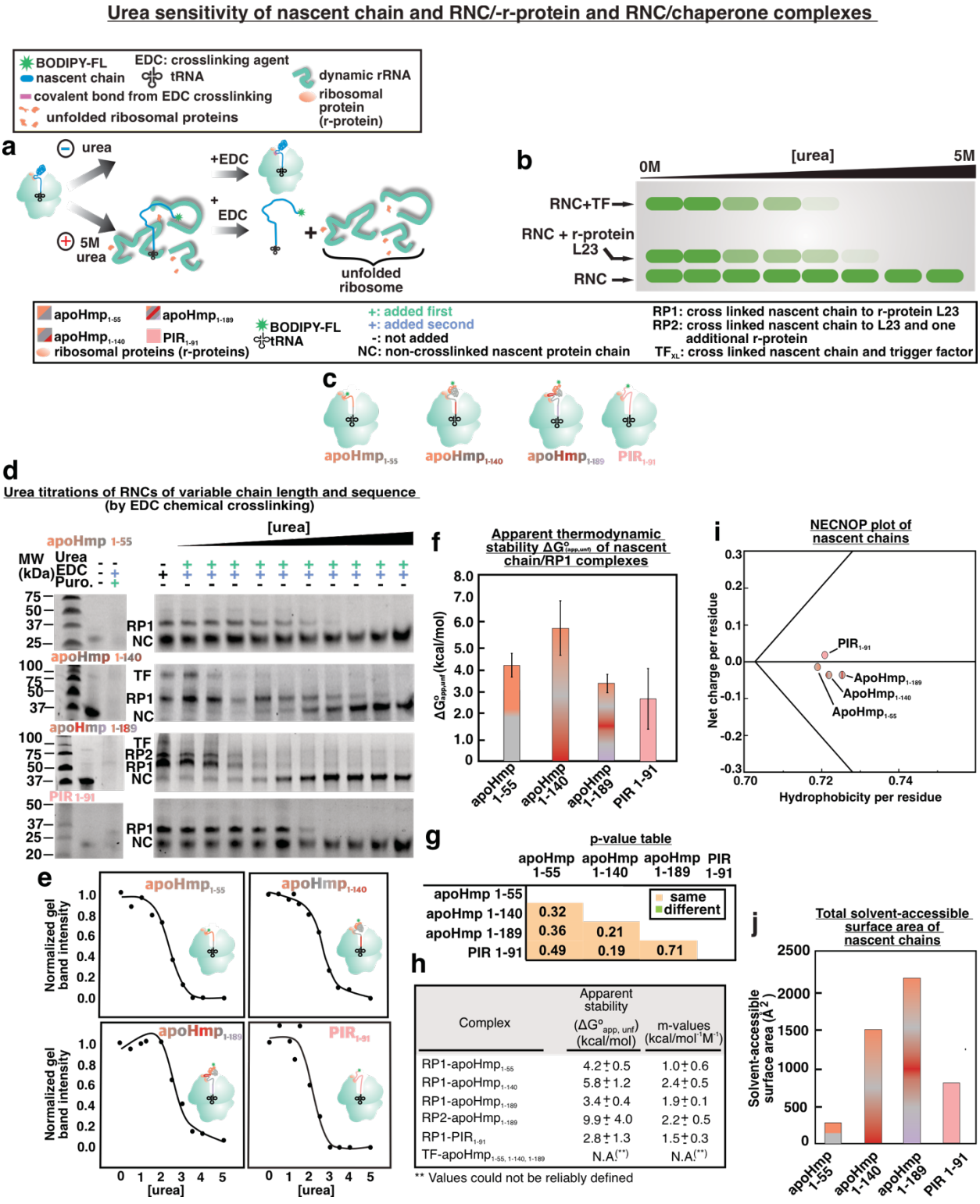


Figure 3-4. EDC-crosslinking-detected urea titrations showing the apparent stability of RNC/-r-protein complexes. **a)** Scheme showing the expected effect of urea addition on

RNCs and **b)** corresponding low-pH SDS-PAGE gels. **c)** Four RNCs were tested in these experiments:apoHmp$_{1-55}$, apoHmp$_{1-140}$, apoHmp$_{1-189}$ or PIR$_{1-91}$. Note that PIR$_{1-91}$ is an intrinsically disordered protein (IDP). **d)** Representative SDS-PAGE analysis. Gel bands are reporters of the apparent stability of RNCs complexes with either L23 or TF. **e)** Representative urea titrations of apoHmp$_{1-55}$, apoHmp$_{1-140}$, apoHmp$_{1-189}$ and PIR complexes. **f)** $\Delta G°_{app,unfold}$ values in the presence of low concentrations of chaperones (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively). Uncertainties are reported as ± SE for n=2-3. **g)** P-value table for a two-tailed Student's T-test, comparing the $\Delta G°_{app,unfold}$ values of RNC/r-protein complexes. Green and orange boxes denote statistically different and statistically equivalent data, respectively, according to a 95% confidence interval. **h)** Table displaying relevant $\Delta G°_{app,unfold}$ and m-values . **i)** NECNOP plot (100) displaying net charge/residue as a function of hydrophobicity/residue of PIR1-91, apoHmp1-55, apoHmp1-140, apoHmp1-189 protein chains. **j)** Estimated total solvent-accessible surface areas of protein chains, assuming fully extended conformations. Values were computed with Surfracer (101).

After collecting gel data on representative apoHmp and PIR nascent chains (Figure 3-4c), we estimated the apparent stability ($\Delta G°_{app,unfold}$) of nascent chain-L23 complexes following procedures similar to the other urea titrations described earlier.[73] Representative EDC-mediated urea titrations are shown in Figure 3-6c. Corresponding plots and apparent-stability data are displayed in Figure 3-4e,f. The matching two-tailed Student's t-test is provided in Figure 3-4g. As shown in Figure 3-4h, the apparent stability values for the apoHmp and PIR nascent-chain/L23 complexes range between ca. 2 and 5.5 kcal/ mol. Remarkably, all complexes display the same urea sensitivity within error. This is true even though the nascent-chain portions that are not buried within the exit-tunnel core have widely different nonpolar and net-charge-per-residue content (Figure 3-4i) and widely different total nonpolar surface accessible surface-area values (Figure 3-4j).

In summary, the urea-titrations in Figure 3-4 show that the urea sensitivity of r-protein-nascent-chain complexes is similar regardless of the nature and length of the nascent chain, across the short and long (55- to 189-residue) chains examined here. In other words, RNC/-r-protein complexes have the same apparent stability although the RNCs have widely different physical properties and compaction, as discussed above and, in the case of PIR lack of compaction, in previous work.[48, 49] A cartoon showing a low-resolution RNC model consistent with the crosslinking and fluorescence-anisotropy data is shown in Figure 3-1c-g. Further, the longer RNCs of apoHmp$_{1\text{-}189}$ also exhibit interactions with an additional yet-unidentified r-protein. The corresponding crosslinked complex is denoted as RP2 in Figure 3-3, 4 and 5. At this juncture, a few additional considerations come to mind. First, the 55- and 140-residue nascent chains of apoHmp lack or include the N-terminal 65-94-residue compact region detected via fluorescence anisotropy decay (Figure 3-1c-g), respectively. Therefore, the non-compact C-

terminal portion of the nascent chain out of the tunnel core seems to be primarily engaged in the detected interactions with the L23 r-protein in the case of apoHmp. Given that the amino-acid sequences of the interacting regions of $apoHmp_{1\text{-}55}$ and $apoHmp_{1\text{-}140}$ must be different yet the interactions are of comparable apparent strength, the interactions seem to be of nonspecific nature, amounting to an overall sufficient affinity the 100% interacting population detected in the gels of Figure 3-5. Second, $apoHmp_{1\text{-}140}$ and $apoHmp_{1\text{-}189}$ are also found to experience interaction of equivalent apparent strength with L23 (see RP1 gel bands in Figure 3-4d and g). On the other hand, $apoHmp_{1\text{-}189}$ also experiences interactions with another r-protein (see RP2 band in Figure 3-4d). This finding supports the idea that the RNC-r-protein contacts are of somewhat non-specific nature and longer RNCs interact with additional ribosomal surface, likely engaging a longer portion of their chain. Finally, previous work showed that the interactions experienced by $PIR_{1\text{-}91}$ are ca. 50% mediated by Mg^{+2} ions.[48] Yet the apparent strength of these interactions is not different from that experienced by the other nascent chains, suggesting that all interactions are relative weak and non-specific (Figure 3-4 f-h). This model is consistent with the fact RNC-r-protein interactions likely need to be continuously remodeled, as nascent chains elongate during translation. In all, the nature (Mg^{+2} dependence, nonpolar *vs* electrostatic, or else) of RNC-r-protein interactions is still poorly understood and clearly awaits future higher-resolution structural investigations.

Nascent chain and r-protein interaction strength does not vary in the presence of one or more molecular chaperone. Next, we explored the effect of molecular chaperones TF and Hsp70 on the RNC-r-protein interactions via the same type of EDC-mediated urea titrations employed in the last two sections. The Hsp70 chaperone was examined holistically as the as DnaK/DnaJ/GrpE chaperone system, denoted here as K/J/E. TF is known to associate with

prokaryotic ribosomes[83] and K/J/E works in cooperation with TF[84] to promote nascent-protein folding and prevent their aggregation.[85-88]

First, we evaluated apoHmp$_{1\text{-}189}$ devoid of both TF and the Hsp70 chaperone system (K/J/E), apoHmp$_{1\text{-}189}$ in the presence of low concentrations of TF only, apoHmp$_{1\text{-}189}$ in the presence of low concentrations of K/J/E only, and apoHmp$_{1\text{-}189}$ in the presence of both chaperones at low concentration (Figure 3-5a). Urea titrations were carried out using increasing concentrations of urea (Figure 3-5b), and the intensity of the crosslinked fraction was plotted as a function of urea (Figure 3-5c). We then obtained a $\Delta G°_{app,\ unfold}$ values for each of these constructs (Figure 3-5d) and evaluated them with a two-tailed Student's t-test (Figure 3-5g). Interestingly, in all construct variations tested, the apparent strength of the L23-nascent chain complex was found to be statistically similar (Figure 3-5f, h). Because this effect is not due to an increase or decrease in the fraction of crosslinked nascent chains to r-proteins (Figure 3-5e, f). Via Western Blot analysis, we were able to conclude that interactions mainly take place with r-protein L23 (Figure 3-3, b and SI Appendix, Figure 3-S11). This finding is consistent with our assessment of interacting proteins from apoHmp$_{1\text{-}140}$ (Figure 3-2 and SI Appendix, Figure 3-S9). This result suggests that in both the presence and absence of one or more molecular chaperones, nascent chains interact with ribosomal L23 in a structurally similar manner, though future work will need to be employed to validate this claim.

To further elucidate the nature of nascent chain and molecular chaperones, we performed experiments using increasing concentrations of TF and K/J/E in the presence of EDC (Figure 3-5i,j). Briefly, interactions with r-protein L23 can be displaced by high concentrations of molecular chaperones (8 μM TF and 10/2/4 μM K/J/E, respectively, Figure 3-5c,d). A low pH SDS-PAGE gel reporting on apoHmp$_{1\text{-}189}$ RNCs shows that at low concentrations of TF and

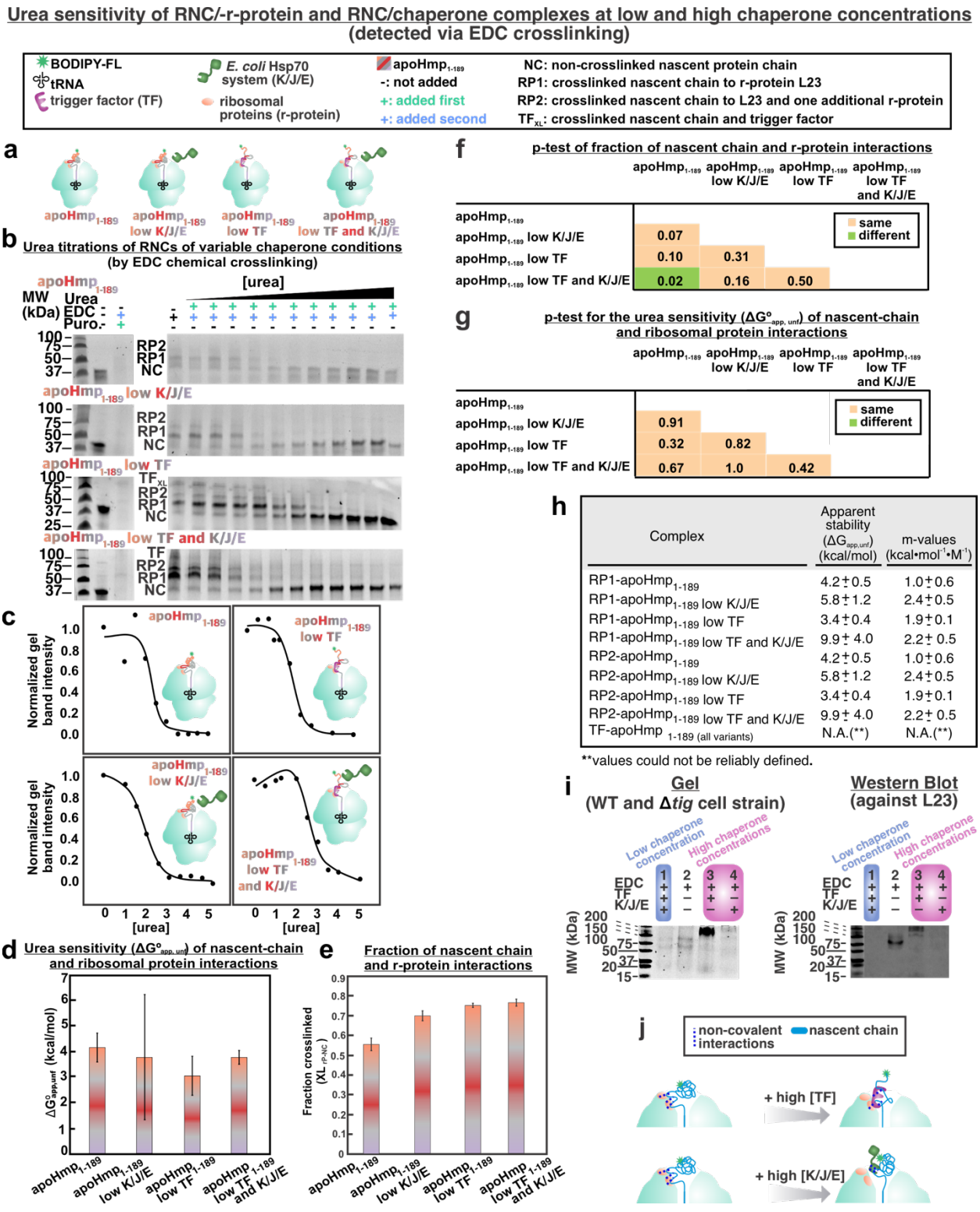


Figure 3-5. Low pH SDS-PAGE and urea-titration analysis of ApoHmp1-189 in the absence and presence of TF and K/J/E chaperones. **a)** Cartoon illustrating the tested

RNCs. **b)** Low pH SDS-PAGE analysis of complexes including apoHmp1-189 RNC and either r-proteins or molecular chaperones as a function of increasing urea concentration. **c)** Representative urea titration curves. **d)** ΔG°app,unfold values in the absence and presence of low concentrations (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively) of molecular chaperones. Error bars denote standard error based on 2-3 experiments. **e)** Fraction of RNC/r-protein complexes relative to total RNCs. Uncertainties are reported as ± SE for n=2-3. **f)** P-value table for two-tailed Student's test assuming unequal variances, comparing ΔG°app,unfold values. Green and orange boxes denote statistically different and statistically equivalent data, respectively, according to a 95% confidence interval. **g)** P-value table comparing fractions of RNC/r-protein complexes. Statistical assessments were similar to those listed in panel f. **h)** Table 3-Showing ΔG°app,unfold and m values of relevant complexes. **i)** Low-pH SDS-PAGE analysis of apoHmp1-189 RNCs treated with EDC in the absence and presence of chaperones at low (lanes 1 and 2, (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively) and higher concentrations (8 μM for TF and 10/2/4 μM for K/J/E respectively, lanes 3 and 4). Western Blot analysis probing interactions with ribosomal protein L23. **j)** Schematic representation of RNC interaction networks at low (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively) and higher (8 μM and 10/2/4 μM K/J/E, respectively) chaperone concentrations.

K/J/E (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively), interactions with two r-proteins are established (Figure 3-5i, gel lanes 1 and 2). Corresponding lanes in the Western Blot of this gel against L23 show that interactions involve r-protein L23 (Figure 3-5i, Western Blot lanes 1 and 2). Interestingly, at high TF concentration (8 μM) Figure 3-5i, gel lane 3), all r-protein interactions are displaced and replaced by contacts with TF (Figure 3-5i, Western Blot lane 3). TF and L23 are known to interact with one another on the ribosome,[78, 80, 88] though we presently cannot explicitly discriminate whether the nascent chains interacts with L23 and TF, or if the nascent chain interacts with TF, which in turn interacts with L23. Here, we propose the simplest scenario namely that RNCs interact with TF only and, in turn, TF interacts with L23, which is known to be the TF docking site on the ribosome.[2]

Similarly, interactions with r-proteins are also displaced by increased concentrations of K/J/E (10/2/4 μM K/J/E, respectively), although no interactions with L23 were detected when K/J/E concentrations are increased (Figure 3-5i Western Blot lane 4), L24 or L29 (SI Appendix, Figure 3-S11), suggesting that the nascent chain likely does not interact with any r-proteins (Figure 3-5i Western Blot lane 4). Similar results are seen for apoHmp$_{1\text{-}140}$ (SI Appendix, Figure 3-S9). Figure 3-5b and c shows model of the interaction interplay between nascent chains (c.a. length of 140-190 amino acids), r-proteins and molecular chaperones assessed in this work.

To summarize, when TF and K/J/E are not present, the nascent chain interacts with r-proteins only. In experiments with low concentrations of molecular chaperones (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively, Figure 3-6i,j), RNCs interact with TF but not K/J/E. At high, physiologically relevant concentrations of TF and K/J/E (8 μM TF and 10/2/4 μM K/J/E, respectively), RNC interactions with r-proteins are

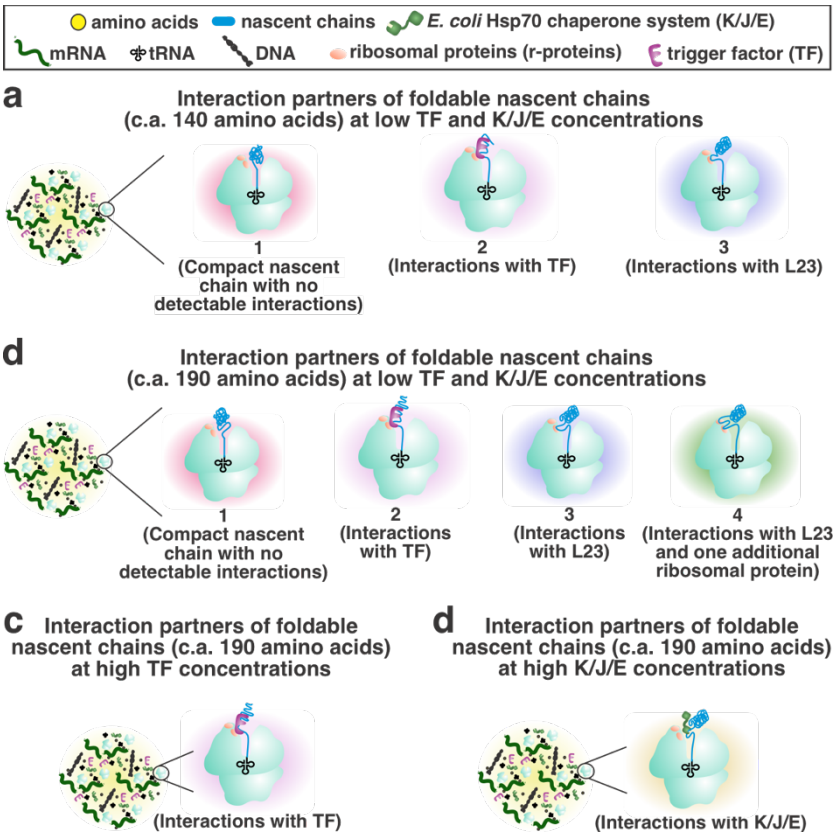


Figure 3-6. Cartoon summarizing the RNC interaction networks detected in this work under different experimental conditions. Interaction partners of **a)** apoHmp$_{1\text{-}140}$ and **b)** apoHmp$_{1\text{-}189}$ RNCs at low chaperone concentrations (2-15 nM of TF and 0.5 μM, 0.04μM and 0.05μM of and DnaK, DnaJ and GrpE respectively). **c)** Interaction partners of apoHmp$_{1\text{-}189}$ RNCs at higher **c)** TF (8 μM) and **d)** K/J/E chaperone concentrations (10/2/4 μM K/J/E, respectively .

displaced by interactions with molecular chaperones. It is worth nothing that TF and K/J/E chaperones are shared with thousands of additional cellular proteins *in vivo,* unlike in the experiments shown here, which include purified resuspended RNCs. Further, our RNC concentrations are only 20-30 nM. Hence, chaperones are in large excess over RNCs even at the low chaperone concentrations employed here. This scenario differs from the cellular environment where both RNCs and molecular chaperones are at comparable concentrations, within the low uM range. Therefore, we propose that the actual cellular milieu likely involves RNC populations that interact in part with TF and K/J/E and in part with r-proteins.

Concluding remarks on the RNC interaction network experienced by the foldable apoHmpH protein. An overall model recapitulating the main features of the RNC interactions network elucidated in this work in shown in Figure 3-6. Briefly, RNCs encompassing the entire N terminal domain of apoHmpH (apoHmp$_{1\text{-}140}$) experience either interactions with the L23 r-protein accompanied by independent tumbling of the N-terminal region or interactions with the TF chaperone (Figure 3-6a). On the other hand, RNCs comprising the entire N terminal domain of apoHmpH and additional 49 residues belonging to the second domain (apoHmp$_{1\text{-}189}$) are subject to either similar interactions with the L23 r-protein, or to interactions with L23 and one additional presently unidentified protein (RP2), or to interactions with TF (Figure 3-6b). Finally, in the presence of an excess of either TF or the Hsp70 chaperone system at total concentrations matching cellular levels, the interactions with ribosomal proteins are replaced by interactions with the respective molecular chaperones (Figure 3-6c-d). The emerging scenario resulting from our work suggests that the ribosome plays an RNC-interacting role entirely like that of molecular chaperones.

In all, our findings highlight the prominent role of the ribosome as an RNC interactor and

suggest that the ribosome may have played a primordial chaperone role in Nature, before the evolution of molecular chaperones.

3.5 Materials and Methods

Preparation of empty ribosomes. Empty ribosomes were generated from an in-house-prepared A19 WT or A19 Δtig *E. coli* S30 cell extract as described.[17, 89] Briefly, cells were grown in Luria-Bertani (LB) broth and harvested at mid-log phase (A_{600} ~ 0.6). The cells were lysed through a French press (thermo Electron Corporation, Waltham, MA) at ~12,000 psi with a single passage. The lysate was subject to centrifugation at 30910 g and 20 °C for 20 min. After centrifugation, the supernatant was incubated in translation buffer (0.75 M Tris-HCl pH 8.2, 7.5 mM DTT, 21 mM $Mg(OAc)_2$, 500 μM amino acids, 6 mM ATP, 67 mM PEP and 160 μg/mL pyruvate kinase) for 80 min to remove any endogenous mRNA from ribosomes. The supernatant was then dialyzed (12-14 kDa MWCO) in buffer (10 mM Tris-HCl pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KOAc and 1 mM DTT) for 12 hrs, with a buffer exchange every 4 hours. The resulting A19 cell extract was used as the empty-ribosome sample.

Preparation of RNCs. RNCs were generated using an in-house prepared A19 *E. coli* transcription-translation coupled cell-free system[17, 89] as described. Cell strains either including (WT) or lacking (*Δtig*) the trigger factor gene were employed.[17, 89] Hsp70 chaperone activity was suppressed via the KLR-70 peptide[90] to a final concentration of 0.2 mM. Transcription-translation proceeded for 30 min at 37 °C in the presence of BODIPY-FL-Met-$tRNA^{f\text{-}Met}$ to specifically label RNCs at the N terminus. BODIPY-FL-Met-$tRNA^{f\text{-}Met}$ was prepared as described.[17] RNCs were stalled at various lengths to generate the desired apoHmp and PIR constructs via oligodeoxynucleotide-directed mRNA cleavage.[17, 91, 92] An anti-ssrA

oligonucleotide[17] was added to a final concentration of 12.83 pmol/µL to prevent premature release of stalled RNCs. RNC pellets were isolated via a sucrose cushion (1.1 M sucrose, 20 mM tris base, 10 mM $Mg(OAc)_2$, 500 mM NH_4Cl, and 0.5 mM EDTA, 1 mM DTT, pH 7.0, as described)[17] and subjected to ultracentrifugation at 160,000 g for 1 hr at 4 °C. The purified pellet was dissolved in resuspension buffer (10 mM tris-HCl, 10 mM $Mg(OAc)_2$, 60 mM NH_4Cl, 0,5 mM EDTA and 1.0 mM DTT, pH 7.0) by shaking in an orbital shaker at 200 rpm on ice for 1 hr.

Other experimental procedures. Details on sucrose gradients, low-pH gels, puromycin assays, chemical crosslinking and urea titrations are available in the SI Appendix.

3.6 Acknowledgments

We are thankful to M. Dalphin for helpful discussions. This work was funded by the National Science Foundation (NSF) grants MCB-1616459 and MCB-0951209 (to S.C). M. M. M. and R.B.H. received NIH TEAM-Science Fellowships from the University of Wisconsin-Madison and M.M.M received the Straka Fellowship from the University of Wisconsin-Madison. A. E. V. received a National Science Foundation GRFP graduate fellowship and a Science and Medicine Graduate Research Scholars Fellowship from the University of Wisconsin-Madison.

3.7 References

(1) 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*, 274-282.

(2) 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.

(3) Fedyukina, D. V.; Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **2011,** *40*, 337-359.

(4) Pechmann, S.; Willmund, F.; Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **2013,** *49*, 411-421.

(5) Liutkute, M.; Samatova, E.; Rodnina, M. V. Cotranslational folding of proteins on the ribosome. *Biomolecules* **2020,** *10*, 97.

(6) Waudby, C. A.; Dobson, C. M.; Christodoulou, J. Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* **2019,** *44*, 914-926.

(7) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **2000,** *289*, 905-920.

(8) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **2001,** *107*, 679-688.

(9) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000,** *289*, 920-930.

(10) Voss, N.; Gerstein, M.; Steitz, T.; Moore, P. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **2006,** *360*, 893-906.

(11) Malkin, L. I.; Rich, A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* **1967,** *26*, 329-346.

(12) Blobel, G.; Sabatini, D. Controlled proteolysis of nascent polypeptides in rat liver cell fractions: I. Location of the polypeptides within ribosomes. *J. Cell Biol.* **1970,** *45*, 130-145.

(13) Wang, S.; Sakai, H.; Wiedmann, M. Nac covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. *J. Cell Biol.* **1995,** *130*, 519-528.

(14) Kramer, G.; Ramachandiran, V.; Hardesty, B. Cotranslational folding—omnia mea mecum porto? *Int. J. Biochem. Cell Biol.* **2001,** *33*, 541-553.

(15) Tsalkova, T.; Odom, O.; Kramer, G.; Hardesty, B. Different conformations of nascent peptides on ribosomes. *J. Mol. Biol.* **1998,** *278*, 713-723.

(16) 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*, 725-736.

(17) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(18) Hutchinson, R. B.; Chen, X.; Zhou, N.; Cavagnero, S. Fluorescence anisotropy decays and microscale-volume viscometry reveal the compaction of ribosome-bound nascent proteins. *J. Phys. Chem. B* **2021,** *125*, 6543-6558.

(19) Lu, J.; Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **2005,** *44*, 8230-8243.

(20) Mingarro, I.; Nilsson, I.; Whitley, P.; Von Heijne, G. Different conformations of nascent polypeptides during translocation across the er membrane. *BMC Cell Biol.* **2000,** *1*, 3.

(21) Bhushan, S.; Gartmann, M.; Halic, M.; Armache, J.-P.; Jarasch, A.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. A-helical nascent polypeptide chains

visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2010,** *17*, 313.

(22) Woolhead, C. A.; McCormick, P. J.; Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the dribosome and in their exposure to ribosomal proteins. *Cell* **2004,** *116*, 725-736.

(23) Agirrezabala, X.; Samatova, E.; Macher, M.; Liutkute, M.; Maiti, M.; Gil-Carton, D.; Novacek, J.; Valle, M.; Rodnina, M. V. A switch from α-helical to β-strand conformation during co-translational protein folding. *EMBO J.* **2022,** *41*, e109175.

(24) Kosolapov, A.; Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **2009,** *16*, 405-411.

(25) Nilsson, O. B.; Nickson, A. A.; Hollins, J. J.; Wickles, S.; Steward, A.; Beckmann, R.; von Heijne, G.; Clarke, J. Cotranslational folding of spectrin domains via partially structured states. *Nat. Struct. Mol. Biol.* **2017,** *24*, 221-225.

(26) Tian, P.; Steward, A.; Kudva, R.; Su, T.; Shilling, P. J.; Nickson, A. A.; Hollins, J. J.; Beckmann, R.; von Heijne, G.; Clarke, J., et al. Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115*, E11284.

(27) Holtkamp, W.; Kokic, G.; Jager, M.; Mittelstaet, J.; Komar, A. A.; Rodnina, M. V. Cotranslational protein folding on the ribosome monitored in real time. *Science* **2015,** *350*, 1104-1107.

(28) Liutkute, M.; Maiti, M.; Samatova, E.; Enderlein, J.; Rodnina, M. V. Gradual compaction of the nascent peptide during cotranslational folding on the ribosome. *Elife* **2020,** *9*.

(29) Ziv, G.; Haran, G.; Thirumalai, D. Ribosome exit tunnel can entropically stabilize α-helices. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 18956.

(30) Marino, J.; von Heijne, G.; Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **2016,** *590*, 655-660.

(31) Bañó-Polo, M.; Baeza-Delgado, C.; Tamborero, S.; Hazel, A.; Grau, B.; Nilsson, I.; Whitley, P.; Gumbart, J. C.; von Heijne, G.; Mingarro, I. Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nat. Commun.* **2018,** *9*, 5246.

(32) Wruck, F.; Tian, P. F.; 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*.

(33) 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.

(34) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(35) Bhushan, S.; Meyer, H.; Starosta, A. L.; Becker, T.; Mielke, T.; Berninghausen, O.; Sattler, M.; Wilson, D. N.; Beckmann, R. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **2010,** *40*, 138-46.

(36) Eisner, G.; Moser, M.; Schäfer, U.; Beck, K.; Müller, M. Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. *J. Biol. Chem.* **2006,** *281*, 7172-7179.

(37) Houben, E. N. G.; Zarivach, R.; Oudega, B.; Luirink, J. *Early encounters of a nascent membrane protein: Specificity and timing of contacts inside and outside the ribosome*. 2005; Vol. 170, p 27-35.

(38) Peterson, J. H.; Woolhead, C. A.; Bernstein, H. D. The conformation of a nascent polypeptide inside the ribosome tunnel affects protein targeting and protein folding. *Mol. Microbiol.* **2010,** *78*, 203-217.

(39) Ullers, R. S.; Houben, E. N.; Raine, A.; ten Hagen-Jongman, C. M.; Ehrenberg, M. n.; Brunner, J.; Oudega, B.; Harms, N.; Luirink, J. Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J. Cell Biol.* **2003,** *161*, 679-684.

(40) Bhushan, S.; Hoffmann, T.; Seidelt, B.; Frauenfeld, J.; Mielke, T.; Berninghausen, O.; Wilson, D. N.; Beckmann, R. Secm-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* **2011,** *9*, 10.

(41) Cruz-Vera, L. R.; Rajagopal, S.; Squires, C.; Yanofsky, C. Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell* **2005,** *19*, 333-343.

(42) Nakatogawa, H.; Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **2002,** *108*, 629-36.

(43) Seidelt, B.; Innis, C. A.; Wilson, D. N.; Gartmann, M.; Armache, J.-P.; Villa, E.; Trabuco, L. G.; Becker, T.; Mielke, T.; Schulten, K., et al. Structural insight into nascent polypeptide chain–mediated translational stalling. *Science* **2009,** *326*, 1412-1415.

(44) Carragher, B.; CHENG, Y.; FROST, A.; GLAESER, R. M.; LANDER, G. C.; Nogales, E.; WANG, H. W. Current outcomes when optimizing 'standard'sample preparation for single-particle cryo-em. *J. Microsc.* **2019,** *276*, 39-45.

(45) Zhang, Y.; Wolfle, T.; Rospert, S. Interaction of nascent chains with the ribosomal tunnel proteins rpl4, rpl17, and rpl39 of saccharomyces cerevisiae. *J. Biol. Chem.* **2013,** *288*, 33697-707.

(46) Burridge, C.; Waudby, C. A.; Wlodarski, T.; Cassaignau, A. M. E.; Cabrita, L. D.; Christodoulou, J. Nascent chain dynamics and ribosome interactions within folded ribosome-nascent chain complexes observed by NMR spectroscopy. *Chem. Sci.* **2021,** *12*, 13120-13126.

(47) Waudby, C. A.; Burridge, C.; Christodoulou, J. Optimal design of adaptively sampled NMR experiments for measurement of methyl group dynamics with application to a ribosome-nascent chain complex. *J. Magn. Reson.* **2021,** *326*.

(48) Guzman-Luna, V.; Fuchs, A. M.; Allen, A. J.; Staikos, A.; Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Commun. Biol.* **2021,** *4*, 1-17.

(49) Knight, A. M.; Culviner, P. H.; Kurt-Yilmaz, N.; Zou, T. S.; Ozkan, S. B.; Cavagnero, S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **2013,** *8*, 1195-1204.

(50) 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. L.; Wentink, A. S., et al. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* **2016,** *23*, 278-285.

(51) Cabrita, L. D.; Hsu, S. T. D.; Launay, H.; Dobson, C. M.; Christodoulou, J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106*, 22239-22244.

(52) Hsu, S. T. D.; Cabrita, L. D.; Fucini, P.; Christodoulou, J.; Dobson, C. M. Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J. Am. Chem. Soc.* **2009,** *131*, 8366-+.

(53) Subramanian, A. R.; Davis, B. D.; Beller, R. J. In *The ribosome dissociation factor and the ribosome-polysome cycle*, Cold Spring Harbor Symp. Quant. Biol., Cold Spring Harbor Laboratory Press: 1969; pp 223-230.

(54) Beller, R. J.; Davis, B. D. Selective dissociation of free ribosomes of Escherichia coli by sodium ions. *J. Mol. Biol.* **1971,** *55*, 477-485.

(55) Edelman, I. S.; Ts'o, P. O.; Vinograd, J. The binding of magnesium to microsomal nucleoprotein and ribonucleic acid. *Biochim. Biophys. Acta* **1960,** *43*, 393-403.

(56) Van Der Saag, P. T.; Vlak, J. M.; De Greef, T. F. Ribosomes from xenopus laevis eggs and embryos in a cell-free protein-synthesizing system: Translational regulation. *Cell Differ.* **1976,** *4*, 385-397.

(57) Ron, E. Z.; Kohler, R. E.; Davis, B. D. Magnesium ion dependence of free and polysomal ribosomes from Escherichia coli. *J. Mol. Biol.* **1968,** *36*, 83-89.

(58) Oppenheim, J.; Scheinbuks, J.; Biava, C.; Marcus, L. Polyribosomes in azotobacter vinelandii: I. Isolation, characterization and distribution of ribosomes, polyribosomes and subunits in logarithmically growing azotobacter. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1968,** *161*, 386-401.

(59) Kelley, W. S.; Schaechter, M. Magnesium ion-dependent dissociation of polysomes and free 70 s ribosomes in bacillus megaterium. *J. Mol. Biol.* **1969,** *42*, 599-602.

(60) Beller, R. J.; Lubsen, N. H. Effect of polypeptide chain length on dissociation of ribosomal complexes. *Biochemistry* **1972,** *11*, 3271-3276.

(61) Spirin, A. S. Structural transformations of ribosomes (dissociation, unfolding and disassembly). *FEBS Lett.* **1974,** *40*, S28-S37.

(62) Spitnik-Elson, P.; Greeman, B.; Abramovitz, R. The influence of 6-m urea on 30-s ribosomes of Escherichia coli. *Eur. J. Biochem.* **1974,** *49*, 87-92.

(63) Roberts, M. E.; Walker, I. O. Structural studies on Escherichia coli ribosomes: III. Denaturation and sedimentation of ribosomal subunits unfolded in urea. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1970,** *199*, 184-193.

(64) Roberts, M. E.; Walker, I. Structural studies on Escherichia coli ribosomes: III. Denaturation and sedimentation of ribosomal subunits unfolded in urea. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1970,** *199*, 184-193.

(65) Ilari, A.; Bonamore, A.; Farina, A.; Johnson, K. A.; Boffi, A. The x-ray structure of ferric Escherichia coli flavohemoglobin reveals an unexpected geometry of the distal heme pocket. *J. Biol. Chem.* **2002,** *277*, 23725-23732.

(66) Bonamore, A.; Boffi, A. Flavohemoglobin: Structure and reactivity. *IUBMB Life* **2008,** *60*, 19-28.

(67) Forrester, M. T.; Foster, M. W. Protection from nitrosative stress: A central role for microbial flavohemoglobin. *Free Radical Biol. Med.* **2012,** *52*, 1620-1633.

(68) Moncoq, K.; Broutin, I.; Larue, V.; Perdereau, D.; Cailliau, K.; Browaeys-Poly, E.; Burnol, A. F.; Ducruix, A. The pir domain of grb14 is an intrinsically unstructured protein: Implication in insulin signaling. *FEBS Lett.* **2003,** *554*, 240-6.

(69) Beechem, J. M.; Gratton, E. Fluorescence spectroscopy data analysis environment: A second generation global analysis program. In *Time-resolved laser spectroscopy in biochemistry*, Lakowicz, J. R., Ed.; SPIE, Bellingham, 1988; pp 70-81.

(70) Jameson, D. M.; Gratton, E.; Hall, R. D. The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry. *Appl. Spectrosc. Rev.* **1984,** *20*, 55-106.

(71) Ross, J. A.; Jameson, D. M. Time-resolved methods in biophysics. 8. Frequency domain fluorometry: Applications to intrinsic protein fluorescence. *Photochem. Photobiol. Sci.* **2008,** *7*, 1301-1312.

(72) Weinreis, S. A.; Ellis, J. P.; Cavagnero, S. Dynamic fluorescence depolarization: A powerful tool to explore protein folding on the ribosome. *Methods* **2010,** *52*, 57-73.

(73) Santoro, M. M.; Bolen, D. W. Unfolding free-energy changes determined by the linear extrapolation method.1.Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **1988,** *27*, 8063-8068.

(74) Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **1990,** *8*, 93-98.

(75) Hermanson, G. T., Bioconjugate techniques. 3rd ed.; Elsevier: 2013.

(76) Hoare, D. G.; Koshland, D. E. A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* **1967,** *242*, 2447-2453.

(77) Kirchdoerfer, R. N.; Huang, J. J. T.; Isola, M. K.; Cavagnero, S. Fluorescence-based analysis of aminoacyl- and peptidyl-tRNA by low-pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* **2007,** *364*, 92-94.

(78) Deckert, A.; Waudby, C. A.; Wlodarski, T.; Wentink, A. S.; Wang, X.; Kirkpatrick, J. P.; Paton, J. F. S.; Camilloni, C.; Kukic, P.; Dobson, C. M., et al. Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113*, 5012-5017.

(79) Raine, A.; Lovmar, M.; Wikberg, J.; Ehrenberg, M. n. Trigger factor binding to ribosomes with nascent peptide chains of varying lengths and sequences. *J. Biol. Chem.* **2006,** *281*, 28033-28038.

(80) Ferbitz, L.; Maier, T.; Patzelt, H.; Bukau, B.; Deuerling, E.; Ban, N. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **2004,** *431*, 590-596.

(81) Lakshmipathy, S. K.; Tomic, S.; Kaiser, C. M.; Chang, H.-C.; Genevaux, P.; Georgopoulos, C.; Barral, J. M.; Johnson, A. E.; Hartl, F. U.; Etchells, S. A. Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* **2007,** *282*, 12186-12193.

(82) Lewis, S. D.; Shafer, J. A. Conversion of exposed aspartyl and glutamyl residues in proteins to asparaginyl and glutaminyl residues. *Biochim. Biophys. Acta, Protein Struct.* **1973,** *303*, 284-291.

(83) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **2011,** *475*, 324-32.

(84) Agashe, V. R.; Guha, S.; Chang, H. C.; Genevaux, P.; Hayer-Hartl, M.; Stemp, M.; Georgopoulos, C.; Hartl, F. U.; Barral, J. M. Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. *Cell* **2004,** *117*, 199-209.

(85) Mecha, M. F.; Hutchinson, R. B.; Lee, J. H.; Cavagnero, S. Protein folding in vitro and in the cell: From a solitary journey to a team effort. *Biophys. Chem.* **2022,** *287*, 106821.

(86) Teter, S. A.; Houry, W. A.; Ang, D.; Tradler, T.; Rockabrand, D.; Fischer, G.; Blum, P.; Georgopoulos, C.; Hartl, F. U. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **1999,** *97*, 755-765.

(87) Wruck, F.; Avellaneda, M. J.; Koers, E. J.; Minde, D. P.; Mayer, M. P.; Kramer, G.; Mashaghi, A.; Tans, S. J. Protein folding mediated by trigger factor and Hsp70: New insights from single-molecule approaches. *J. Mol. Biol.* **2018,** *430*, 438-449.

(88) Deuerling, E.; Schulze-Specking, A.; Tomoyasu, T.; Mogk, A.; Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **1999,** *400*, 693-696.

(89) Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S. In vitro expression and characterization of native apomyoglobin under low molecular crowding conditions. *Protein Expr. Purif.* **2006,** *45*, 381-392.

(90) Dalphin, M. D.; Stangl, A. J.; Liu, Y.; Cavagnero, S. KLR-70: A novel cationic inhibitor of the bacterial Hsp70 chaperone. *Biochemistry* **2020,** *59*, 1946-1960.

(91) Behrmann, M.; Koch, H.-G.; Hengelage, T.; Wieseler, B.; Hoffschulte, H. K.; Müller, M. Requirements for the translocation of elongation-arrested, ribosome-associated OmpA across the plasma membrane of Escherichia coli. *J. Biol. Chem.* **1998,** *273*, 13898-13904.

(92) Donis-Keller, H. Site specific enzymatic cleavage of rna. *Nucleic Acids Res.* **1979,** *7*, 179-192.

3.8 Appendix

Supporting Text

Sucrose-gradient analysis of 70S empty ribosomes and ribosome-peptidyl-tRNA complexes. In order to explore the effect of nascent-chain characteristics on the bacterial ribosome, we examined whether the incorporation of aminoacyl initiator tRNA (Met-tRNA^{fMet}) or very short peptidyl tRNAs affects the ribosomal complex in the presence of urea. We collectively denote these species, which were generated via oligodeoxynucleotide-directed mRNA cleavage (see Methods),[1-3] as tRNAs carrying short nascent chains, or snc-tRNAs. Note that the antisense DNA construct used in the oligonucleotide-directed mRNA cleavage approach was designed to generate ribosome stalling after the first N-terminal residue (Met) of the nascent chain only. It is known that the *E. coli* RNAse H enzyme, employed here in conjunction with antisense oligodeoxynucleotides, typically establishes well-defined sharp cleavage sites, characterized by a site-specific distribution of the cleavage site of ca. 1-3 nucleotides.[3] Hence a dominant population of very short (1-3 residues) nascent chains in the ribosomal complexes encompassing snc-tRNAs is expected, consistent with the observed sharp gel bands (SI Appendix, Figure 3-S1). It follows that the nascent chains belonging to snc-tRNAs are not sufficiently long to interact with ribosomal proteins (r-proteins) across the ribosomal exit tunnel.[4] In addition, the specific thermodynamic stabilization imparted by RNA-oligodeoxynycleotide complexation was designed to be at least 12.4 kcal/mole, with a corresponding antisense-oligodeoxynucleotide length ranging from 11 to 38 DNA bases. The above free-energy value is significantly larger than what was used in the original report of the oligodeoxynucleotide-directed RNA cleavage approach.[3] Therefore, ribosomal complexes that include snc-tRNAs are expected to bear no contributions due to interactions of nascent chains with ribosomal proteins. It is worth noting,

however, that in the case of some specific RNA sequences under non-optimal conditions, the cleavage-site distribution width was found to be larger than 1-2 residues.[5] Therefore, in our case, any hypothetical longer RNCs than 1-3 residues are expected to be poorly populated.

Once appropriate RNCs harboring snc-tRNAs were made, we set out to test the response of the ribosome against exposure to urea. Specifically, we compared the urea sensitivity of empty 70S ribosomes (SI Appendix, Figure 3-S1a-b) to that of ribosomes harboring snc-tRNAs (SI Appendix Figure 3-S1c) by sucrose gradient-detected urea titrations. Sucrose gradients are able to resolve ribosomal subunits and entire 70S ribosomal particles.[6] Further, these gradients have previously been employed to monitor the unfolding of whole ribosomes or ribosomal subunits in the absence and presence of targeted structure-perturbing buffers[7-14] (e.g., containing EDTA or other related agents) or classical denaturing agents like urea.[15-17] Conveniently, high concentrations of urea do not perturb elution-profile integrity through line-broadening or other effects.[18] Therefore, urea titrations of the 70S ribosome including sucrose-gradient detection are a powerful tool to explore ribosomal-subunit dissociation and unfolding.

Sucrose-gradient elution profiles were monitored at 260 (SI Appendix Figure 3-S1) and 280 nm (SI Appendix, Figure 3-S2) in separate experiments, to probe for any potential differences in the response of rRNA and r-proteins. Our data show that rRNA and r-proteins belonging to 70S ribosomal particles devoid of tRNA and nascent chains, denoted here as empty ribosomes, are overall stable up to 1.0 M urea, with only a small extent of dissociation of the 70S particle into its 30S and 50S subunits (SI Appendix Figure 3-S1a). At 2 M urea, empty ribosomes undergo subunit dissociation accompanied by extensive line-broadening. Due to the dominant presence of intact 16S and 23S rRNA band in ethidium bromide-stained agarose gels (SI Appendix Figure 3-S3), line broadening is not ascribed to rRNA degradation. Therefore, we interpret the broad

sucrose-gradient peaks observed at 2 M urea as diagnostic of rRNA and(or) r-protein unfolding accompanied by conformational heterogeneity. At > 2 M urea, both the empty ribosome, and ribosomes harboring snc-tRNA undergo severe line-broadening (SI Appendix Figure 3-S1). At the highest urea concentration tested in this work (4 M), line-broadening is so extensive that it is difficult to deconvolute the contribution of individual ribosomal components. A similar scenario is supported by the empty-ribosome data at 280 nm, suggesting that rRNA and r-protein unfolding proceeds in concert (SI Appendix Figure 3-S2).

Ribosomes carrying tRNAs linked to very short nascent chains (snc-tRNA) are also characterized by progressive disassembly, as urea concentration increases (SI Appendix Figure 3-S1 and S2). Unlike empty ribosomes, however, the ribosomes harboring snc-tRNAs are still ca. 50% intact at 2 M urea. Further, at this urea concentration the 30S and 50S subunits (dissociated from 70S intact ribosomes) have not yet undergone any unfolding, given that no line broadening is observed.

Notably, the sucrose gradient profiles of ribosomes harboring snc-tRNAs also display two additional peaks eluting after the 70S ribosome (SI Appendix Figure 3-S1 and S2). These features are either due to polysomes (i.e., nearby ribosomes linked via the same mRNA strand) or to self-associated ribosomes brought together by through-space noncovalent surface contacts. The mRNA encoding short nascent-chains (sncs) is predominantly cleaved only 11 ribonucleotides away from the mRNA ribosome binding site, i.e. the Shine-Dalgarno sequence.[19] Ribosome profiling[20] and structural considerations reveal that each translating bacterial ribosome spans a length corresponding ca. 24 ribonucleotides.[21] Further, computer simulations suggested that polysomes have c.a. 24 residues between neighboring ribosomes.[22] Therefore, geometrical factors render polysome formation very unlikely. To gain further insights, transmission electron

microscope (TEM) negative stain images of ribosomes carrying snc-tRNAs were acquired in the presence of 2% methyl-tungstate.[23] The representative TEM image displayed in SI-Appendix Figure 3-S1f shows that, in addition to isolated ribosomes (within dashed blue squares), some closely spaced ribosomes (within dashed red squares) comprising 2 or more particles are also present. The spatial arrangement of these particles renders it impossible to establish where these species are polysomes or other forms of self-associated ribosomes. While polysomes seem unlikely due to the above-listed geometrical arguments, it is possible that some longer-than expected nascent chains may be populated in these samples, preventing the ruling out of polysomes. In all, the origin of the late-eluting peaks found in ribosome samples harboring snc-tRNAs remains unestablished and further future investigations are required to shed further light on this matter. On the other hand, the disassembly pattern of these peaks is entirely like that of 70S ribosomes harboring snc-tRNAs. Therefore, the late-eluting peaks do not add any new information nor modify the conclusions reached for the 70S particles.

The urea sensitivity of a representative ribosome harboring a longer nascent chain (32-residue long) derived from *E. coli* $Hmp_{108\text{-}140}$ has also been analyzed. The results are shown in SI-Appendix Figure 3-S1 and S2. These ribosomes behave in the same way as ribosomes harboring snc-tRNAs, suggesting that nascent-chain length does not affect the apparent stability of the 70S ribosome, and that the enhanced apparent stability may be dominated by the role of the tRNA. Additional experiments probing this topic in further detail are described in some sections of the main manuscript.

In summary, the data in SI-Appendix Figures 3-S1 to 3-S3 show that empty-intact 70S ribosomes are more sensitive to urea-induced denaturation than ribosomes carrying tRNAs linked to nascent chains of variable lengths.

Nascent-chain sequence and length (beyond 32 residues) does not affect the PTC.

Puromycin is a small-molecule antibiotic that induces premature release of nascent polypeptides from the ribosome. It mimics the adenosine-Phe portion of the CCA 3' end of Phe-tRNA^{Phe} (SI-Appendix, Figure 3-S5). Puromycin gets incorporated into nascent polypeptides as a result of nucleophilic attack of the carboxyl C -terminus of the nascent chain.[24] On the other hand, this antibiotic works properly only if the A and P sites, which are entirely located within the 50S subunit, are intact.[25] At high urea concentrations, the PTC is denatured and puromycin is no longer able to promote the release of nascent protein chains from the ribosome.

The main features of the puromycin ribosome-release assay and its urea-concentration dependence are illustrated in SI-Appendix Figure 5a, b. In short, RNC reactivity to puromycin is monitored via low-pH gels[26] as a function of increasing urea concentration.

We performed low-pH SDS-PAGE analysis on a variety of apoHmp RNCs comprising variable chain-length values (SI-Appendix, Figure 3-S5), net charge and hydrophobicity, and degree of folding. The target nascent protein chains included snc (i.e., a 1-3 amino acid chain derived from apoHmp), apoHmp$_{1\text{-}32}$, apoHmp$_{108\text{-}140}$, apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$. These constructs specifically enabled us to probe differences between nascent-chain sequences located within the ribosomal exit tunnel core, as well as partially and fully folded nascent-chain domains. Low-pH SDS-PAGE was used to generate peptidyl tRNA (PT) bands whose intensities were quantified as a function of increasing urea concentration. The peptidyl-puromycin band (PP) was not used to track ribosome release because of its highly environmentally sensitive fluorescence intensity. Titration curves were generated based on the relative band intensities of the PT bands of each construct, reporting on the extent of puromycin reactivity (SI-Appendix Figure 3-S5c). The data were then fit to an equation relating the experimental observable at variable urea

concentrations to the apparent stability ($\Delta G°_{app,\ unfold}$) of each construct, following the general procedures by Santoro and Bolen (see Methods).[27, 28]

As shown in SI-Appendix Figure 3-S5d-f, the results indicate that the apparent stability of the PTC center, monitored via puromycin activity assays, is very similar for all nascent-chain construct. Moreover, the two-tailed Student's test assuming unequal variances (Welch's test) shows that most constructs display equivalent behavior (SI-Appendix Figure 3-S5f). In other words, most constructs contribute to a similar extent to the apparent stability of the PTC, amounting to ca. 2.5 – 4.5 kcal/mol. Therefore, the urea sensitivity of the PTC does not depend on nascent-chain properties. and may be dominated by the stabilizing effect of the tRNA. The results also suggest that the specific tRNA sequence does not have an effect either. An exception is provided by snc and apoHmp$_{1\text{-}140}$ nascent proteins, given that the apoHmp$_{1\text{-}140}$ construct displays a statistically significant PTC-stabilizing role. This feature is likely not due to the compact partially folded state of apoHmp$_{1\text{-}140}$, given that fluorescence anisotropy shows that nascent apoHmp$_{1\text{-}189}$ has a comparable degree of compaction.

Note that ribosome disassembly and unfolding due to denaturing agents is an irreversible process.[29] Therefore, true thermodynamic $\Delta G°_{unfold}$ values cannot be obtained, upon treating the ribosome with denaturing agents. Thus, we refer to the $\Delta G°_{unfold}$ values derived in this and the following sections of this work as *apparent stability* values. The mere function of these quantities, which were derived from data fitting of urea-titrations, is to describe the urea sensitivity – and not the thermodynamic stability – of specific portions of *E. coli* ribosomes.

Sequence of events leading to ribosome disassembly: role of tRNA and nascent proteins. Finally, a proposed equilibrium ribosome disassembly mechanism based on the data of SI-Appendix Figures 3-S1 through 3-S6 is shown in SI-Appendix Figure 3-S8. The simple steps

displayed in panel a of SI-Appendix Figure 3-S8 pertain to empty 70S ribosomes and are consistent with the sucrose gradient data. The ribosome starts disassembling into its component 50S and 30S subunits at 1 M urea. More extensive subunit disassembly together with subunit unraveling follows at higher urea concentrations. This process is accompanied by pervasive r-protein and rRNA conformational heterogeneity, likely due to disruption of secondary and tertiary structure leading to r-RNA and r-protein unfolding. Panel b of the same figure shows how the process gets modified if the ribosome carries aminoacyl or peptidyl tRNA. Briefly, in this case the ribosomal-subunit disassembly occurs at higher urea concentrations (1 -2 M). Given the weak dependence of the sucrose-gradient and puromycin assays on the nature of the nascent chain, we deduce that the tRNA likely dominates the effect and that the length, hydrophobicity, net charge and foldability of apoHmp nascent chains do not play a stabilizing role in ribosome stability.

Supporting Materials and Methods

Denaturation of ribosome-nascent-chain complexes and empty ribosomes. A 10 M stock solution of 0.22 μm-filtered urea was prepared in resuspension buffer and the refractive index was measured with an Abbe Refractometer (Thermo Spectronic, Fisher Scientific) to derive actual urea concentrations as described.[30] RNCs subject to sucrose cushion ultracentrifugation (see section on RNC preparation) or empty ribosomes obtained from crude S30 were incubated in the presence of variable concentrations of urea for 1 hr at ambient temperature in the dark.

Assessment of urea sensitivity of 70S ribosome and RNCs via sucrose gradient analysis. An in-house prepared A19 *E. coli* mixture (with 70S ribosomes) was incubated for 1 hr at variable urea concentrations at ambient temperature. Samples were loaded onto a 5-45% sucrose gradient

and centrifuged using a Beckman L-70 Ultracentrifuge with a SW41 rotor at 288, 000 x g for 1.5 hrs at 4 °C. Gradient profiles were obtained on a Biocomp Fractionator at 0.2 mm/sec. The absorbance was measured at 254 nm and 280 nm using a Triax flow cell from BioComp to check for intactness of the rRNA and r-proteins, respectively. RNCs were treated in a similar manner after denaturation in urea for 1 hr (see RNC denaturation in Methods). The 30S, 50S and 70S subunits were collected in both instances and loaded on a 2% agarose gel (0.02 M Tris Base, 0.01 M acetic acid and 0.0005 M EDTA, or 0.5x TAE). Ethidium bromide was added to a final concentration of 0.5 μg/mL. Samples were run for 100 min at 3.92 V/cm. The gel was imaged on a GE FLA 9500 Laser Imager at PMT values between 500-700.

Imaging of RNCs via negative staining. RNC samples prepared in the presence of Met-$tRNA^{fMet}$ (2 μL) were placed onto a glow-discharged copper 300-mesh formvar-carbon grid (made in house by Medical Sciences Electron Microscopy staff at UW-Madison), blotted with filter paper and allowed to dry. A Nano-W staining solution (Nanoprobes) was placed on the grid in equal volume, blotted with filter paper and allowed to dry. Images were collected on a CM120 transmission electron microscope (Philips) at 140000x magnification and 80 keV using a BIOSPR12 camera.

Assessment of apparent stability of RNC/-r-protein and RNC/chaperone complexes via a crosslinking assay. The EDC crosslinker is capable of capturing interactions involving RNCs and r-proteins or chaperones within the ribosomal exit-tunnel vestibule and outside the ribosomal exit-tunnel core.[31] Either ribosome-bound or ribosome-released control samples (with EDC added after RNC ribosome release) were used to probe RNC production and to test the lack of ribosome-released nascent-chain crosslinking to other species. After incubation in urea for 1 hr, a 10x concentration of EDC solution (800 mM EDC, pH 6.8-7.0) was added to RNCs to a final

concentration of 1x. Samples were incubated for 30 min at 30 °C and then quenched with a 10x concentration of Quenching Buffer (1.0 M Tris-HCl pH 7.0, 1.0 M Glycine, 1.0 M KOAc) to a final concentration of 1x (0.1 M Tris-HCl pH 7.0, 0.1 M Glycine, 0.1M KOAc). Samples were loaded onto a low-pH SDS-PAGE gels using either a 10% acrylamide gel (apoHmp$_{1\text{-}55}$, apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$) or a 9% acrylamide gel (PIR) in a 1:1 ratio with loading buffer. Gels ran at 3.92 V/cm for either 4 hours (apoHmp$_{1\text{-}55}$, apoHmp$_{1\text{-}140}$ and apoHmp$_{1\text{-}189}$) or 2.5 hrs (PIR) and were imaged on a GE FLA 9500 Laser Imager. Fluorophores were excited at 473 nm, with a PMT value within the 700-950 range.

Crosslinked band intensities were evaluated via the ImageJ software.[32] The normalized intensity of each band was assessed via relation $I_{XL} = \frac{I_{1,XL}}{I_{2,XL}}$, where I_{XL} is the normalized intensity of an individual crosslinked species, $I_{1,\,XL}$ is the intensity of the individual crosslinked species, $I_{2,\,XL}$ is the intensity of the species at 0 M urea. The fractional band intensities were then plotted as a function of urea concentration and then plotted to fit pre- and post- transition region baselines. A 2-state unfolding expression taking pre- and post-transition baseline slopes into account[27] was used to fit the raw urea-titration data and deduce m-value and $\Delta G°_{H2O}$ values via the Kaleidagraph software.[28]

Western Blot analysis of r-protein interactions with L23, L24 and L29. Western Blots were performed as described.[31] Aliquots of the rabbit anti-uL23 antibody were kindly donated by Shu-ou Shan (California Institute of Technology). The anti-uL23 antibody was generated by GenScript, using the CGKVKRHGQRIGRRS peptide as epitope and has been validated in previous work.[33] Rabbit anti-uL17, -uL18/L22, -uL24, -uL29, and -uL32 antibodies were kindly facilitated by Bryan W. Davies (University of Texas-Austin) and Melanie Oakes (University of California, Irvine) who obtained them from Masayasu Nomura (University of Wisconsin-

Madison). The antibodies were generated using the purified *E. coli* ribosomal proteins L17, L23, L18+L22, L24, L29, and L32.[34]

Assessment of apparent stability of PTC via a puromycin-assisted nascent-chain release assay. To confirm that the polypeptide was attached to the ribosome, a low-pH SDS-PAGE[26] using a 9% acrylamide gel was conducted before (positive control) and after treatment of 1 mM puromycin, which reacted with the samples for 30 minutes at 37 °C. Samples were loaded on gels in a 1:1 ratio with loading buffer. Identical samples of the positive control and puromycin-released samples at 2.23 M urea were loaded on each gel to control for intrinsic gel differences (i.e., gel crosslinking, which may affect the fluorophore quantum yield). Prior to gel loading, samples were heated at 37 °C for five min and allowed to sit at room temperature for 5 min. Gels ran at 3.92 V/cm for 3.5 hr and imaged on a GE FLA 9500 Laser Imager. Fluorophores were excited at 473 nm and a PMT value between 315-700.

Due to the number of samples, 2 gels were needed for the positive control samples and 2 gels were needed for the puromycin-released samples per experiment. Fluorescence intensity adjustments were made between gels by normalizing the 2.23 M urea samples from each gel via Equation 1

$$I_{\text{intergel}} = \frac{I_{\text{2.23 M urea,gel 1}}}{I_{\text{2.23 M urea,gel 2}}} * I_{\text{x,gel 2}} \qquad \text{(S1)}$$

where $I_{\text{2.23 M urea, gel1}}$ and $I_{\text{2.23 M urea, gel2}}$ refer to the band intensities of the 2.23 M urea sample in each gel. I_{gel2} is the band intensity of the sample loaded onto the second gel and $I_{\text{x,intergel}}$ is any given band intensity on the second gel in comparison to the band intensity of the first gel.

To control for differences in RNC concentrations and compare band intensities between the bound and released sample gels, Equation 2 was used

$$I_{\text{relative}} = \frac{I_{+\text{puro}}}{I_{-\text{puro}}}, \qquad \text{(S2)}$$

where I_{+puro} is the band intensity of the puromycin-released sample and I_{-puro} is the band intensity of the positive control. Band intensities were then divided by the band intensity of the sample containing the highest urea concentration to normalize intensities between 0 and 1. The resulting intensities were plotted to fit pre- and post- transition region baselines. Using an extrapolation method.[27], the transition region slope and y-intercept were deduced to obtain the m-value and $\Delta G°_{H2O}$ with the Kaleidagraph software.[28]

Ribosomal-protein (r-protein) stability assessment via a Trp fluorescence-emission assay. The *E. coli* r-proteins collectively contain 32 Trp residues. Trp is sensitive to changes in its environment and exposure to polar solvents causes a red-shift in its excitation spectra, which was monitored as a function of increasing urea concentrations.[35-37] RNC samples were excited at 285 nm (bandpass of 4 nm) and the fluorescence was monitored from 295-500 nm (bandpass of 4 nm) on a Photon Counting Spectrofluorimeter (ISS) since the indole group of tryptophan is the major component of UV absorbance in that region.[38]

To generate a titration curve for the RNC complex, the buffer spectra was first subtracted from the produced emission spectra. Then a baseline correction was done on the resulting spectra.

The spectral center of mass of the resulting spectra was obtained using the emission spectra between 300-385 nm (to omit the scatter peak and Raman peak) and Equation 3

$$\text{Spectral center of mass (nm)} = \frac{\Sigma(\lambda) \times I_{\lambda}}{\Sigma(I)} \ , \qquad \text{(S3)}$$

where λ is the wavelength and I_{λ} is the intensity at a specific wavelength.

The fraction of unfolded protein at each concentration of urea was determined according to

$$\Delta\text{Fraction of unfolded ribosome} = \left(1 + Q\left(\frac{\lambda_x - \lambda_{unfold}}{\lambda_{fold} - \lambda_x}\right)\right)^{-1}, \quad \text{(S4)}$$

where λ_{fold} is the shortest wavelength calculated from Equation 3 for the folded species and λ_{unfold} is the longest wavelength calculated from Equation 3 for the unfolded species. Q denotes the ratio between the quantum yields of folded and unfolded states. Quantum-yield changes were calculated by taking highest intensity values from each folded sample (0.0 M, 0.13 M and 0.45 M urea) as well as each unfolded sample (5.5 M, 6.0 M and 6.5 M urea). The three folded and unfolded values were averaged amongst their respective groups to determine Q. ΔCoM was plotted as a function of urea concentration and the pre- and post- transition baselines were determined in Microsoft Excel. Free energy of unfolding curves were generated with the software Kaleidagraph (Synergy Software) using a known extrapolation method.

Supporting Figures

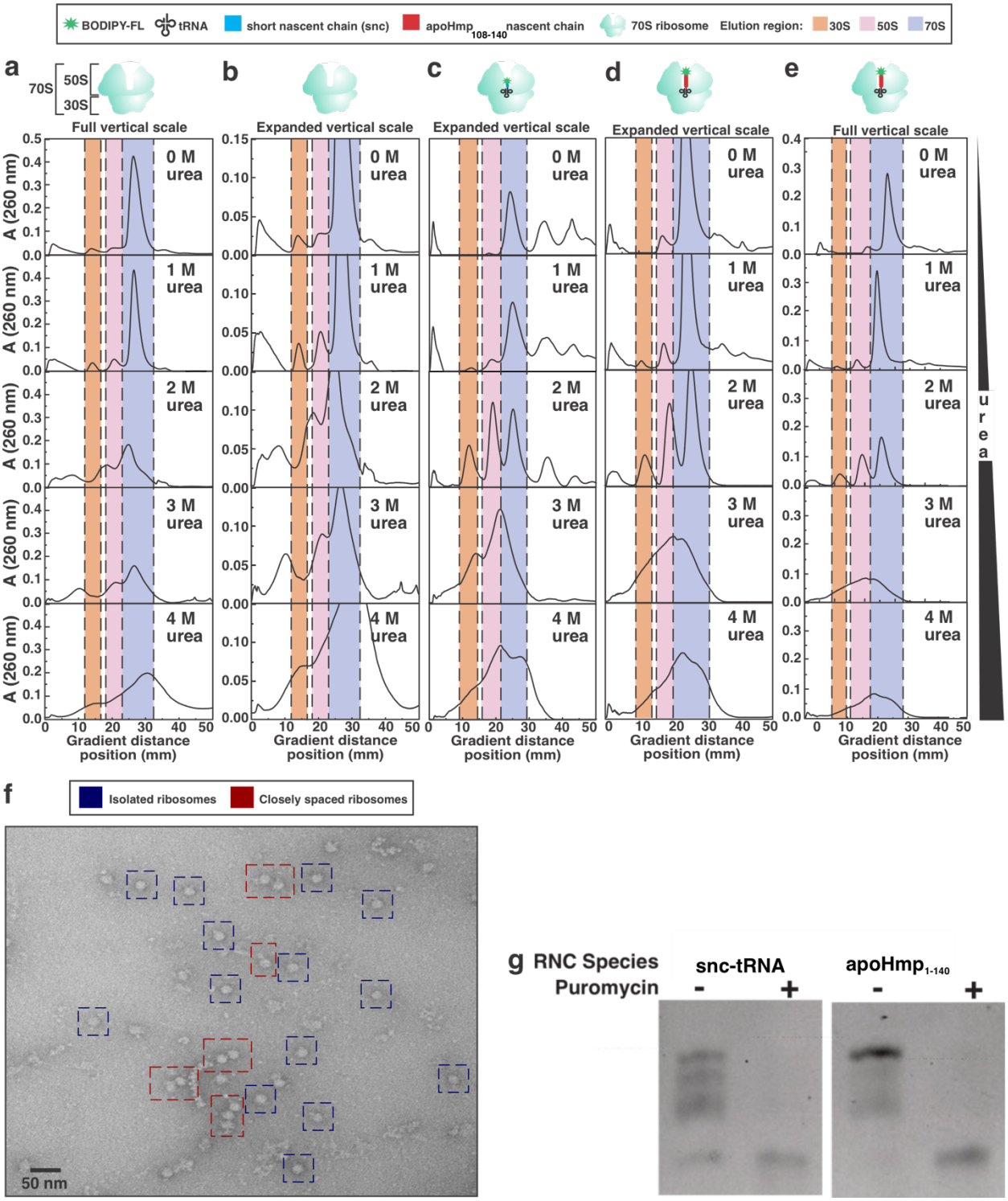


Figure 3-S1. Sucrose gradient profiles for both empty 70S ribosomes (a), the same 70S ribosome profile is set to the same y-axis scale as the short (1-2 amino acid) nascent chains (snc-tRNA), (b) ribosomes harboring short (1-2 amino acid) nascent chains (c), ribosomes bearing a longer nascent-peptide chain (d), and the same ribosomes bearing a longer nascent-peptide chain at the full vertical scale (e) are shown for increasing concentrations of urea. Profiles were collected at 260 nm. (f) A TEM image of the snc-tRNA is shown. The dark blue squares show what are likely isolated 70S ribosomes and the red squares show closely spaced ribosomes.
(g) Bound and released snc-tRNA and apoHmp$_{1\text{-}140}$ nascent chains with the addition of puromycin (released) or without (bound) on a low pH 10% SDS-PAGE gel.

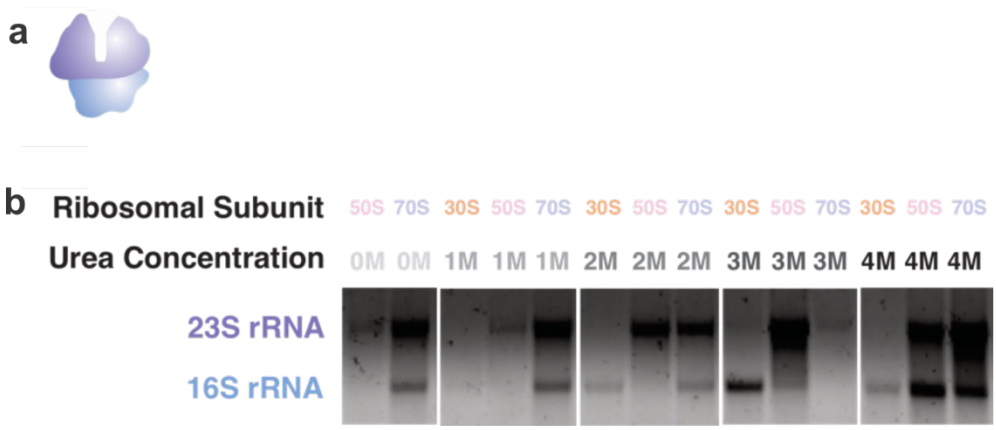


Figure 3-S2. Sucrose gradient profiles for both empty 70S ribosomes (a), the same 70S ribosome profile is set to the same y-axis scale as the short (1-2 amino acid) nascent chains (b), and ribosomes harboring short (1-2 amino acid) nascent chains (c), ribosomes bearing a longer nascent-peptide chain (d), and the same ribosomes bearing a longer nascent-peptide chain at the full vertical scale (e) are shown for increasing concentrations of urea. Profiles were collected at 280 nm.

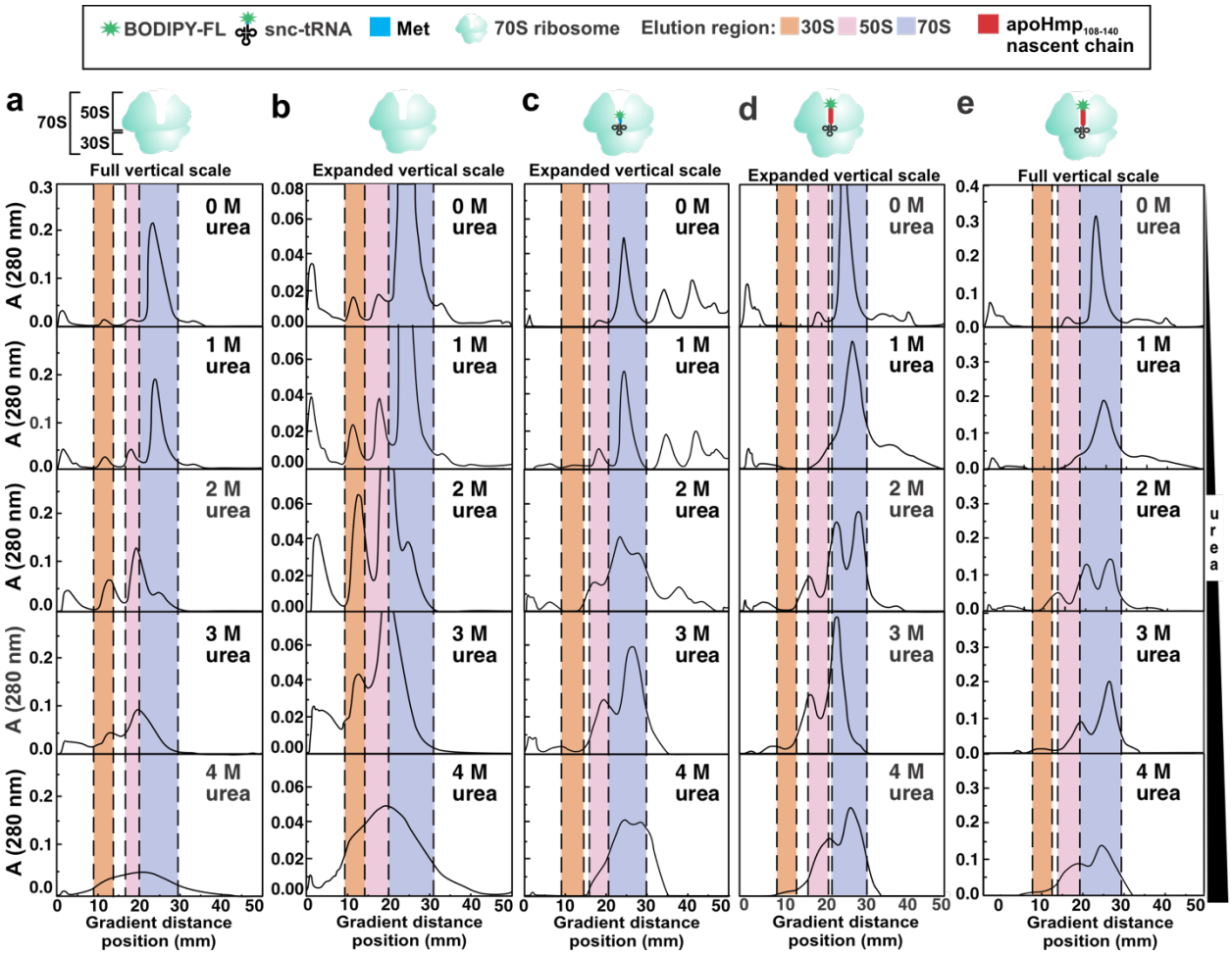


Figure 3-S3. (a) The 70S bacterial ribosome contains two subunits, the 30S (blue) and the 50S (purple), which contain the 16S and 23S rRNA respectively. (b) Representative agarose gel stained with ethidium bromide for the 30S, 50S and 70S of ribosome samples at increasing concentrations of urea. Subunit peak corresponds to the area in which the sample was collected.

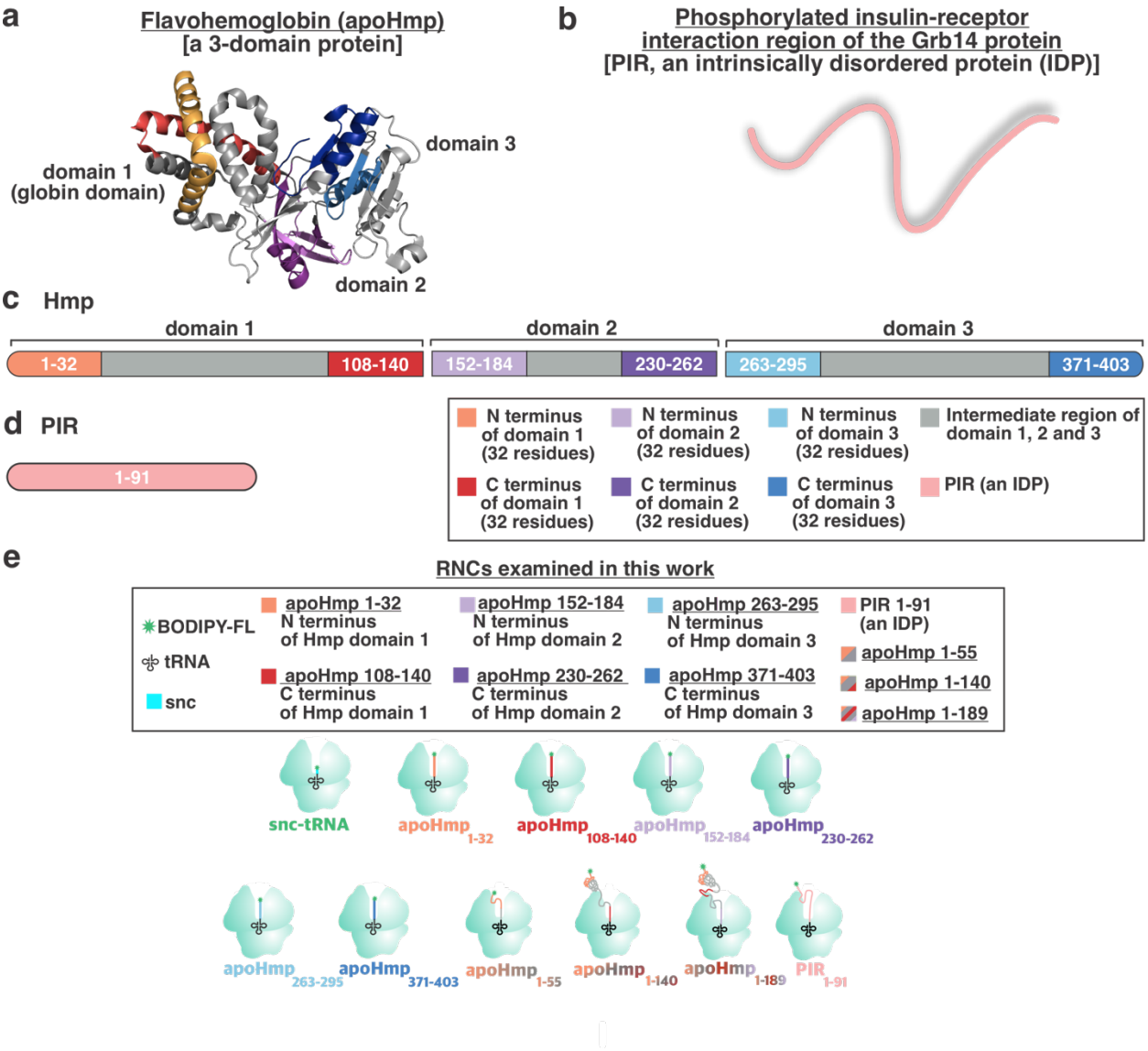


Figure 3-S4. a) *E. coli* flavohemoglobin (apoHmp) is the model foldable protein used in this study. apoHmp has three domains (shown in red, purple and blue). PDB code: 1GVH. **b)** The phosphorylated insulin receptor interacting region of the Grb14 protein from rat (PIR) is the model intrinsically disordered protein used in this work. **c)** A pictorial representation of the three domains of ApoHmp and their corresponding N-and C-termini are shown. **d)** The pictorial sequence length of PIR is shown to scale. **e)** Eleven constructs were derived from apoHmp to determine whether nascent chain length or sequence had any effect on the stability of r-proteins, and four representative constructs were chosen to assess peptidyl-transferase center (PTC) site

stability. The construct denoted “snc-tRNA” had a nascent chain length of one to three amino acids that were used to determine the relative contribution of tRNA to the stability of the ribosomal complex. apoHmp$_{108-140}$, apoHmp$_{1-32}$, apoHmp$_{230-262}$, apoHmp$_{152-184}$, apoHmp$_{371-403}$ and apoHmp$_{263-295}$ denoted constructs contained either the N- or C- termini of the respective domain with a chain length of 32 amino acids. apoHmp$_{1-55}$ contains all the apoHmp$_{1-32}$ amino acids, plus the next native 23 amino acids in the apoHmp sequence. apoHmp$_{1-140}$ denotes the entire first domain of apoHmp (140 amino acids). apoHmp$_{1-189}$ contains the entire first domain of apoHmp outside of the ribosomal exit tunnel. PIR was used to elucidate any differences in stability between foldable proteins and IDPs (91 amino acids).

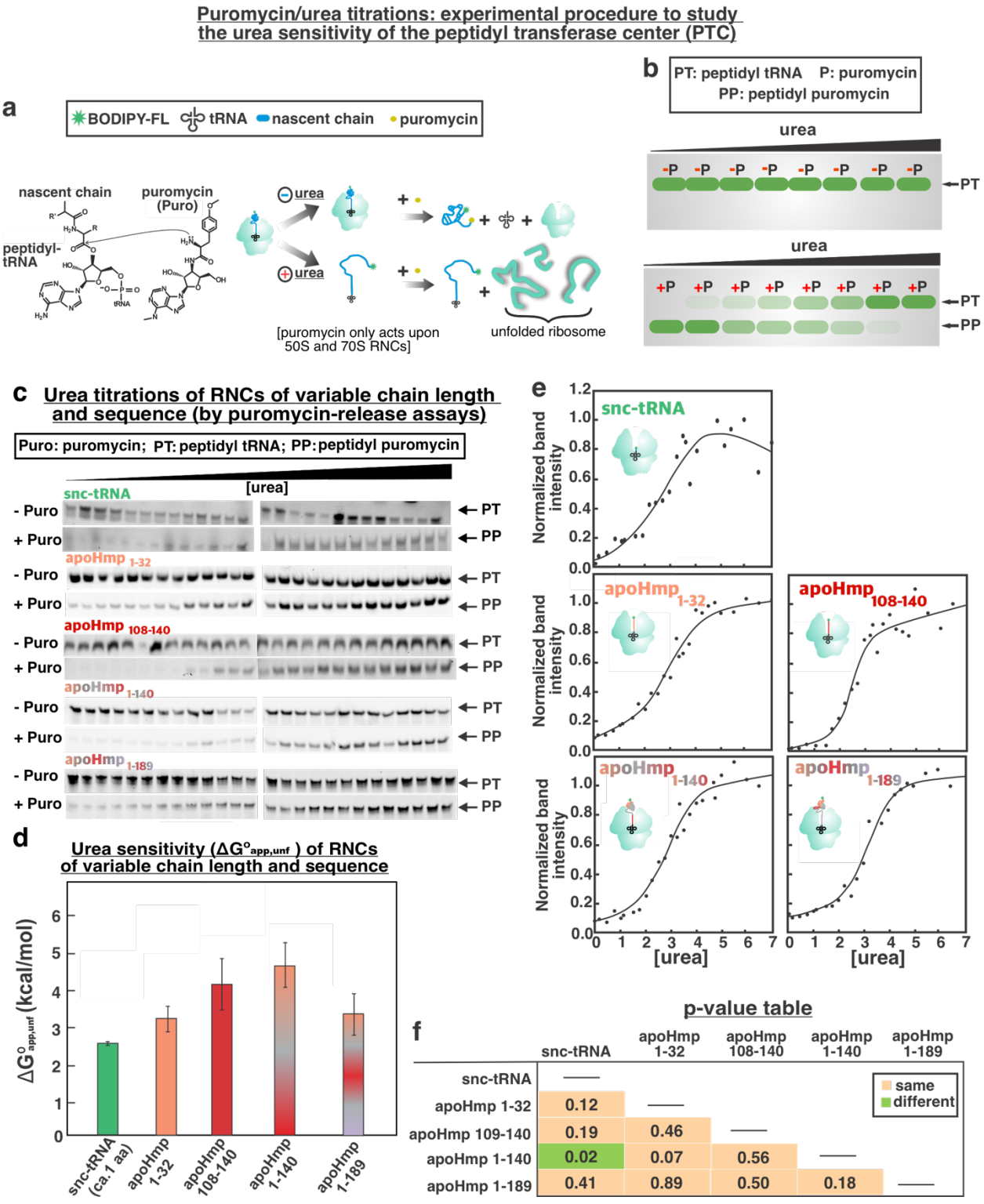
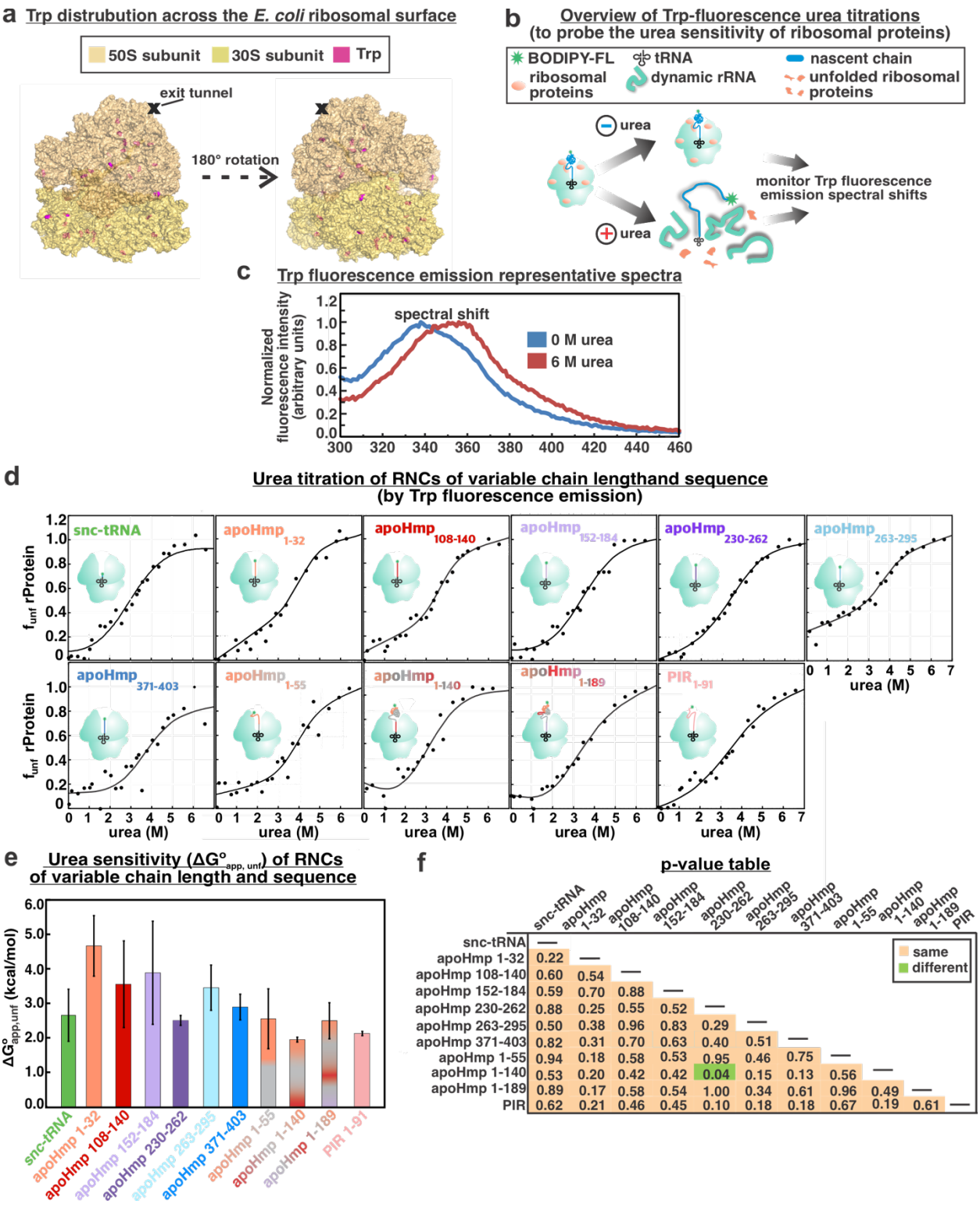


Figure 3-S5. a) Puromycin mimics the 3' end of the amino acyl tRNA and covalently attaches to

nascent chains, promoting their release from peptidyl tRNA. This process is accompanied by a significant decrease in molecular weight. **b)** Low-pH SDS-PAGE is employed to quantify peptidyl-tRNA (PT) band intensities as a function of urea concentration. Puromycin is unable to perform its function at high urea concentration because of pervasive PTC unfolding (SI Appendix, Figure 3-S1). **c)** Representative low-pH gels showing that peptidyl tRNA bands are unaffected by high urea concentrations. Addition of puromycin (1 mM) results in a decrease in peptidyl-tRNA (PT) band intensities, at low urea concentration. **d)** Block diagram mapping apparent unfolding free energy values ($\Delta G°_{app,unfold}$). Error bars denote standard errors based on 2-7 experiments. **e)** Representative urea titration curves corresponding to the raw data in panel c. Data were fit to a two-state unfolding expression (see Methods section for details). **f)** P-value table for a two-tailed Student's test assuming unequal variances (Welch's t-test). Green and orange boxes denote statistically different and statistically equivalent data, respectively, according to a $\geq$ 95% confidence interval.



	snc-tRNA	apoHmp 1-32	apoHmp 108-140	apoHmp 152-184	apoHmp 230-262	apoHmp 263-295	apoHmp 371-403	apoHmp 1-55	apoHmp 1-140	apoHmp 1-189	PIR
snc-tRNA	—										
apoHmp 1-32	0.22	—									
apoHmp 108-140	0.60	0.54	—								
apoHmp 152-184	0.59	0.70	0.88	—							
apoHmp 230-262	0.88	0.25	0.55	0.52	—						
apoHmp 263-295	0.50	0.38	0.96	0.83	0.29	—					
apoHmp 371-403	0.82	0.31	0.70	0.63	0.40	0.51	—				
apoHmp 1-55	0.94	0.18	0.58	0.53	0.95	0.46	0.75	—			
apoHmp 1-140	0.53	0.20	0.42	0.42	0.04	0.15	0.13	0.56	—		
apoHmp 1-189	0.89	0.17	0.58	0.54	1.00	0.34	0.61	0.96	0.49	—	
PIR	0.62	0.21	0.46	0.45	0.10	0.18	0.18	0.67	0.19	0.61	—

Figure 3-S6. a) The r-proteins of the *E. coli* 70S ribosome include 32 tryptophans (Trp, magenta). 16 and 16 Trps are located within the 50S and 30S subunits, respectively (PDB IDs:

2WWL and 2WWQ). **b)** Scheme illustrating the methodology followed to assess r-protein stability. **c)** Trp fluorescence-emission bands become red-shifted at increasing urea concentrations due to changes towards a more polar medium. **d)** Representative urea titration curves from steady-state fluorescence of RNCs of snc-tRNA, apoHmp1-32, apoHmp108-140, apoHmp152-184, apoHmp230-262, apoHmp263-295, apoHmp371-403, apoHmp1-55, apoHmp1-140, apoHmp1-189 and PIR respectively. The symbol f_{unf} denotes change in the fraction of unfolded r-proteins (eq. 4). **e)** Apparent unfolding free energies ($\Delta G°_{app,unfold}$). Uncertainties are reported as ± SE for n=2-3. **f)** P-value table for the two-tailed Student's T-test (Welch's test) comparing $\Delta G°_{app,unfold}$ values. Green and orange boxes denote statistically different and statistically equivalent data, respectively, according to a 95% confidence interval.

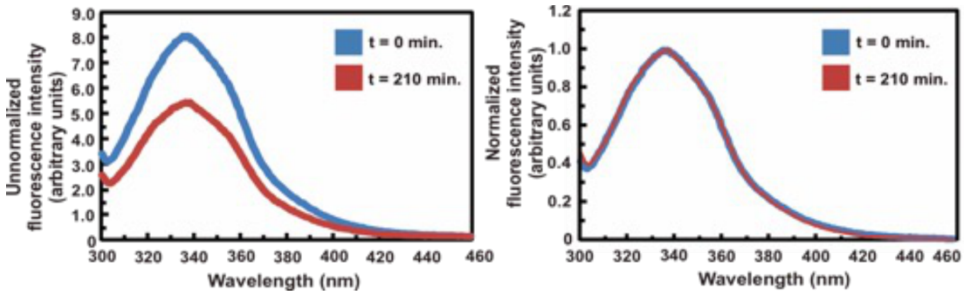


Figure 3-S7. Time-course experiments for apoHmp$_{1\text{-}55}$ are shown at 0.45 M urea. There is no change in spectral shift as incubation time is dramatically increased.

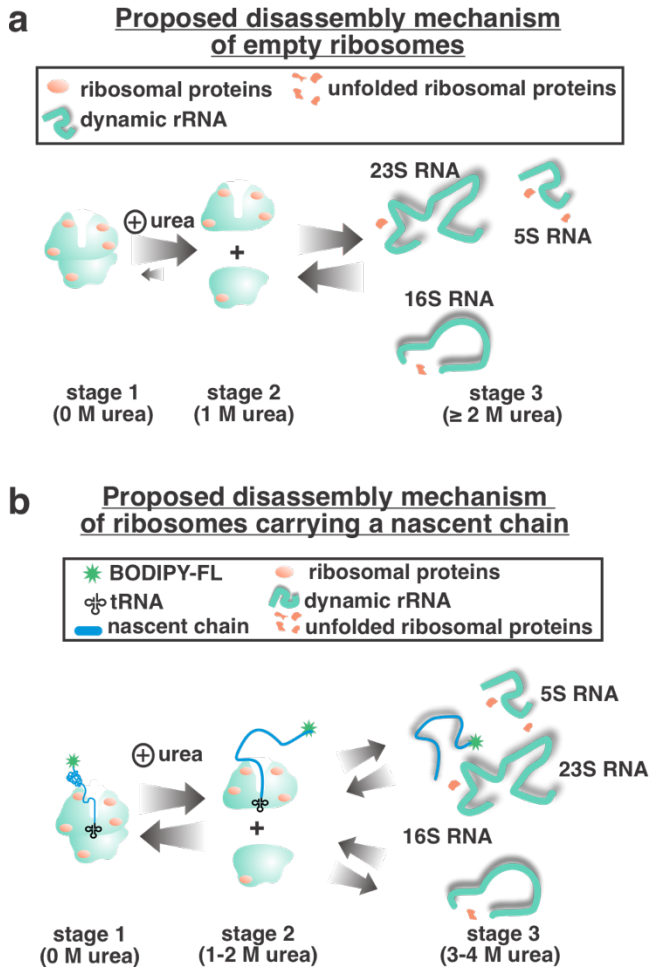


Figure 3-S8. A proposed disassembly model for empty ribosomes and ribosome-bound nascent chains. (a) Briefly, empty ribosomes begin to dissociate from one another at c.a. 1 M urea. At > 2 M urea empty ribosomes experience r-protein and rRNA unfolding. (b) Briefly, RNCs begin to dissociate from one another at c.a. 1-2 M urea. At > 3 M urea empty ribosomes experience r-protein and rRNA unfolding.

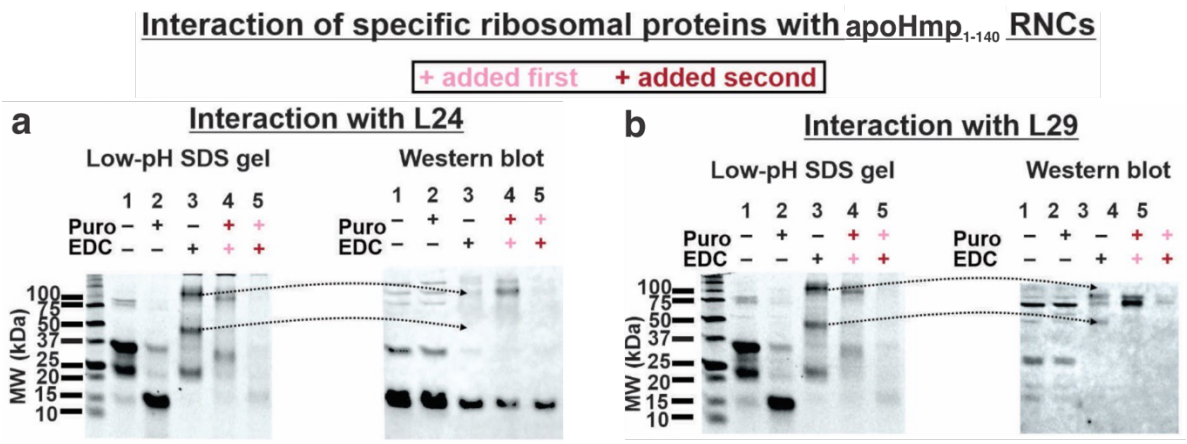


Figure 3-S9. Identification of ribosomal proteins crosslinked to apoHmp$_{1\text{-}140}$ RNCs via Western blotting. Left, low-pH 10% SDS-PAGE analysis of N-terminal fluorescently-labeled apoHmp$_{1\text{-}189}$ RNCs generated via transcription-translation in an *E. coli Dtig* S30 cell-free system followed by Western blotting (right) employing antibodies against ribosomal proteins (a) L24 and (b) L29. See Methods for Western blotting and antibodies details, representative data, out of n = 2, are displayed.

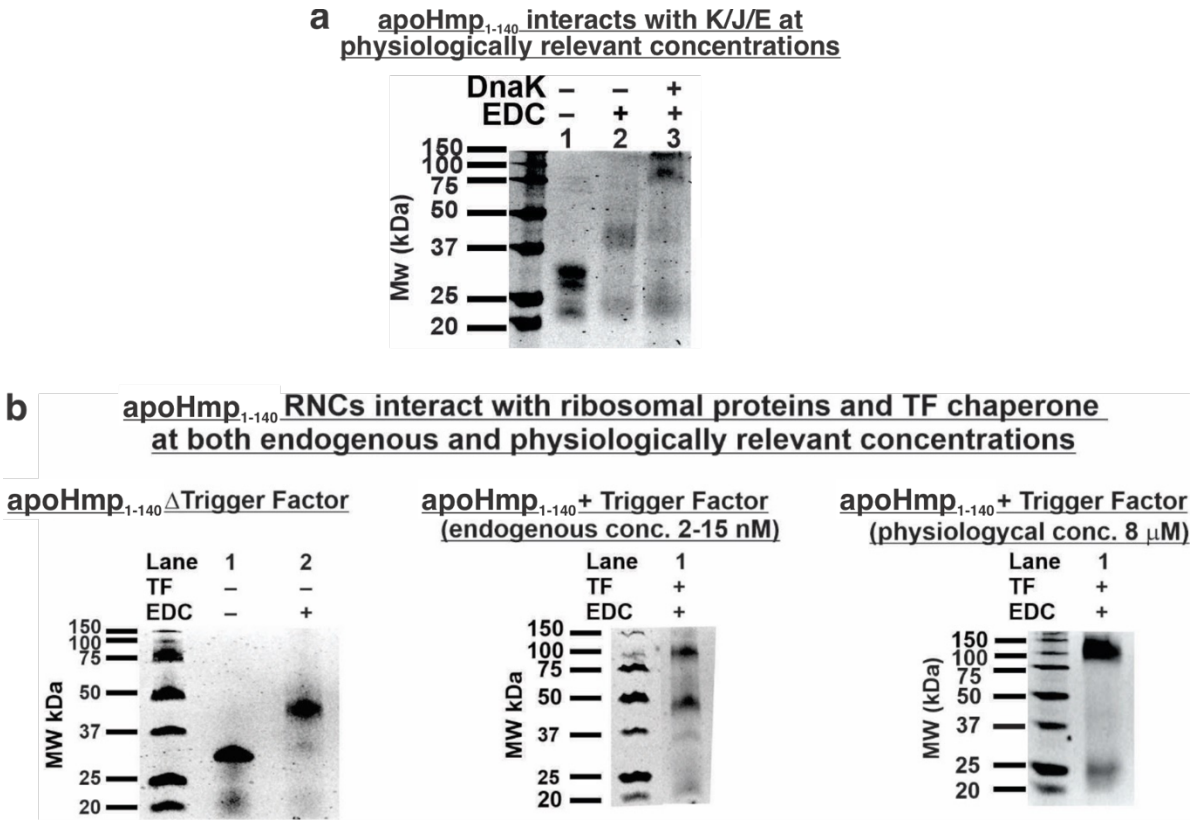


Figure 3-S10. (a) Low pH 10% SDS-PAGE gel showing apoHmp$_{1\text{-}140}$ with no Hsp70 chaperones added (K/J/E) and physiologically relevant concentrations, respectively. (b) ApoHmp$_{1\text{-}140}$ at various concentrations of trigger factor molecular chaperone. Concentration of trigger factor increases from left gel to right gel. See methods for more information about chaperone concentration.

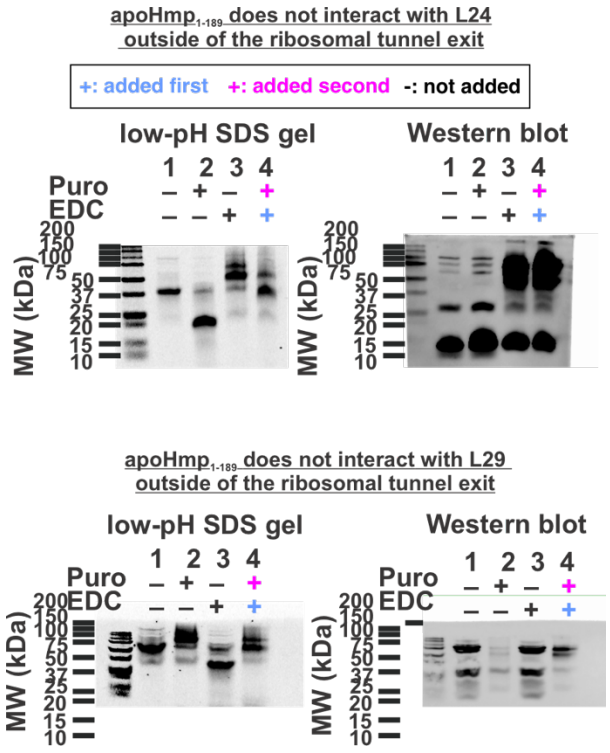


Figure 3-S11. Identification of ribosomal proteins crosslinked to apoHmp$_{1\text{-}189}$ RNCs via Western blotting. Left, low-pH 10% SDS-PAGE analysis of N-terminal fluorescently-labeled apoHmp$_{1\text{-}189}$ RNCs generated via transcription-translation in an *E. coli Dtig* S30 cell-free system followed by Western blotting (right) employing antibodies against ribosomal proteins L24 and L29. See Methods for Western blotting and antibodies details, representative data, out of n = 2, are displayed.

Supporting Tables

Table 3-S1. Summary of published experimental evidence on interactions between nascent protein chains (RNCs) and the ribosome based on single-particle cryo-EM or chemical crosslinking data.

RNC	Technique	Interacting ribosomal protein	Reference
Proteins with an N-terminal signal sequence			
Leader peptidase	photo-crosslinking	L4, L22 and L23	[37]
Two regulatory ribosome stalling peptides	cryo-EM	L4 and L17	[35]
Transmembrane segment of 111p membrane protein	photo-crosslinking	not determined	[39]
Signal anchor of FtsQprotein	photo-crosslinking	L23 and L29	[40]
pOmt	photo-crosslinking	L23 and L24	[41]
EsP1-25	photo-crosslinking	L23 and L24	[42]
Proteins with a C-terminal ribosome-stalling sequence			

TnaC	photo-crosslinking	L22 and L24	[43]
TnaC	cryo-EM	L22	[44]
SecM	cryo-EM	L22 and L23	[45]
SecM	mutagenesis	L22	[46]
SecM	cryo-EM	L23	[33]
Signal anchor of Dap2 Protein	photo-crosslinking	RpI4, Rp17 and RpI29	[47]
SecM	NMR	not determined	[48]

Proteins with no N- or C-terminal tag

Phosphorylated insulin receptor (PIR)	chemical crosslinking	L23	[31]

Table 3-S2. Name, nucleotide sequence, and calculated ΔG° of unfolding for each oligo nucleotide used in this work. ΔG° was calculated using an online calculator (biosyn.com). Sequences used for calculation are as shown.

Construct	Oligo-nucleotide sequence	ΔG°unf
Anti-ssrA	TTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCGACTA	**54.2**
snc-tRNA	GCGTCAAGCAT	**12.4**
ApoHmp$_{1-32}$	CATACGGTCGTAGAAATGGGCGGTTAACTTTGGCCCCG	**57.9**
ApoHmp$_{108-140}$	TTAGTTATAGATTTCCGCCTCGCGATTGATAAATACAT	**48.9**
ApoHmp$_{152-184}$	TATTCTGCCACTGCGCCACCGTCGACCGGC	**49.2**
ApoHmp$_{230-262}$	CGGAGCGACCAGTTTCACGACATCGCCACCA	**49.0**
ApoHmp$_{263-295}$	ATGGAACCAGTTCACTTGTGCTGTGTGGCCTGCTTTTG	**54.9**
ApoHmp$_{371-403}$	CAGCACCTTATGCGGGCCAAAGCATTCGTAATG	**47.9**
ApoHmp$_{1-55}$	AAGTGCAACTAAGCGGTAATGCGGACCAATGAG	**49.3**
ApoHmp$_{1-140}$	TTAGTTATAGATTTCCGCCTCGCGATTGATAAATACAT	**48.9**
ApoHmp$_{1-189}$	GAGATATTGCCCCGGACGGTATTCTGCCACTGCGCCAC	**59.2**
PIR$_{1-91}$	GTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCT	**63.9**

Supporting References

(1) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(2) Behrmann, M.; Koch, H.-G.; Hengelage, T.; Wieseler, B.; Hoffschulte, H. K.; Müller, M. Requirements for the translocation of elongation-arrested, ribosome-associated OmpA across the plasma membrane of Escherichia coli. *J. Biol. Chem.* **1998,** *273*, 13898-13904.

(3) Donis-Keller, H. Site specific enzymatic cleavage of rna. *Nucleic Acids Res.* **1979,** *7*, 179-192.

(4) 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*, 274-282.

(5) Günzl, A.; Palfi, Z.; Bindereif, A. Analysis of rna–protein complexes by oligonucleotide-targeted RNase H digestion. *Methods* **2002,** *26*, 162-169.

(6) Dos Santos, R. F.; Arraiano, C. M.; Andrade, J. M. Isolation and analysis of bacterial ribosomes through sucrose gradient ultracentrifugation. In *Rna chaperones*, Springer: 2020; pp 299-310.

(7) Robinson, A.; Sykes, J. A comparison of the unfolding and dissociation of the large ribosome subunits from rhodopseudomonas· spheroides ncib 8253 and Escherichia coli mre 600. *Biochem. J.* **1973,** *133*, 739-747.

(8) Hosokawa, K. Binding of 5 s ribosomal ribonucleic acid to the unfolded 50 S ribosomes of Escherichia coli II. *J. Biol. Chem.* **1970,** *245*, 5880-5887.

(9) Tal, M. Metal ions and ribosomal conformation. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1969,** *195*, 76-86.

(10) Natori, S.; Maruta, H.; Mizuno, D. i. Unfolding of Escherichia coli ribosomes by phosphate ion in the presence of oligonucleotides. *J. Mol. Biol.* **1968,** *38*, 109-119.

(11) Weller, D. L.; Shechter, Y.; Musgrave, D.; Rougvie, M.; Horowitz, J. Conformational changes in Escherichia coli ribosomes at low magnesium ion concentrations. *Biochemistry* **1968,** *7*, 3668-3675.

(12) Beller, R. J.; Lubsen, N. H. Effect of polypeptide chain length on dissociation of ribosomal complexes. *Biochemistry* **1972,** *11*, 3271-3276.

(13) Belitsina, N.; Spirin, A. Studies on the structure of ribosomes: Iv. Participation of aminoacyl-transfer rna and peptidyl-transfer rna in the association of ribosomal subparticles. *J. Mol. Biol.* **1970,** *52*, 45-55.

(14) Ron, E. Z.; Kohler, R. E.; Davis, B. D. Magnesium ion dependence of free and polysomal ribosomes from Escherichia coli. *J. Mol. Biol.* **1968,** *36*, 83-89.

(15) Roberts, M. E.; Walker, I. Structural studies on Escherichia coli ribosomes: III. Denaturation and sedimentation of ribosomal subunits unfolded in urea. *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.* **1970,** *199*, 184-193.

(16) 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.

(17) Spitnik-Elson, P.; Greenman, B. The detachment of ribosomal proteins by urea: Evidence for non-electrostatic rna-protein interaction in the ribosome. *FEBS Lett.* **1971,** *17*, 187-192.

(18) Hung, H.-C.; Chang, G.-G. Multiple unfolding intermediates of human placental alkaline phosphatase in equilibrium urea denaturation. *Biophys. J.* **2001,** *81*, 3456-3471.

(19) Shine, J.; Dalgarno, L. The 3′-terminal sequence of Escherichia coli 16s ribosomal rna: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* **1974,** *71*, 1342-1346.

(20) Mohammad, F.; Green, R.; Buskirk, A. R. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife* **2019,** *8*, e42591.

(21) Yusupova, G.; Jenner, L.; Rees, B.; Moras, D.; Yusupov, M. Structural basis for messenger rna movement on the ribosome. *Nature* **2006,** *444*, 391-394.

(22) Brandt, F.; Etchells, S. A.; Ortiz, J. O.; Elcock, A. H.; Hartl, F. U.; Baumeister, W. The native 3d organization of bacterial polysomes. *Cell* **2009,** *136*, 261-271.

(23) Oliver, R. M. Negative stain electron microscopy of protein macromolecules. In *Methods enzymol.*, Elsevier: 1973; Vol. 27, pp 617-672.

(24) Yarmolinsky, M. B.; Gabriel, L. Inhibition by puromycin of amino acid incorporation into protein. *Proc. Natl. Acad. Sci. U.S.A.* **1959,** *45*, 1721-1729.

(25) Wohlgemuth, I.; Beringer, M.; Rodnina, M. V. Rapid peptide bond formation on isolated 50S ribosomal subunits. *EMBO Rep.* **2006,** *7*, 699-703.

(26) Kirchdoerfer, R. N.; Huang, J. J. T.; Isola, M. K.; Cavagnero, S. Fluorescence-based analysis of aminoacyl- and peptidyl-tRNA by low-pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* **2007,** *364*, 92-94.

(27) Santoro, M. M.; Bolen, D. W. Unfolding free-energy changes determined by the linear extrapolation method.1.Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **1988,** *27*, 8063-8068.

(28) Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **1990,** *8*, 93-98.

(29) Bonincontro, A.; Cinelli, S.; Mengoni, M.; Onori, G.; Risuleo, G.; Santucci, A. Differential stability of E. coli ribosomal particles and free rna towards thermal degradation studied by microcalorimetry. *Biophys. Chem.* **1998,** *75*, 97-103.

(30) Warren, J. R.; Gordon, J. A. On the refractive indices of aqueous solutions of urea. *J. Phys. Chem.* **1966,** *70*, 297-300.

(31) Guzman-Luna, V.; Fuchs, A. M.; Allen, A. J.; Staikos, A.; Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Commun. Biol.* **2021,** *4*, 1-17.

(32) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9*, 671-675.

(33) Wang, S.; Jomaa, A.; Jaskolowski, M.; Yang, C.-I.; Ban, N.; Shan, S.-o. The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat. Struct. Mol. Biol.* **2019,** *26*, 919-929.

(34) Lindahl, L.; Post, L.; Zengel, J.; Gilbert, S.; Strycharz, W.; Nomura, M. Organization of ribosomal-protein genes in Escherichia-coli. 7. Mapping of ribosomal-protein genes by in vitro protein-synthesis using DNA fragments of lambda-fus3 transducing phage DNA as templates. *J. Biol. Chem.* **1977,** *252*, 7365-7383.

(35) Beechem, J. M.; Brand, L. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.* **1985,** *54*, 43-71.

(36) Demchenko, A. P. *Ultraviolet spectroscopy of proteins*. Springer Science & Business Media: 2013.

(37) Weber, G. Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds. *Biochem. J.* **1960,** *75*, 335.

(38) Teale, F.; Weber, G. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.* **1957,** *65*, 476.

(39) Woolhead, C. A.; McCormick, P. J.; Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the dribosome and in their exposure to ribosomal proteins. *Cell* **2004,** *116*, 725-736.

(40) Ullers, R. S.; Houben, E. N.; Raine, A.; ten Hagen-Jongman, C. M.; Ehrenberg, M. n.; Brunner, J.; Oudega, B.; Harms, N.; Luirink, J. Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the Escherichia coli ribosome. *J. Cell Biol.* **2003,** *161*, 679-684.

(41) Eisner, G.; Moser, M.; Schäfer, U.; Beck, K.; Müller, M. Alternate recruitment of signal recognition particle and trigger factor to the signal sequence of a growing nascent polypeptide. *J. Biol. Chem.* **2006,** *281*, 7172-7179.

(42) Peterson, J. H.; Woolhead, C. A.; Bernstein, H. D. The conformation of a nascent polypeptide inside the ribosome tunnel affects protein targeting and protein folding. *Mol. Microbiol.* **2010,** *78*, 203-217.

(43) Cruz-Vera, L. R.; Rajagopal, S.; Squires, C.; Yanofsky, C. Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression. *Mol. Cell* **2005,** *19*, 333-343.

(44) Seidelt, B.; Innis, C. A.; Wilson, D. N.; Gartmann, M.; Armache, J.-P.; Villa, E.; Trabuco, L. G.; Becker, T.; Mielke, T.; Schulten, K., et al. Structural insight into nascent polypeptide chain–mediated translational stalling. *Science* **2009,** *326*, 1412-1415.

(45) Bhushan, S.; Meyer, H.; Starosta, A. L.; Becker, T.; Mielke, T.; Berninghausen, O.; Sattler, M.; Wilson, D. N.; Beckmann, R. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol Cell* **2010,** *40*, 138-46.

(46) Nakatogawa, H.; Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **2002,** *108*, 629-36.

(47) Zhang, Y.; Wolfle, T.; Rospert, S. Interaction of nascent chains with the ribosomal tunnel proteins rpl4, rpl17, and rpl39 of saccharomyces cerevisiae. *J. Biol. Chem.* **2013,** *288*, 33697-707.

(48) Burridge, C.; Waudby, C. A.; Wlodarski, T.; Cassaignau, A. M. E.; Cabrita, L. D.; Christodoulou, J. Nascent chain dynamics and ribosome interactions within folded ribosome-nascent chain complexes observed by NMR spectroscopy. *Chem. Sci.* **2021,** *12*, 13120-13126.

Chapter 4

Critical Beginnings: Selective Tuning of Solubility and Structural Accuracy of Newly-Synthesized Proteins by the Hsp70 Chaperone System

This chapter reproduces an article submitted to *The Journal of Physical Chemistry B* by Rayna M. Addabbo, Rachel B. Hutchinson, Heather J. Allaman, Matthew D. Dalphin, Miranda F. Mecha, Yue Liu, and Silvia Cavagnero. R.M.A. and S.C. designed the project. All coauthors except for S.C. collected and interpreted the data. R.M.A. collected fluorescence lifetime and anisotropy decay data. R.B.H. performed fluorescence lifetime and anisotropy decay measurements, viscosity measurements, and the experiments aimed at further characterization of soluble aggregates. M.D.D. and Y.L. collected the circular dichroism and urea titration data. Y.L. and H.J.A. performed the solubility assay experiments. H.J.A. performed the incubation solubility experiments. M.F.M. collected and analyzed the NMR data. R.M.A., R.B.H., and S.C. participated in most of the data organization, discussions, and manuscript writing.

4.1 Abstract

Proteins are particularly prone to aggregation immediately after release from the ribosome, and it is therefore important to elucidate the role of chaperones during these key steps of protein life. The Hsp70 and trigger factor (TF) chaperone systems interact with nascent proteins during biogenesis and immediately post-translationally. It is unclear, however, whether these chaperones can prevent formation of soluble and insoluble aggregates. Here, we address this question by monitoring the solubility and structural accuracy of globin proteins biosynthesized in an *E. coli* cell-free system containing different concentrations of the bacterial Hsp70 and TF chaperones. We find that Hsp70 concentrations required to grant solubility to newly synthesized proteins are extremely sensitive to client-protein sequence. Importantly, Hsp70 concentrations yielding soluble client proteins are insufficient to prevent formation of soluble aggregates. In fact, for some aggregation-prone protein variants, avoidance of soluble-aggregate formation demands Hsp70 concentrations that exceed cellular levels in *E. coli*. In all, our data highlight the prominent role of soluble aggregates upon nascent-protein release from the ribosome, and show the limitations of the Hsp70 chaperone system in the case of highly aggregation-prone proteins. These results demonstrate the need to devise better strategies to prevent soluble-aggregate formation upon release from the ribosome.

4.2 Introduction

After biosynthesis, most *E. coli* proteins and the model protein sperm whale apomyoglobin are kinetically trapped relative to a variety of aggregates (e.g., non-amyloid and amyloid[1, 2]), in cell-like media under physiologically relevant conditions.[2, 3] However, transient non-native conformations generated immediately after nascent-protein release from the ribosome are not kinetically trapped relative to aggregates, presumably due to their high free energy and lower barriers to folding/aggregation. Indeed, these conformations experience a particularly high propensity to aggregate.[3, 4] Recent studies showed that, upon release from the ribosome, kinetic channeling of nascent proteins toward either native or aggregate states is extremely sensitive to protein sequence. For instance, single-point mutations that perturb folding/unfolding/aggregation kinetic barriers upon release from the ribosome, but not native-state thermodynamic stability, cause proteins to be released from the ribosome as insoluble aggregates (Figure 4-1a).[3]

Molecular chaperones are known to facilitate protein folding, prevent aggregation, and perform a variety of additional critical cellular functions that enable cell survival and growth.[5-8] The Hsp70 system and trigger factor (TF) are the primary bacterial chaperones that act during the early co- and post-translational periods in protein life.[9, 10] *E. coli* cells can tolerate the deletion of either Hsp70 (known as DnaK in *E. coli*) or TF. On the other hand, simultaneous deletion of both chaperones is lethal to cells at temperatures higher than 20 °C.[9, 11, 12] While TF and Hsp70 have partially overlapping roles, Hsp70 enables bacterial cell viability across a wider temperature range than TF.[9, 12]

Hsp70 and TF not only are important for wild-type cell viability but also serve as “buffers” against mutations in bacteria. For instance, overexpression of TF increases the *in vivo* activity and solubility of deleterious mutants of the *E. coli* controller of cell-death protein B (CcdB).[13, 14]

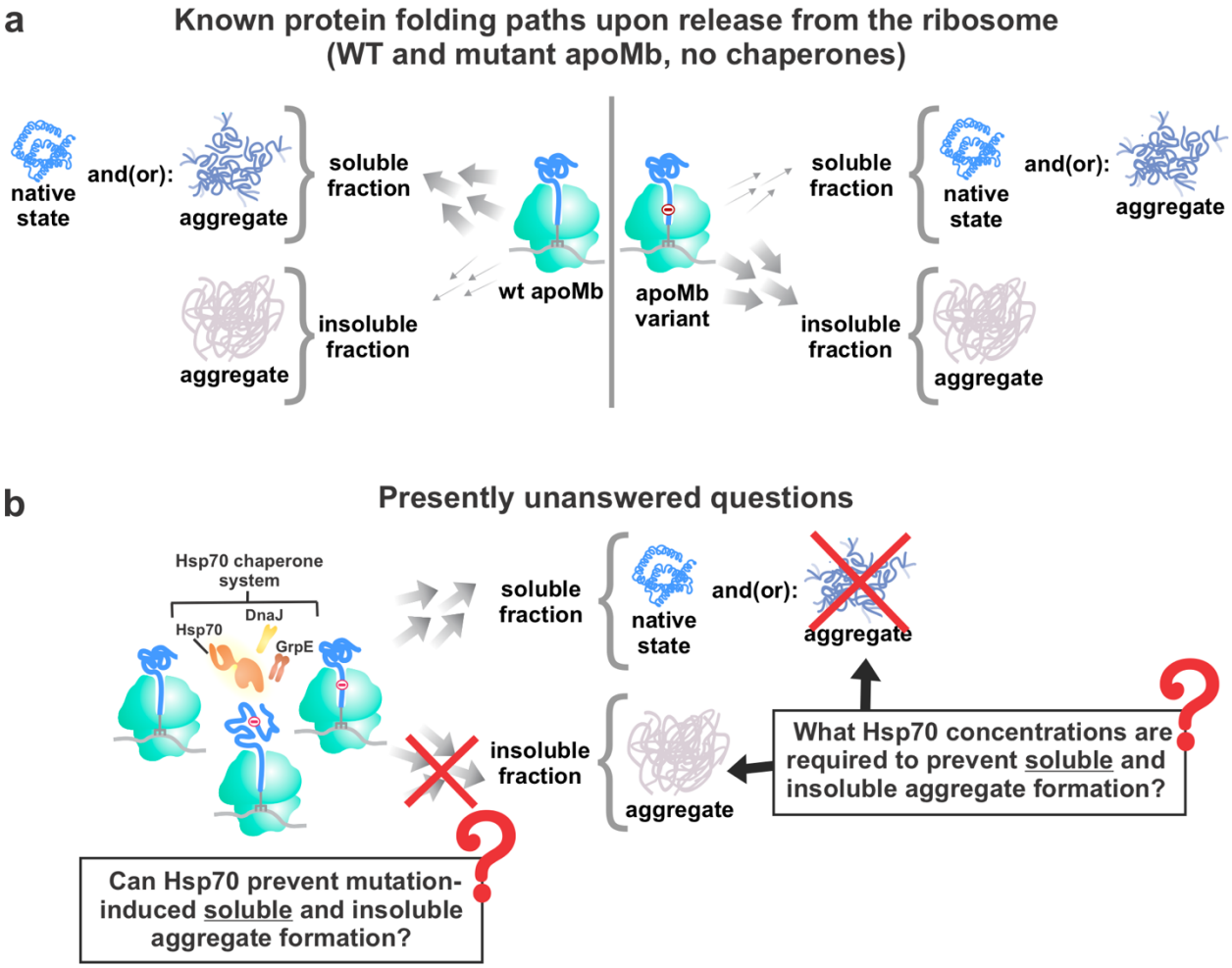


Figure 4-1. Unanswered questions this work addresses (a) Upon protein biosynthesis of WT apoMb at low concentrations (0.3 ± 0.2 μM), in the absence of molecular chaperones, apoMb forms a soluble species. A single point mutation (M131D) disrupts folding kinetics and causes apoMb to form an insoluble aggregate.[3] (b) Unanswered questions this work addresses.

In addition, in evolving *E. coli* bacteria, overexpression of Hsp70 prevents cell-line extinction from the accumulation of random mutations across several proteins.[15] Further, *E. coli* proteins that interact more frequently with Hsp70[16] evolve faster than clients that interact less frequently with this chaperone.[15, 17] The latter studies show that the Hsp70 and TF chaperones are capable of preventing harmful mutation-related phenotypes.

Despite the above knowledge, the types of mutations requiring chaperone assistance (e.g., surface *vs* core, stabilizing *vs* destabilizing) and the structural characteristics of ribosome-released client-protein variants (e.g., native states, soluble or insoluble aggregates) generated in the absence and presence of chaperones are not known.

Interestingly, one of the ways the Hsp70 and TF chaperones operate is by granting solubility to a variety of proteins, upon release from the ribosome. For instance, in the absence of any chaperone, 72% of *E. coli* proteins (excluding membrane proteins) are insoluble (<80% solubility) when synthesized on the ribosome via a cell-free transcription-translation system.[4] These studies were performed at 37 °C, and *E. coli* proteins were expressed at an average concentration of 33 μg/ml.[4] TF grants solubility (i.e., > 50% increase in solubility) to only 2% of these aggregation-prone proteins.[18] On the other hand, Hsp70 renders 52% of these proteins soluble.[18] Further, in addition to increasing the solubility of *E. coli* homologous proteins,[18] Hsp70 and TF also perform a similar action on heterologous[19] proteins. Importantly, it is imperative to note that a protein's soluble status does not necessarily imply native conformation. For instance, soluble proteins, including members of the globin fold, sometimes populate soluble aggregates under physiologically relevant conditions.[3, 20] Soluble aggregates are not bioactive, and reduce the amount of functional gene product produced upon recombinant protein expression.[21] Further,

in higher organisms including mammals, it is becoming clear that soluble aggregates contribute to a variety of deadly maladies.[22-24]

Despite the above progress, there is a crucial gap of knowledge about whether Hsp70 or TF can prevent the formation of soluble aggregates. In addition, it is not known what specific chaperone concentrations are required to prevent soluble and insoluble aggregate formation upon release form the ribosome (Figure 4-1b).

In this study, we focus on the well-biochemically characterized model protein apomyoglobin.[25-32] We explore the nature of its native state and its soluble/insoluble aggregates immediately upon release from the ribosome in the absence and presence of the Hsp70 and trigger factor (TF) chaperones.

Specifically, we employ apoMb as a benchmark to address two key unanswered questions, as schematically illustrated in Figure 4-1. Namely, (1) Can the Hsp70 chaperone system prevent the harm caused by single-point mutations that disrupt the apoMb nonpolar-core? and (2) what Hsp70 concentrations are needed to prevent soluble and insoluble aggregate formation of WT and mutant apoMb upon release from the ribosome? We employ a combination of site-specific fluorophore labeling, solubility assays, frequency-domain fluorescence lifetime and anisotropy decays, and multidimensional NMR. These synergistic approaches offer an excellent opportunity to identify both soluble and insoluble gene products and, importantly, to characterize the nature of the soluble fraction of newly synthesized proteins. We focus on the role of Hsp70 during the last stages of *de novo* protein biosynthesis. We evaluate two apoMb variants bearing single-point mutations designed to disrupt the kinetics of co- and immediately post-translational folding. We show that physiologically relevant concentrations of Hsp70 prevent insoluble aggregate formation induced by single point mutations. However, under non-stress conditions, higher than

physiologically attainable Hsp70 concentrations[33] are required to prevent generation of soluble aggregates. The wild-type protein also forms soluble aggregates with different characteristics, upon release from the ribosome. Therefore, different proteins have distinct Hsp70 chaperone requirements to prevent formation of insoluble and soluble aggregates, and the *E. coli* Hsp70 chaperone system is incapable of reversing the deleterious effect of some single-point mutations responsible for incorrect kinetic channeling upon release from the ribosome. In all, this work demonstrates the crucial importance of controlling kinetic channeling upon release from the ribosome, and it shows that the Hsp70 chaperone system is capable of modulating this crucial step of protein life, but only to a limited extent.

4.3 Results and Discussion

Design Criteria. The convenient properties of sperm whale apomyoglobin (apoMb), the model protein selected for these studies, are briefly detailed below. First, apoMb belongs to the ubiquitous globin fold,[34-36] which is also highly represented in bacteria,[36] and its three-dimensional structure has been thoroughly characterized.[37] Second, the size and hydrophobicity/charge pattern of apoMb[38] are typical of those encountered across the *E. coli* proteome.[39] It is worth noting that the codon usage of our wild-type and mutant apoMb-encoding plasmids have been optimized for *E. coli*.[40] Third, apoMb is a well biochemically characterized protein[26, 31, 41, 42] and has been an important model system for detailed experimental investigations on protein folding paths upon refolding into buffered solution[27, 28, 32, 43-45] and, importantly, within the cellular environment.[3, 25, 38, 46-51] Fourth, apoMb is a quintessential example of the many proteins that are unable to reach their native state at 100% level upon refolding from

denaturant.[52] Indeed, purified chemically denatured wild-type apoMb forms some soluble aggregates upon refolding into buffer.[53] Upon biosynthesis in an *E. coli* transcription-translation cell-free system (at ca. 0.3 μM expression level) lacking chaperones, wild-type apoMb also gives rise to large slowly tumbling species that were recently attributed to soluble aggregates.[3] This adduct may include homo- or hetero-aggregates, associated via either specific or non-specific interactions.[3] On the other hand, the nature of the apoMb soluble fraction and the role of the Hsp70 chaperone upon transcription-translation are still very poorly understood. Further, at very high expression levels upon recombinant overexpression, apoMb gives rise to insoluble aggregates, a.k.a. inclusion bodies.[47, 54] Fifth, apoMb has several Hsp70 binding sites,[55] and it interacts with this chaperone transiently upon refolding from denaturant, but not in its fully folded state, in buffered solution.[56] The effect of Hsp70 in a cell-like environment is not clear, especially in regards to the prevention of soluble-aggregate formation. In summary, apoMb and its variants are excellent model proteins to explore the effect of the Hsp70 chaperone system on protein solubility and structural accuracy upon nascent-protein release from the ribosome.

Newly synthesized proteins are generated via an in-house-produced *E. coli* cell-free system[57] and are *in situ* site-specifically labeled with the BODIPY- FL fluorophore at the N terminus.[48] This selective labeling procedure enables specific detection of the desired newly synthesized proteins by fluoroimaging and fluorescence lifetimes and anisotropies (see below). The high sensitivity of fluorescence enables detection of a wide range of chaperone-to-client-protein ratios (>100-fold).

As schematically illustrated in Figure 4-2, we discriminate soluble from insoluble fractions of fluorophore-labeled newly synthesized apoMb by SDS-PAGE. Then we proceed with further analysis of the composition and dynamic characteristics of the soluble fraction. Namely, we

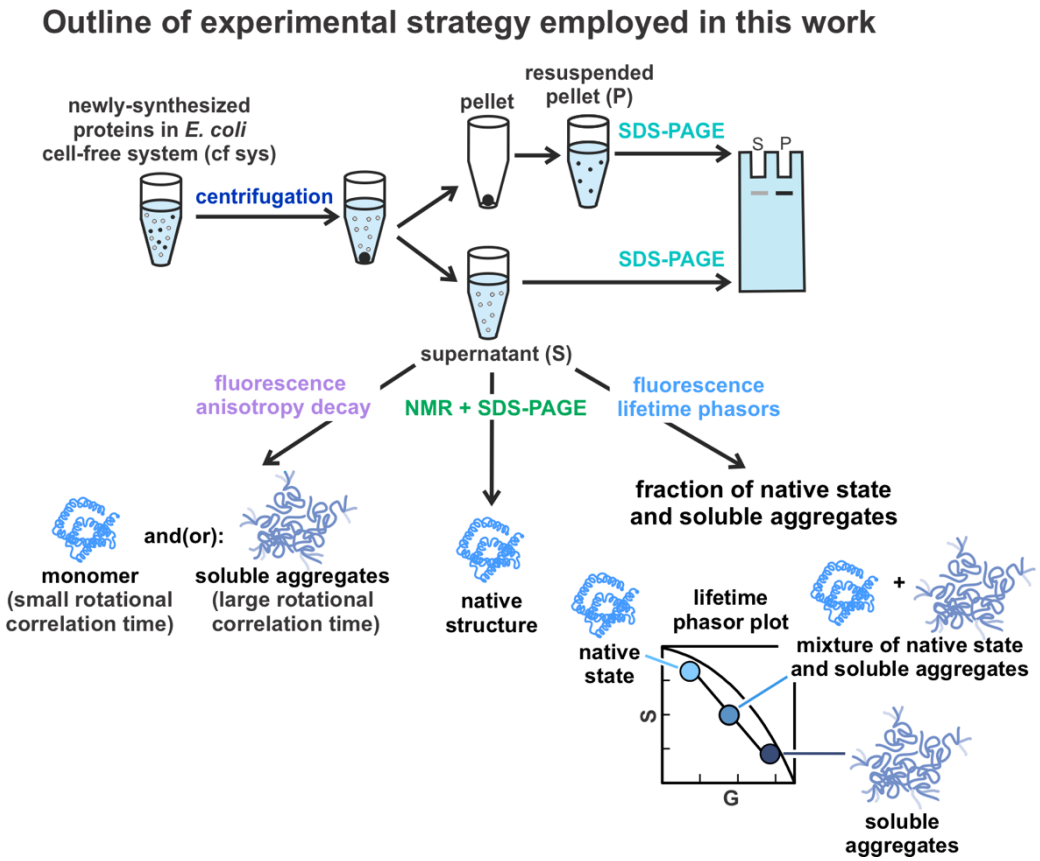


Figure 4-2. Overview of techniques employed in this work. Solubility assays were used to separate insoluble and soluble fractions. NMR spectroscopy in combination with SDS-PAGE was employed to identify the percent of newly synthesized protein in its native state. Fluorescence anisotropy decays revealed the size of local and global tumbling units and the amplitude of local motions. Fluorescence-emission frequency shifts showed changes in environment polarity. Fluorescence lifetime identified variations in local environment. Fluorescence lifetime phasors served to quantify fractions of monomeric protein and(or) soluble aggregates. All measurements targeted the characterization of newly synthesized proteins within the cell-free-system environment.

discriminate native states from very slow-tumbling species and quantify their relative amounts via a combination of 2D NMR and SDS-PAGE in the presence of apoMb calibrants of known concentration. To assess environmental differences, particle size and amplitude of local motions, we employ frequency-domain fluorescence spectroscopy and measure fluorescence lifetimes and anisotropy decays.

Design and Biophysical Characterization of I28D and M131D ApoMb Variants. We designed, produced and analyzed single-point apoMb variants with the goal of selectively slowing down the kinetics of native structure formation. Towards this end, we replaced single nonpolar residues residing in the central region of the native protein core[37] with aspartic acid, given that this type of mutation is known to slow down the acquisition of the native state.[58] We designed the I28D and M131D variants of apoMb, with the goal of selectively probing co- and post-translational events. Namely, when native apoMb has been fully synthesized but before it is released from the ribosome, the I28D mutation site resides outside of the ribosomal exit tunnel, while the M131D mutation resides inside it, as illustrated in Figure 4-3a. These geometrical arguments are supported by our knowledge of the ribosome structure,[59-61] including the conformational features of the ribosomal exit-tunnel[62] and the known number of nascent-chain residues buried inside the tunnel (ca. 30-35) within known proteomes.[59, 63] More specifically, given the size of the known compact N-terminal domain of apoMb (ca 60-80 N-terminal residues[38, 48]), the negatively charged aspartic acid of the I28D variant is expected to perturb the compact domain co-translationally, given that the compact domain resides outside the ribosomal-tunnel core.[49, 64] Of course this variant may also influence folding/aggregation processes post-translationally. Conversely, the M131D mutation is located only 23 residues away from the

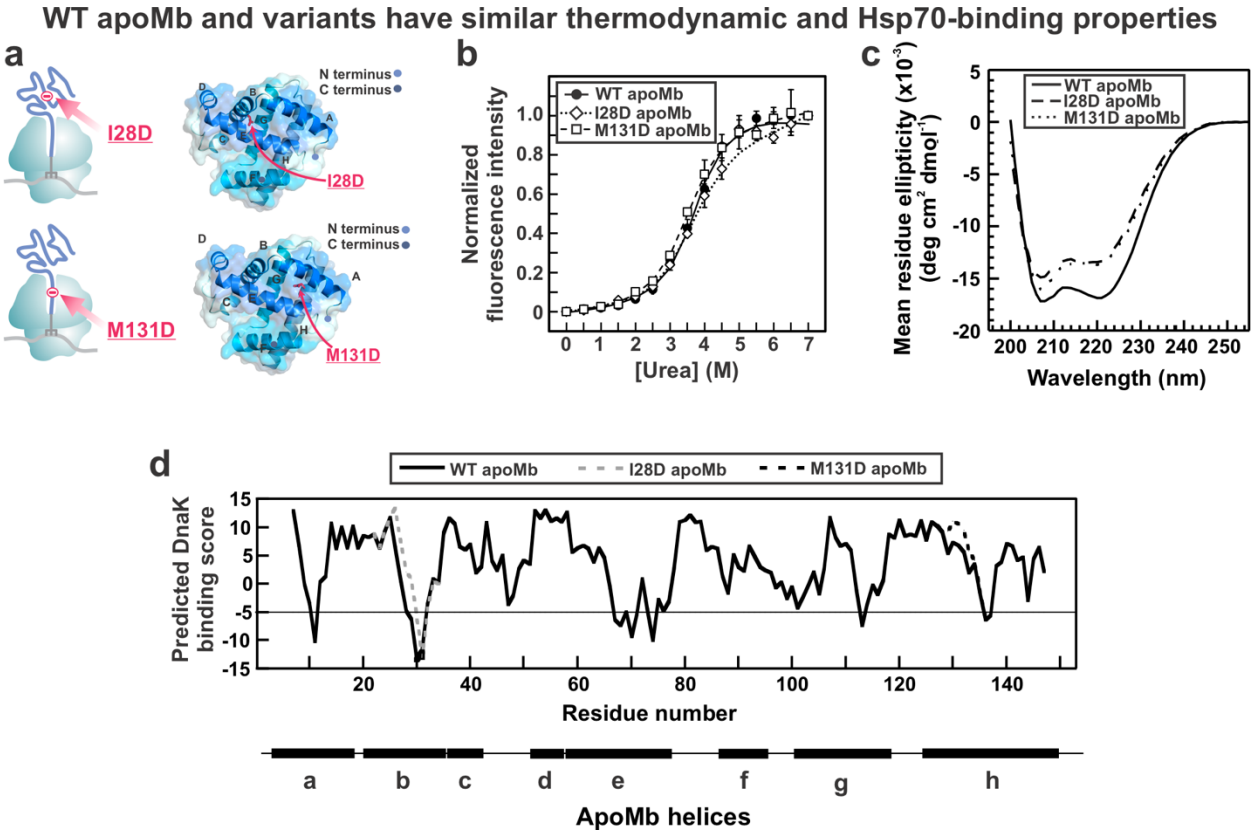


Figure 4-3. Biophysical characterization of I28D and M131D apoMb variants. (a) Location of the I28D and M131D point mutations when the protein has been fully synthesized on the ribosome (left). I28D and M131D mutation sites within the folded myoglobin crystal structure (right, PDB code 1MBC[65]). (b) Representative tryptophan fluorescence urea titration curves for n=3 experiments for WT apoMb (black circles, continuous line), I28D apoMb (open squares, dashed line), and M131D apoMb (open diamonds, dashed line). (c) Far-UV circular dichroism spectra of purified WT apoMb (solid line), I28D apoMb (dashed line) and M131D apoMb (dotted line) (n=3 experiments). Data for WT and M131D apoMb in b and c is reproduced from Addabbo *et al*[3]. (d) Probability scores for Hsp70 binding to WT apoMb (solid black), I28D apoMb (gray dashed line), and M131D apoMb (black dashed line) using the algorithm developed by Rüdiger *et al*,[66] where negative values denote greater binding probability.

C-terminus of the full-length nascent protein, i.e., close to the peptidyl transferase center (PTC) and within the core of the ribosomal exit tunnel. In summary, the I28D and M131D mutation sites are expected to primarily affect co- and post-translational folding and aggregation events, respectively.

Figure 4-3b-d shows the biophysical characterization of the I28D and M131D apoMb variants. Urea titrations performed on the purified proteins (Figure 4-3b) indicate that the unfolding free energies of WT, I28D and M131D apoMb are all the same within error (Table 4-1). Therefore, these mutations, while designed to slow down the folding kinetics, do not influence the overall protein stability. Far-UV circular dichroism (CD) spectra demonstrate that the I28D and M131D variants experience only a small loss in the α-helical content in their native states relative to WT apoMb (Figure 4-3c). Importantly, Figure 4-3d also shows that the I28D and M131D mutations are not predicted to interfere with any the Hsp70 binding sites present in the WT protein.[66]

In all, our data show that the designed variants do not cause significant changes in structure, thermodynamic stability, or Hsp70 binding site distribution and affinity. Therefore, our results, discussed in the next sessions, can be interpreted by ruling out the contributions of each of these parameters.

Hsp70 and TF Chaperones Prevent Insoluble-Aggregate Formation by Disruptive Point Mutations, upon Release from the Ribosome. WT apoMb and the two variants described in the previous sections were employed to probe the effect of the Hsp70 and TF molecular chaperones on nascent-protein release from the ribosome. In this part of our study, we focused on the newly synthesized protein solubility. WT apoMb and its variants were biosynthesized in

Table 4-1. Standard-state unfolding free energy (ΔG°) values determined from urea titrations.

Protein	ΔG°
WT apoMb	5 ± 1
I28D apoMb	4.6 ± 0.2
M131D apoMb	4 ± 1

ΔG° values are the average and standard error for three independent experiments.

an *E. coli* cell-free transcription-translation system lacking or including increasing amounts of the Hsp70 and TF molecular chaperones, up to physiologically relevant concentrations of ca 50 μM.[33] The data, displayed in Figure 4-4, show that WT apoMb attains greater than 90% solubility even in the absence of chaperones. Thus, WT apoMb does not need molecular chaperones for solubility. Indeed, as previously demonstrated,[3] the ribosome alone is fully responsible for the generation of a soluble WT apoMb gene product in the absence of any chaperones,[3] at total expressed-protein concentration of 0.3 μM.

In contrast, as shown in Figure 4-4a, the I28D and M131D apoMb variants are partially insoluble when biosynthesized in the absence of chaperones. Next, we analyzed the effect of adding increasing amounts of chaperones to independently biosynthesized WT and apoMb variants. We directed our initial focus on the Hsp70 chaperone system, which comprises *E. coli* Hsp70 or DnaK (K), and the cochaperones DnaJ (J) and GrpE (E). We found that an increase in the concentration of the Hsp70 system, in the absence of TF, increases the solubility of the I28D and M131D apoMb mutants (Figure 4-4a). Interestingly, only 6.5 μM Hsp70 is required to increase the solubility of both mutants to a >80% solubility level. Physiologically relevant concentrations of Hsp70 (50 μM) restore solubility of both mutants to near wild-type levels (87-90% soluble).

Then, we explored the effect of the TF chaperone on the solubility of newly synthesized WT apoMb and its variants. We found that TF increases the solubility of both variants to a similar extent as Hsp70 (Figure 4-4b). This result is consistent with the known overlapping roles of the two chaperones.[9, 12]

In summary, the restored solubility of both variants shows that Hsp70 and TF can independently prevent biological harm (i.e., formation of insoluble species) generated by the

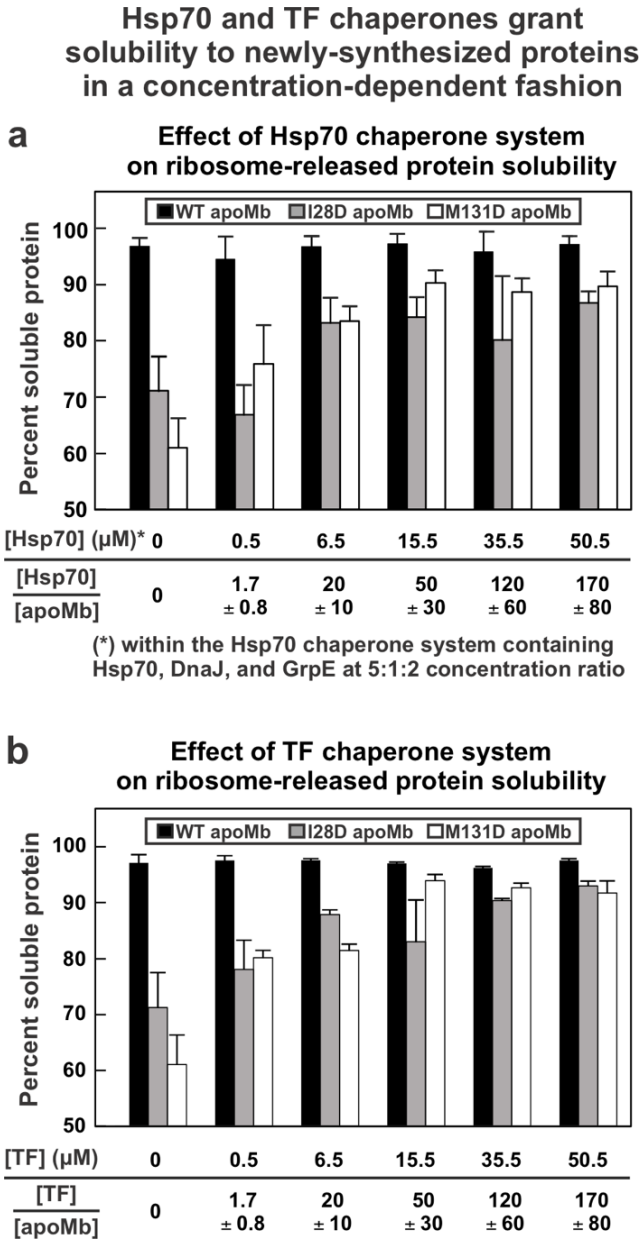


Figure 4-4. The Hsp70 chaperone system and TF grant solubility to I28D and M131D apoMb variants. (a) Solubility of WT, I28D, and M131D apoMb produced with increasing amounts of Hsp70 within the Δtig *E. coli* cell free system. DnaJ and GrpE were added so that the relative ratio of K:J:E was 5:1:2, expect for the 0.5 μM Hsp70 point which has 0.04 μM DnaJ and 0.05 μM GrpE. The apoMb protein concentration is 0.3 ± 0.2 μM. (b) Solubility of WT, I28D, and

M131D apoMb produced with increasing amounts of TF in the presence of an Hsp70 inhibitor within a Δtig *E. coli* cell-free system.

I28D and M131D "kinetic variants". These species disrupt the kinetics of co-/post-translational (I28D) or only post-translational (M131D) folding. Importantly, the fact that the results obtained for the two variants are equivalent suggests that there are no relevant co-translational events potentially experienced by the I28D variant. Therefore, the key steps leading to apoMb insoluble aggregate formation take place mostly post-translationally. A proposed free-energy landscape matching the latter key concept is discussed in a later section of this work. Further, these events are likely related to post-translational interactions with chaperones while the newly released nascent chain kinetically partitions between folding and aggregation paths.

The soluble fraction of both I28D and M131D apoMb variants remains soluble upon long-term incubation (Figure 4-5). Therefore, importantly, these soluble fractions are kinetically trapped from forming insoluble aggregates, at ~0.3 μM nascent-protein concentration. This value is within the range of most protein concentrations in *E. coli* (0.11-4.30 μM).[67] Alternatively, the critical concentration for formation of the insoluble phase of these proteins has not been reached yet, under the tested temperature, pressure, pH and environmental conditions.

In summary, the Hsp70 chaperone system and TF chaperone act during a crucial immediately post-translational time span, to grant solubility to newly synthesized proteins. Note that the TF chaperone is known to bind nascent proteins cotranslationally and this expectation is corroborated by previous experiments on ribosome-bound WT apoMb nascent chains (apoMb RNCs) in the presence of anti-TF antibodies.[48] On the other hand, chaperones may bind the nascent chain cotranslationally and then post-translationally address aggregation issues that may arise upon release from the ribosome. In support of this idea, TF was found to continue interacting with nascent chains even after reaching a non-ribosome bound state.[68] Further, the soluble states appear to be kinetically trapped from forming additional aggregates.[69-72]

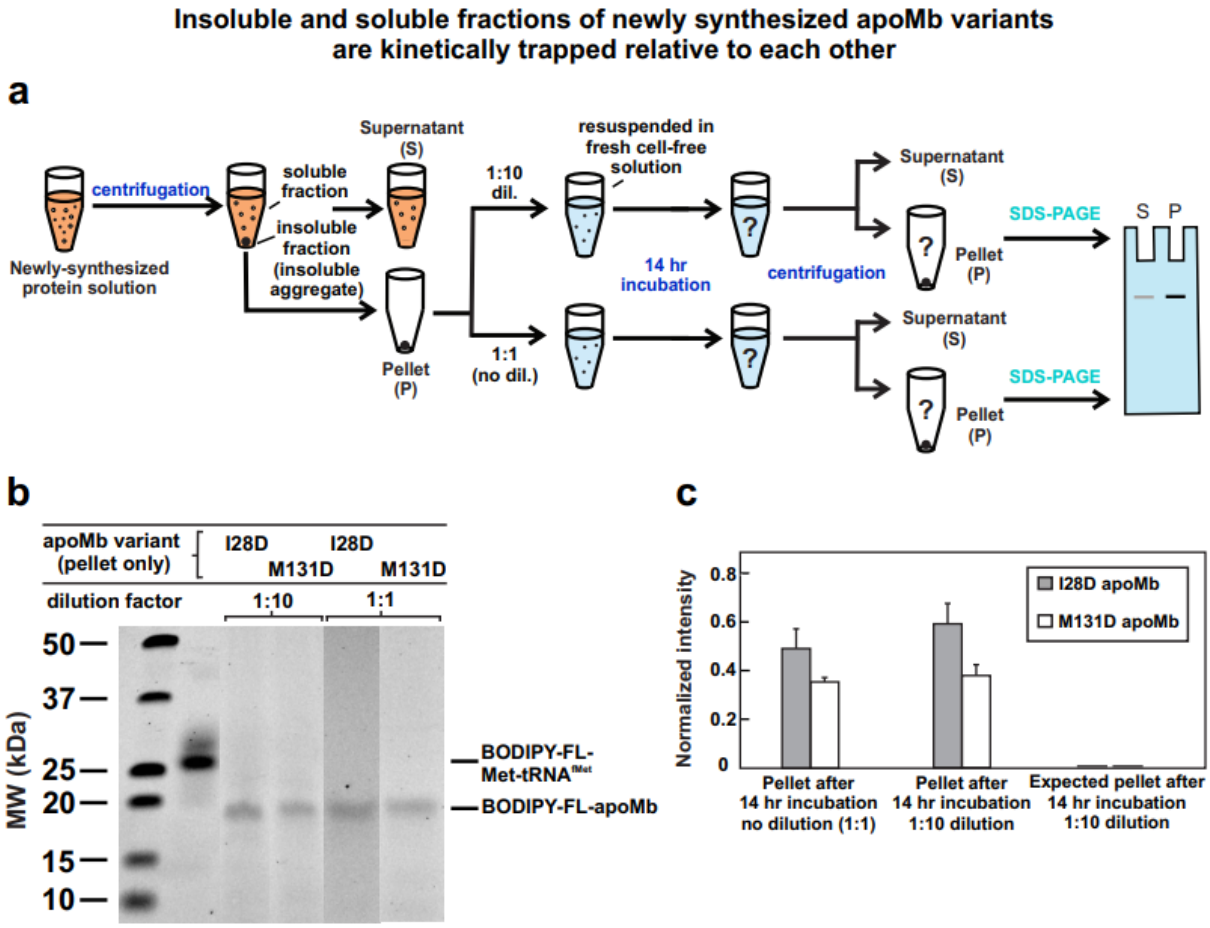


Figure 4-5. Insoluble apoMb variants are kinetically trapped from soluble states. (a) Schematic of experimental procedure. N-terminally fluorophore-labeled protein-containing solutions are centrifuged followed by separation of soluble and insoluble fractions. The insoluble pellet (P) is resuspended in fresh cell-free solution, and the sample is incubated for about 14 hours. After centrifugation, supernatant (S) and pellet (P) are isolated and visualized by low-pH SDS-PAGE and fluoroimaging. (b) Gel data shows the signal from the insoluble fractions of I28D and M131D apoMb upon dilution and incubation. The fluorophore-labeled apoMb and tRNA were visualized with a 488 nm excitation laser and 500-550 nm band-pass emission filter (BPB1). Molecular weight markers were visualized with 635 nm excitation laser and a 675 nm

long-pass filter (LPR). (c) The intensity of the insoluble fraction of apoMb variants does not change upon incubation or dilution.

One major missing piece of information is the fact that the composition of the soluble fraction is not known. For instance, this fraction may be entirely composed by native protein, or it may include additional, potentially self-associated soluble species. The sections below detail our efforts to further characterized the soluble fraction of ribosome-released newly synthesized protein chains.

Insoluble and Soluble Fractions Generated upon Nascent-Chain Release from the Ribosome Are Kinetically Trapped relative to Each Other. Next, we examined whether the soluble and insoluble fractions populated by newly synthesized I28D and M131D apoMb are thermodynamically linked. Briefly, Figure 4-5a shows the simple experimental design that was adopted to test this concept. In essence, we employed resuspension of the isolated insoluble fraction into a fresh cell-free solution devoid of newly synthesized variant-apoMb gene products, followed by long-term incubation and centrifugation. If the soluble and insoluble fractions were thermodynamically linked, the insoluble fraction is expected to generate an overlaying equilibrium population of soluble fraction at the critical concentration, upon long-term incubation. In fact, we estimated that most of the pellet would dissolve, if the soluble and insoluble fractions of newly synthesized proteins interconverted according to a thermodynamic equilibrium (see Methods and SI for additional details). Given that no pellet dissolution was observed (Figure 4-5 b,c), we conclude that the soluble and insoluble fractions of I28D and M131D apoMb are kinetically trapped relative to each other. Note that the presence or absence of ATP does not change the kinetic trapping (Supporting Figure 4-S1).

Solubility Is Not Sufficient for Structural Accuracy: Newly Synthesized WT ApoMb Generated in the Absence of Hsp70 and TF Chaperones Populates Soluble Aggregates, in Addition to Some Native State. Soluble proteins are not necessarily structurally accurate, given

that they may attain soluble misfolded states and(or) soluble aggregates, in addition to native monomers, upon release from the ribosome. Therefore, we set out to further explore the nature of the soluble fraction of newly synthesized apoMb in the absence and presence of the Hsp70 chaperone. Previous work employed 2D NMR in combination with SDS-PAGE and fluorescence to characterize the soluble fraction of WT apoMb generated in the cell-free-system environment in the absence of chaperones.[3] NMR in combination with SDS-PAGE showed that only 23.5% of the protein is biosynthesized in its native state, with the remaining 76.5% in slow exchange with the native conformation and spectroscopically undetectable.[3] This result revealed the presence of either (a) a very slow-tumbling species or (b) additional monomeric conformations in mutual intermediate exchange on the NMR chemical-shift timescale and in slow exchange with the native state.[3]

Fluorescence anisotropy decay in the frequency domain is a powerful tool to study protein conformation and dynamics,[73-76] including events that take place upon release from the ribosome.[48, 49] Application of this methodology to newly synthesized WT apoMb (Table 4-2, item 1) shows that the soluble fraction of newly synthesized wild-type apoMb includes very slow-tumbling species in the absence of chaperones, at even lower apoMb concentration than the NMR studies. This result lends support to interpretation (a) of the NMR results.

Next, we compared data collected via fluorescence anisotropy decays in the frequency domain under different conditions. We identified rotational correlation times (τ_c) corresponding to local and global motions, as well as order parameters and matching amplitudes of local and global motions. As shown in Figure 4-6a, τ_c values are proportional to the size of the corresponding tumbling species, either the entire particle or portions of it that are covalently linked to the N-terminal BODIPY-FL fluorophore. This technique can resolve up to three

different τ_C values within a sample, as long as the motions are independent[49] and bear sufficiently different timescales (typically by ≥ 5-fold).[79] Figure 4-6b displays representative fluorescence anisotropy decay data for WT apoMb made in the absence of the Hsp70 chaperone. Newly synthesized WT apoMb produced in the absence of molecular chaperones within the cell-free system has a large rotational correlation time (τ_c) for global tumbling. Namely: $\tau_c = 260 \pm 180$ in cell-free systems and $\tau_c = 700 \pm 200$ ns in cell-free systems diluted 10-13-fold (items 1 and 2 in Table 4-2). Importantly, the latter rotational correlation time is significantly larger than the corresponding value obtained for independently purified apoMb ($\tau_c = 10 \pm 2$ ns, item 8 of Table 4-2) mixed in a solution containing ribosomes at comparable concentration to ribosome concentration of crude cell-free systems (one-tailed student t-test, Figure 4-S4). Therefore, WT apoMb made within the *E. coli* cell-free environment in the absence of chaperones includes a slow-tumbling species with a large τ_c of 200-700 ns, which we define here as soluble aggregate. The volume of this species is estimated to be 28-90-fold larger than monomeric apoMb (Figure 4-S1). Note that the large τ_c uncertainty of newly synthesized apoMb in the absence of chaperones reflects the fact that it is difficult to resolve rotational motions substantially larger than the fluorescence lifetime of the fluorophore (~5 ns for BODIPY-FL).

As pictorially illustrated in Figure 4-6c, the term soluble aggregate encompasses either homo-aggregates formed from multiple mutually interacting apoMb chains, hetero-aggregates composed of apoMb in complex with the ribosome, or hetero-aggregates featuring apoMb interacting with other cellular proteins. Additional work is necessary to shed further light on the structural details of the soluble aggregate. Importantly, the size of this species decreases when protein expression is carried out in the presence of molecular chaperones (see next section).

Table 4-2. Frequency-domain fluorescence anisotropy decay parameters of cell-free-expressed WT and mutant apoMb in the absence and presence of molecular chaperones.

System	Timescale of motion (anisotropy decays)[c]								
	slow (≥10 ns)	intermediate (0.5-2 ns)			fast (0-0.5 ns)			reduced χ^2	
	$\tau_{c,S}$ (ns)	$\tau_{c,I}$ (ns)	S_I	$\theta_{c,I}$ (deg)	$\tau_{c,F}$ (ns)	S_F	$\theta_{c,F}$ (deg)	2-component[d]	3-component[d]
WT apoMb									
Item 1: Δtig cf sys. + Hsp70 inhib. diluted[a]	700 ± 200				0.122 ± 0.004	0.25 ± 0.002	68.5 ± 0.8	**0.47 ± 0.03**	0.44 ± 0.04
Item 2: Δtig cf sys. + Hsp70 inhib.[a]	260 ± 180				0.104 ± 0.005	0.38 ± 0.01	59.6 ± 0.7	**0.57 ± 0.06**	0.37 ± 0.02
Item 3: WT cf sys. + 21 μM Hsp70[b]	220 ± 140	0.91 ± 0.07	0.78 ± 0.01	32.5 ± 0.9	0.074 ± 0.008	0.57 ± 0.01	47.0 ± 0.8	1.5 ± 0.2	**0.30 ± 0.03**
Item 4: Δtig cf sys + 75 μM Hsp70[b]	70 ± 21	0.7 ± 0.1	0.65 ± 0.05	42 ± 3	0.02 ± 0.02	0.79 ± 0.07	31 ± 6	5.0 ± 0.8	**1.0 ± 0.2**
Item 5: WT cf sys. + 75 μM Hsp70[a,b]	62 ± 9	1.28 ± 0.09	0.76 ± 0.02	33 ± 2	0.080 ± 0.005	0.629 ± 0.006	43.2 ± 0.4	2.8 ± 0.4	**0.03 ± 0.03**
Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. Added post-Translationally[a,b]	44 ± 5	1.25 ± 0.07	0.72 ± 0.02	37 ± 1	0.07 ± 0.01	0.62 ± 0.01	43.9 ± 0.9	3.2 ± 0.2	**0.39 ± 0.07**
Item 7: WT cf sys. + 75 μM Hsp70, diluted[a,b]	34 ± 2	1.3 ± 0.1	0.80 ± 0.01	30.6 ± 0.1	0.108 ± 0.008	0.58 ± 0.01	46.8 ± 0.7	1.65 ± 0.07	**0.4 ± 0.1**

Item 8: purified apoMb resuspended in RNC solution[e]	10 ± 2	0.103 ± 0.006	0.67 ± 0.04	48 ± 3	**1.3 ± 0.3**	0.7 ± 0.1
I28D apoMb WT cf sys + 75 μM Hsp70						
Item 9: Spatially biased/long lifetime species	240 ± 80	0.17 ± 0.01	0.63 ± 0.01	43.0 ± 0.5	1.1 ± 0.5	0.5 ± 0.2[f]
Item 10: Dynamic/short lived species		0.19 ± 0.01	0	180	1.1 ± 0.5	0.5 ± 0.2[f]

Symbols are as follows: Δtig, strain lacking gene encoding trigger factor (TF) chaperone; WT, wild-type; cf sys., cell-free system; Hsp70 inhib., peptide that inhibits Hsp70.[77]

[a] anisotropy decay parameters for items 1,2,5, 6, and 7 reproduced with permission from Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S., Complementary Role of Co- and Post-Translational Events in De Novo Protein Biogenesis. *J. Phys. Chem. B* **2020,** *124* (30), 6488-6507.

[b] For the 21 μM and 75 μM Hsp70 conditions, 0.04 -12 μM DnaJ and 19 μM GrpE were also added.

[c] The subscripts S, I, and F refer to slow global motions (μs), intermediate local motions (ns) and fast local motions (sub-ns). The symbols $\tau_{c,S,}$ $\tau_{c,I,}$ and $\tau_{c,F}$ denote the rotational correlation times. S_I, and S_F are the order parameters, and $\theta_{c,I}$ and $\theta_{c,F}$ are the cone semi-angles that describe the spatial confinement of the tumbling motions. Uncertainty values for τ_c, S, and θ_c are reported as either propagated error or standard error (SE, whichever value was larger), out of 2−7 independent experiments.

[d] The reduced χ^2 values listed in the table are the average and SE for all independent experiments. Three-component models were

selected instead of two-component models if the reduced χ^2 for the three-component fit was ≥ 2.5 times smaller than the reduced χ^2 for the two-component model. Accordingly, a three-component model was selected for WT apoMb produced in the presence of chaperones, and a two-component model was selected for WT apoMb produced in the absence of chaperones. The chosen number of components for any given experimental item is shown in bold.

[e] Data generated from reprocessing of anisotropy decay data originally reported in Ziehr *et al.*[99] Reprocessing procedure described in Supporting Methods.

[f] Reduced χ^2 values for associative model.

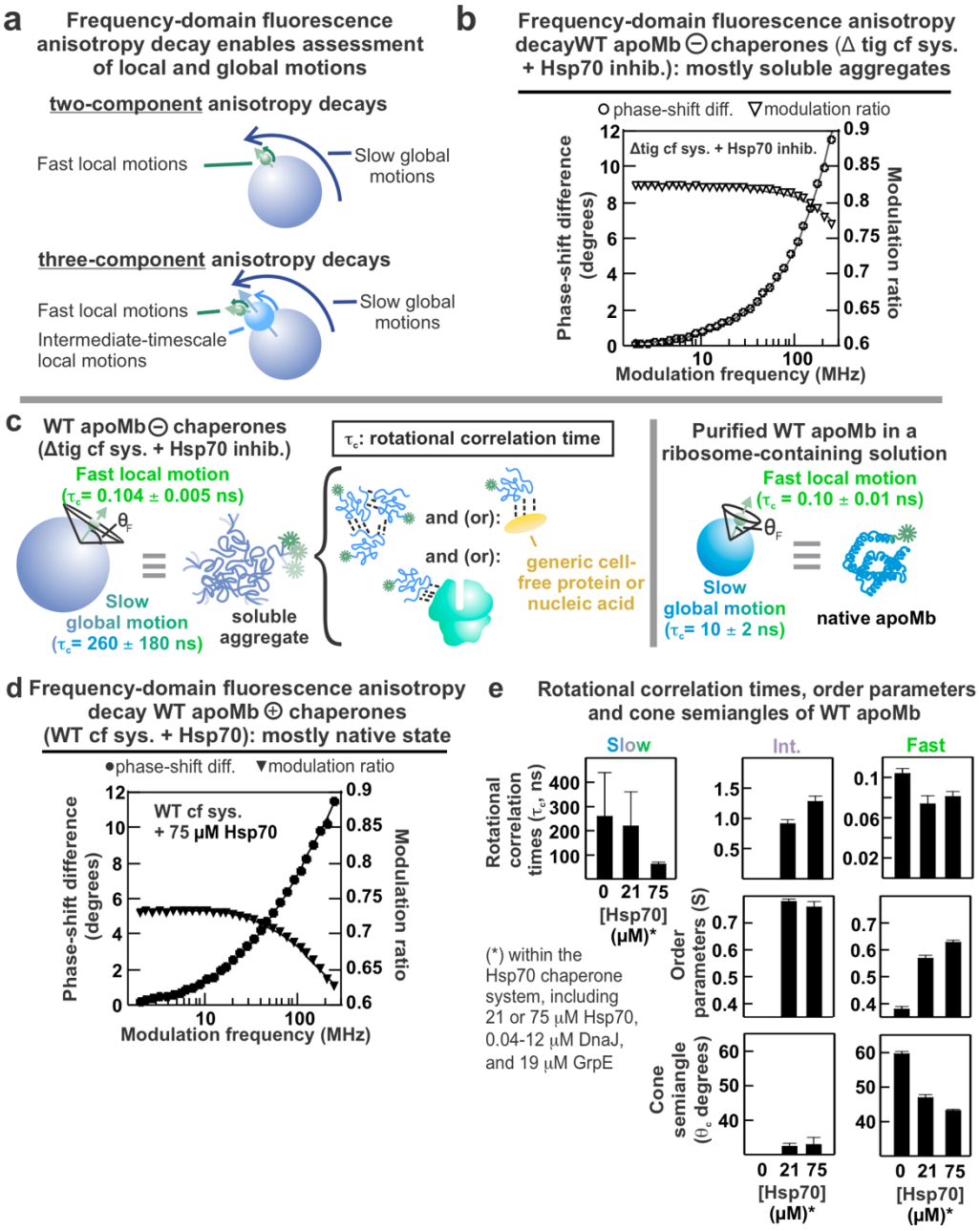


Figure 4-6. ApoMb generated in the absence or presence of Hsp70 has different structural features. (a) Frequency-domain fluorescence anisotropy decays can resolve up to three rotational motions. (b) Representative anisotropy decay data for apoMb (0.3 ± 0.2 μM) synthesized in the absence of TF and KJE chaperones. (c) τ_c values for apoMb synthesized in the absence of

chaperones are consistent with soluble aggregates (left) and larger than τ_c values for purified monomeric protein (right). (d) Representative anisotropy decay data for apoMb (0.3 ± 0.2 μM) synthesized in the presence of 75 μM Hsp70 (e) Rotational correlation times (top), order parameters (S, middle) and cone-semiangles (bottom) for the global and local motions of apoMb synthesized in the presence of different chaperone concentrations. Error bars denote propagated error or the standard error for n = 2-7 (whichever value was larger). Some of these data are also presented in Table 4-2 of Addabbo et al.[3] For the 21 and 75 μM Hsp70 conditions, 0.04 -12 μM DnaJ and 19 μM GrpE were also present. For 0 μM chaperones, a Δtig cell-free was used with 0.4 mM Hsp70 inhibitor.[77]

Therefore, the large τ_c observed for soluble aggregates is not merely the result of non-specific binding to cell-free components, though this effect may partially contribute to the observed τ_c. Note that the τ_c of WT apoMb does not significantly change upon addition of the KLR-70 Hsp70 peptide inhibitor[77] after translation (Table 4-2 item 6). Therefore, WT apoMb is not bound to Hsp70 at equilibrium.

Experiments employing fluorophores with long lifetimes, including DAOTA (lifetime = 16.2 ns) and ADOTA (lifetime = 19.2 ns)[80, 81] were also carried out (data not shown). These investigations, however, were unsuccessful due to poor protein labeling efficiencies and(or) due to non-specific binding of the long-lifetime probes to cell-free-system components. Additional strategies to better understand the nature of the soluble aggregate, including mass photometry and ribosome immunoprecipitation assays to selectively eliminate 70S ribosomes, led to no further usable information (SI text and Figure 4-S6).

At this juncture, it is helpful to recapitulate the results of NMR/SDS-PAGE and fluorescence anisotropy. As already discussed, previous 2D NMR data acquired on cell-free expressed WT apoMb within a chaperone-free cell-free environment under the same experimental conditions show that some native WT apoMb is present in solution.[3] On the other hand, fluorescence anisotropy decay data collected at even lower concentration show that a slow-tumbling species is also present in solution. Therefore, we conclude that cell-free expression of WT apoMb in the absence of chaperones yields a mixture of native state and soluble aggregates. Phasor-plot analysis, shown in a later section of this work, sheds additional light about the relative populations of these two components.

Increasingly High Hsp70 Chaperone Concentrations Lead to Improved Structural Accuracy of Newly Synthesized WT ApoMb. Previous fluorescence-anisotropy-decay data

collected in the presence of the Hsp70 chaperone showed that the observed global size of the tumbling particles gets smaller in the presence of the Hsp70 chaperone system.[3] Here, we share the same fluorescence-anisotropy data but display them in a more comprehensive fashion, including representative raw phase-modulation profiles, and add some new results on data collected at one additional Hsp70 concentration (21 mM). In this way, we more directly address trends related to the presence of increasing concentrations of the Hsp70 chaperone. Some representative raw data are shown in Figure 4-6d and chaperone-dependence trends are illustrated in Figure 4-6 and Table 4-2.

Items 4-7 of Table 4-2 show that apoMb biosynthesized in the presence of 75 μM Hsp70 has a smaller global τ_c than apoMb generated in the absence of chaperones. Interestingly, the latter species fits best to a 3-component anisotropy decay. We tentatively ascribe the 1.3 ns intermediate-timescale motion to a small compact N-terminal region, likely including some independently tumbling A-helix residues.

In addition, the samples generated in the absence and presence of Hsp70 show different local dynamics. The fast (sub-ns, $\tau_C \sim 0.1$ ns) local motions of WT apoMb are diagnostic of motions of the first N-terminal residues of the newly synthesized protein chain, based on tumbling-rate arguments and on the Stokes-Einstein-Debye equation.[48] Interestingly, in the presence of the Hsp70 chaperone, newly synthesized apoMb exhibits confined dynamics characterized by a relatively small cone semi-angle for the fast sub-ns local motions and by a large order parameter (Figure 4-6e and Table 4-2 items 3-7).[49] Conversely, apoMb synthesized in the absence of Hsp70 exhibits N-terminal local motions of slightly wider amplitude (Figure 4-6e and Table 4-2 items 1-2). These differences in the amplitude of the motions of the protein N termini are likely related

to the different local microstructure of WT apoMb in its monomeric and soluble-aggregate forms.

As shown in Figure 4-6e and Table 4-2, the presence of increasing concentrations of the Hsp70 chaperone system leads to an apparent decrease in the observed size of newly synthesized apoMb, as testified by the increasingly lower τ_c values. However, the observed τ_c at physiologically relevant concentrations of the Hsp70 chaperone system (34-62 ns, Table 4-2) is still larger than expected for a 100% population of folded monomeric protein (Table 4-2, item 8). This result suggests that either smaller-size aggregates and(or) a lower fraction of same-size aggregates are still present in the system, even at 75 μM Hsp70 concentration.

The overall idea that chaperones increase the percent native state, but do not fully prevent soluble aggregate formation is supported by the 2D NMR spectroscopy/SDS-PAGE data in the presence of Hsp70 and TF (Figure 4-7), in combination with known data under chaperone-free conditions.[3] As summarized in Table 4-3, the percent of NMR-detected native state increases at higher chaperone concentrations. As already discussed in the previous section, NMR/SDS-PAGE is a powerful approach because it provides a direct assessment of the concentration of newly synthesized native apoMb in the cell-free environment, even in the presence of some spectroscopically undetectable slow-tumbling apoMb.

The combination of fluorescence anisotropy decays and NMR/SDS-PAGE suggests that apoMb produced in the presence of Hsp70 comprises a mixture of native protein and a slow-tumbling version of the same protein. Additional data derived from fluorescence-lifetime phasor plots further support this interpretation and shed additional details on the relative proportions of the two species (see next section).

Table 4-3. Percent of native state of ribosome-released wild-type apoMb proteins detected by 2D NMR spectroscopy in combination with SDS-PAGE at different chaperone concentrations.

apoMb concentration (μM)	Hsp70 concentration (μM)	TF concentration (μM)	Percent native state[a]
59.8	0	0	23.5 ± 0.9
7.3	0.5	0.2	30 ± 2

[a]Data is average ± SE calculated from seven clearly resolved apoMb resonances.

In principle, instead of being an authentic soluble aggregate, newly synthesized apoMb may simply undergo weak transient interactions with other species that slow down its tumbling motion within the complex cell-free mixture. However, weak and fast interactions on the NMR chemical-shift timescale are ruled out because these interactions would broaden the native-state apoMb resonances, and no line-broadening was observed (Figure 4-7).

The presence of soluble aggregates is also consistent with fluorescence emission spectra of newly synthesized WT apoMb in the cell-free environment. These spectra show a small reproducible blue shift for the protein produced in the absence of chaperones, relative to chaperone-free conditions (Supporting Figure 4-S2). This result supports the presence of a more nonpolar environment, possibly due to a larger fraction of soluble aggregate, for the newly synthesized protein produced in the absence of chaperones.

Interestingly, the τ_c of the sample generated in the presence of chaperones gets reproducibly smaller upon sample dilution (Table 4-2 items 5 and 7). The decrease in viscosity upon dilution is partially responsible for this change in τ_c (33% of the effect), as testified by the experimental microscale-volume viscosity measurements reported in Table 4-4. Small changes in the degree of soluble-aggregate levels may also contribute to the observed τ_c variations. The estimated protein hydrated volume is similar for both samples (Figure 4-S1). Further, as shown in Table 4-5 (items 5 and 7), fluorescence lifetime changes are negligible upon dilution, suggesting similar fractions of native state and soluble aggregates, in the presence of Hsp70 and TF chaperones (see also phasor analysis in next section). We conclude that undiluted and diluted newly synthesized apoMb samples prepared in Hsp70- and TF-containing solutions include similar fractions of monomeric folded protein.

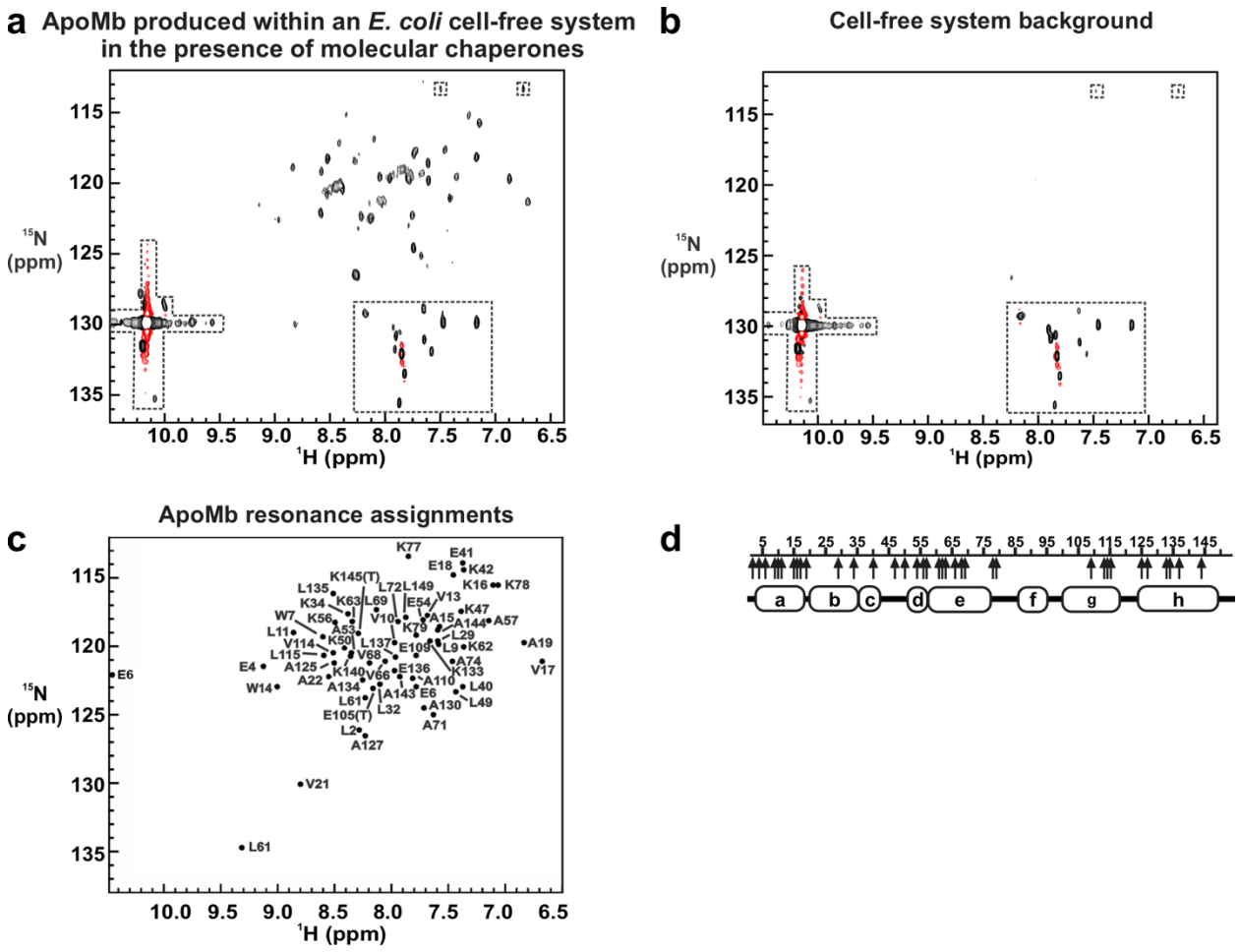


Figure 4-7. WT apoMb is folded and soluble aggregates are minimized when produced off ribosome in the presence of chaperones. ^{1}H, ^{15}N HSQC NMR spectra of selectively labeled apoMb (^{15}N-labeled A, E, L, K, V, and ^{13}C- and ^{15}N-labeled W) generated by transcription-translation in an *E. coli* cell-free system[25]. All data were immediately collected following transcription-translation and contained levels of chaperones naturally present in the cell-free system ([apoMb] = 7.3 μM, [Hsp70] = 0.5 μM, [DnaJ] = 0.04 μM, [GrpE] = 0.05 μM, [TF] = 0.2 μM).[3] All dashed boxes indicate resonances present in the negative control, denoting products of amino acid metabolism[82] and free tryptophan (a) Spectrum of cell-free expressed apoMb. Black, unboxed resonances are consistent with resonances of apoMb native state and have been

tentatively assigned.[83] (b) Cell-free reaction in the absence of T7 RNA polymerase (negative control). (c) Simulated 2D ^{1}H, ^{15}N HSQC spectrum of apoMb based on known assignments.[83] (d) Labeled amino acids (arrows) and α-helical secondary structure of native WT apoMb.

Table 4-4. Viscosity of undiluted and diluted cell free systems.

System	Viscosity (mPa-s)[a]
Item 1: Δtig cf sys.	**0.940 ± 0.005**
+ Hsp70 inhib. diluted	
Item 2: Δtig cf sys.	**1.13 ± 0.02**
+ Hsp70 inhib.	
Item 5: WT cf sys.	**1.11 ± 0.01**
+ 75 μM Hsp70[b]	
Item 7: WT cf sys.	**0.940 ± 0.004**
+ 75 μM Hsp70, diluted[b]	

Symbols are as follows: Δtig, strain lacking gene encoding trigger factor (TF) chaperone; WT, wild-type; cf sys., cell-free system; Hsp70 inhib., peptide that inhibits Hsp70.[77]

[a] The viscosity values are the average and SE for three independent experiments at shear rates of 1500 s^{-1} and 8000 s^{-1}.

[b] For the 75 μM Hsp70 conditions, 0.04 -12 μM DnaJ and 19 μM GrpE were also added.

Table 4-5. Time-resolved fluorescence lifetime of apoMb produced in the presence of variable chaperone concentrations.

	Fluorescence lifetimes and related parameters[a]				Avg fluorescence lifetime (ns)[c]
	Long		Short		
	Fraction[b]	L_1 (ns)	Fraction[b]	L_2 (ns)	
WT ApoMb					
Item 1: Δtig cf sys. + Hsp70 inhib. diluted	0.51 ± 0.02	5.08 ± 0.07	0.47 ± 0.02	1.77 ± 0.03	3.4 ± 0.1
Item 2: Δtig cell strain + Hsp70 inhib.	0.59 ± 0.01	4.82 ± 0.04	0.39 ± 0.01	1.76 ± 0.04	3.52 ± 0.05
Item 3: WT cf sys. + 21μM Hsp70[d]	0.67 ± 0.02	5.13 ± 0.04	0.30 ± 0.02	1.79 ± 0.05	4.0 ± 0.1
Item 4: Δtig cf sys + 75 μM Hsp70[d]	0.75 ± 0.03	4.9 ± 0.2	0.17 ± 0.03	1.3 ± 0.2	3.9 ± 0.2
Item 5: WT cf sys. + 75 μM Hsp70[d]	0.76 ± 0.01	5.15 ± 0.04	0.22 ± 0.01	1.79 ± 0.04	4.29 ± 0.04
Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. added post-translationally[d]	0.76 ± 0.01	5.07 ± 0.03	0.21 ± 0.01	1.74 ± 0.08	4.25 ± 0.07
Item 7: WT cf sys. + 75 μM	0.75 ± 0.01	5.37 ± 0.03	0.22 ± 0.01	1.85 ± 0.05	4.47 ± 0.07

Hsp70, diluted[d]					
Item 8: purified apoMb resuspended in RNC solution	0.95 ± 0.02	5.7 ± 0.1	0.04 ± 0.02	2.0 ± 0.1	5.53 ± 0.08
I28D ApoMb					
Items 9-10: WT cf sys. + 75 μM Hsp70	0.67 ± 0.01	5.5 ± 0.2	0.31 ± 0.01	1.9 ± 0.1	4.2 ± 0.2
BODIPY FL fluorophore controls					
Item 11: BODIPY-FL dye in Δtig cf sys. + Hsp70 inhib.	0.971 ± 0.007	5.22 ± 0.02	0.022 ± 0.006	1.8 ± 0.4	5.12 ± 0.04
Item 12: BODIPY-FL dye in WT cf sys. + 75 μM Hsp70	0.978 ± 0.002	5.36 ± 0.006	0.016 ± 0.002	1.66 ± 0.01	5.27 ± 0.01

Symbols are as follows: Δtig, strain lacking gene encoding trigger factor (TF) chaperone; WT, wild-type; cf sys., cell-free system; Hsp70 inhib., peptide that inhibits Hsp70[77]; L_1 and L_2, fluorescence lifetimes

[a]Fluorescence lifetime data were fit to two fluorescence lifetimes (L_1 and L_2) and a third fictitious lifetime (0.001ns) to account for small contributions from scattering.

[b]Fraction denotes the fractional contribution of each lifetime to the total fluorescence decay.

[c]An average (avg) fluorescence lifetime, weighted by the fractional contribution of L_1 and L_2 is further displayed. Error bars are the propagated error or standard error (SE) for 2-7 independent experiments. The larger of the two errors was selected.

[d]For the 21 μM and 75 μM Hsp70 conditions, 0.04 -12 μM DnaJ and 19 μM GrpE were also added.

Lifetime Phasors Enable Estimating the Fraction of Newly Synthesized WT ApoMb Native State and Soluble Aggregates at Different Chaperone Concentrations. Fluorescence lifetimes are powerful probes of environment. Representative raw data for fluorescence lifetimes of WT apoMb generated in the absence and presence of chaperones are shown in Figure 4-8a. The N-terminal BODIPY-FL fluorophore within the protein has different fluorescence lifetimes in the absence and presence of chaperones, as shown in Figure 4-8b and Table 4-5. These differences are far greater than those observed for the free dye in the absence or presence of chaperones (Figure 4-8b). This result points to a different environment experienced by WT apoMb in the absence and presence of chaperones.

Phasor analysis of fluorescence lifetime data provides additional key information. Lifetime phasors are powerful probes of relative fractions of different components in mixtures.[84, 85] If different experimental conditions generate phasor points that lie along a single line, lifetime phasors support the presence of different populations of mixture components.[84, 85] S and G phasor values were computed from phase shift and modulation data at 23.08 MHz modulation frequency data collected for apoMb expressed with different molecular-chaperone levels. First, we found that the phasors of the free BODIPY-FL dye (in the cell-free environment) lie very close to the universal circle (Figure 4-8c, see also fluorescence lifetimes of the free dye in Table 4-5). In contrast, all phasors of cell-free synthesized WT apoMb produced in the presence of different concentrations of TF and Hsp70 chaperones lie along a straight line within the universal circle

(Figure 4-8c). This result strongly supports the presence of different relative amounts of two non-interacting populations, under the different chaperone conditions.[84] Given the fluorescence anisotropy and NMR/SDS-PAGE results, we ascribe these variable populations to native WT apoMb monomers and soluble aggregates. By explicitly plotting the phasors of 100% monomeric

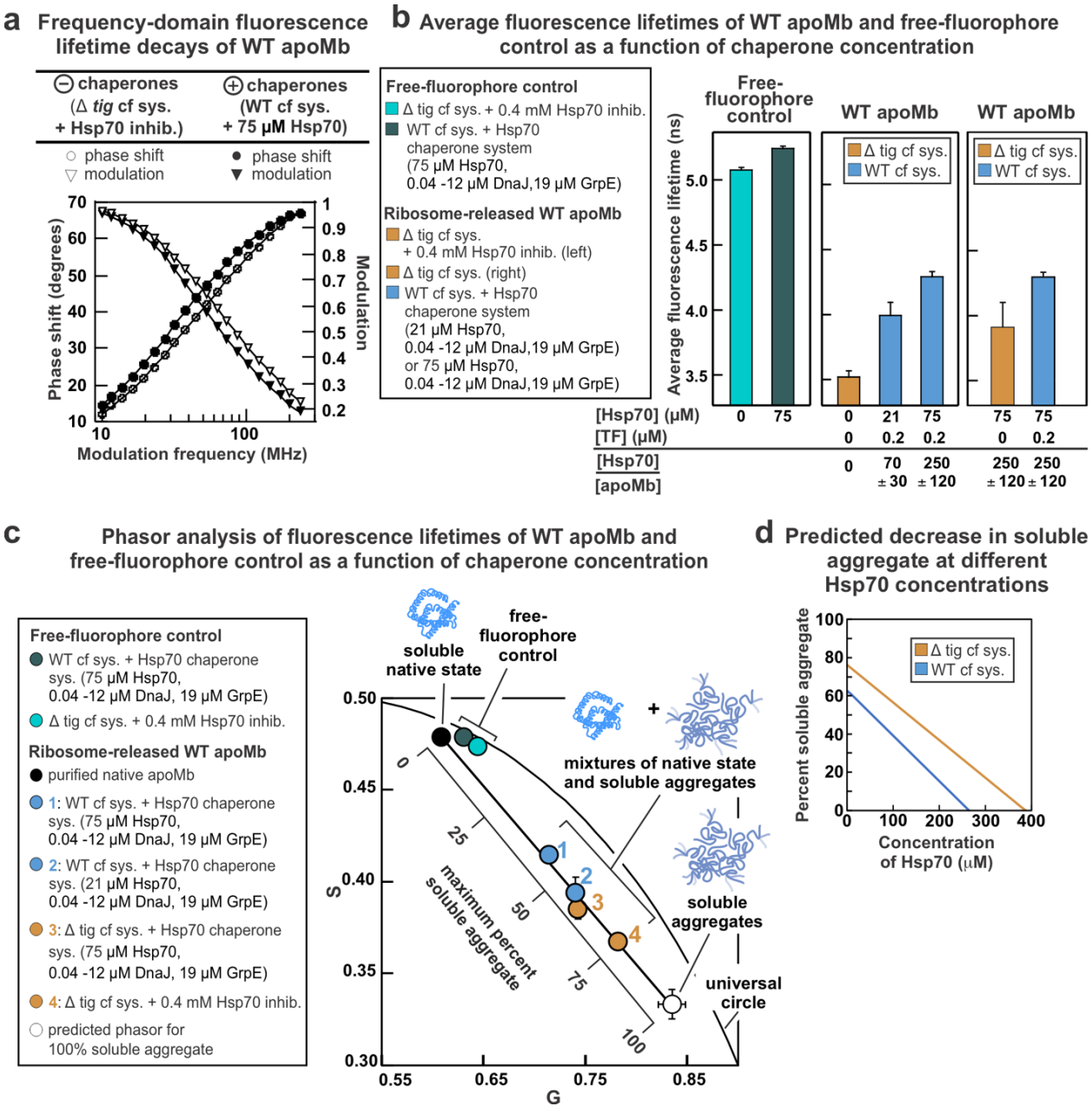


Figure 4-8. The Hsp70 chaperone system increases the structural accuracy of WT apoMb.

(a) Representative frequency-domain fluorescence lifetime decay data for apoMb generated in the absence and presence of KJE and TF chaperones (b) Average fluorescence lifetimes of apoMb (0.3 ± 0.2 μM) and free BODIPY-FL fluorophore in cell-free systems containing different chaperone concentrations. Error bars denote either propagated error or standard error for n = 2 – 7 (whichever value was larger). (c) Fluorescence lifetime phasor plot at 23.08 MHz

modulation frequency for free BODIPY-FL fluorophore and apoMb in cell-free systems containing different KJE and TF chaperone concentrations. Error bars denote propagated error or standard error for n = 2 – 7. All errors in G and S not seen are too small to be visualized on the plot. Some phasor plot points were also shown in the supporting material of Addabbo et al.[3] (d) Predicted percent of WT apoMb soluble aggregate generated with different chaperone concentrations. For the 21 and 75 μM Hsp70 conditions, 0.04 -12 μM DnaJ and 19 μM GrpE were also added.

apoMb and by estimating the percent of soluble aggregates from the NMR experiments in the absence of chaperones,[3] we were able to estimate the maximum fractions of soluble aggregates populated under a variety of different conditions, including different concentrations of TF and Hsp70 chaperones, as shown in Figure 4-8c and Table 4-6. The emerging picture is extremely informative. Namely, larger concentrations of TF and Hsp70 chaperones lead to a decrease in the fraction of soluble aggregates. The presence of TF in solution decreases the required concentration of Hsp70 to reach any given fraction of native protein in the cell-free mixture.

Interestingly, physiologically relevant concentrations of the Hsp70 chaperone system are not sufficient to generate 100% native protein. As shown in Figure 4-8d, the projected concentrations of Hsp70 required to eliminate the presence of WT apoMb soluble aggregates (x-axis intercept) exceeds by far physiologically relevant concentrations of this chaperone, even in the presence of TF. This important result shows that cell-relevant concentration of TF and Hsp70 chaperones are not sufficient to prevent formation of soluble aggregates of all proteins. In the specific case of the WT sperm-whale apoMb model protein analyzed here, this result may be a consequence of the fact that this protein is produced in the presence of heme, in Nature. This cofactor is expected to contribute beyond the action of molecular chaperones, to generate native state devoid of soluble aggregates.[86, 87] Given that our cell-free systems were heme-free, this additional form of folding assistance was absent in our experiments.

Hsp70 Acts during Early Co-Translational or Immediately Post-Translational Folding Steps to Grant Structural Accuracy to Newly-Synthesized Proteins. Hsp70 acts on nascent proteins immediately post-translationally because the τ_c and phasor coordinates of WT apoMb do not significantly change upon addition of the KLR-70 Hsp70 peptide inhibitor[77] after translation (Table 4-2 item 6, Supporting Figure 4-S3). Therefore, WT apoMb is not bound to Hsp70 after

Table 4-6. Summary of estimated populations of native and aggregated states for WT and I28D apoMb made with different chaperone concentrations.

Protein	Hsp70 concentration (μM)[a]	Estimated percent native and aggregated states[b]	
WT apoMb	0	Insoluble aggregate	3 ± 1
		Soluble aggregate	74 ± 1
		Native	23 ± 1
WT apoMb	21	Insoluble aggregate	3 ± 2
		Soluble aggregate	57 ± 11
		Native	40 ± 8
WT apoMb	75	Insoluble aggregate	3 ± 1
		Soluble aggregate	44 ± 6
		Native	53 ± 7
I28D	75	Insoluble aggregate	13 ± 2
		Soluble aggregate	59 ± 2
		Dynamic conformation	28 ± 1

[a] For the 21 μM and 75 μM Hsp70 conditions, a WT cell-free system was employed and 0.04 -12 μM DnaJ and 19 μM GrpE were added. For the 0 μM Hsp70 condition, a Δtig cell-free was employed and 0.4 mM Hsp70 inhibitor[77] was added.

[b] Error bars are the propagated error.

translation, and primarily interacts with newly-synthesized proteins before they complete their folding. In addition, further incubation of either WT[3] apoMb or its I28D variant (see fluorescence anisotropy decay section) for up to ca. 1 hr does not lead to any changes in the rotational properties of the newly synthesized protein.

A Working Model: Hsp70 and TF Lead to a Decrease in the Fraction of Soluble Aggregates Generated upon Newly Synthesized Protein Expression. The results from the previous sections show that increasing concentrations of the Hsp70 and TF chaperone systems decrease the extent of soluble-aggregate formation in a concentration-dependent manner. This concept is schematically illustrated in panels a and b of Figure 4-9. Note that the chaperone concentrations required to prevent soluble aggregate formation are higher than those required to prevent insoluble aggregate formation, as shown upon comparing the data in Figures 4, 8 and 10. More details on this item are provided in the next section.

Different Protein Sequences Have Distinct Hsp70-Chaperone Requirements for the Prevention of Soluble-Aggregate Formation. Next, we focused on the newly synthesized I28D apoMb variant to gain additional insights into the nature of the soluble fraction populated by this species upon release from the ribosome. Fluorescence depolarization in the frequency domain of newly synthesized proteins was carried out directly within the cell-free environment. The WT and I28D variants were compared side-by-side, and representative fluorescence lifetime and fluorescence depolarization decay profiles are shown in Figure 4-10a and b, respectively. Fluorescence anisotropy decay parameters deduced from curve fitting are also shown in Table 4-2. All data were collected in the presence of high concentrations of the Hsp70 chaperone (75 μM) and its cochaperones (at K/J/E 5:1:2 molar ratios). The ribosome-released I28D variant fits best to an associative fluorescence anisotropy decay model.[88-91] Several associative models were

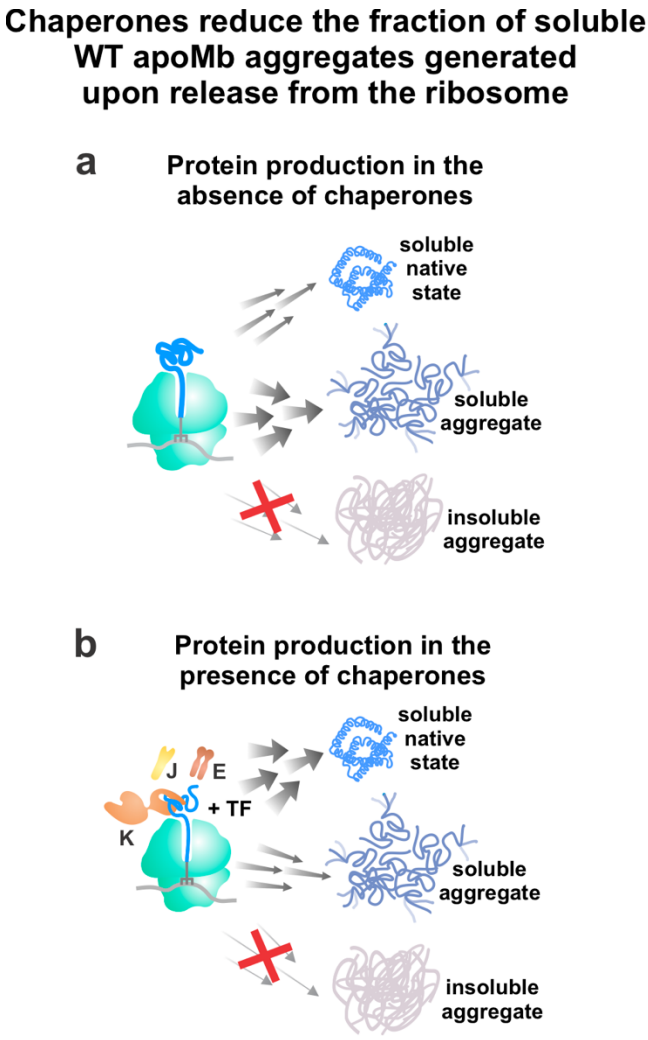


Figure 4-9. Quantitative concentrations of KJE are required to increase structural accuracy of apoMb. (a) Synthesis in the absence of KJE (Δtig cell-free system plus 0.4 mM Hsp70 inhibitor[77]) results in 73.5% ± 0.9% soluble aggregate when apoMb concentration is 59.8 μM. (b) Protein is channeled to its native state when synthesized in the presence of KJE (0.3 ± 0.2 μM apoMb, 75 μM Hsp70, 0.04 -12 μM DnaJ, 19 μM GrpE, and 0.2 μM TF), enabling production of at least 53% ± 7% native state. While apoMb can be synthesized in the absence of chaperones and remain fully soluble, there is a quantitative requirement for KJE concentrations for structural accuracy.

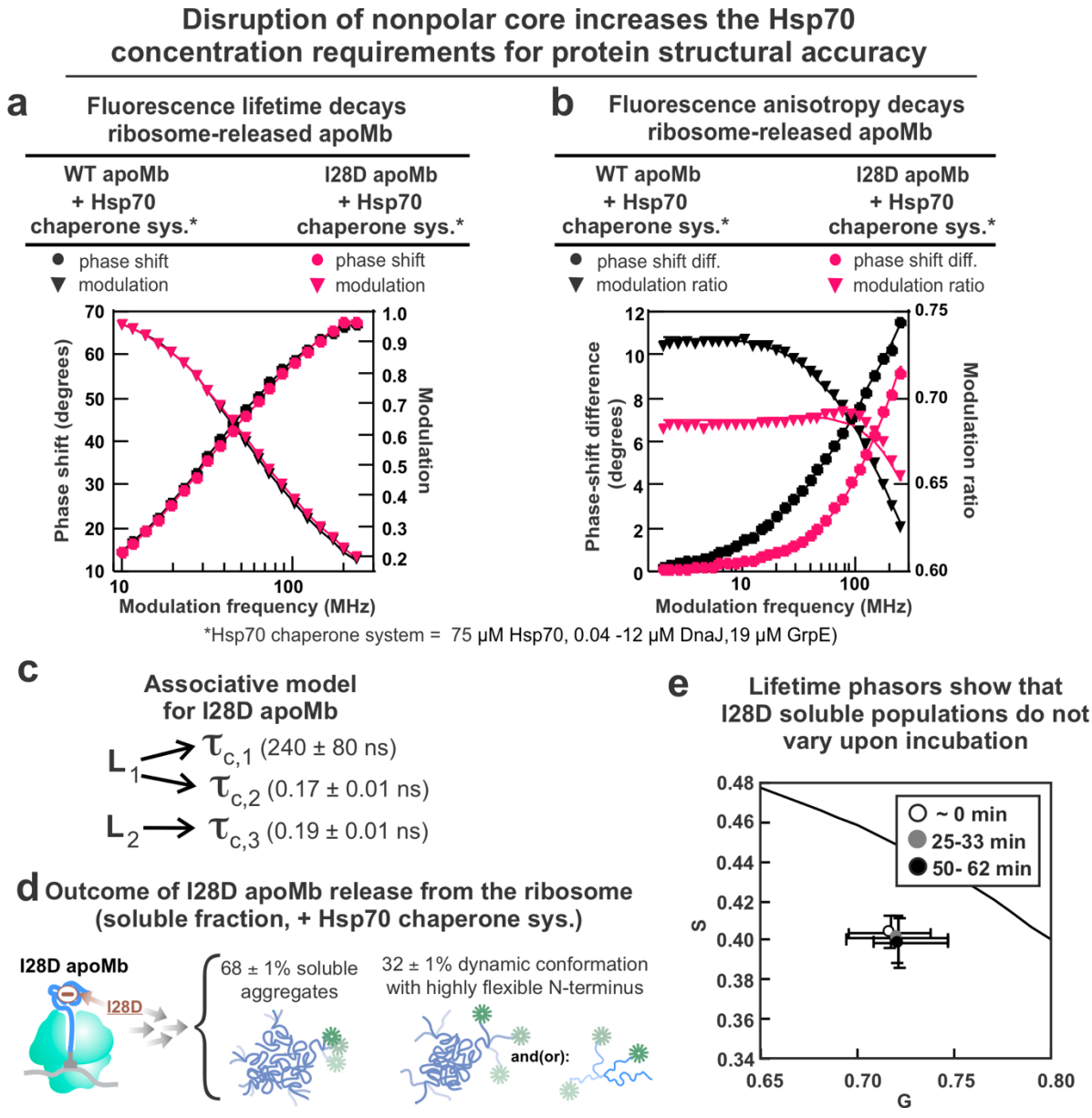


Figure 4-10. Different protein sequences have different Hsp70 chaperone concentration requirements for structural accuracy. (a) Frequency-domain fluorescence lifetime decays of WT apoMb (black) and I28D apoMb (pink) produced in the WT cell-free system in the presence of 75 μM Hsp70. WT apoMb data is the same data shown in Figure 8a. (b) Frequency-domain fluorescence anisotropy decays of WT apoMb (black) and I28D apoMb (pink) produced in the presence of 75 μM Hsp70. WT apoMb data is the same data shown in Figure 6d. (c) Simplified

representation of associative model used to fit I28D anisotropy data in panel d. (d) Cartoon illustrations of I28D apoMb made in the presence of 75 μM Hsp70. The I28D data fits best to an associative model where there are two detectable populations of species in solution. One population has a large τ_c value, suggesting the sample contains soluble aggregates. The global τ_c of the other population cannot be resolved. It has an isotropic fast local motion. (e) Fluorescence lifetime phasors for I28D apoMb made in the presence of 75 μM Hsp70 and incubated for 1 hr after protein synthesis.

tried and the simplest one that provides excellent data fitting is shown in Figure 4-10c (see also Table 4-2). Briefly, this model comprises a slow-tumbling aggregate (τ_{c1} = 240 ± 80 ns) with a flexible N terminus (τ_{c2} = 0.17 ± 0.01 ns) as well as an additional species with an extremely dynamic isotropically tumbling N terminus (τ_{c3} = 0.19 ± 0.01 ns). The latter species may well be an aggregate with an extremely flexible N terminus or an overall highly flexible monomer. This concept is pictorially illustrated in Figure 4-10d, which also lists the relative populations of the two species.

Interestingly, the large aggregate with τ_{c1} = 240 ± 80 ns formed by I28D apoMb has different characteristics from the soluble aggregate detected in the case of WT apoMb, in terms of N-terminal amplitude of motions (Table 4-2) and lifetime values (Table 4-5) from those of the soluble aggregate of WT apoMb. Therefore, the soluble aggregates of I28D and WT apoMb likely have a different structure as well, in addition to different dynamic properties.

Finally, the data of the fluorescence lifetime phasor plot of Figure 4-10e shows that incubation of the I28D apoMb within the cell-free-system environment leads to no changes in solution composition as function of time. Therefore, the two I28D apoMb soluble species detected in solution (soluble aggregate and highly dynamic species) are likely kinetically trapped relative to each other.

In all, the data in Figure 4-10 demonstrate that different protein sequences have different chaperone requirements for solubility and structural accuracy. Importantly, we show that even single point mutations can change the chaperone concentration required to prevent insoluble and soluble aggregation. While I28D is mostly soluble at ~75 μM Hsp70, this chaperone concentration is not sufficient to prevent soluble aggregation. This result is consistent with observations for WT apoMb. While WT apoMb does not need Hsp70 to prevent the formation of

insoluble aggregates at 0.2 μM apoMb, it needs ~300-400 μM Hsp70 to prevent the formation of soluble aggregates (Figure 4-8c,d). WT and I28D apoMb have distinct requirement for Hsp70-mediated prevention of soluble-aggregate formation. Further, in the case of I28D apoMb, we were unable to identify any native state formation, upon release from the ribosome. In essence, physiologically relevant concentrations of the Hsp70 chaperone system are insufficient to prevent soluble aggregation of the I28D variant.

In summary, proteins that are soluble are not necessarily well-folded. Surprisingly, higher concentrations of Hsp70 are required to prevent soluble aggregate formation compared to insoluble aggregation, indicating that the Hsp70 concentration requirement for structural accuracy is distinct from the concentration requirement for solubility. Further, different protein sequences have different chaperone requirements for solubility and structural accuracy. For some protein variants, physiologically relevant chaperone concentrations may be insufficient to prevent aggregation.

Release of Newly Synthesized Proteins from the Ribosome Generates Classes of Conformations That are Kinetically Trapped Relative to Each Other. The data discussed so far illustrate the fact that ribosome release of *de-novo* synthesized proteins generates two or three distinct classes of conformations. Importantly, all these conformations are kinetically trapped relative to each other. This concept is pictorially illustrated in the cartoons of Figure 4-11 and is justified by the following arguments.

Data in Figures 6-8 show that, upon release of the ribosome, WT apoMb is generated as a combination of soluble aggregates and native state protein, which is typically in reversible exchange with intermediates and unfolded states.[43] Lifetime phasors of WT apoMb[3] (Figure 4-8c) demonstrate that the native state is kinetically trapped from soluble aggregates. Two results

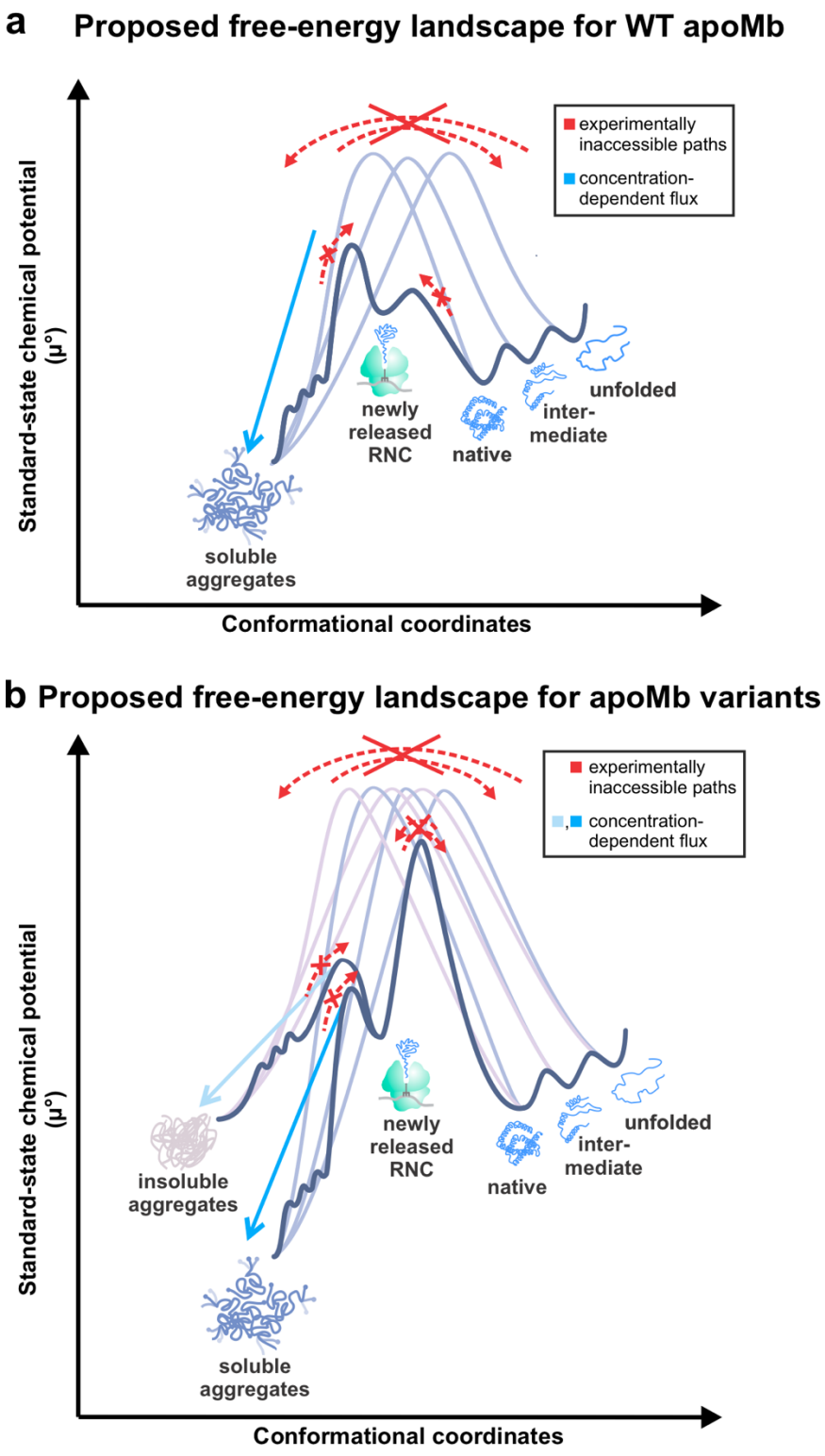


Figure 4-11. Standard-state chemical potential energy landscapes of WT and mutated

apoMb. Energy landscapes for (a) WT apoMb and (b) apoMb variants at protein concentration lower (left) and higher (right) than the critical protein concentration for insoluble aggregate formation. Single point mutations in the apoMb variants increase the kinetic energy barrier between the ribosome-released conformation and the native state.

provide evidence for kinetic trapping. First, apoMb generated in the presence of different chaperone concentrations results in phasor points that are located at different positions along a straight line. Therefore, these samples contain different concentrations of the native state and aggregates. Second, the phasor points do not shift upon 1 hr incubation, demonstrating that the relative concentrations of these two species (aggregates and native) do not change over time.[3] These results show that the native state and soluble aggregates are kinetically trapped relative to each other over a timescale longer than typical *E. coli*'s doubling time. Note that if the protein concentration was below the critical concentration for soluble aggregate formation, we would also not observe variations with time in phasor plots. However, the presence of soluble aggregates demonstrates that the protein concentration is above the critical concentration for aggregate formation. Phasor coordinates will also not shift in time if equilibrium is established rapidly prior to the first fluorescence measurement. However, if the aggregates and native state already reached equilibrium, then all samples would show the same concentrations of these two species. The cartoon in Figure 4-11a illustrates this important concept.

Therefore, due to the above kinetic trapping, co- and immediately post-translational steps assume an outmost importance for WT apoMb, in that they route the newly synthesized protein towards a region of the free-energy landscape that either supports (in the case of native, intermediates and unfolded states) or does not support (in the case of soluble aggregates) biological activity. Therefore, co- and immediately post-translational events effectively determine the entire fate of proteins that, like apoMb, give rise to aggregated species upon release from the ribosome. Further, the role of helpers like the Hsp70 chaperone system is of outmost importance, given that chaperones modulate the relative ratios of native-state and soluble-aggregate components via transient co- and immediately post-translational interactions

with newly synthesized proteins. Once release from the ribosome and transient interactions with molecular chaperones are complete, native protein fractions can be biologically active, presumably with no harm from soluble aggregates, which remain kinetically trapped relative to the native state. Soluble aggregates, however, decrease the yield of native state and are often harmful or toxic to cells.[21-23] Therefore, they are generally undesirable.

Second, kinetic trapping of newly synthesized native states and soluble and insoluble aggregates relative to each other is also applicable to newly synthesized I28D and M131D variants of apoMb. The cartoon in Figure 4-11b illustrates that fact that newly synthesized native states, soluble aggregates and insoluble aggregates of the two apoMb variants, which are all experimentally observed (Figs. 4-7, 10 and Addabbo *et al.*[3]), are trapped relative to each other. Kinetic trapping of insoluble aggregates relative to the soluble fraction is proven in Figure 4-5. In addition, kinetic trapping of soluble aggregates relative to the native state is proven in Figure 4-10e for the I28D apoMb variant.

In summary, release from the ribosome can generate undesirable soluble and insoluble aggregates of newly synthesized proteins, in addition to the native state. These three classes of species are kinetically trapped relative to each other, in the case of apoMb and its variant model systems. Therefore, it is crucially important to avoid the formation of soluble and insoluble aggregates at the outset. The correctly folded species generated upon release from the ribosome were indeed shown to persist over long timescales. This scenario may of course be subject to change, but only in case covalent damage or changes in environmental conditions were to modify the free-energy landscape and perturb the extent of kinetic trapping.

We also showed that if the concentrations and client-protein amino-acid sequences are sufficiently high and within physiologically relevant ranges, the Hsp70 and TF molecular

chaperones contribute to reduce the extent of aggregation upon release from the ribosome. This is unfortunately not always the case, as seen for instance in Figure 4-8c,d.

Finally, it is worth noting that *in vivo* overexpression of WT apoMb takes place at concentrations much higher than the cell-free concentrations analyzed in this work. Under these conditions, mostly insoluble aggregates known as inclusion bodies are routinely formed.[47, 54] Therefore, upon *in vivo* overexpression of WT apoMb, a free-energy landscape similar to the one shown here for the variants (Figure 4-11b) likely also applies to the WT protein.

Biological Implications. Our studies highlight the importance of soluble aggregate formation upon release of nascent protein chains from the ribosome. In contrast, prior investigations analyzing protein expression in the absence and presence of chaperones focused primarily on the native state alone or on insoluble aggregates.[4, 18]

Further, the data in this work indicate the need to examine client-protein dependence on Hsp70 in a quantitative manner. Different proteins have different Hsp70 concentration needs depending on their sequence and likely other physical properties. In biological research, there is a significant need to be able to produce large quantities of soluble folded protein by overexpression. The co-overexpression of molecular chaperones during protein overexpression in bacteria has shown variable success, to date.

The flow chart in Figure 4-12 highlights the key parameters that influence the generation of soluble native state, soluble aggregates and insoluble aggregates within the *E. coli* cellular environment. Our work shows that at sufficient concentrations, chaperones can play a protective role and prevent aggregation caused by mutations. However, physiological concentrations of chaperones are sometimes insufficient to prevent aggregation of some protein variants.

In addition, as highlighted in Figure 4-13, apoMb variants require higher chaperone concentrations to prevent soluble and insoluble aggregate formation, relative to the wild-type protein. Our findings show that, depending on the specific protein sequence, molecular chaperones at higher concentrations than physiologically attainable may be necessary to generate 100% native states upon protein expression. This observed sequence dependence of chaperone requirements is likely a consequence of the sequence dependence of the relative rates of protein folding and aggregation upon release from the ribosome, as previously shown in the case of WT and M131D apoMb.[3]

Previous work in mammalian cell, yeast, and mouse model systems indicated that heterologous expression of molecular chaperones and upregulation of chaperones by chemical induction of the heat shock response reduce protein aggregation.[92-99] However, more investigations are necessary to identify whether the necessary chaperone concentrations can in practice always be achieved for the complete prevention of soluble and insoluble aggregate formation upon release from the ribosome.

4.4 Methods

Cell-Free Transcription-Translation. Ribosome-released protein were produced as previously described.[48] In summary, a pET-Blue1 plasmid containing genes encoding WT sperm whale apoMb, or I28D apoMb [47] were expressed in either a WT or Δtig (lacking endogenous TF) A19 *E.coli* transcription-translation system prepared *in-house*.[25] The endogenous concentration of Hsp70, DnaJ, and GrpE in the cell-free systems are approximately 0.5 μM, 0.04 μM, and 0.05 μM, respectively. The endogenous concentration of TF in the WT cell-free system is 0.2 μM. Cotranslational site-specific N-terminal labeling of expressed apoMb was achieved by addition

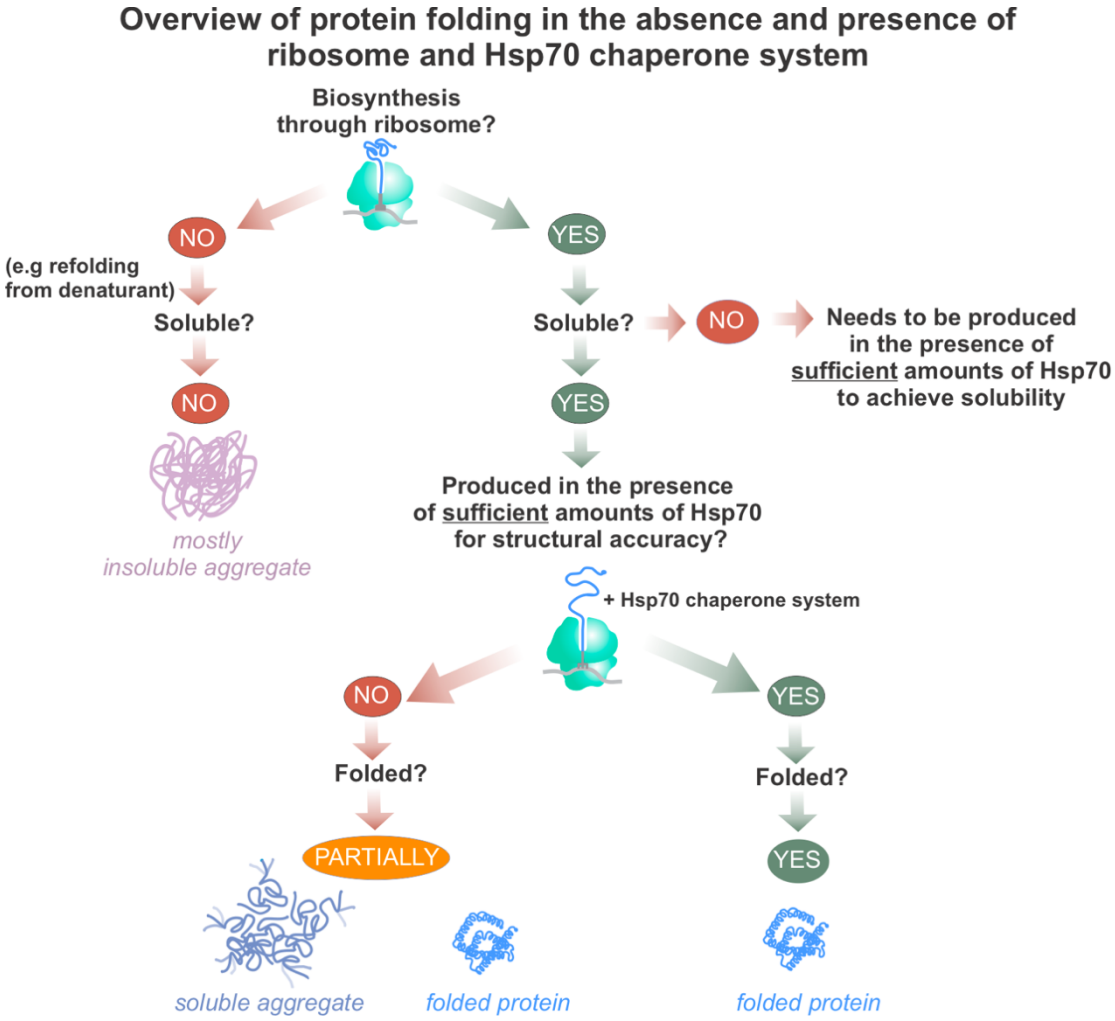


Figure 4-12. Flowchart summarizing the key parameters affecting the generation of soluble and folded proteins in the cellular environment. If a protein has not been biosynthesized through the ribosome, it refolds mainly into an insoluble aggregate, with only very small amounts of soluble folded species. If the protein is biosynthesized through the ribosome in the absence of sufficient Hsp70, it is soluble but not structurally accurate. Only when a protein is synthesized by the ribosome in the presence of sufficient Hsp70, is it both soluble and well-folded.

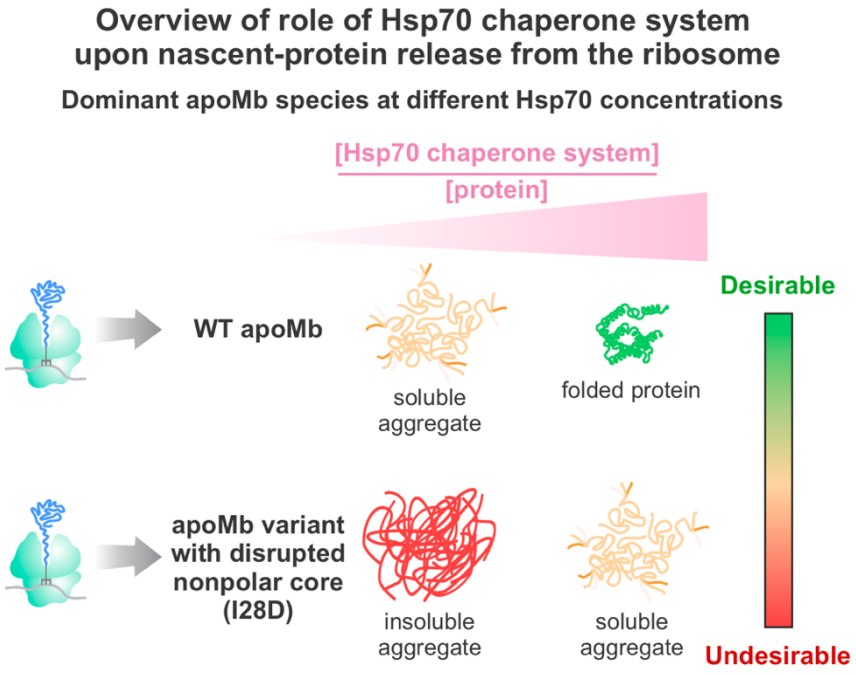


Figure 4-13. Client protein requirements for Hsp70 chaperone system concentration. The Hsp70 concentration requirement for protein structural accuracy is greater than the concentration requirement for protein solubility. The Hsp70 concentration requirements for solubility and structural accuracy depend on the protein sequence. Single point mutations increase the Hsp70 concentration requirement for protein solubility (I28D and M131D) and structural accuracy (I28D).

of *in-house* prepared BODIPY-FL-Met-tRNA^{fMet}.[48] When applicable, an *in-house* inhibitor to Hsp70 was added to a final concentration of 400 μM to produce apoMb in the absence of Hsp70. All cell-free transcription-translation reactions were carried out at 37°C for 20 minutes. Full-length ribosome-released WT apoMb was produced in the presence of added purified Hsp70, DnaJ and GrpE. The final concentrations of Hsp70 were approximately 21μM or 75 μM, as listed. DnaJ was present at final concentrations of 0.04 μM - 12 μM. GrpE was present at approximately 19 μM.

Generation of I28D and M131D ApoMb Plasmids. A Quikchange II Site-Directed Mutagenesis Kit was used to mutate the pET-Blue1 plasmid containing the WT sperm whale apoMb gene to generate a plasmids containing the I28D and M131D point mutations.

SDS-PAGE Gel Solubility Assays. Cell-free expressed apoMb with an N-terminal BODIPY-FL fluorophore was centrifuged at 15,800 rcf and 4°C for 20 minutes. The resulting pellet was separated from the supernatant and resuspended in resuspension buffer (10 mM TrisCl pH 7.0, 10 mM $Mg(OAc)_2$, 60 mM ammonium chloride and 0.5 mM EDTA). The resulting supernatant and pellet fractions were then run on a low-pH SDS-PAGE gel as described previously. [100] The fluorophore-labeled apoMb in each fraction was subsequently visualized using an FLA 9500 Typhoon Gel Imager with a 488 nm excitation laser and 500-550 nm band-pass emission filter (BPB1). Molecular weight markers were visualized with 635 nm excitation laser and a 675 nm long-pass filter (LPR). The apoMb band intensities in both fractions were then analyzed using the Image J software package.[101, 102] The percent of soluble protein in each sample was then calculated by the following expression, where $I_{soluble}$ *and* $I_{insoluble}$ are the fluorescent band intensities of the soluble and insoluble fractions:

$$\% \text{ Solubility} = \frac{I_{soluble}}{I_{soluble} + I_{insoluble}} \times 100 \quad (1)$$

When calculating the percent solubility of apoMb RNCs, the soluble and insoluble band intensities were calculated as the difference in intensity between samples treated without and with puromycin.

Post-Translational Kinetic Trapping of Soluble and Insoluble States of Newly Synthesized Protein. Experiments shown in Figure 4-5 were performed as follows. Ribosome-released M131D and I28D apoMb were generated as described above in the absence of chaperones, in a Δtig S30 system, with Hsp70 chaperone inhibitor (150 μM). The process was quenched by placement on ice. Soluble and insoluble protein fractions were separated by centrifugation at 15 800 rcf at 4 °C for 20 min. The pellet was resuspended, either in the same volume as the original reaction or diluted 1:10, in a cell-free transcription−translation solution, either with or without ATP, containing all components except for T7 RNA polymerase, BODIPY FL-Met-tRNA^{fMet} and plasmid DNA, to prevent further protein production. Additionally, puromycin (1 mM, in 10 mM Tris, 11 mM $Mg(OAc)_2$, 60 mM NH_4Cl, 0.5 mM EDTA at pH 7) was added to both the supernatant and resuspended pellet to further ensure that no additional protein synthesis would occur. Supernatant and resuspended pellet were incubated at 37 °C for ca. 14 h, and then placed on ice. The solutions were then centrifuged at 15 800 rcf at 4 °C for 20 min. Pellets were resuspended in loading buffer (83mM TrisBase, 3.8M glycerol, 116mM SDS, pH 5.7). All samples were subject to low-pH SDS-PAGE, and gel-band intensities were quantified with ImageJ. Intensities from samples on different gels were normalized using the tRNA intensity, where the tRNA concentration is about 11.67μM, first diluted 0.1:25 in RNase free water and then 1:11 in loading buffer.

Fluorescence Lifetime and Anisotropy Decay. An ISS Chronos spectrophotometer was used to collect lifetime and depolarization data as previously reported.[88] A 477nm laser diode was

used to excite fluorescently labeled samples. A 480 nm bandpass excitation filter and a 495nm long pass emission filter (HHQ495lp, Chroma) were used. The instrument has calcite prism polarizers. For both lifetime and depolarization experiments, samples were excited with the excitation polarizer in the vertical position. For lifetime experiments, the emission polarizer was set to the magic angle, 54.7°. The spectrophotometer was temperature-regulated using a water bath set to 25°C. Samples were temperature preequilibrated in the instrument before data collection for ≥30min. Fluorescence depolarization experiments were corrected using an independent G experiment, performed for each experimental day, as described by Knight *et al.*[88]

Lifetime and depolarization experiments were fit in the Globals Software Suite (LFD). X^2 was calculated for all fits, assuming an instrumental error in the modulation of 0.004 and an instrumental error in the phase of 0.2°.[75] All lifetime data were fit to three exponential decay expressions. The first two discrete components were determined by the fit and are the fluorescence lifetimes of the labeled samples. A third fictitious 1ps lifetime was fixed in the fit, to account for an extremely small amount of scattering in the sample. Lifetime phasors[84] were calculated at 23.08 MHz for full-length WT apoMb produced in the cell-free system without stalling (Figure 4-6b). The expressions below were used,

$$G = M\cos\phi \tag{2}$$

$$S = M\sin\phi \tag{3}$$

Where ϕ is the phase shift between the excitation and emission light intensity profile and M is the frequency-domain modulation, which is defined as

$$\mathrm{M} = \frac{AC_{EM}/DC_{EM}}{AC_{EX}/DC_{EX}} \tag{4}$$[76]

where AC is the amplitude of the light intensity wave, and DC is the average intensity. The EM and EX subscripts refer to the emission and excitation light, respectively. Fluorescence

depolarization experiments on WT apoMb and M131D apoMb were fit to 3-component exponential decay expressions, which resulted in at least 2-fold reduction in the calculated X^2 of the fit, relative to the X^2 of the 2-component exponential decay fits. Ribosome-released proteins were fit to either a 2-component or 3-component exponential decay fit. 3-component exponential decay fits were only used if they resulted in a 2-fold or greater reduction in the calculated X^2 relative to the 2- component exponential decay fits. Order parameters and cone semi-angles were calculated according to *Ellis et al.*[49]

Lifetime Phasor Analysis to Estimate the Mole Fraction of Soluble Aggregate Produced in the Absence of DnaK. This section describes the method to calculate the percent soluble aggregate reported in Figure 4-8. NMR experiments were only able to detect 23.5% of apoMb protein made in the absence of chaperones. Therefore, we estimated that this sample, which contained 59.8 μM protein contained 76.5% soluble aggregates. The samples used for fluorescence lifetime and anisotropy experiments contained 0.3 ± 0.2 μM apoMb protein. Since the samples made without chaperones for fluorescence experiments contained a lower concentration than the samples used for NMR, we concluded that these samples would have at most 76% soluble aggregates. The fluorescence lifetime phasor of WT apoMb shifts along a straight line depending on the relative concentration of soluble aggregate and native protein (Figure 4-8c).[3] The lifetime phasor for purified apoMb was generated from fluorescence data of E41C apoMb that was overexpressed in *E. coli,* and labeled with BODIPY FL at the mutated cysteine position via an ester bond. The purified and labeled protein was refolded in a solution containing unlabeled ribosome-bound nascent chains.[78] The concentration of ribosomes in this experiment was similar to the concentration of ribosomes in the fluorescence experiments described in the main text of this work. In this work, we reanalyzed the previously published

fluorescence data to determine the phasor plot point for the pure native state with no soluble aggregates. The phasors for all samples were calculated from the fluorescence lifetime data using eqs. 2-3. G is plotted on the x-axis and S is plotted on the y-axis of the phasor plot. For this work, all phasors were calculated using the M and ϕ values at the 23 MHz modulation frequency. The universal circle is defined by,

$$\left(G - \frac{1}{2}\right)^2 + S^2 = \frac{1}{4}. \quad (5)$$

We then used the experimental phasors to predict the phasor for a sample that contained 100% soluble aggregates. The predicted phasor was calculated with eqs. 6-8.

$$G_{mix} = f_{SA}G_{SA} + f_N G_N \quad (6)$$

$$S_{mix} = f_{SA}S_{SA} + f_N S_N. \quad (7)$$

$$f_{SA} + f_N = 1, \quad (8)$$

Where G_{mix} and S_{mix} are the experimental phasor coordinates for apoMb produced in the absence of Hsp70. The fractions f_{SA} *and* f_N are the fraction of the total fluorescence intensity of the mixture due to the soluble aggregate and native state, respectively. G_{SA}, S_{SA}, G_N and S_N are the phasor coordinates for the pure soluble aggregate and native state. We assumed that the purified protein sample contained no soluble aggregates, and we assumed that the sample made without chaperones had ≤76.5% soluble aggregates. We also assumed that all samples contained only two species: the native state and soluble aggregates. After calculating the predicted phasor for pure soluble aggregates, we used these values to estimate the percent soluble aggregates for the other chaperone conditions. We calculated the percent soluble aggregate for each condition by using the experimental phasor coordinates for G_{mix} and S_{mix}, the predicted phasor coordinates for the soluble aggregates for G_{SA} and S_{SA}, and the experimental phasor coordinates for purified apoMb for G_N and S_N. From these equations, we solved for f_{SA} and f_N.

The maximum percent of soluble aggregates for each chaperone concentration was used to plot the percent soluble aggregates vs the concentration of Hsp70. The linear fit for these plots was used to estimate the amount of Hsp70 required to eliminate soluble aggregates in WT and Δtig S30.

***In vivo* apoMb Expression and Purification.** Large-scale overexpression and purification of WT, M131D, and I28D apoMb were carried out as described.[3] A pET-Blue1 plasmid containing the I28D mutation was used to generate this variant. Purified proteins were identified with MALDI mass spectrometry. The experimental molecular weights matched the expected molecular weights for all three species (WT apoMb: 17,331 amu; M131D apoMb: 17,315 amu; I28D: 17,333 amu).

NMR Sample Preparation. Selectively-labeled apoMb containing ^{15}N A, E, L, K, V, and ^{15}N–^{13}C W analyzed in NMR experiments was generated via cell-free expression and the same dialysis procedure described by Addabbo *et al*.[3] The only difference in the dialysis procedure for this work was that apoMb was generated with WT S30 containing endogenous chaperone concentrations (0.5 μM Hsp70, 0.04 μM DnaJ, 0.05 μM GrpE, and 0.2 μM TF). Prior to NMR analysis, 40 μL of D2O were added to 360 μL of sample the cell-free mixture and loaded onto a 5-mm-diameter Shigemi tube (Wilmad, BMS-005B). A negative control sample was prepared as above, except no T7 RNA polymerase was added to prevent gene expression. SDS PAGE gel analysis was used to determine the apoMb concentration of NMR samples. A calibration curve was generated from the gel intensity of cell-free reactions containing different known concentrations of apoMb. The calibration curve was used to calculate the apoMb sample concentration used for NMR experiments ([total apoMb]). Gels were stained with Coomassie blue and imaged on a Typhoon FLA 9500 (GE Healthcare) upon excitation with a 532 nm laser

in the presence of a 575 nm long-pass filter. ApoMb gel band intensities were quantified with the ImageJ software.[101, 102]

NMR Data Collection, Processing, and Volume Analysis. NMR spectra were collected and processed as described previously.[3] Data in parts a and b of Figure 4-8 were collected with 740 and 700 scans per row, respectively. Volume analysis of NMR resonances was carried out to estimate the fraction of spectroscopically undetectable apoMb protein due to extremely slow tumbling, as described previously.[3] Briefly, the 2D NMR volume of the 1H−15N resonance corresponding to the free-Trp indole (Vol_{Trp}) was measured with the NMR View J software for a sample containing 800 μM of free-Trp. The volume of the following clearly resolved apoMb resonances were measured (Vol_{res}): V17, A19, A22, A57, K62, K78, or A130. Then, the apoMb concentration was calculated with eq. 9:[3]

$$\frac{Conc_{Trp}}{Vol_{Trp}} \text{ x } Vol_{res} = Conc_{res} \quad (9)$$

The seven separate $Conc_{res}$ values were averaged together to calculate the concentration of native apoMb detected with NMR ($[Native]_{NMR}$). The percent native protein was calculated with eq. 10:

$$\% \text{ native} = \frac{[Native]_{NMR}}{[\text{total apoMb}]} \text{ x } 100\% \quad (10)$$

Where [total apoMb] is the total apoMb protein concentration determined by SDS-PAGE gel analysis with a calibration curve generated with apoMb standards of known concentration.

Characterization of WT ApoMb and ApoMb Variants. The secondary structure of WT apoMb and apoMb variants was assessed with circular dichroism, as described previously.[3] The thermodynamic stability of the WT and variant apoMb proteins was assessed by Trp fluorescence upon denaturation with urea, as described in Addabbo *et al.*[3] The Hsp70 binding sites for the proteins were identified with the algorithm developed by Rüdiger *et al*.[66]

4.5 Conclusions

In summary, this work highlights an important role of the bacterial Hsp70 chaperone system during *de novo* protein biogenesis. Namely, increasing Hsp70 concentration not only decreases the fraction of insoluble translation product (whenever relevant), but also increases the structural accuracy of the soluble gene product by augmenting the fraction of protein native state over soluble aggregates. Small changes in amino-acid sequence (single point mutations) can dramatically perturb the solubility of nascent proteins upon release from the ribosome. Hsp70 is able to restore this solubility. On the other hand, the concentration of Hsp70 that is sufficient to prevent wild-type and variant client proteins from forming insoluble aggregates is insufficient to prevent soluble-aggregate formation. In essence, Hsp70 concentration requirements for solubility and structural accuracy within the soluble state are distinct, with structural accuracy requiring higher levels of Hsp70 than mere solubility. In the case of the M131D and I28D apoMb variants, the concentration of Hsp70 chaperone system required to generate a fully native client protein upon release from the ribosome exceeds chaperone solubility. This finding suggests that the evolution of other chaperone systems may have arisen to overcome the imperfect nature of the inherent limitations of the Hsp70 chaperone system.

4.6 Acknowledgments

We thank the National Science Foundation (NSF) for funding (grants MCB-2124672 and CBET 1912259 to S.C.). R.M.A. was the recipient of a National Science Foundation Graduate Research Fellowship Program (Grant No. DGE-1256259). Support for this research was also

provided by the Graduate School and by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Dorothy Powelson Scholarship Fund. R.H. thanks the National Institute of General Medical Sciences of the National Institutes of Health for a TEAM-Science Fellowship (award number R25GM083252). We are also grateful to Prof. David Jameson for sharing useful insights. We thank Aaron Hoskins for allowing us to use the Refeyn Two^{MP} instrument and Ye Liu and Brenda Watt for helping us with data collection and analysis with this instrument. Finally, we thank Justin Dang for helpful discussions.

4.7 References

(1) Varela, A. E.; England, K. A.; Cavagnero, S. Kinetic trapping in protein folding. *Protein Eng. Des. Sel.* **2019,** *32*, 103-108.

(2) Varela, A. E.; Lang, J. F.; Wu, Y. F.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S. Kinetic trapping of folded proteins relative to aggregates under physiologically relevant conditions. *J. Phys. Chem. B* **2018,** *122*, 7682-7698.

(3) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(4) Niwa, T.; Ying, B. W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 4201-4206.

(5) Carra, S.; Alberti, S.; Benesch, J. L. P.; Boelens, W.; Buchner, J.; Carver, J. A.; Cecconi, C.; Ecroyd, H.; Gusev, N.; Hightower, L. E., et al. Small heat shock proteins: Multifaceted proteins with important implications for life. *Cell Stress Chaperones* **2019,** *24*, 295-308.

(6) Dahiya, V.; Buchner, J. Functional principles and regulation of molecular chaperones. In *Molecular chaperones in human disorders*, Donev, R., Ed.; 2019; Vol. 114, pp 1-60.

(7) Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **2013,** *82*, 323-355.

(8) Rosenzweig, R.; Nillegoda, N. B.; Mayer, M. P.; Bukau, B. The Hsp70 chaperone network. *Nat. Rev. Mol. Cell Biol.* **2019,** *20*, 665-680.

(9) Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., Bukau, B. Trigger factor and DnaK cooperate in folding newly synthesized proteins. *Nature* **1999,** *400*, 693-696.

(10) Koubek, J.; Schmitt, J.; Galmozzi, C. V.; Kramer, G. Mechanisms of cotranslational protein maturation in bacteria. *Front. Mol. Biosci.* **2021,** *8*.

(11) Teter, S. A.; Houry, W. A.; Ang, D.; Tradler, T.; Rockabrand, D.; Fischer, G.; Blum, P.; Georgopoulos, C.; Hartl, F. U. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **1999,** *97*, 755-765.

(12) Genevaux, P.; Keppel, F.; Schwager, F.; Langendijk-Genevaux, P. S.; Hartl, F. U.; Georgopoulos, C. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* **2004,** *5*, 195-200.

(13) Tripathi, A.; Gupta, K.; Khare, S.; Jain, P. C.; Patel, S.; Kumar, P.; Pulianmackal, A. J.; Aghera, N.; Varadarajan, R. Molecular determinants of mutant phenotypes, inferred from saturation mutagenesis data. *Mol. Biol. Evol.* **2016,** *33*, 2960-2975.

(14) Tripathi, A.; Swaroop, S.; Varadarajan, R. Molecular determinants of temperature-sensitive phenotypes. *Biochemistry* **2019,** *58*, 1738-1750.

(15) Aguilar-Rodríguez, J.; Sabater-Muñoz, B.; Montagud-Martínez, R.; Berlanga, V.; Alvarez-Ponce, D.; Wagner, A.; Fares, M. A. The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol. Evol.* **2016,** *8*, 2979-2991.

(16) Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U. DnaK functions as a central hub in the e. coli chaperone network. *Cell Rep.* **2012,** *1*, 251-264.

(17) Kadibalban, A. S.; Bogumil, D.; Landan, G.; Dagan, T. DnaK-dependent accelerated evolutionary rate in prokaryotes. *Genome Biol. Evol.* **2016,** *8*, 1590-1599.

(18) Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H. Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 8937.

(19) Fatima, K.; Naqvi, F.; Younas, H. A review: Molecular chaperone-mediated folding, unfolding and disaggregation of expressed recombinant proteins. *Cell Biochem. Biophys.* **2021,** *79*, 153-174.

(20) Ario de, M.; Laszlo, V.; Sophia, D.; Pierre, G. Native folding of aggregation-prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol–overexpressed molecular chaperones. *Cell Stress Chaperones* **2005,** *10*, 329-339.

(21) Haacke, A.; Fendrich, G.; Ramage, P.; Geiser, M. Chaperone over-expression in Escherichia coli: Apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. *Protein Expr. Purif.* **2009,** *64*, 185-193.

(22) Cascella, R.; Bigi, A.; Cremades, N.; Cecchi, C. Effects of oligomer toxicity, fibril toxicity and fibril spreading in synucleinopathies. *Cell. Mol. Life Sci.* **2022,** *79*, 174.

(23) Cline, E. N.; Bicca, M. A.; Viola, K. L.; Klein, W. L. The amyloid-β oligomer hypothesis: Beginning of the third decade. *J. Alzheimer's Dis.* **2018,** *64*, S567-S610.

(24) Mallucci, G. R.; Klenerman, D.; Rubinsztein, D. C. Developing therapies for neurodegenerative disorders: Insights from protein aggregation and cellular stress responses. *Annu. Rev. Cell Dev. Biol.* **2020,** *36*, 165-189.

(25) Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S. In vitro expression and characterization of native apomyoglobin under low molecular crowding conditions. *Protein Expr. Purif.* **2006,** *45*, 381-392.

(26) Breslow, E.; Beychok, S.; Hardman, K. D.; Gurd, F. R. N. Relative conformations of sperm whale metmyoglobin and apomyoglobin in solution. *J. Biol. Chem.* **1965,** *240*, 304-&.

(27) Cavagnero, S., Nishimura, C., Schwarzinger, S., Dyson, H.J., Wright, P.E. Conformation and dynamic characterization of the molten globule state of an apomyoglobin mutant with an altered folding pathway. *Biochemistry* **2001,** *40*, 14459-14467.

(28) Cavagnero, S.; Dyson, H. J.; Wright, P. E. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. *J. Mol. Biol.* **1999,** *285*, 269-282.

(29) Eliezer, D.; Jennings, P. A.; Dyson, H. J.; Wright, P. E. Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR. *FEBS Lett.* **1997,** *417*, 92-96.

(30) Jungbauer, L. M.; Bakke, C. K.; Cavagnero, S. Experimental and computational analysis of translation products in apomyoglobin expression. *J. Mol. Biol.* **2006,** *357*, 1121-1143.

(31) Postnikova, G. B.; Yumakova, E. M. Fluorescence study of the conformational properties of myoglobin structure .3. Ph-dependent changes in porphyrin and tryptophan fluorescence of the complex of sperm whale apomyoglobin with protoporphyrin-ix - the role of the porphyrin macrocycle and iron in formation of native myoglobin structure. *Eur. J. Biochem.* **1991,** *198*, 241-246.

(32) Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H.J., Wright, P.E. Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* **1999,** *8*, 45-49.

(33) Hesterkamp, T.; Bukau, B. Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in e-coli. *EMBO J.* **1998,** *17*, 4818-4828.

(34) Lecomte, J. T. J.; Vuletich, D. A.; Lesk, A. M. Structural divergence and distant relationships in proteins: Evolution of the globins. *Curr. Opin. Struct. Biol.* **2005,** *15*, 290-301.

(35) Vinogradov, S. N.; Hoogewijs, D.; Bailly, X.; Arredondo-Peter, R.; Gough, J.; Dewilde, S.; Moens, L.; Vanfleteren, J. R. A phylogenomic profile of globins. *BMC Evol. Biol.* **2006,** *6*.

(36) Wu, G. H.; Wainwright, L. M.; Poole, R. K. Microbial globins. In *Advances in microbial physiology, vol 47*, Poole, R. K., Ed.; 2003; Vol. 47, pp 255-310.

(37) Lecomte, J. T. J.; Sukits, S. F.; Bhattacharya, S.; Falzone, C. J. Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci.* **1999,** *8*, 1484-1491.

(38) Hutchinson, R. B.; Chen, X.; Zhou, N. K.; Cavagnero, S. Fluorescence anisotropy decays and microscale-volume viscometry reveal the compaction of ribosome-bound nascent proteins. *J. Phys. Chem. B* **2021,** *125*, 6543-6558.

(39) Yaeger-Weiss, S. K.; Jennaro, T. S.; Mecha, M.; Becker, J. H.; Yang, H. M.; Winkler, G. L. W.; Cavagnero, S. Net charge and nonpolar content guide the identification of folded and prion proteins. *Biochemistry* **2020,** *59*, 1881-1895.

(40) Springer, B. A.; Sligar, S. G. High-level expression of sperm whale myoglobin in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1987,** *84*, 8961-8965.

(41) Barrick, D.; Baldwin, R. L. 3-state analysis of sperm whale apomyoglobin folding. *Biochemistry* **1993,** *32*, 3790-3796.

(42) Shen, L. L.; Hermans, J. Kinetics of conformation change of sperm-whale myoglobin. 3. Folding and unfolding of apomyoglobin and suggested overall mechanism. *Biochemistry* **1972,** *11*, 1845-&.

(43) Dyson, H. J.; Wright, P. E. How does your protein fold? Elucidating the apomyoglobin folding pathway. *Acc. Chem. Res.* **2017,** *50*, 105-111.

(44) Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H.J., Wright, P.E. Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. *Biochemistry* **2000,** *39*, 11227-11237.

(45) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pthway of apomyoglobin. *Science* **1993,** *262*, 892-896.

(46) Cavagnero, S.; Kurt, N. Folding and misfolding as a function of polypeptide chain elongation: Conformational trends and implications for intracellular events. In *Misbehaving proteins: Protein (mis)folding, aggregation and stability*, Murphy, R.; Tsai, A., Eds.; Kluwer Academic Press: 2006; pp 217-246.

(47) Chow, C. C.; Chow, C.; Raghunathan, V.; Huppert, T. J.; Kimball, E. B.; Cavagnero, S. Chain length dependence of apomyoglobin folding: Structural evolution from misfolded sheets to native helices. *Biochemistry* **2003,** *42*, 7090-9.

(48) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(49) Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci.* **2009,** *18*, 2003-2015.

(50) Jungbauer, L. M., Bakke, C., Cavagnero S. Experimental and computational analysis of translation products in apomyoglobin expression *J. Mol. Biol.* **2006,** *357*, 1121-1143.

(51) Kurt, N.; Cavagnero, S. Nonnative helical motif in a chaperone-bound protein fragment. *Biophys. J.* **2008,** *94*, L48-L50.

(52) 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.

(53) Chow, C.; Kurt, N.; Murphy, R. M.; Cavagnero, S. Structural characterization of apomyoglobin self-associated species in aqueous buffer and urea solution. *Biophys. J.* **2006,** *90*, 298-309.

(54) Jennings, P. A.; Stone, M. J.; Wright, P. E. Overexpression of myoglobin and assignment of its amide, C-alpha and C-beta resonances. *J. Biomol. NMR* **1995,** *6*, 271-276.

(55) Vega, C. A.; Kurt, N.; Chen, Z.; Rüdiger, S.; Cavagnero, S. Binding specificity of an α-helical protein sequence to a full-length Hsp70 chaperone and its minimal substrate-binding domain. *Biochemistry* **2006,** *45*, 13835-13846.

(56) Kurt, N.; Rajagopalan, S.; Cavagnero, S. Effect of Hsp70 chaperone on the folding and misfolding of polypeptides modeling an elongating protein chain. *J. Mol. Biol.* **2006,** *355*, 809-820.

(57) Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S. *In vitro* expression and characterization of native apomyoglobin under low molecular crowding conditions. *Prot. Expr. Purif.* **2006,** *45*, 381-392.

(58) Spudich, G. M.; Miller, E. J.; Marqusee, S. Destabilization of the Escherichia coli RNase H kinetic intermediate: Switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* **2004,** *335*, 609-618.

(59) 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*, 274-82.

(60) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **2005,** *310*, 827-34.

(61) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. Structure of the 70s ribosome complexed with mRNA and tRNA. *Science* **2006,** *313*, 1935-1942.

(62) 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.

(63) Malkin, L. I.; Rich, A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J. Mol. Biol.* **1967,** *26*, 329-346.

(64) Guzman-Luna, V.; Fuchs, A. M.; Allen, A. J.; Staikos, A.; Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Commun. Biol.* **2021,** *4*.

(65) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* **1986,** *192*, 133-154.

(66) Rüdiger, S.; Germeroth, L.; Schneider-Mergener, J.; Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **1997,** *16*, 1501-1507.

(67) Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F. U.; Kerner, M. J.; Frishman, D. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **2008,** *9*.

(68) Kaiser, C. M.; Chang, H. C.; Agashe, V. R.; Lakshmipathy, S. K.; Etchells, S. A.; Hayer-Hartl, M.; Hartl, F. U.; Barral, J. M. Real-time observation of trigger factor function on translating ribosomes. *Nature* **2006,** *444*, 455-460.

(69) Beck, K.; Wu, L. F.; Brunner, J.; Muller, M. Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *EMBO J.* **2000,** *19*, 134-143.

(70) De Geyter, J.; Portaliou, A. G.; Srinivasu, B.; Krishnamurthy, S.; Economou, A.; Karamanou, S. Trigger factor is a bona fide secretory pathway chaperone that interacts with SecB and the translocase. *EMBO Rep.* **2020,** *21*.

(71) Liu, C. P.; Perrett, S.; Zhou, J. M. Dimeric trigger factor stably binds folding-competent intermediates and cooperates with the DnaK-DnaJ-GrpE chaperone system to allow refolding. *J. Biol. Chem.* **2005,** *280*, 13315-13320.

(72) Yang, F.; Chen, T. Y.; Krzeminski, L.; Santiago, A. G.; Jung, W.; Chen, P. Single-molecule dynamics of the molecular chaperone trigger factor in living cells. *Mol. Microbiol.* **2016,** *102*, 992-1003.

(73) Jameson, D. M. *Introduction to fluorescence*. CRC Press: 2014.

(74) Jameson, D. M.; Gratton, E.; Hall, R. D. The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry. *Appl. Spectrosc. Rev.* **1984,** *20*, 55-106.

(75) Ross, J. A.; Jameson, D. M. Time-resolved methods in biophysics. 8. Frequency domain fluorometry: Applications to intrinsic protein fluorescence. *Photochem. Photobiol. Sci.* **2008,** *7*, 1301-1312.

(76) Weinreis, S. A.; Ellis, J. P.; Cavagnero, S. Dynamic fluorescence depolarization: A powerful tool to explore protein folding on the ribosome. *Methods* **2010,** *52*, 57-73.

(77) Dalphin, M. D.; Stangl, A. J.; Liu, Y.; Cavagnero, S. KLR-70: A novel cationic inhibitor of the bacterial Hsp70 chaperone. *Biochemistry* **2020**, doi: 10.1021/acs.biochem.0c00320.

(78) Ziehr, D. R.; Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Production of ribosome-released nascent proteins with optimal physical properties. *Anal. Chem.* **2010,** *82*, 4637-4643.

(79) Lakowicz, J. R.; Cherek, H.; Laczko, G.; Gratton, E. Time-resolved fluorescence emission-spectra of labeled phospholipid vesicles as observed using multi-frequency phase modulation fluorometry. *Biochim. Biophys. Acta* **1984,** *777*, 183-193.

(80) Bosson, J.; Gouin, J.; Lacour, J. Cationic triangulenes and helicenes: Synthesis, chemical stability, optical properties and extended applications of these unusual dyes. *Chem. Soc. Rev.* **2014,** *43*, 2824-2840.

(81) Maliwal, B. P.; Fudala, R.; Raut, S.; Kokate, R.; Sørensen, T. J.; Laursen, B. W.; Gryczynski, Z.; Gryczynski, I. Long-lived bright red emitting azaoxa-triangulenium fluorophores. *PLoS ONE* **2013,** *8*, e63043.

(82) Ozawa, K.; Headlam, M. J.; Schaeffer, P. M.; Henderson, B. R.; Dixon, N. E.; Otting, G. Optimization of an Escherichia coli system for cell-free synthesis of selectively n-labelled proteins for rapid analysis by NMR spectroscopy. *Eur. J. Biochem.* **2004,** *271*, 4084-93.

(83) Eliezer, D.; Wright, P. E. Is apomyoglobin a molten globule? Structural characterization by NMR. *J. Mol. Biol.* **1996,** *263*, 531-8.

(84) Štefl, M.; James, N. G.; Ross, J. A.; Jameson, D. M. Applications of phasors to in vitro time-resolved fluorescence measurements. *Anal. Biochem.* **2011,** *410*, 62-69.

(85) James, N. G.; Ross, J. A.; Štefl, M.; Jameson, D. M. Applications of phasor plots to in vitro protein studies. *Anal. Biochem.* **2011,** *410*, 70-76.

(86) Komar, A. A.; Kommer, A.; Krasheninnikov, I. A.; Spirin, A. S. Cotranslational heme binding to nascent globin chains. *FEBS Lett.* **1993,** *326*, 261-263.

(87) Komar, A. A.; Kommer, A.; Krasheninnikov, I. A.; Spirin, A. S. Cotranslational folding of globin. *J. Biol. Chem.* **1997,** *272*, 10646-10651.

(88) Knight, A. M.; Culviner, P. H.; Kurt-Yilmaz, N. e.; Zou, T.; Ozkan, S. B.; Cavagnero, S. Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **2013,** *8*, 1195-1204.

(89) Szmacinski, H.; Jayaweera, R.; Cherek, H.; Lakowicz, J. R. Demonstration of an associated anisotropy decay by frequency-domain fluorometry. *Biophys. Chem.* **1987,** *27*, 233-241.

(90) Hazlett, T. L.; Jameson, D. M. In *Time-resolved fluorescence studies on components of the prokaryotic protein elongation system*, Time-Resolved Laser Spectroscopy in Biochemistry, International Society for Optics and Photonics: 1988; pp 412-419.

(91) Bialik, C. N.; Wolf, B.; Rachofsky, E. L.; Ross, J. A.; Laws, W. R. Dynamics of biomolecules: Assignment of local motions by fluorescence anisotropy decay. *Biophys. J.* **1998,** *75*, 2564-2573.

(92) Klaips, C. L.; Jayaraj, G. G.; Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* **2018,** *217*, 51-63.

(93) Muchowski, P. J.; Schaffar, G.; Sittler, A.; Wanker, E. E.; Hayer-Hartl, M. K.; Hartl, F. U. Hsp70 and Hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 7841-7846.

(94) Sittler, A.; Lurz, R.; Lueder, G.; Priller, J.; Hayer-Hartl, M. K.; Hartl, F. U.; Lehrach, H.; Wanker, E. E. Geldanamycin activates a heat shock response and inhibits huntingtin aggregation in a cell culture model of huntington's disease. *Hum. Mol. Genet.* **2001,** *10*, 1307-1315.

(95) Nagy, M.; Fenton, W. A.; Li, D.; Furtak, K.; Horwich, A. L. Extended survival of misfolded G85R SOD1-linked ALS mice by transgenic expression of chaperone Hsp110. *Proc. Natl. Acad. Sci. USA* **2016,** *113*, 5424-5428.

(96) Chaudhury, S.; Keegan, B. M.; Blagg, B. S. J. The role and therapeutic potential of Hsp90, Hsp70, and smaller heat shock proteins in peripheral and central neuropathies. *Med. Res. Rev.* **2021,** *41*, 202-222.

(97) Klucken, J.; Shin, Y.; Masliah, E.; Hyman, B. T.; McLean, P. J. Hsp70 reduces alpha-synuclein aggregation and toxicity. *J. Biol. Chem.* **2004,** *279*, 25497-25502.

(98) Shen, H. Y.; He, J. C.; Wang, Y. M.; Huang, Q. Y.; Chen, J. F. Geldanamycin induces heat shock protein 70 and protects against mptp-induced dopaminergic neurotoxicity in mice. *J. Biol. Chem.* **2005,** *280*, 39962-39969.

(99) Flower, T. R.; Chesnokova, L. S.; Froelich, C. A.; Dixon, C.; Witt, S. N. Heat shock prevents alpha-synuclein-induced apoptosis in a yeast model of Parkinson's disease. *J. Mol. Biol.* **2005,** *351*, 1081-1100.

(100) Kirchdoerfer, R. N.; Huang, J. J.; Isola, M. K.; Cavagnero, S. Fluorescence-based analysis of aminoacyl- and peptidyl-tRNA by low-pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* **2007,** *364*, 92-4.

(101) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012,** *9*, 671-675.

(102) Abràmoff, M. D.; Magalhães, P. J.; Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **2004,** *11*, 36-42.

4.8 Appendix

Description of Supporting Information

Supporting Information. Evidence for kinetic trapping of soluble and insoluble states; Additional experiments aiming at further characterization of soluble aggregates; Estimation of the concentration of TF; Concentration of apoMb in cell-free system; Reprocessing anisotropy data reported in Ziehr *et al*; Experimental volume of apoMb samples; Statistical p values for solubility assays; Statistical p values for rotational correlation times; Statistical p values for lifetimes; Fluorescence emission spectra for BODIPY-FL-labeled WT apoMb; Phasor plot of apoMb made in the presence of Hsp70 + Hsp70 inhib. added post-translationally; The following files are available free of charge.

Supporting Text

Evidence for kinetic trapping of soluble and insoluble states. Figure 4-5b shows that upon incubation of the insoluble fraction of the cell free reaction, there is no significant change in pellet concentration for any of the samples. This result provides evidence that the soluble and insoluble states are kinetically trapped from each other. Here, we further explain why this experimental result demonstrates kinetic trapping. The approximate apoMb protein concentration in the cell-free system in 300 nM, and the percent of soluble M131D protein in the absence of chaperones is 60% (Figure 4-4), so the concentration of the soluble fraction is approximately 180 nM. As shown in Figure 4-8c, WT apoMb generated in the absence of chaperones contains 76% soluble aggregates. If we assume M131D expressed in the absence of chaperones contains a similar percentage of soluble aggregates, then the concentration of soluble aggregates for M131D apoMb is approximately 137 nM. This sample also contains insoluble aggregates, demonstrating that the sample concentration is higher than the critical concentration for insoluble

aggregation, assuming that soluble and insoluble aggregate states can interconvert, and their relative concentrations are governed by thermodynamic equilibrium. If this sample is governed by thermodynamic equilibrium, then 137 nM is the critical concentration for insoluble aggregation, which is maximum possible concentration of soluble aggregates. Any additional protein added to the solution will become insoluble if the soluble and insoluble fractions are in thermodynamic equilibrium. The number of moles of protein in the insoluble fraction is calculated from the percent solubility data (Figure 4-4), assuming a 15 μL initial sample volume. Using number of moles in the insoluble fraction, and the critical concentration (137 nM), the volume required to fully resuspend the pellet and reach the critical concentration was calculated to be approximately 13 μL. This is the minimum volume that would dissolve the entirety of the pellet. Assuming the I28D mutant behaves similarly, the volume required to completely dissolve the I28D pellet would be the same or less than 13 μL. In the experiments shown in figure 5, the pellets were dissolved in volumes of either 15 μL or 150 μL, and after incubation and centrifugation, there was still protein present in the insoluble fraction (Figures 5b and 5c). If the soluble and insoluble states were in thermodynamic equilibrium, the concentration of resuspended pellet would be below the critical concentration for insoluble aggregate formation. The presence of a pellet even at large dilution volumes and upon incubation for 14 hr demonstrates that the soluble and insoluble fractions are kinetically trapped.

Additional experiments aiming at further characterization of soluble aggregates. Time-resolved fluorescence anisotropy and lifetime experiments revealed the presence of soluble states that have a large τ_c, have a different fluorescence lifetime than native apoMb, and are kinetically trapped from the native state. However, the fluorescence experiments cannot distinguish between homo-aggregates (multiple apoMb proteins interacting with each other) or a hetero-aggregate

(apoMb interacting with the ribosome or other components within the cell-free system). Several additional experiments were carried out to further characterize the nature of the soluble aggregate, but these experiments were unable to provide additional information about the nature of these soluble aggregates. The results from these experiments are summarized below and in Table 4-S6. The experimental methods are described in the supplementary methods.

Centrifugation to remove ribosomes. After synthesizing WT apoMb, ultracentrifugation was used to remove ribosomes from the cell-free system.[1] The purpose of removing ribosomes was to compare the τ_c of apoMb in the presence and absence of ribosomes to determine if released apoMb interacts with ribosomes in the cell-free system. Based on the absorbance at 260 nm, approximately 56% of ribosomes remained in the supernatant after centrifugation. However, the solution likely contains other species that absorb at 260 nm (e.g. DNA, RNA, protein). Therefore, this method may not be a reliable method to determine whether the ribosomes are removed from the solution. Additionally, centrifugation removed 61 ± 6% of protein from the solution (Figure 4-S4). As a result, the protein concentration was too low to give sufficient signal for accurate fluorescence anisotropy decay measurements.

Immunoprecipitation to remove ribosomes. A second method to remove ribosomes from the solution was immunoprecipitation with antibodies for ribosomal proteins L23 and L29. We tested this method with RNCs and used SDS-PAGE analysis to compare the RNC concentration before and after immunoprecipitation. As shown in Figure 4-S5, immunoprecipitation did not decrease the RNC concentration. This result was confirmed by quantifying the intensity of the gel bands with ImageJ. Therefore, this method was not effective for removing ribosomes from the solution.

Crosslinking of WT apoMb released from purified ribosome-bound nascent chains (RNCs).

Chemical crosslinking was used to identify interactions between multiple wt apoMb proteins and between apoMb and other ribosomal proteins present after purification of RNCs. The crosslinker 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC) was used because it is water soluble, it is sensitive to intra- and inter-molecular interactions in proteins, and it has previously been used to identify nascent-chain interactions.[2-4] There was no evidence of crosslinking occurring with WT apoMb released from purified RNCs (Figure 4-S6). Therefore, apoMb released in the presence of chaperones does not show soluble aggregates that are sensitive to crosslinking. The interactions that give rise to the large τ_c values in the fluorescence anisotropy decay experiments may not be sensitive to crosslinking because they may be transient and occur on timescales faster than the reaction time for the EDC crosslinker (<6 s).[3] Additionally, the species generated from release from stalled ribosomes may differ from the species generated when apoMb is synthesized normally without stalling the ribosome. Therefore, we then repeated the crosslinking experiment in cell-free systems containing different chaperone concentrations.

Crosslinking of ribosome-released WT apoMb in the cell-free systems containing different chaperone concentrations. As shown in Figure 4-S7, all apoMb samples made with different Hsp70 concentrations showed similar levels of crosslinking. The percent of WT apoMb that underwent crosslinking was quantified with eq. 1:

$$\text{Percent crosslinked apoMb} = 1 - \frac{I_{\text{apoMb with EDC}}}{I_{\text{apoMb without EDC}}} \quad (1)$$

The percent crosslinked apoMb is not statistically different when apoMb is made with different concentrations of Hsp70 (Figure 4-S7c). This result differs from the fluorescence anisotropy decay results which showed different τ_c values when apoMb was made in the presence and absence of chaperones. One possible explanation for this difference is that the interactions that

give rise to the large τ_c values in the fluorescence anisotropy decay experiments may occur on timescales faster than the reaction time for the EDC crosslinker (<6 s)[3] and thus may not be detected with these experiments. Alternatively, there may be crosslinking-sensitive interactions other than the soluble aggregates observed with fluorescence anisotropy decays, such as interactions between apoMb and chaperones. These additional crosslinked species could cause the percent of crosslinked apoMb to be the same in samples with different concentrations of soluble aggregates. Interestingly, EDC-crosslinked samples with different chaperone concentrations all showed a large molecular weight species (>500 kDa) that was not able to run through the gel. This species may include the soluble aggregates detected with fluorescence anisotropy decay as well as other crosslinked species.

Mass photometry. Mass photometry was used to characterize the masses of apoMb species made in the presence of absence of chaperones. Mass photometry measures the light scattering of particles in a solution. The degree of scattering is proportional to the particle's mass, so the scattering signal is then converted to molecular mass of the species in the sample. This method does not require fluorescent labels, and it can measure the masses of single molecules, making it a powerful approach for resolving multiple species and populations within a sample.[5] Additionally, it has been previously used to assess protein-protein interactions.[6-7] The goal of this experiment was to identify if apoMb made in the absence of chaperones contained species with different masses than the apoMb species made in the presence of chaperones. The presence of larger molecular weight species could indicate the presence of soluble aggregates. However, WT apoMb made and released in the presence and absence of chaperones showed peaks in the mass photometry spectrum with similar masses and relative abundance (Figure 4-S8). Therefore, we were unable to detect soluble aggregates with this method. The soluble aggregates may have a

mass similar to the Hsp70 chaperone (Figure 4-S9) or other components in the cell-free system (Figure 4-S8). Therefore, peaks from soluble aggregates may be difficult to identify and resolve from peaks present from other species in the sample. ApoMb samples analyzed by mass photometry were prepared by releasing the protein from purified RNCs, to reduce the number of non-apoMb species present in the sample. However, the apoMb generated from release from stalled ribosomes may have different conformations than apoMb that is synthesized and released without stalling the ribosome.

Supporting Methods

Estimation of the concentration of TF. The concentration of TF in WT cell free reactions was estimated as 0.2 μM, by consideration of calculated free and ribosome-bound TF concentrations. *Ellis et al.,* determined the concentration of TF (90 nM) present in purified RNC samples derived from the WT cell free system.[8] The authors observed a nascent protein chain-length dependence to the amount of TF isolated upon sucrose cushion-purification of nascent chains. The above observation implies that most TF isolated is RNC-bound when pulled down during the ultracentrifugation-based RNC purification procedure[8]. After correcting for reaction volume, the value determined by *Ellis et al,* was therefore used as an estimate for the concentration of ribosome-bound TF (18 nM) within the unpurified WT cell free system.

An estimate for the concentration of free TF in the WT cell free system was determined by considering the equilibrium dissociation constant between TF and nascent chains.[9] The concentration of total RNCs produced in the WT cell free system was determined independently and corrected by cell extract volume (31 nM). The concentration of free RNCs (13 nM) unbound

by TF was determined by taking the difference between the total RNC concentration and the concentration of RNCs bound to TF. A dissociation constant (K_D = 138 nM)[9] for TF and nascent chains was used to calculate the concentration of free TF in the unpurified cell free system (191 nM). The sum of free TF and TF bound to RNCs yields an approximate total TF concentration of 0.2 μM.

Concentration of apoMb in cell-free system. The concentration of apoMb produced in the cell-free system was previously determined to be 300 nM with western blotting.[10] The standard error in total apoMb fluorescence intensity (soluble plus insoluble fractions) in the SDS-PAGE gels for solubility experiments (Figure 4-4) was used to calculate the uncertainty for the apoMb concentration. The gel band intensity was quantified with ImageJ. The standard error and percent error for the apoMb total fluorescence intensities are listed in Table 4-S7. The average percent error for all three species was 49%. Therefore, the uncertainty for the apoMb concentration was calculated to be 0.15 μM, and the apoMb concentration is reported as 0.3 ± 0.2 μM. This uncertainty was propagated to determine the uncertainty in the chaperone to apoMb protein ratios listed in Figures 4 and 8.

Reprocessing anisotropy data reported in Ziehr *et al.* Ziehr *et al.* measured the fluorescence anisotropy decay of purified apoMb resuspended in a solution of unlabeled apoMb RNCs.[11] When fitting this data in GLOBALS to determine the τ_c of this species, the authors allowed the r_0 value to vary. Their resulting global τ_c for this species was 11 ± 2 ns. For this work, we reprocessed this data, fixing the r_0 value to 0.37. The resulting τ_c was 10 ± 2 ns, which is the value reported in this work.

Calculation of volume of and number of monomers contained in apoMb samples. The hydrated volume of the apoMb samples was calculated from the rotational correlation time,

assuming the protein rotates as an oblate ellipsoid with an axial ratio (ρ) of 0.45 ± 0.04. The oblate ellipsoid shape and axial ratio were chosen based on measurements of the Pymol structure of sperm whale myoglobin (PDB ID: 1mbc). To calculate the hydrated volume, τ_c was first converted to the rotational diffusion coefficient $D_{sphere.}$ Assuming apoMb rotates as an oblate ellipsoid, τ_c is related to D_{sphere} according to eqs. (2-3).[12]

$$\tau_c = (2D_{\parallel} + 4D_{\perp})^{-1} = \left[\frac{3\rho(\rho-b)}{(\rho^2-1)}D_{sphere} + \left(\frac{6\rho[(2\rho^2-1)b-\rho]}{(\rho^4-1)}\right)D_{sphere}\right]^{-1} \quad (2)$$

$$b = (1-\rho^2)^{-\frac{1}{2}}\arctan\left[\frac{(1-\rho^2)^{1/2}}{\rho}\right] \quad (3)$$

The derivation for eq. 2 is reported in Hutchinson *et al*.[13] D_{sphere} was then converted to the hydrated volume (V_h) with eq. 4.

$$V_h = \frac{RT}{6\eta D_{sphere}} \quad (4)$$

Where R is the universal gas constant, T is the temperature (298 K), and η is the solution viscosity.

The experimental volume of the apoMb species was then used to calculate the number of apoMb monomers contained in these species, assuming they are homo-aggregates. V_h for monomeric apoMb was estimated from the measured radii of the Pymol structure of apoMb (PDB ID: 1mbc, radii dimensions = 0.8 ± 0.1 nm and 1.8 ± 0.2 nm). The unhydrated volume was calculated to be 12.0 ± 0.2 nm^3. The hydrated volume was calculated with eq. 5[11]

$$V_h = \bar{v} + \delta h \quad (5)$$

Where $\bar{v}$ is the specific volume of the protein, δ is the specific volume of water which equals $1.0035 \times 10^{-6}\ \frac{m^3_{H_2O}}{g_{H_2O}}$, and h is the specific protein hydration which equals 0.5 $\frac{g_{H_2O}}{g_{protein}}$. The hydrated volume of an apoMb monomer was calculated to be 26 ± 6 nm^3/protein, using the known

molecular weight for apoMb (17,331 Da).[1] The hydrated volume for each sample was divided by 26 to estimate the number of monomers contained in that species. The calculated volumes and number of apoMb monomers contained in each species are listed in SI Table 4-S1.

Centrifugation to remove ribosomes. Full-length WT apoMb ribosome-bound nascent chains (RNCs) were generated as described previously.[8, 14] The RNCs were made with WT S30 and added Hsp70 chaperones (75 μM Hsp70, 12 μM DnaJ, 19 μM GrpE). Oligodeoxynucleotide-directed mRNA cleavage was used to generate stalled RNCs carrying the full-length apoMb protein. RNCs were purified via centrifugation (160,000 rcf for 60 min at 4 °C) over a sucrose cushion (1.1 M sucrose, 20 mM tris−HCl, 10 mM Mg^{2+} acetate, 500 mM NH_4Cl, 0.5 mM EDTA, and 1 mM DL-dithiothreitol, pH adjusted to 7.0). For purification, 15 μL of cell-free reaction was centrifuged over 150 μL of sucrose cushion. Pellets containing RNC complexes were dissolved in 15 μL of resuspension buffer (10 mM Tris HCl, 10 mM Mg(OAc)2, 60 mM NH_4Cl, 0.5 mM EDTA, and 1 mM dithiothreitol, pH 7.0). ApoMb protein was released from RNCs with 1.8 M hydroxylamine (2.5 h, 37 °C, with periodic mixing) and quenched with 20% acetic acid to pH 7.0. The solution was centrifuged a second time for 1 h at 160,000 g to remove ribosomes. The pellet was resuspended with resuspension buffer to the same volume as the supernatant. The supernatant and pellets were analyzed with SDS PAGE (Figure 4-S4) and the ribosome concentration was estimated from the absorbance at 260 nm (A_{260}) using the extinction coefficient for the *E. coli* 70S ribosome ($3.91 \times 10^7\ M^{-1}\ s^{-1}$).[15] The percent of ribosomes removed was calculated with eq. 6:

$$\text{Percent ribosomes in supernatant} = \frac{A_{260,\text{supernatant}}}{A_{260,\text{ supernatant}} + A_{260,\text{ pellet}}} \quad (6)$$

The percent of apoMb protein removed with centrifugation was calculated with eq. 7:

$$\text{Percent ribosomes in supernatant} = \frac{I_{pellet}}{I_{supernatant} + I_{pellet}} \quad (7)$$

Where I_{pellet} and $I_{supernatant}$ are the intensities of the gel bands from the resuspended pellet and supernatant, respectively, quantified with ImageJ. The pellet was resuspended with resuspension buffer to the same volume as the supernatant.

Immunoprecipitation to remove ribosomes. Immunoprecipitation was then tested as method to remove ribosomes from the cell-free system. This method was tested with RNCs to measure the percent of ribosomes removed. Immunoprecipitation was performed with Dynabeads labeled with Protein A (Thermo Fisher). Eight full-length WT apoMb RNC samples were generated, as described previously in this work. One sample was not subjected to immunoprecipitation, and it was flash frozen immediately after protein synthesis was completed. Six samples were subjected to immunoprecipitation with either L23 antibody or L29 antibody. One sample was a negative control and underwent the immunoprecipitation protocol in the absence of antibodies. For each sample, fifty μL of Dynabeads were added to a 1.5 μL Eppendorf tube. The beads were separated from the solution with a magnet, and the supernatant was removed. The tube was removed from the magnet. Thirty μL of either L23 (0.1 μg/μL) or L29 (1 μg/μL) in Ab binding and washing buffer were added to the Dynabeads. For the negative control, 30 uL of Ab binding and washing buffer without antibodies was added. The beads were incubated for 30 min at room temperature. The tube was placed on the magnet and the supernatant was removed. The beads were resuspended in 30 uL of Ab binding and washing buffer. Then, the tube was placed on the magnet and the supernatant was removed. Fifteen uL of RNC sample were added to the tube. The RNCs were incubated with the beads for 1.5 hr either at room temperature, on ice, or at 37 °C. The samples on ice and at room temperature were placed on a gel rocker to shake them during incubation. The sample at 37 °C was mixed every 30 min. The tube was then placed on

the magnet and the supernatant was removed. The supernatants were analyzed with SDS-PAGE to determine the percent of RNCs removed with immunoprecipitation (Figure 4-S5).

Origin of antibodies used in immunoprecipitation experiments. Rabbit anti-uL23 antibody was kindly donated by Shu-ou Shan (California Institute of Technology). The anti-uL23 antibody was generated by GenScript, using the CGKVKRHGQRIGRRS peptide as epitope.[16] We obtained rabbit anti-uL29 antibody with help from Bryan W. Davies (University of Texas-Austin) and Melanie Oakes (University of California, Irvine) who obtained them from Masayasu Nomura (University of Wisconsin-Madison). This antibody was generated using the purified *E. coli* ribosomal protein L29. Previously published studies successfully applied these antibodies.[3, 17]

Crosslinking of WT apoMb released from purified RNCs. Four samples of full-length WT apoMb RNCs were generated, as described previously, in a WT cell-free system containing 75 μM Hsp70, 12 μM DnaJ, and 19 μM GrpE. Samples 1 and 2 were resuspended in resuspension buffer (10 mM Tris HCl, 10 mM $Mg(OAc)_2$, 60 mM NH_4Cl, 0.5 mM EDTA, and 1 mM dithiothreitol, pH 7.0). Samples 3 and 4 were resuspended in cell-free system containing 75 μM Hsp70, 12 μM DnaJ, and 19 μM GrpE. released from ribosomes with 1.8 M hydroxylamine (2.5 hr, 37 °C) and quenched with 20% acetic acid to pH 7.0. Then, samples 2 and 4 were subjected to EDC crosslinking, as described above. Samples were then analyzed with SDS PAGE (Figure 4-S6).

Crosslinking of ribosome-released WT apoMb in the cell-free systems containing different chaperone concentrations. Ribosome-released WT apoMb was generated with three different conditions: 1. No chaperones (Δtig S30, 150 μM Hsp70 inhibitor) 2. endogenous chaperones (WT S30 which contains [Hsp70] = 0.5 μM, [DnaJ] = 0.04 μM, [GrpE] = 0.05 μM, [TF] = 0.2

μM. No additional chaperones were added.) and 3. 75 μM Hsp70 (WT S30 plus 75 μM Hsp70, 12 μM DnaJ, and 19 μM GrpE). All three reactions were incubated for 30 min at 37 °C then placed on ice for 15 min. All three apoMb samples within the crude cell-free system were subjected to EDC crosslinking, as described previously.[3] EDC crosslinker ((1-ethyl-3-(3 dimethylaminopropyl)carbodiimide hydrochloride), Thermo Fisher Scientific) was dissolved with DNase/RNase free water (Corning) to 800 mM and 1M KOH was added to adjust the pH to 6.8–7.0. EDC was added cell-free system samples to a final concentration of 80 mM and the reactions were incubated at 30 °C for 30 min. Crosslinking reactions were promptly quenched with quenching buffer (10X QB: 1 M Tris-HCl pH 7.0, 1 M glycine, and 1 M KOAc) and placed on ice. After quenching, the reactions were analyzed with SDS-PAGE (Figure 4-S7).

Mass photometry. Mass photometry measurements were performed with the Refeyn Two^{MP} mass photometer. Glass coverslips (24 x 50 mm, 1.5H, 170 ± 5 μm) were washed first with MilliQ water, then isopropanol, and then water. Washing was repeated three times, and coverslips were dried with nitrogen gas. Adhesive gaskets (3 mm x 1 mm wells, 3-10 μL well capacity, Grace Bio-Labs) were attached to the coverslip. The coverslip was placed on the instrument objective. Fifteen μL of resuspension buffer were added to the gasket well, and the objective height was adjusted to find the focus. One μL of sample was added to the gasket and the sample was mixed. Final dilution of protein samples listed in Figure 4-S8. A 60 s measurement was recorded. This process was repeated for all samples. Samples included Hsp70 chaperone alone, Δtig S30, WT S30, WT apoMb made in the presence of chaperones (WT S30, 75 μM Hsp70, 12 μM DnaJ, and 19 μM GrpE) and released from purified RNCs in the presence of chaperones, and WT apoMb made in the absence of chaperones (Δtig S30, 390 μM Hsp70 inhibitor[18]) and released from purified RNCs in the absence of chaperones. ApoMb proteins were

released with 1.8 M hydroxylamine. Contrast was converted to mass using a calibration curve generated from measuring a mixture of beta-amylase and bovine thyroglobulin. Note that this calibration curve was not collected on the same day as the apoMb protein experiments, so masses reported here are approximate. Calibrations must be collected on the same experiment day to give accurate sample masses.

Supporting Figures

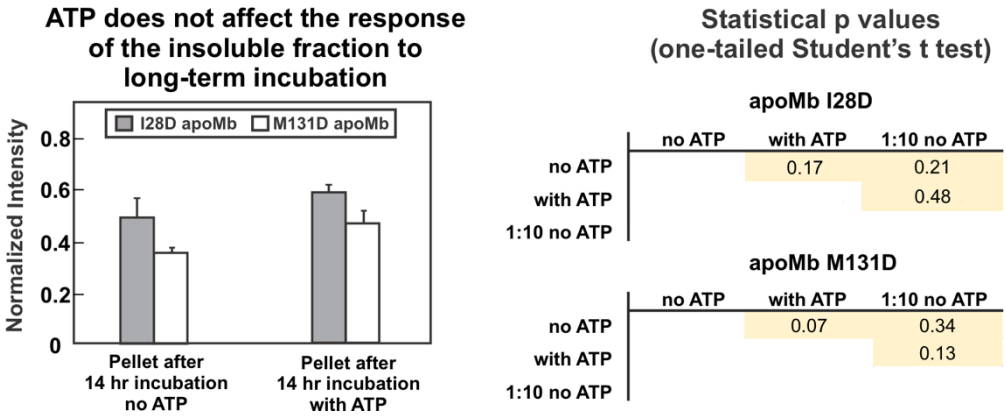


Figure 4-S1. The presence of ATP does not affect kinetic trapping of soluble and insoluble fractions. The fluorescence intensity of the insoluble fraction of apoMb variants after 14 hr incubation is not statistically different in the absence or presence of ATP. These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are statistically the same.

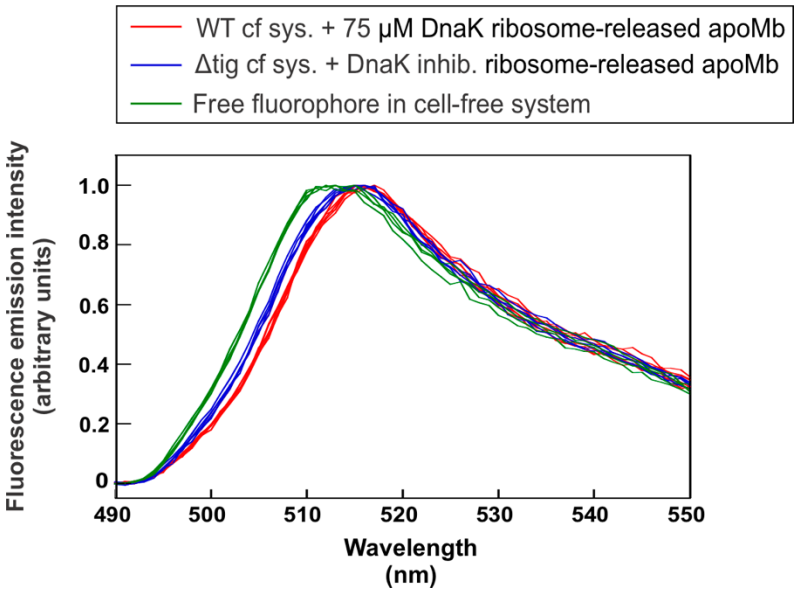


Figure 4-S2. Fluorescence emission spectra for BODIPY-FL labeled WT apoMb produced in the WT cell free system in the presence of 75 μM DnaK (red) or in the Δtig cell free system in the presence of the DnaK inhibitor (blue). There is a minor but highly reproducible blue shift in the absence of chaperones, suggesting that the BODIPY-FL probe is in a more non-polar environment, consistent with soluble aggregation. The green traces show free BODIPY-FL dye in the cell free system. Data is shown for n = 4 − 7 experiments.

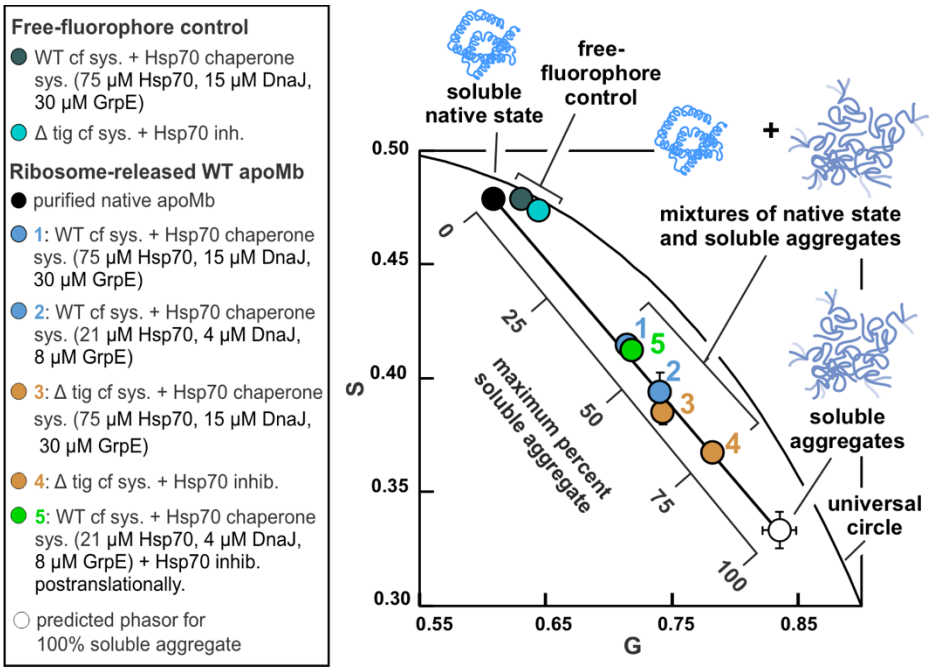


Figure 4-S3. Phasor plot with additional phasor point for WT apoMb made in WT S30 in the presence of 75 μM Hsp70, 12 μM DnaJ, 19 μM GrpE, and 0.2 μM TF. Hsp70 inhibitor[18] (inhib.) (400 μM) and ATP (1 mM) were added postranslationally. The addition of the inhibitor after translation does not significantly shift the phasor for this species, indicating that Hsp70 is likely not remain bound to the protein after it folds.

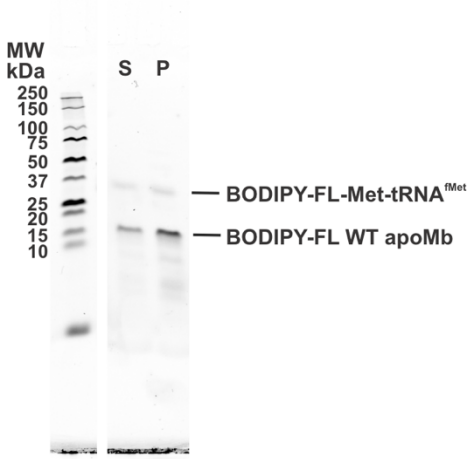


Figure 4-S4. SDS-PAGE results showing WT apoMb protein present in supernatant (S) and resuspended pellet (P) after second centrifugation to remove ribosomes. A 488 nm excitation laser and a 500-550 nm band-pass filter was used to image BODIPY-FL labeled protein samples. A 635 nm laser and a 675 nm long-pass filter were used to image gel markers.

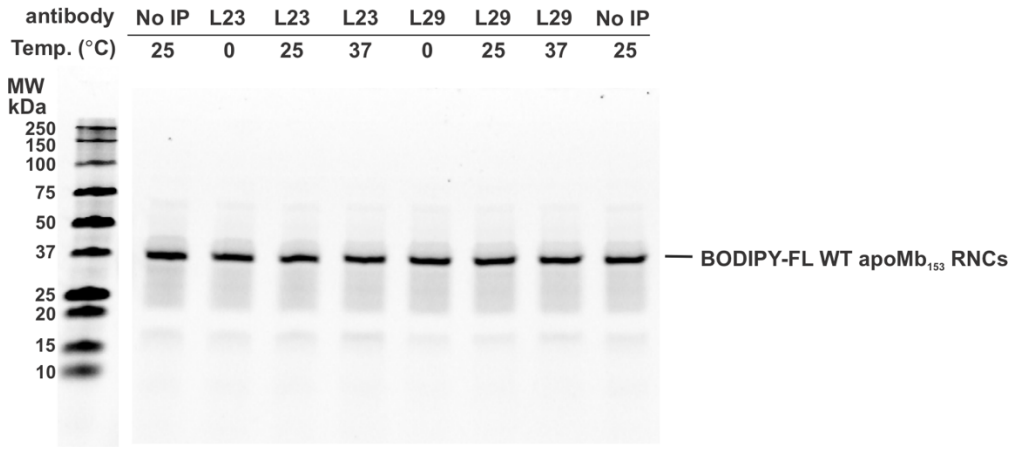


Figure 4-S5. SDS PAGE analysis of supernatant from RNCs after immunoprecipitation (IP) with antibodies for the ribosomal proteins L23 and L29. The BODIPY-FL-labeled apoMb was visualized with a 488 nm excitation laser and 500-550 nm band-pass emission filter (BPB1). Molecular weight markers were visualized with 635 nm excitation laser and a 675 nm long-pass filter (LPR). Immunoprecipitation did not remove RNCs from the solution.

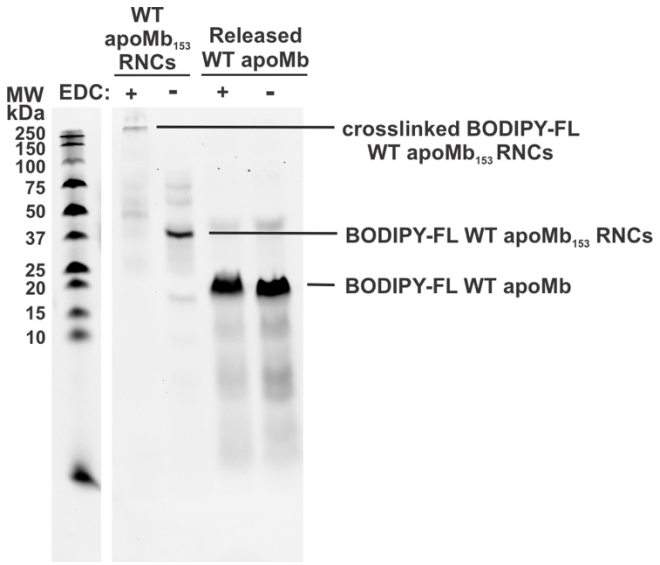


Figure 4-S6. SDS PAGE analysis of EDC crosslinking with ribosome bound and released WT apoMb. The BODIPY-FL-labeled apoMb was visualized with a 488 nm excitation laser and 500-550 nm band-pass emission filter (BPB1). Molecular weight markers were visualized with 635 nm excitation laser and a 675 nm long-pass filter (LPR). Released apoMb was released with hydroxylamine (1.8 M, 2.5 hr, 37 °C) from purified RNCs in the presence of chaperones (75 μM Hsp70, 12 μM DnaJ, and 19 μM GrpE). ApoMb RNCs experience crosslinking. ApoMb released from purified RNCs shows no evidence of crosslinking. Soluble aggregates may not be sensitive to crosslinking because they may be transient and occur on timescales faster than the reaction time for the EDC crosslinker (<6 s).[3]

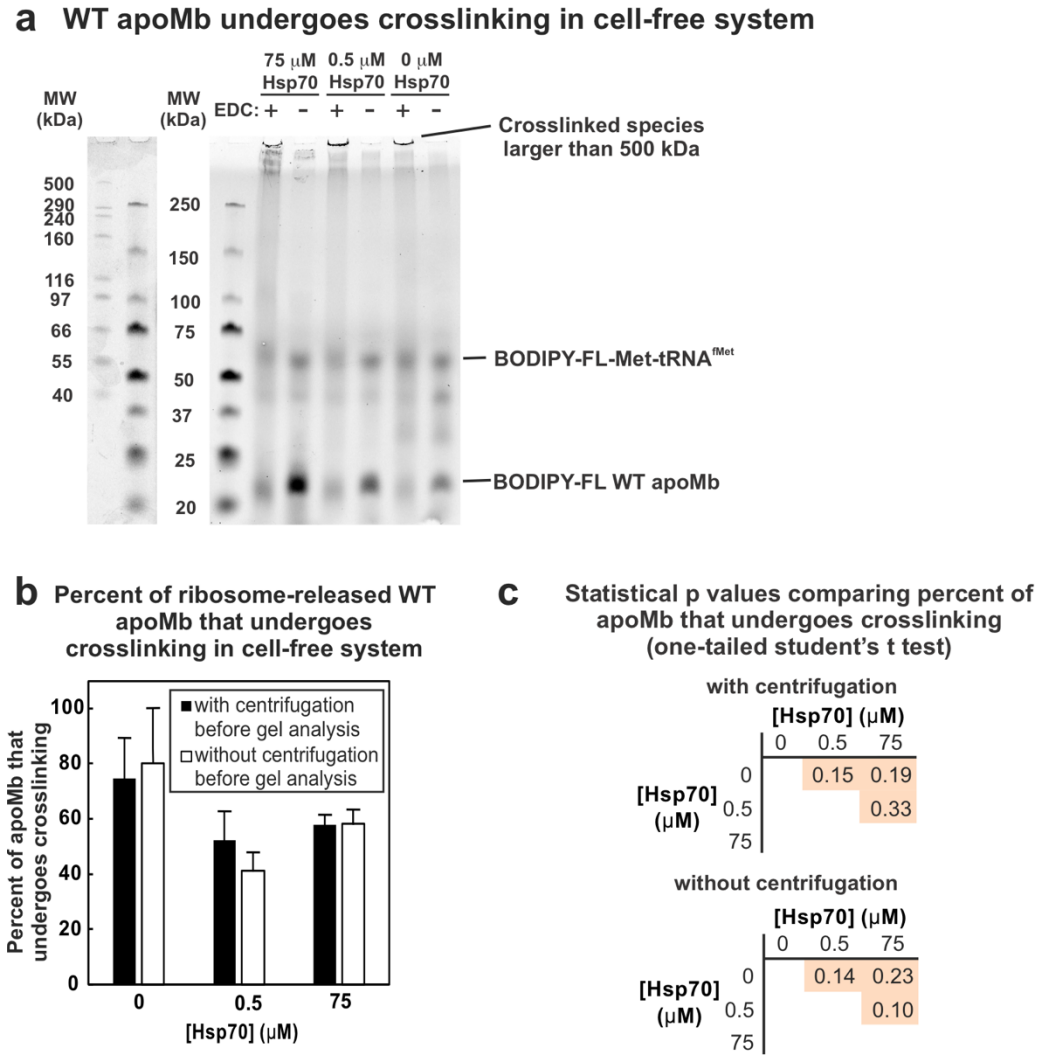


Figure 4-S7. Released WT apoMb undergoes crosslinking in the cell-free system (cf sys.). The sample with 75 μM Hsp70 also had 12 μM DnaJ, 19 μM GrpE, and 0.2 μM TF. The sample with 0.5 μM Hsp70 also had 0.04 μM DnaJ, 0.05 μM GrpE, and 0.2 μM TF. (a) SDS PAGE gel analysis of ribosome-released WT apoMb in cell-free systems with different chaperone concentrations. The BODIPY-FL-labeled apoMb was visualized with a 488 nm excitation laser and 500-550 nm band-pass emission filter (BPB1). Molecular weight markers were visualized with 635 nm excitation laser and a 675 nm long-pass filter (LPR). Gel was imaged with a GE Typhoon fluoroimager. A 488 nm excitation laser and a 500-550 nm band-pass filter was used to

image BODIPY-FL labeled protein samples. A 635 nm laser and a 675 nm long-pass filter was used to image gel markers. (b) ApoMb samples with different chaperone concentrations showed similar levels of crosslinking. (c) Percent of crosslinked apoMb was not statistically different under different chaperone conditions. These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are not statistically different.

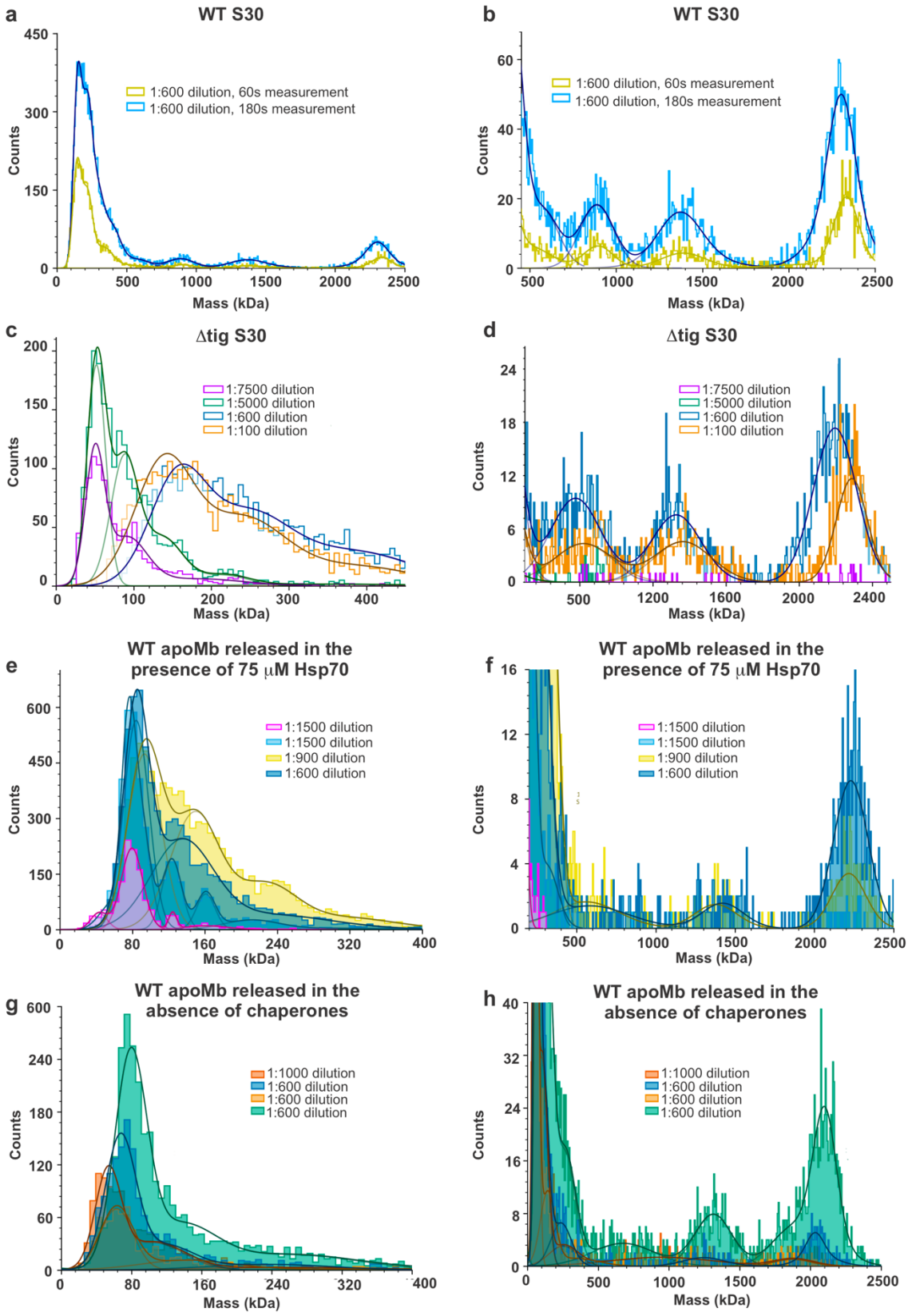

a
WT S30
1:600 dilution, 60s measurement
1:600 dilution, 180s measurement
Counts
Mass (kDa)
b
WT S30
1:600 dilution, 60s measurement
1:600 dilution, 180s measurement
Counts
Mass (kDa)
c
Δtig S30
1:7500 dilution
1:5000 dilution
1:600 dilution
1:100 dilution
Counts
Mass (kDa)
d
Δtig S30
1:7500 dilution
1:5000 dilution
1:600 dilution
1:100 dilution
Counts
Mass (kDa)
e
WT apoMb released in the presence of 75 μM Hsp70
1:1500 dilution
1:1500 dilution
1:900 dilution
1:600 dilution
Counts
Mass (kDa)
f
WT apoMb released in the presence of 75 μM Hsp70
1:1500 dilution
1:1500 dilution
1:900 dilution
1:600 dilution
Counts
Mass (kDa)
g
WT apoMb released in the absence of chaperones
1:1000 dilution
1:600 dilution
1:600 dilution
1:600 dilution
Counts
Mass (kDa)
h
WT apoMb released in the absence of chaperones
1:1000 dilution
1:600 dilution
1:600 dilution
1:600 dilution
Counts
Mass (kDa)

Figure 4-S8. Mass photometry data for cell-free systems (cf sys.) and released WT apoMb made in the presence and absence of chaperones. ApoMb was released from purified full-length RNCs. Panels that are side by side are from the same data but show a different mass range for the x-axis. All data shows were from a 60s measurement time, unless otherwise indicated in the figure. (a) and (b) WT S30 cf system. (c) and (d) Δtig cf sys. (e) and (f) WT apoMb made and released in WT S30 in the presence of 75 μM Hsp70, 12 μM DnaJ, 19 μM GrpE, and 0.2 μM TF. (g) and (h) WT apoMb made and released in Δtig S30 in the presence of 390 μM Hsp70 inhibitor.[18] WT apoMb made in the presence and absence of chaperones shows peaks at similar masses. No peaks could be identified to be soluble aggregates. Peaks from soluble aggregates may overlap with peaks from chaperones, ribosomal subunits, or ribosomal proteins.

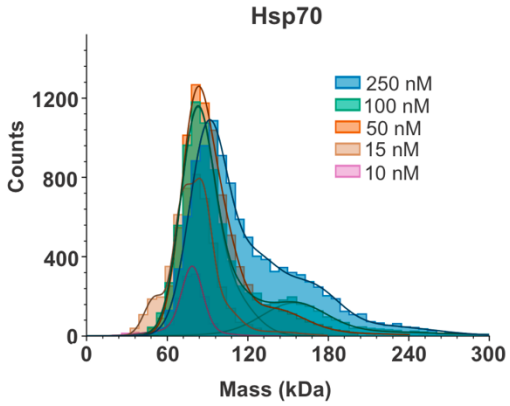


Figure 4-S9. Mass photometry data for different Hsp70 concentrations. Hsp70 was diluted with resuspension buffer.

Supporting Tables

Table 4-S1. Experimental volume of apoMb samples.

System	Rotational correlation time for slow-timescale motion (ns)	Viscosity (mPa-s)[a]	Hydrated volume (nm^3)[b]	Estimated number of apoMb proteins contained in aggregate
Item 1: Δtig cf sys. + Hsp70 inhib., diluted	700 ± 200	0.940 ± 0.005	2400 ± 700	90 ± 30
Item 2: Δtig cell strain + Hsp70 inhib.	260 ± 180	1.13 ± 0.02	700 ± 500	28 ± 21
Item 3: WT cf sys.+ 21μM Hsp70	220 ± 140	1.12 ± 0.02[c]	600 ± 400	24 ± 16
Item 4: Δtig cf sys + 75 uM Hsp70	70 ± 21	1.14 ± 0.01	200 ± 60	7 ± 3
Item 5: WT cf sys.+ 75 μM Hsp70	62 ± 9	1.11 ± 0.01	180 ± 30	7 ± 2
Item 6: WT cf sys. + 75 uM Hsp70 + Hsp70 inhib. added post-translationally	44 ± 1	1.11 ± 0.01[f]	130 ± 20	5 ± 1
Item 7: WT cf sys. + 75 uM Hsp70, diluted	34 ± 2	0.940 ± 0.004	115 ± 8	4 ± 1
Item 8: purified apoMb resuspended in RNC solution	10 ± 2	1.08 ± 0.02[d]	30 ± 3	1.1 ± 0.3
Items 9-10: I28D in WT cf sys.+ 75 μM Hsp70	240 ± 80	1.11 ± 0.01[e]	700 ± 200	30 ± 10

Symbols are as follows: Δtig, strain lacking gene encoding trigger factor (TF) chaperone; WT, wild-type; cf sys., cell-free system; inhib., peptide that inhibits Hsp70[18]

[a]The viscosity values are the average and SE for three independent experiments at shear rates of 1500 s^{-1} and 8000 s^{-1}.

[b]Volume calculated from experimental rotational correlation time and viscosity with eq. 21 which assumes the sample was an oblate ellipsoid.

[c]Viscosity for item 3 predicted from viscosities of items 2 and 5.

[d] Viscosity for purified apoMb resuspended in RNC solution estimated from Hutchinson *et al.*[13]

[e]Viscosity estimated from viscosity of Item 5.

Table 4-S2. Statistical p values for the solubility of apoMb made in the presence of different Hsp70 concentrations (Figure 4-4a). These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are statistically the same (yellow). P values ≤ 0.05 indicate data are statistically different (green).

		Protein construct	wild-type apoMb						I28D apoMb						M131D apoMb					
		[TF]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Protein construct	[TF]	[DnaK]	0.0	0.5	6.5	15.5	35.5	50.5	0.0	0.5	6.5	15.5	35.5	50.5	0.0	0.5	6.5	15.5	35.5	50.5
wild-type apoMb	0.0	0.0		0.31	0.48	0.44	0.41	0.44	0.00	0.01	0.04	0.03	0.14	0.01	0.00	0.04	0.01	0.04	0.03	0.05
	0.0	0.5			0.32	0.29	0.41	0.29	0.01	0.01	0.07	0.07	0.17	0.09	0.00	0.05	0.05	0.21	0.15	0.19
	0.0	6.5				0.43	0.42	0.44	0.00	0.01	0.04	0.03	0.14	0.01	0.00	0.04	0.01	0.05	0.03	0.05
	0.0	15.5					0.38	0.49	0.00	0.01	0.04	0.02	0.14	0.01	0.00	0.04	0.01	0.04	0.03	0.04
	0.0	35.5						0.38	0.01	0.01	0.05	0.04	0.15	0.06	0.00	0.04	0.03	0.14	0.09	0.12
	0.0	50.5							0.00	0.01	0.04	0.03	0.14	0.01	0.00	0.04	0.01	0.03	0.03	0.04
I28D apoMb	0.0	0.0								0.31	0.08	0.05	0.26	0.03	0.12	0.31	0.06	0.01	0.02	0.01
	0.0	0.5									0.04	0.03	0.18	0.02	0.23	0.18	0.03	0.02	0.02	0.02
	0.0	6.5										0.44	0.41	0.26	0.01	0.22	0.48	0.13	0.18	0.15
	0.0	15.5											0.38	0.29	0.01	0.18	0.44	0.12	0.18	0.15
	0.0	35.5												0.31	0.11	0.38	0.40	0.23	0.27	0.25
	0.0	50.5													0.00	0.13	0.19	0.15	0.29	0.22
M131D apoMb	0.0	0.0														0.08	0.01	0.00	0.00	0.00
	0.0	0.5															0.20	0.08	0.10	0.09
	0.0	6.5																0.06	0.11	0.09
	0.0	15.5																	0.32	0.43
	0.0	35.5																		0.40
	0.0	50.5																		

Table 4-S3. Statistical p values for the solubility of apoMb made in the presence of different TF concentrations (Figure 4-4b). These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are statistically the same (yellow). P values ≤ 0.05 indicate data are statistically different (green).

		Protein construct	wild-type apoMb						I28D apoMb						M131D apoMb					
		[TF]	0.0	0.5	6.5	15.5	35.5	50.5	0.0	0.5	6.5	15.5	35.5	50.5	0.0	0.5	6.5	15.5	35.5	50.5
Protein construct	[TF]	[DnaK]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
wild-type apoMb	0.0	0.0		0.24	0.23	0.29	0.40	0.23	0.04	0.02	0.03	0.11	0.07	0.18	0.00	0.00	0.01	0.27	0.15	0.14
	0.5	0.0			0.48	0.31	0.13	0.49	0.04	0.02	0.00	0.10	0.01	0.01	0.00	0.00	0.00	0.03	0.01	0.05
	6.5	0.0				0.10	0.02	0.48	0.04	0.02	0.00	0.10	0.00	0.01	0.00	0.00	0.00	0.04	0.01	0.06
	15.5	0.0					0.03	0.11	0.04	0.02	0.00	0.10	0.00	0.02	0.00	0.00	0.00	0.05	0.02	0.07
	35.5	0.0						0.02	0.04	0.02	0.00	0.11	0.00	0.03	0.00	0.00	0.00	0.09	0.03	0.09
	50.5	0.0							0.04	0.02	0.00	0.10	0.00	0.01	0.00	0.00	0.00	0.04	0.01	0.06
I28D apoMb	0.0	0.0								0.35	0.10	0.23	0.08	0.06	0.12	0.27	0.23	0.05	0.06	0.07
	0.5	0.0									0.08	0.31	0.05	0.03	0.02	0.36	0.28	0.03	0.03	0.04
	6.5	0.0										0.29	0.04	0.01	0.00	0.01	0.01	0.01	0.01	0.10
	15.5	0.0											0.21	0.16	0.04	0.37	0.43	0.14	0.16	0.18
	35.5	0.0												0.04	0.01	0.01	0.01	0.04	0.06	0.29
	50.5	0.0													0.00	0.00	0.00	0.28	0.40	0.32
M131D apoMb	0.0	0.0														0.01	0.01	0.00	0.00	0.00
	0.5	0.0															0.24	0.00	0.00	0.01
	6.5	0.0																0.00	0.00	0.01
	15.5	0.0																	0.22	0.22
	35.5	0.0																		0.36
	50.5	0.0																		

Table 4-S4. Statistical p values for the slow rotational correlation times in Table 4-1. These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are statistically the same (yellow). P values ≤ 0.05 indicate data are statistically different (green).

	Item 1: Δtig cf sys. + Hsp70 inhib., diluted	Item 2: Δtig cell strain + Hsp70 inhib.	Item 3: WT cf sys.+ 21 μM Hsp70	Item 4: Δtig cf sys + 75 μM Hsp70	Item 5: WT cf sys.+ 75 μM Hsp70	Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. added post-translationally	Item 7: WT cf sys. + 75 μM Hsp70, diluted	Item 8: purified apoMb resuspended in RNC solution	Items 9-10: I28D in WT cf sys.+ 75 μM Hsp70
Item 1: Δtig cf sys. + Hsp70 inhib., diluted		0.09	0.06	0.03	0.03	0.03	0.03	0.03	0.07
Item 2: Δtig cell strain + Hsp70 inhib.			0.43	0.18	0.17	0.15	0.14	0.12	0.46
Item 3: WT cf sys.+ 21μM Hsp70				0.18	0.16	0.14	0.13	0.11	0.45
Item 4: Δtig cf sys. + 75 μM Hsp70					0.37	0.17	0.11	0.05	0.13
Item 5: WT cf sys.+ 75 μM Hsp70						0.05	0.01	0.00	0.13
Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. added post-translationally							0.00	0.01	0.12
Item 7: WT cf sys. + 75 μM Hsp70, diluted								0.00	0.11
Item 8: purified apoMb resuspended in RNC solution									0.10
Items 9-10: I28D in WT cf sys.+ 75 μM Hsp70									

Table 4-S5. Statistical p values for the average fluorescence lifetime values in Table 4-4. These p values were determined with the one-tailed Student's test assuming unequal variances. P values > 0.05 indicate data are statistically the same (yellow). P values ≤ 0.05 indicate data are statistically different (green).

	Item 1: Δtig cf sys. + Hsp70 inhib., diluted	Item 2: Δtig cell strain + Hsp70 inhib.	Item 3: WT cf sys.+ 21 μM Hsp70	Item 4: Δtig cf sys + 75 μM Hsp70	Item 5: WT cf sys.+ 75 μM Hsp70	Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. added post-translationally	Item 7: WT cf sys. + 75 μM Hsp70, diluted	Item 8: purified apoMb resuspended in RNC solution	Items 9-10: I28D in WT cf sys.+ 75 μM Hsp70
Item 1: Δtig cf sys. + Hsp70 inhib., diluted		0.19	0.00	0.01	0.00	0.00	0.00	0.00	0.04
Item 2: Δtig cell strain + Hsp70 inhib.			0.01	0.02	0.00	0.00	0.00	0.00	0.05
Item 3: WT cf sys.+ 21μM Hsp70				0.25	0.03	0.04	0.01	0.00	0.10
Item 4: Δtig cf sys. + 75 μM Hsp70					0.02	0.02	0.01	0.00	0.08
Item 5: WT cf sys.+ 75 μM Hsp70						0.26	0.01	0.00	0.33
Item 6: WT cf sys. + 75 μM Hsp70 + Hsp70 inhib. added post-translationally							0.02	0.00	0.28
Item 7: WT cf sys. + 75 μM Hsp70, diluted								0.00	0.37
Item 8: purified apoMb resuspended in RNC solution									0.03
Items 9-10: I28D in WT cf sys.+ 75 μM Hsp70									

Table 4-S6. Summary of additional experiments aiming at further characterization of soluble aggregates.

Experiment	Goal	Results
Centrifugation to remove ribosomes	Remove ribosomes and then measure τ_c of apoMb protein in absence of ribosomes.	1. Centrifugation also removed protein 2. These experiments were done with protein released with hydroxylamine from purified RNCs. Therefore, these experiments may not be good representation for how protein will behave when released normally in the crude cell free system.
Immunoprecipitation to remove ribosomes	Remove ribosomes and measure τ_c of apoMb protein in absence of ribosomes.	1. Immunoprecipitation did not remove ribosomes from the solution.
Crosslinking of WT apoMb released from purified RNCs.	Identify soluble aggregate as crosslinked band on gel, or monitor decrease in intensity of monomer band	1. ApoMb released in the presence of chaperones (75 μM Hsp70, 12 μM DnaJ, 19 μM GrpE, and 0.2 μM TF) showed no evidence of crosslinking. 2. Soluble aggregates may not be sensitive to crosslinking because they may be transient and occur on timescales faster than the reaction time for the EDC crosslinker (<6 s).[3] 3. This experiment was done with protein released with hydroxylamine from

		purified stalled full-length RNCs. However, apoMb generated from release from stalled ribosomes may have different conformations than apoMb that is synthesized and released without stalling the ribosome.
Crosslinking of ribosome-released WT apoMb in the cell-free systems containing different chaperone concentrations.	Identify soluble aggregate as crosslinked band on gel, or monitor decrease in intensity of monomer band	1. ApoMb samples made with different Hsp70 concentrations showed similar levels of crosslinking. 2. Soluble aggregates may not be sensitive to crosslinking because they may be transient and occur on timescales faster than the reaction time for the EDC crosslinker (<6 s).[3] 3. Crosslinked species may include the soluble aggregates detected with fluorescence anisotropy decay as well as other crosslinked species.

Mass photometry	Identify peak for soluble aggregate and compare amount with and without chaperones	1. WT apoMb made and released in the presence and absence of chaperones showed peaks in the mass photometry spectrum with similar masses. 2. This experiment was done with protein released with hydroxylamine from purified stalled full-length RNCs. However, apoMb generated from release from stalled ribosomes may have different conformations than apoMb that is synthesized and released without stalling the ribosome. 3. Monomer is too small to detect with this method.

Table 4-S7. Standard error and percent error for concentration of apoMb made in cell-free system.

Sample	Average gel fluorescence intensity of soluble + insoluble fractions (counts)	Standard error in total gel fluorescence intensity of soluble + insoluble fractions (counts)	Percent error (%)
WT apoMb	13,000	4,500	35
I28D apoMb	10,000	6,000	60
M131D	13,000	7,000	54
Average	12,000	5,800	49

Supporting References

(1) Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S. Kinetic trapping of folded proteins relative to aggregates under physiologically relevant conditions. *J. Phys. Chem. B* **2018,** *122*, 7682-7698.

(2) Hermanson, G. T. *Bioconjugate techniques*. 3rd ed. London : Academic Press, c2013.: 2013.

(3) Guzman-Luna, V.; Fuchs, A. M.; Allen, A. J.; Staikos, A.; Cavagnero, S. An intrinsically disordered nascent protein interacts with specific regions of the ribosomal surface near the exit tunnel. *Commun. Biol.* **2021,** *4*, 1236.

(4) Hoare, D. G.; Koshland, D. E., Jr. A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* **1967,** *242*, 2447-2453.

(5) Young, G.; Hundt, N.; Cole, D.; Fineberg, A.; Andrecka, J.; Tyler, A.; Olerinyova, A.; Ansari, A.; Marklund, E. G.; Collier, M. P., et al. Quantitative mass imaging of single biological macromolecules. *Science* **2018,** *360*, 423-427.

(6) Wu, D.; Piszczek, G. Measuring the affinity of protein-protein interactions on a single-molecule level by mass photometry. *Anal. Biochem.* **2020,** *592*, 113575.

(7) Häußermann, K.; Young, G.; Kukura, P.; Dietz, H. Dissecting foxp2 oligomerization and DNA binding. *Angew. Chem. Int. Ed.* **2019,** *58*, 7662-7667.

(8) Ellis, J. P.; Bakke, C. K.; Kirchdoerfer, R. N.; Jungbauer, L. M.; Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **2008,** *3*, 555-566.

(9) Rutkowska, A.; Mayer, M. P.; Hoffmann, A.; Merz, F.; Zachmann-Brand, B.; Schaffitzel, C.; Ban, N.; Deuerling, E.; Bukau, B. Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **2008,** *283*, 4124-4132.

(10) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(11) Ziehr, D. R.; Ellis, J. P.; Culviner, P. H.; Cavagnero, S. Production of ribosome-released nascent proteins with optimal physical properties. *Anal. Chem.* **2010,** *82*, 4637-4643.

(12) Lakowicz, J. R. *Principles of fluorescence spectroscopy*. 3rd ed.; New York : Plenum Press: 2006.

(13) Hutchinson, R. B.; Chen, X.; Zhou, N.; Cavagnero, S. Fluorescence anisotropy decays and microscale-volume viscometry reveal the compaction of ribosome-bound nascent proteins. *J. Phys. Chem. B* **2021,** *125*, 6543-6558.

(14) Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S. In vitro expression and characterization of native apomyoglobin under low molecular crowding conditions. *Protein Expr. Purif.* **2006,** *45*, 381-392.

(15) Moore, S. D.; Baker, T. A.; Sauer, R. T. Forced extraction of targeted components from complex macromolecular assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 11685-11690.

(16) Wang, S.; Jomaa, A.; Jaskolowski, M.; Yang, C.; Ban, N.; Shan, S. The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat. Struct. Mol. Biol.* **2019,** *26*, 919-929.

(17) Fahnestock, S. R.; Strycharz, W. A.; Marquis, D. M. Immunochemical evidence of homologies among 50 S ribosomal proteins of bacillus stearothermophilus and Escherichia coli. *J. Biol. Chem.* **1981,** *256*, 10111-10116.

(18) Dalphin, M. D.; Stangl, A. J.; Liu, Y.; Cavagnero, S. KLR-70: A novel cationic inhibitor of the bacterial Hsp70 chaperone. *Biochemistry* **2020,** *59*, 1946-1960.

Chapter 5

Distribution and Solvent Exposure of Hsp70 Chaperone Binding Sites Across the *E. coli* Proteome

This chapter reproduces an in-press article in the journal *PROTEINS: Structure, Function, and Bioinformatics* by Xi Chen*, Rachel B. Hutchinson*, and Silvia Cavagnero (*co-first authors). X.C., R.B.H., and S.C. designed the project. X.C. collected the data. R.B.H and X.C. analyzed the data. X.C., R.B.H., and S.C. wrote the manuscript.

5.1 Abstract

Many proteins must interact with molecular chaperones to achieve their native state in the cell. Yet, how chaperone binding-site characteristics affect the folding process is poorly understood. The ubiquitous Hsp70 chaperone system prevents client-protein aggregation by holding unfolded conformations and by unfolding misfolded states. Hsp70 binding sites of client proteins comprise a nonpolar core surrounded by positively charged residues. However, a detailed analysis of Hsp70 binding sites on a proteome-wide scale is still lacking. Further, it is not known whether proteins undergo some degree of folding while chaperone bound. Here, we begin to address the above questions by identifying Hsp70 binding sites in 2,258 *E. coli* proteins. We find that most proteins bear at least one Hsp70 binding site and that the number of Hsp70 binding sites is directly proportional to protein size. Aggregation propensity upon release from the ribosome correlates with number of Hsp70 binding sites only in the case of large proteins. Interestingly, Hsp70 binding sites are more solvent-exposed than other nonpolar sites, in protein native states. Our findings show that the majority of *E. coli* proteins are systematically enabled to interact with Hsp70 even if this interaction only takes place during a fraction of the protein lifetime. In addition, our data suggest that some conformational sampling may take place within Hsp70-bound states, due to the solvent exposure of some chaperone binding sites in native proteins. In all, we propose that Hsp70-chaperone-binding traits have evolved to favor Hsp70-assisted protein folding devoid of aggregation.

5.2 Introduction

Molecular chaperones are essential for cell viability and protein solubility. Simultaneous deletion of trigger factor and Hsp70 chaperones in live cells is deadly, at temperatures higher than 20 ºC.[1-3] In addition, chaperones grant solubility to 67% of aggregation-prone *E. coli* proteins, upon expression in transcription-translation cell-free systems.[4] Chaperones act both co- and post-translationally and are known to promote protein folding by preventing aggregation[5-9] and sometimes even by facilitating protein disaggregation.[10-12] Importantly, it has been shown that 72% of *E. coli* proteins are insoluble when synthesized in an *in vitro* translation system that lacks molecular chaperones.[13] Remarkably, the presence of molecular chaperones in the medium increases the solubility of 97% of the above aggregation-prone pool.[4] Chaperones are also necessary for protein solubility in other organisms. For instance, the eukaryotic luminal Hsp70 chaperone BiP maintains client-protein solubility in the yeast endoplasmic reticulum, allowing retrotranslocation and degradation processes.[14] ClpB, a prokaryotic Hsp100 chaperone, disassembles protein aggregates either alone or with assistance by the Hsp70 chaperone system.[15-16] Further, the downregulation or knockout of molecular chaperones in eukaryotes has often been associated with increased formation of aggregates (e.g., soluble oligomers and amyloid fibrils) that are linked to the pathology of deadly proteinopathies.[17-19] In some cases, these aggregates were even proven to be causative for the respective diseases.[20-24]

The 70-kDa heat shock protein (Hsp70) is one of the most ubiquitous chaperones in Nature. The genes encoding Hsp70 are highly conserved among prokaryotes, eukaryotes, and most archaea.[25-27] The *E. coli* Hsp70 chaperone, also known as DnaK, interacts with at least 700 client proteins in the *E. coli* cellular environment.[28] DnaK consists of two domains: a 45 kDa nucleotide-binding domain (NBD) and a 25 kDa substrate-binding domain (SBD), as shown in

Figure 5-1A.[29] The two domains are allosterically regulated via nucleotide binding. When the chaperone is nucleotide-free or ADP-bound, the two domains function separately. The binding of ATP induces a conformational change, leading NBD and SBD to dock to each other (Figure 5-1A). The SBD contains an α-helical lid (SBDα) and a β-sheet pocket (SBDβ). Within the SBDβ, two β-sheets and two loops form a pocket where the client protein binds. The binding cleft consists of hydrophobic and negatively charged amino acids.[30] These residues interact with nonpolar and positively charged amino acids of client proteins via a combination of van der Waals and electrostatic interactions.[31] Here, we define client proteins as proteins that bear one or more Hsp70 binding sites.

The Hsp70 chaperone system interacts with client proteins through a four-step functional cycle that involves two co-chaperones: Hsp40 and a nucleotide exchange factor (NEF) (Figure 5-1B).[32-34] During the first step, ATP-bound Hsp70 interacts weakly with client proteins via a combination of hydrogen bonds, van der Waals, electrostatic and other noncovalent interactions. Client proteins can either directly bind Hsp70, or they can be delivered to the chaperone upon interaction with the Hsp40 cochaperone (i.e., DnaJ in *E. coli*). During the second step, ATP hydrolysis induces a conformational change and causes the α-helical lid to close on top of the client protein. ADP-bound Hsp70 has a high affinity and a slow dissociation rate for the client proteins. During the third step, a NEF (i.e., GrpE in *E. coli*) promotes the release of ADP. During the fourth and final step, the chaperone binds ATP, causing the α-helical lid to open. ATP-bound Hsp70 has a low affinity and a fast exchange rate for the client protein so that the latter gets released, and the chaperone cycle can start anew.

Hsp70 promotes protein folding upon binding to either unfolded or misfolded client-protein states, thus exerting a “hold-only” or “fold-promoting” behavior, respectively

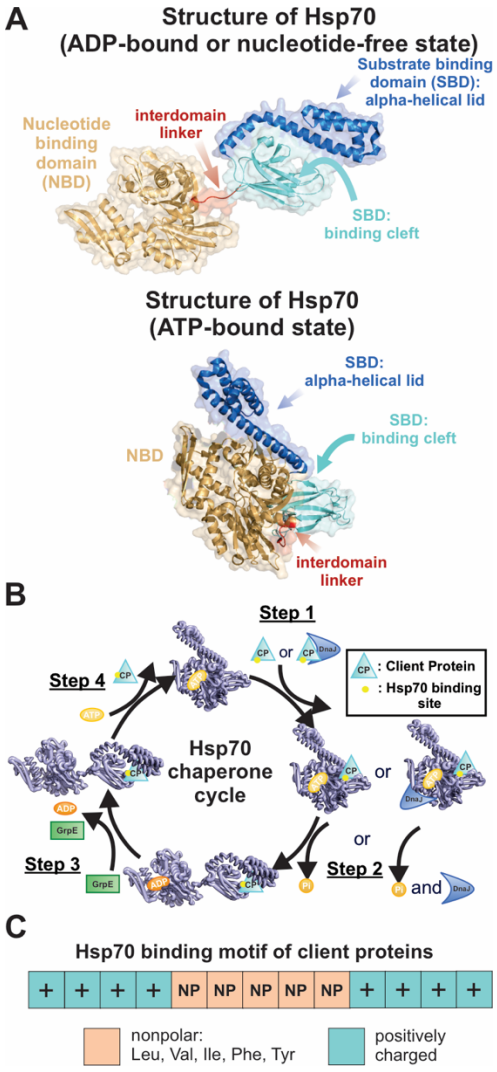


Figure 5-1. The Hsp70 chaperone: structure, functional cycle and client-protein binding motif. (A) Structure of ADP-bound (or nucleotide-free) Hsp70 (PDB ID: 2KHO, top), and structure of ATP-bound Hsp70 chaperone (PDB ID: 4B9Q, bottom). (B) Diagram illustrating the main steps of the Hsp70 chaperone cycle. Hsp70, together with its co-chaperones DnaJ and GrpE, promotes client-protein folding via an ATP-dependent cycle. (C) Binding motif of *E. coli* client proteins for interaction with the Hsp70 chaperone. The positively charged residues surrounding the five central nonpolar amino acids are increasingly less important for chaperone binding, as their distance from the binding-site core increases.[31]

(Mecha *et al*. 2022, in preparation).[35] According to the "hold-only" action, Hsp70 transiently reduces the concentration of free unfolded client in solution. This binding effectively slows down misbehavior because the rates of aggregation nucleation and elongation decrease at lower concentration of aggregation-prone unfolded precursors. Both experimental and computational studies provide evidence that Hsp70 employs the "hold-only" mechanism upon interaction with RNase H^D, a slow-folding and non-aggregation-prone client protein (Mecha *et al*. 2022, in preparation).[36] The Hsp70 chaperone system adopts a "fold-promoting" role in the case of aggregation-prone client proteins that populate transient misfolded intermediates. In this case, Hsp70 binds the client misfolded state, and it is believed to exploit the free energy of ATP hydrolysis to unfold the misfolded intermediate. For instance, the firefly luciferase (fluc) client achieves its native state much more quickly in the presence of the Hsp70 chaperone and ATP.[37] Single-molecule experiments showed that, in the presence of ATP and in the absence of GrpE cochaperone, the Hsp70-bound misfolded state of fluc is converted to an unfolded state.[37] Other experimental studies found that a single Hsp70 molecule consumes five ATPs before it can efficiently refold one misfolded luciferase.[38] Therefore, several cycles of Hsp70 chaperone binding and release may be required for the client protein to reach its native conformation.[38]

Client proteins may have more than one Hsp70 binding site. Multiple Hsp70 binding sites are compatible with multiple simultaneously bound Hsp70 molecules. Single-molecule studies by Kellner *et al.* showed that one rhodanese protein can bind four to seven Hsp70 molecules.[39] Imamoglu *et al.* showed that ~12 Hsp70 molecules can bind a single fluc protein.[37] When multiple Hsp70 chaperones interact simultaneously with a single client protein, release of chaperone molecules is likely asynchronous. NMR studies on the hTRF1 client protein support this view,[40-41] given that hTRF1 bound to multiple chaperones is conformationally

heterogeneous. Asynchronous chaperone release likely reduces chances of misfolding by allowing different client-protein regions to fold independently, before establishing long-range contacts.

Hsp70 interacts with client-proteins via a binding motif that comprises three to six nonpolar residues flanked by positively charged amino acids (Figure 5-1C). Rüdiger and coworkers[31] identified this motif upon scanning 4,360 partially overlapping polypeptides (13-residue each) encompassing the full sequences of 37 distinct proteins. The peptides were immobilized on a cellulose support, which was then incubated in the presence of an overlaying Hsp70 solution lacking ATP. Unbound Hsp70 was then removed upon exposure to tris-buffered saline (pH 7.6), and Hsp70-associated peptides were identified and quantified via a colorimetric immuno-assay.[31] The authors then proceeded to exploit the binding patterns identified via the above procedure to assign amino-acid-specific Hsp70 binding frequencies. The latter were converted into effective binding free energies ($\Delta\Delta G°$), which were then evaluated across adjacent partially overlapping 13-mers spanning any desired protein sequence.[31] The resulting algorithm is capable of predicting Hsp70 binding sites of any given protein sequence with a high accuracy of ca. 95%. As long as the predicted effective binding-site free energies are below $\Delta\Delta G° = -5$ kcal/mol, the 95% accuracy applies to any protein sequence, regardless of its total number of Hsp70 binding sites.[31] It is worth noting that this algorithm assigns amino acid contribution to the total binding energy based upon weights decreasing with distance from the nonpolar core.[31] In all, the computational tool by Rüdiger *et al.* is extremely powerful because it is reliable, accurate and easy to use. While this algorithm is not capable of identifying structural details including substrate binding registry and orientation,[42] these two characteristics are not relevant to the present study.

More recently, other Hsp70 binding site prediction algorithms were developed. A review by Nordquist *et al*. summarizes and compares all available Hsp70 binding-site prediction methods, including sequence-based, structure-based, and molecular dynamics-based approaches.[42-46] For instance, the Limbo algorithm combines sequence information based on experimental binding assays with structural information from molecular modeling via the FoldX force field.[44] Importantly, Limbo binding profiles are consistent with the Hsp70 binding preferences reported by Rüdiger *et al.* (Figure 5-1C). Unlike the tool by Rudiger *et al.*, however, Limbo provides some information on binding register. Another algorithm, known as ChaperISM, is a structure-based model based on a position-independent scoring matrix.[45] ChaperISM also yields binding-site predictions consistent with those by Rüdiger *et al.*

All three above approaches yield similar scores for Hsp70 binding at the residue-specific level, and are comparably reliable.[42] The sequence-based algorithm by Rüdiger *et al.,* however, is simpler. Although it does not yield information on binding-site registry and orientation, these details are not necessary here. Therefore, the correct Hsp70 binding-site predictions achievable via the tool by Rüdiger *et al.* render it suitable for our studies, and we elected to adopt it here. Interestingly, the number of Hsp70 binding sites is proportional to client-protein size for the 37 client proteins analyzed by Rüdiger *et al*. On average, the authors found that there is one Hsp70 binding site every 36 protein residues.[31] However, the analyzed data set comprised only a few proteins, and it is presently unknown whether this trend also applies on a proteome-wide scale.

Protein size is also known to correlate with aggregation propensity upon release from the ribosome. Niwa *et al.* synthesized over 2,000 individual *E. coli* soluble proteins, one by one, in an *in vitro* cell-free system lacking molecular chaperones and measured the resulting protein solubility.[13] The average molecular weight of the soluble (i.e., with > 70% solubility) and

insoluble (i.e., with < 30% solubility) proteins was found to be 25.5 and 41.1 kDa, respectively.[13, 47] In a later study, Niwa *et al.* identified 409 aggregation-prone proteins whose solubility increased when synthesized in the presence of the Hsp70 chaperone system.[4] In addition, Calloni *et al.* identified individual Hsp70-interacting *E. coli* proteins at their natural abundance, during the exponential growth phase of this bacterium.[28] These authors found that, for Hsp70 interactors, proteins with many Hsp70 binding sites are less soluble than proteins with few binding sites.[28]

While the latter investigations on Hsp70 client-protein solubility were very informative, they focused exclusively on Hsp70 interactors at 37º C. These studies did not include proteins that interact with Hsp70 under heat-related stress[48] and other types of stress conditions, e.g., acidic environment or starvation.[49] However, Hsp70 binding sites that are not used at 37º C under physiologically relevant conditions may be utilized under different environmental conditions, e.g., under stress. Therefore, an assessment of all Hsp70 binding sites on a proteome-wide scale is sorely needed to better understand how this chaperone operates in Nature. Yet this analysis is still lacking. Further, Hsp70 interacts with nonpolar regions of client proteins, often favoring a globally expanded client-protein conformation.[50-52] Yet, it is not known whether Hsp70 client proteins are able to reach their native state while bound to their chaperone host.

In this work, we addressed the above questions by analyzing the relation between client-protein size, solubility, and number of Hsp70 binding sites. We focused on *E. coli* cytoplasmic proteins and Hsp70. Upon employing the sequence-based algorithm developed by Rüdiger *et al.*, we predicted the location of Hsp70 binding sites of 2,258 *E. coli* cytoplasmic proteins and found that Hsp70 binding sites occur every ca. 44 ± 23 residues. Our results show that proteins with zero or one Hsp70 binding site are largely soluble even in the absence of chaperones, while

proteins with multiple chaperone binding sites display a wide range of solubilities upon release from the ribosome, in the absence of Hsp70. We also find that Hsp70 binding sites of *E. coli* soluble proteins span only a moderate fraction of all nonpolar regions. In addition, we determined the solvent exposure of Hsp70 binding sites across protein native states. While Hsp70 binding sites tend to be significantly buried within the native structure, they are more solvent-accessible than other nonpolar sites in the native structure. This result suggests that some regions of client proteins may start folding while they are chaperone bound. In all, our study provides insights into how the frequency and location of Hsp70 binding sites enable organisms to effectively exploit molecular chaperones in living systems.

5.3 Results and Discussion

***E. coli* proteins display a wide distribution of solubilities upon release from the ribosome, in the absence of molecular chaperones.** As discussed in the previous section, the Hsp70 chaperone is of seminal importance for protein life of the cell, not only in *E. coli* but also in higher organisms.[7, 53] In *E. coli*, it has been shown that Hsp70 increases the solubility of a much larger fraction of proteins than other chaperones.[4] First, we explored the relation between protein solubility and number of amino acids across the *E. coli* proteome. In the entirety of this work, we define protein solubility as the ability to remain water-soluble upon release from the ribosome in an *in vitro* cell-free translation system lacking chaperones. This type of solubility was experimentally assessed for each *E. coli* protein at approximately constant total-protein expression levels in a study by Niwa and coworkers.[13]

We started from the raw data by Niwa *et al.*[13] and replotted them as percent solubility upon release from the ribosome as a function of number of amino acids. The results, shown in panels

A and D of Figure 5-2, highlight that there is only a weak inverse correlation between protein size and solubility upon release from the ribosome in the absence of chaperones.[13] A global assessment of the plots in Figure 5-2A,D reveal that small and mid-size proteins have a broad distribution of solubilities while very large proteins (> 900 residues) are more likely to be insoluble, upon release from the ribosome. Therefore, poor intrinsic solubility upon release from the ribosome is definitely an issue for very large proteins, but it is also a severe challenge for some small- and mid-size proteins.

Proteins with zero or one Hsp70 binding site are highly soluble while proteins with multiple Hsp70 binding sites have broadly distributed solubilities. Next, we focused on correlating protein solubility to the number and distribution of binding sites of the Hsp70 chaperone. To accomplish this goal, we determined the number and sequence-location of Hsp70 binding sites of 2,258 *E. coli* cytoplasmic proteins via a known algorithm, which was developed upon experimental analysis of the Hsp70 binding affinity of peptide libraries.[31] We then grouped the proteins into three categories: proteins with zero, one, or multiple Hsp70 binding sites.

Panels A and D of Figure 5-2 (orange and light-blue dots, respectively) reveal that proteins with zero or one Hsp70 binding site are intrinsically soluble upon release from the ribosome, in the absence of chaperones. Further, panels B and E of Figure 5-2 reveal that these proteins tend to be small, with an average size of ca. 100 residues. Hence, small proteins with zero or one Hsp70 binding site are soluble upon release from the ribosome. The orange proteins in Figure 5-2A, lacking Hsp70 binding sites, have evolved to achieve independent solubility with no Hsp70 chaperone assistance. While this is a remarkable property, Figure 5-2A shows that there are only very few proteins displaying this characteristic, in *E. coli*. Interestingly, these sequences are enriched in redox-active proteins (see supporting information excel data sheet).

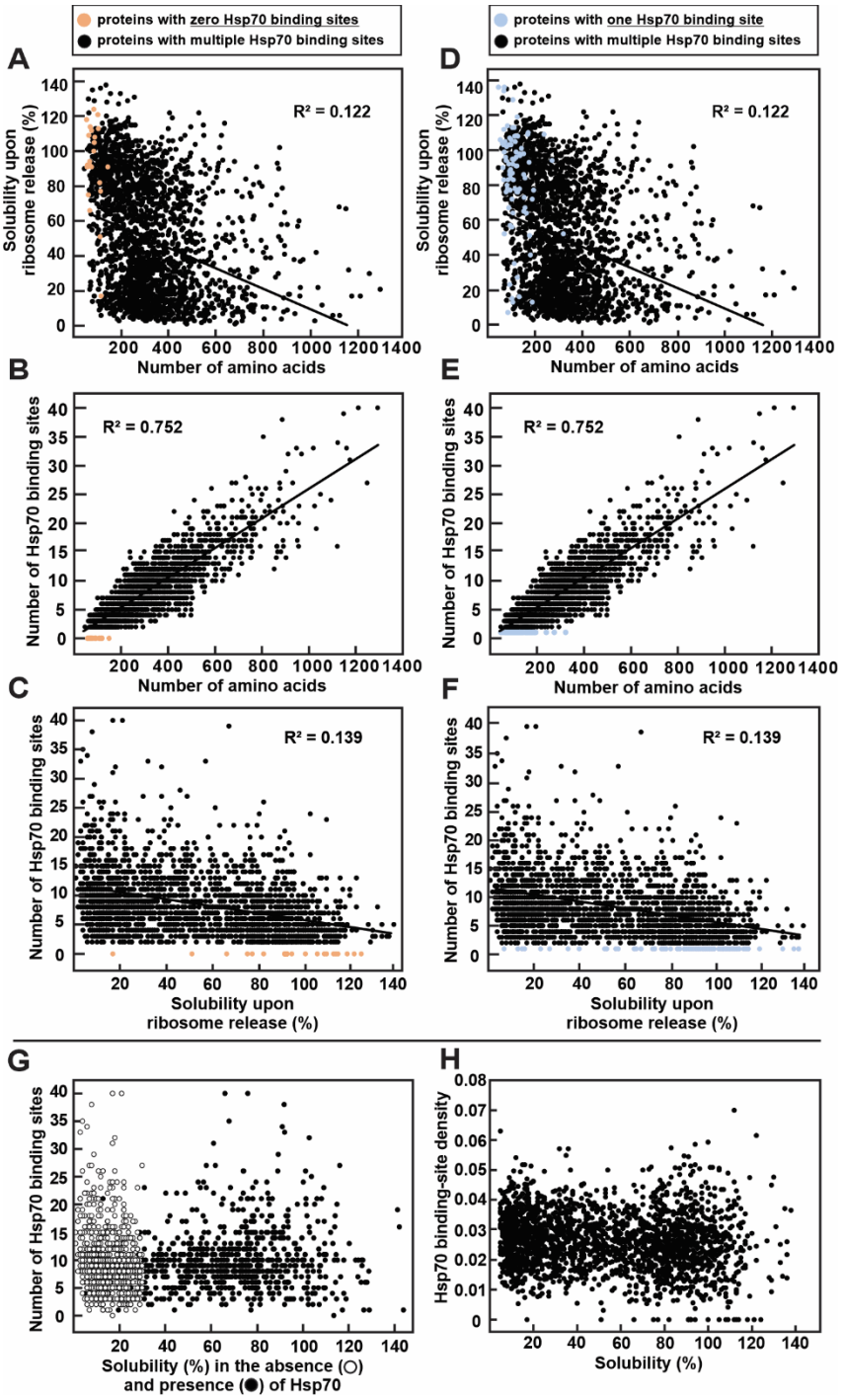


Figure 5-2. Relation between client-protein size or solubility and number of Hsp70 binding sites. (A), (D) Client-protein solubility is weakly correlated with protein size. (B), (E) There is a linear relation between the number of Hsp70 binding sites and client-protein size. (C), (F) The

number of Hsp70 binding sites is weakly correlated with client-protein solubility. (G) Effect of addition of 5 μM Hsp70 to *E. coli* cell-free transcription-translation medium on protein solubility.[4] (H) Relation between Hsp70 binding-site density (defined as the number of Hsp70 binding sites / number of amino acids) and percent solubility upon release from the ribosome in the absence of molecular chaperones. In panels A-F, light orange and light-blue dots denote proteins with zero and one Hsp70 binding sites, respectively.

We then examined the distribution of protein solubility of proteins with zero, one, or multiple binding sites (Figure 5-3). We found that 87% of proteins with zero binding sites and 84% of proteins one binding site are soluble (>70% solubility) in the absence of chaperones. On the other hand, client proteins with multiple binding sites show a large range in solubility. These proteins had an average solubility of 51.2 ± 33.5 % and only 35% of these proteins are greater than 70% soluble (Figure 5-3C). In fact, the average number of Hsp70 binding sites for insoluble proteins (<30% solubility) is 9.99 ± 0.20 compared to 6.27 ± 0.52 for soluble proteins (>70% solubility, $p=3.28\times 10^{-10}$ for a two-tailed t test). Thus, insoluble proteins tend to have more Hsp70 binding sites than soluble proteins. In all, the results of Figure 5-2A,D and Figure 5-3 suggest that Hsp70 binding sites have evolved to augment the solubility of proteins that are intrinsically insoluble upon release from the ribosome.

The number of Hsp70 binding sites is proportional to *E. coli* protein size. Then, we identified the relation between number of Hsp70 binding sites and protein size across all *E. coli* proteins. The results are shown in Figure 5-2B,E. The plots show that there is a good linear correlation between *E. coli* protein size and number of Hsp70 binding sites ($R^2 = 0.752$). On average, we found that there is one Hsp70 binding site every 44 ± 23 amino acids. The binding sites are distributed along protein sequence with a somewhat uneven frequency.

A detailed analysis of this frequency distribution is beyond the scope of this work. On the other hand, it is interesting to note that most soluble (i.e., non-membrane-associated proteins) consist of alternating polar and nonpolar stretches.[54-56] The nonpolar regions that extend over ca. 5 residues correspond to the average size of the core of Hsp70 binding sites.[31] Nonpolar regions of this length are not very common in soluble proteins,[54] yet they are highly represented within Hsp70 binding sites.[31] This feature is presumably due to the high aggregation propensity of long

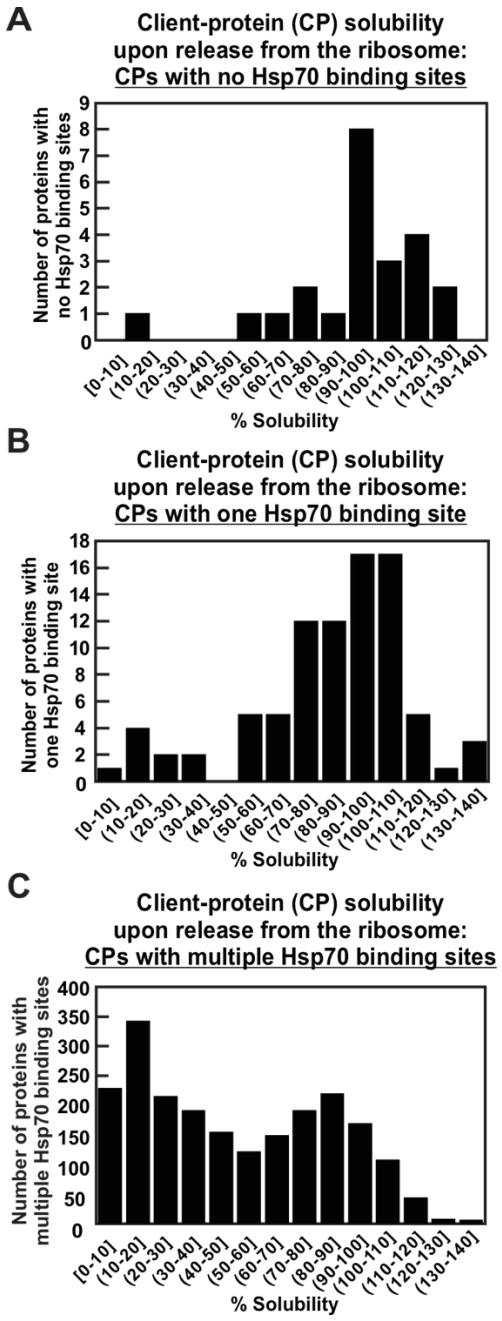


Figure 5-3. Distribution of client-protein solubility for proteins with different numbers of Hsp70 binding sites. Solubility distribution for client proteins with (A) zero Hsp70 binding sites (23 proteins), (B) one binding site (86 proteins), and (C) multiple binding sites (2,149 proteins).

nonpolar regions. Now, the observed Hsp70 binding-site separation of 44 ± 23 residues has a large standard deviation. We propose that this wide uncertainty (± 23 amino acids) is simply related to the fact that alternating nonpolar and polar regions in soluble proteins have randomly distributed sizes.[54]

The above result is consistent with the findings by Rüdiger and coworkers, who identified one binding site every ca. 36 residues out of a small database of 37 proteins from different organisms.[31] On the other hand, the much larger number of proteins (2,258 *E. coli* cytoplasmic proteins) analyzed here broadens the scope of our analysis. In addition, the focus of the present study on *E. coli* significantly increases our knowledge on the Hsp70 binding sites of this specific organism. The findings in Figure 5-2B,E show that the overwhelming majority of *E. coli* proteins bears one or more chaperone binding sites. This result suggests that *E. coli* proteins are posed to receive folding and solubility help from the Hsp70 chaperone.

There is a weak correlation between the number of Hsp70 binding sites and protein solubility. Next, we explored the relation between number of Hsp70 binding sites and solubility of *E. coli* proteins, upon release from the ribosome in the absence of chaperones. Our results (Figure 5-2C,F) show that proteins with the largest number of Hsp70 binding sites tend to be less soluble, suggesting that these proteins are particularly in need of the Hsp70 chaperone. However, the overall correlation between solubility and number of binding sites is weak, as testified by the low R^2 value ($R^2 = 0.1386$). The graphs of Figure 5-3C,F also show that proteins with the most common number of binding sites, spanning from 2 to 20, span a wide range of solubilities.

At this juncture, a multiple linear regression model was used to better characterize the dependence of number of binding sites on client-protein size and solubility. The variance inflation factor for protein size and solubility is 1.14, demonstrating that these two variables are

not highly correlated. We performed a multiple linear-regression analysis with number of binding sites as the dependent variable and protein size and solubility as the independent variables. In this analysis, protein solubility was defined as the percent soluble protein upon biosynthesis in the absence of chaperones according to Niwa *et al,*[13] and size was defined as the number of residues. We found that both size and solubility are significantly correlated with number of Hsp70 binding sites according to best-fit regression equation

$$\text{Number of Hsp70 binding sites} = 1.0737 + 0.0250(\text{protein size}) - 0.0122(\text{percent solubility}) \quad (1)$$

The probabilities that the 0.0250 and -0.0122 coefficients in eq. 1 are significantly different from zero (according to a two-tailed t test) are: $p_{size} < 0.000$, $p_{solubility} = 1.84 \times 10^{-12}$, respectively. Therefore, the correlation between number of Hsp70 binding sites and protein solubility is significant even upon accounting for the effect of protein size. However, the R^2 value for the multiple linear regression is 0.7574, which is only slightly higher than the R^2 value of 0.752 obtained for the linear fit of number of binding sites as a function of protein size.

Addition of Hsp70 chaperone to the transcription-translation medium dramatically improves protein solubility upon release from the ribosome. To visually estimate whether and how the presence of the Hsp70 chaperone in the transcription-translation medium helps enhancing the solubility of proteins bearing one or more Hsp70 binding sites, we plotted the number of Hsp70 chaperone binding sites as a function of solubility upon release from the ribosome in the absence and presence of Hsp70. We focused on the most insoluble *E. coli* proteins and took advantage of the raw data by Niwa *et al.*[4] The results are plotted in Figure 5-2G and show that addition of Hsp70 leads to a dramatic and widespread enhancement in solubility upon release from the ribosome. Consistent with the data originally reported by Niwa *et al.,*[4] which lacked a direct relation to Hsp70 binding sites, Hsp70 chaperone binding sites have

a dramatic beneficial effect for *E. coli* proteins that would otherwise be insoluble upon release from the ribosome. This plot highlights the direct effect of Hsp70 chaperone binding sites for insoluble proteins upon release from the ribosome. These data are in striking contrast with the proteins of Figure 5-2C,F that are soluble even in the absence of Hsp70, yet bear a plethora of Hsp70 binding sites.

Hsp70 binding-site density is not correlated to protein solubility upon release from the ribosome. In order to investigate whether more intrinsically insoluble proteins have a higher density of Hsp70 binding sites (number of Hsp70 binding sites/number of amino acids) along their primary structure, we plotted the Hsp70 binding-site density as a function of protein solubility upon release from the ribosome for 2,258 *E. coli* cytoplasmic proteins. The results, show in in Figure 5-2H, point to no specific trends. Therefore, Nature's strategies to increase protein solubility upon release from the ribosome does not include the evolution of closely spaced Hsp70 chaperone binding sites across the protein sequence.

In summary, both soluble and insoluble *E. coli* proteins have a similar distribution of Hsp70 binding-site density regardless of their solubility.

Why do client proteins include Hsp70 binding sites even when they are not required for solubility upon release from the ribosome? In previous sections, we showed that proteins with one Hsp70 binding site tend to be soluble upon release from the ribosome (Figure 5-2A,D and Figure 5-3). While most proteins with multiple Hsp70 binding sites require chaperones for solubility, 35% of these proteins are soluble even in the absence of chaperones. We propose that evolution has only partially optimized the distribution of Hsp70 chaperone binding sites. Another explanation is that while these proteins may not require chaperones for solubility, they may still need chaperones to fold into their native state. For instance, the model protein sperm whale

apomyoglobin forms a soluble aggregate when synthesized in the absence of chaperones.[57] Biosynthesis in the presence of the Hsp70 chaperone enables a higher fraction of apomyoglobin proteins to reach the native state.[57] Future experimental studies are needed to further evaluate the structural accuracy of soluble proteins from *E. coli*.

Hsp70 primarily interacts with client proteins during and immediately after translation under non-stress conditions. Previous computational studies based on experimental kinetic parameters[58] determined that client proteins are only able to interact with Hsp70 if they are fairly thermodynamic unstable, with a standard-state folding free energy $\Delta G°_{UN} \geq -2$ kcal mol^{-1}.[58] According to Robertson and Murphy[59] and Ghosh and Dill,[60-61] all *E. coli* proteins have a standard-state folding free energy of less than –3 kcal mol^{-1}.[60] Therefore, given the considerable thermodynamic stability of *E. coli* proteins, none of them are expected to interact with Hsp70 at equilibrium.

Calloni *et al.* showed that out of 294 experimentally examined Hsp70 client proteins, 93% interact with Hsp70 in a non-equilibrium state that corresponds to either co- or immediately post-translational events.[28] Further, Teter *et. al* also showed that most Hsp70 clients interact with this chaperone either cotranslationally or immediately post-translationally.[2] In all, predictions and experiments are in full agreement: Hsp70 interacts primarily with client proteins before the native/unfolded equilibrium has been established in the cell.

Additional insights are provided by recent studies that showed that protein solubility upon release from the ribosome is under kinetic control, for apomyoglobin and many *E. coli* proteins.[57, 62] Therefore, the early stages of protein life are critical for reaching the native state.[57]

Proteins are also known to interact with Hsp70 under stress conditions, which are usually accompanied by an increase in aggregation propensity. Environmental stress including variations

in pH and ionic strength or mutations may accelerate the kinetic flux towards formation of insoluble aggregates. In addition, raising temperature by only a few degrees significantly destabilizes the proteome,[60] causing more proteins to require chaperone assistance even after proteins have been released from the ribosome and folded. Even if Hsp70 is not essential for cell survival at 30-37 °C,[63] it becomes essential under heat shock conditions (≥ 42 °C).[48, 64]

Further, Hsp70 binding sites that are not needed for solubility at birth, detected in Figure 5-2C,F, may be useful to assist protein disaggregation processes later in life, when environmental stress becomes more probable. Some proteins are also known to interact with Hsp70 while partially or globally unfolded during translocation within cellular compartments.[50]

In summary, Hsp70 mainly interacts with client proteins during and immediately after translation, during translocation, or under stress conditions. Therefore, the Hsp70 binding sites discussed in this work are relevant within this context.

Only a small portion of client-protein nonpolar regions contains Hsp70 binding sites. A critical prerequisite for protein solubility is the burial of nonpolar residues inside the protein core.[65-67] Burial of nonpolar residues prevents nonpolar interactions with other proteins leading to aggregation. Additionally, the Hsp70 binding motif contains several nonpolar residues. Therefore, we explored the relation between client-protein nonpolar regions and chaperone binding sites for 286 randomly selected client proteins.

As shown in Figure 5-4, Hsp70 binding sites make up only a small fraction of the total number of nonpolar sites. For proteins with one Hsp70 binding site, only 20% of all hydrophobic regions are chaperone binding sites. For proteins with multiple binding sites, Hsp70 binding sites make up 56% of the total nonpolar patches (Figure 5-4A, Table 5-1). Interestingly, about 22% of Hsp70 binding sites did not classify as nonpolar sites. This result may be due to contribution of

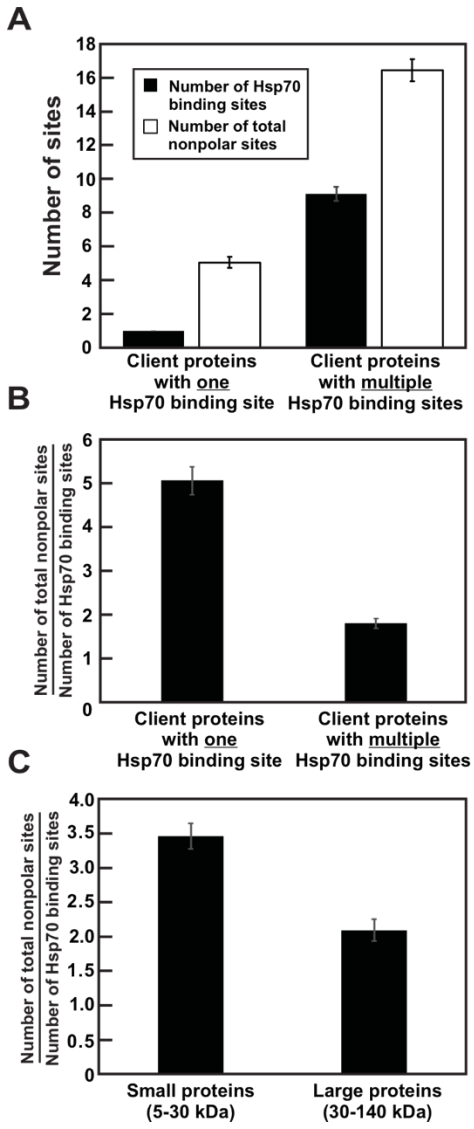


Figure 5-4. Correspondence between nonpolar sites and Hsp70 binding sites. (A) On average, only a small portion of the total nonpolar sites are Hsp70 binding sites. (B) The ratio of number of total nonpolar sites to the number of Hsp70 binding sites is much greater for proteins with one binding site than proteins with multiple binding sites. (C) Likewise, the ratio of number of total nonpolar sites to the number of Hsp70 binding site is higher for smaller proteins (5-30 kDa) than for larger proteins (30-140 kDa). Error bars denote standard error.

Table 5-1. Statistical p values comparing number of Hsp70 binding sites and number of nonpolar sites for the subsets of client proteins in Figure 5-4A. These p values were determined with the two-tailed Student's test assuming unequal variances. P values ≤ 0.05 indicate statistically different data and are shown in green.

	Number of Hsp70 binding sites in client proteins with one Hsp70 binding site	Number of nonpolar sites in client proteins with one Hsp70 binding site	Number of binding sites in client proteins with multiple Hsp70 binding sites	Number of nonpolar sites in client proteins with multiple Hsp70 binding sites	Number of Hsp70 binding sites in all client proteins	Number of nonpolar sites in all client proteins
Number of Hsp70 binding sites in client proteins with one Hsp70 binding site		3.10×10^{-21}	1.49×10^{-48}	1.48×10^{-59}	3.14×10^{-40}	2.01×10^{-61}
Number of nonpolar sites in client proteins with one Hsp70 binding site			1.56×10^{-13}	1.45×10^{-39}	9.33×10^{-4}	1.88×10^{-29}
Number of Hsp70 binding sites in client proteins with multiple Hsp70 binding sites				6.13×10^{-19}	1.13×10^{-5}	3.85×10^{-8}
Number of nonpolar sites in client proteins with multiple Hsp70 binding sites					2.20×10^{-31}	8.38×10^{-5}
Number of Hsp70 binding sites in all client proteins						1.06×10^{-19}
Number of nonpolar sites in all client proteins						

positively charged residues in the flanking regions of the binding sites. For instance, a peptide sequence that has multiple core nonpolar residues and only one positively charged amino acid along flanking regions may be characterized as both a nonpolar and a binding site, while a sequence enriched in Arg and Lys in the surrounding segments and only a few central nonpolar amino acids may be classified only as a binding site. Nevertheless, our results demonstrate that only a fraction (20-56%) of nonpolar protein regions are Hsp70 binding sites.

Many hydrophobic sites are not Hsp70 binding sites, suggesting that these regions do not require chaperone assistance to be properly folded. For instance, SH3 has multiple nonpolar sites, but it can fold correctly and has no tendency to aggregate even in the absence of chaperones.[68-69] However, proteins whose nonpolar residues correspond to Hsp70 binding sites likely rely on chaperones to fold correctly. Interestingly, the ratio of total nonpolar sites to Hsp70 binding sites is much higher for proteins with one binding site than proteins with multiple binding sites (Figure 5-4B, Table 5-2). In addition, larger proteins (30-140 kDa) tend to have more binding sites relative to the number of nonpolar sites than smaller proteins (5-30 kDa) (Figure 5-4C, Table 5-2).

Experimental studies demonstrated that proteins with multiple domains are more likely to misfold than single-domain proteins.[70] Therefore, we hypothesize that, due to their increased folding complexity, large proteins need more chaperones to provide folding assistance, thus requiring more chaperone binding sites. On the contrary, small proteins have relatively simple structures and tend to be soluble, hence minimal chaperones are needed to grant solubility and correct native structure. Again, our results show that proteins have evolved to optimize their amino-acid sequence and most productively exploit Hsp70 chaperone binding sites.

Table 5-2. Statistical p values comparing ratio of the and number of nonpolar sites to the number of Hsp70 binding sites for the subsets of client protein in Figure 5-4B,C. These p values were determined with the two-tailed Student's test assuming unequal variances. P values ≤ 0.05 denote statistically different data and are displayed in green.

	$\frac{\text{Number of nonpolar sites}}{\text{Number of Hsp70 binding sites}}$
Comparing proteins with one Hsp70 binding site to proteins with multiple Hsp70 binding sites (Figure 4b)	3.11×10^{-15}
Comparing small proteins (5-30 kDa) to large proteins (30-140 kDa) (Figure 4c)	8.7×10^{-6}

Hsp70 binding sites are more solvent-exposed than other nonpolar regions in the native structure. Next, we evaluated and compared the solvent exposure of the Hsp70 binding sites and other nonpolar regions. To determine the degree of burial of the Hsp70 binding sites within the native structure, we calculated the ratio of solvent accessible surface area (SASA) of the Hsp70 binding sites in the native and unfolded protein structures (Table 5-3). The SASA ratio of a sequence of 13 amino acids ranges from 0.001 for fully buried residues to 0.85 for fully exposed residues. The average SASA ratio for Hsp70 binding sites is 0.22 ± 0.01, indicating that the binding sites are mostly buried within the native structure (Figure 5-5A, Tables 5-3,5-4). Interestingly, the solvent-exposed nonpolar and positively charged residues within the Hsp70 binding sites of client proteins may interact with the nonpolar and negatively charged regions of the Hsp70 chaperone. These interactions may take place both near the primary Hsp70 substrate binding region and elsewhere on the chaperone surface (Figure 5-5B,C). The burial of binding sites within the native structure may be important for preventing chaperones from binding to client proteins that are already folded and do not require chaperone assistance. In addition, these results suggest that proteins cannot fully fold while chaperone bound. These results are consistent with previous studies that reported that most Hsp70 binding sites are buried within protein cores.[31] Our findings are also consistent with previous work that showed that Hsp70 generally binds unfolded or non-native states, instead of fully folded, native states of client proteins.[39, 52, 71-75]

Next, we compared the solvent accessibility of surface area of binding sites to that of other nonpolar regions. The average SASA ratio of Hsp70 binding sites (0.22 ± 0.01) is slightly higher than that of other nonpolar sites (0.17 ± 0.01), demonstrating that the binding sites are more solvent-exposed than other nonpolar regions in the structure (Figure 5-5, Tables 5-3,5-4, and

Supplementary Figure 5-S1). An overview of representative Hsp70 client proteins bearing significantly solvent-exposed chaperone binding sites is shown in Figure 5-6. The greater solvent exposure of Hsp70 binding sites may stem from the positive charges in the flanking segments of the binding motif, which render it less hydrophobic. This finding provides clues into the Hsp70 chaperone potential involvement during the folding. We hypothesize that some proteins may be able to partially fold while bound to Hsp70. The protein may first bury nonpolar regions that are not Hsp70 binding sites before it is released from the chaperone. The protein may then complete its folding, including burial of chaperone binding sites, upon release from the chaperone. This folding mechanism could explain why chaperone binding sites tend to be more solvent accessible than other hydrophobic patches in the fully folded conformation. This potential folding mechanism also reduces the amount of nonpolar surface area the protein needs to bury at one given time. Hsp70 binding sites are still mostly buried within the protein. As a result, client proteins are unlikely to fold completely on the chaperone, and they may still require significant burial of the nonpolar surface of the binding sites upon release from the chaperone. The above hypotheses will be experimentally tested in the future.

5.4 Conclusions

In summary, our results show that the number of Hsp70 binding sites within proteins depends on both protein size and solubility. This feature may enable organisms to make the most effective use of chaperones. We demonstrate that there is a strong linear relation between the number of Hsp70 binding sites and client-protein size for all *E. coli* proteins, with binding sites appearing on average every 44 amino acids. Interestingly, our results show that proteins with zero and one Hsp70 binding site generally do not need chaperone assistance for solubility. Proteins that have

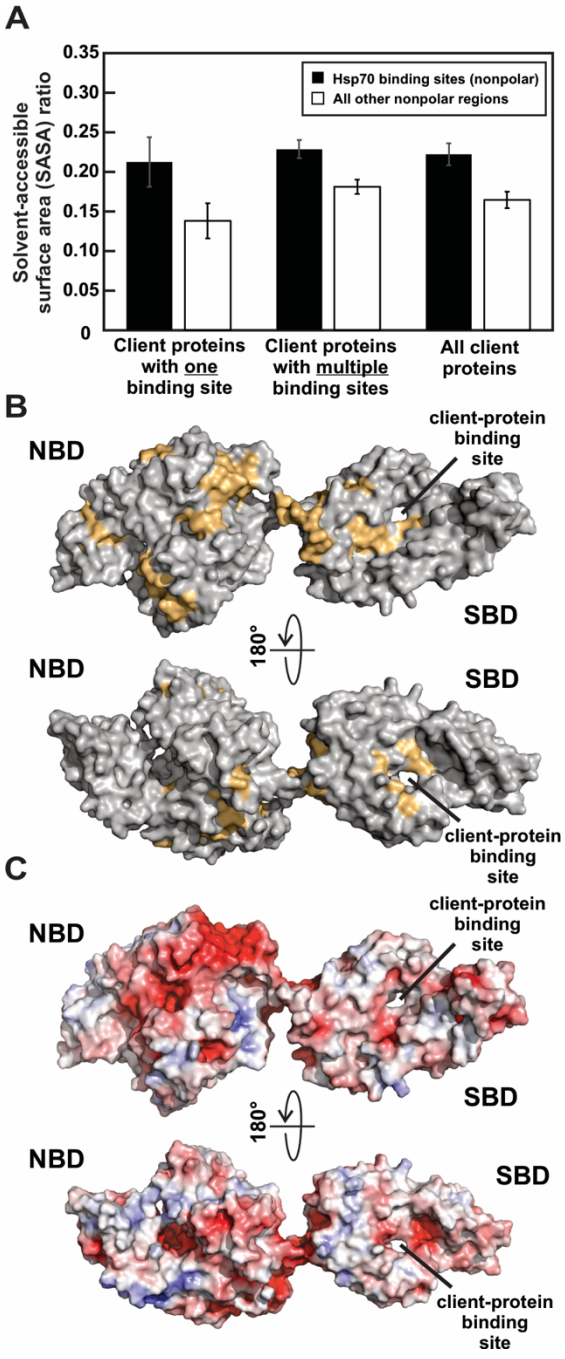


Figure 5-5. Hsp70 binding sites are more solvent-exposed than other nonpolar sites. (A) The average SASA ratio is greater for Hsp70 binding sites than for other nonpolar regions. Error bars denote standard error. (B) Space filling representation of ADP-bound or nucleotide-free Hsp70 (PDB: 2KHO). Hydrophobic chains with a fraction of mean buried area > 0.775 are

shown in light orange. (C) Electrostatic surface potential map of ADP-bound or nucleotide-free Hsp70 (PDB: 2KHO). Blue denotes regions of positive potential (>+1 k_BT/e, where k_B and T denote the Boltzmann constant and the temperature, respectively), and red denotes regions of negative potential (<−1 k_BT/e). Electrostatic surface potentials were generated with the APBS algorithm in the PyMOL software package. SBD and NBD denote the substrate-binding and nucleotide-binding domains of Hsp70, respectively.

Table 5-3. Solvent exposure of Hsp70 binding sites and other nonpolar sites. Hsp70 binding sites have a higher average solvent accessible surface area (SASA) and SASA ratio (relative to the unfolded state) than other nonpolar sites. Values are listed as average ± standard error. Standard deviations and standard errors were evaluated across the set of analyzed proteins.

	Hsp70 binding sites	Nonpolar sites that are not Hsp70 binding sites
Average solvent accessible surface area (SASA) ($Å^2$)	36.98 ± 1.83	26.52 ± 1.33
SASA ratio $\left(\frac{SASA_{folded}}{SASA_{unfolded}}\right)$	0.22 ± 0.01	0.17 ± 0.01

Table 5-4. Statistical p values for solvent accessible surface area ratios (SASA ratios) of the subsets of Hsp70 client proteins in Table 5-3 and Figure 5-5A. These p values were determined via the two-tailed Student's test assuming unequal variances. P values > 0.05 indicate statistically equivalent data and are shown in orange. P values ≤ 0.05 indicate statistically different data and are displayed in green.

	Binding sites of client proteins with one Hsp70 binding site	Nonpolar sites of client proteins with one Hsp70 binding site	Binding sites of client proteins with multiple Hsp70 binding sites	Nonpolar sites of client proteins with multiple Hsp70 binding sites	Binding sites of all client proteins	Nonpolar sites of all client proteins
Binding sites of client proteins with one Hsp70 binding site		0.0621	0.6307	0.3481	0.7735	0.1608
Nonpolar sites of client proteins with one Hsp70 binding site			0.0009	0.0868	0.0027	0.2953
Binding sites of client proteins with multiple Hsp70 binding sites				0.0012	0.7262	0.0001
Nonpolar sites of client proteins with multiple Hsp70 binding sites					0.0135	0.2323
Binding sites of all client proteins						0.0012
Nonpolar sites of all client proteins						

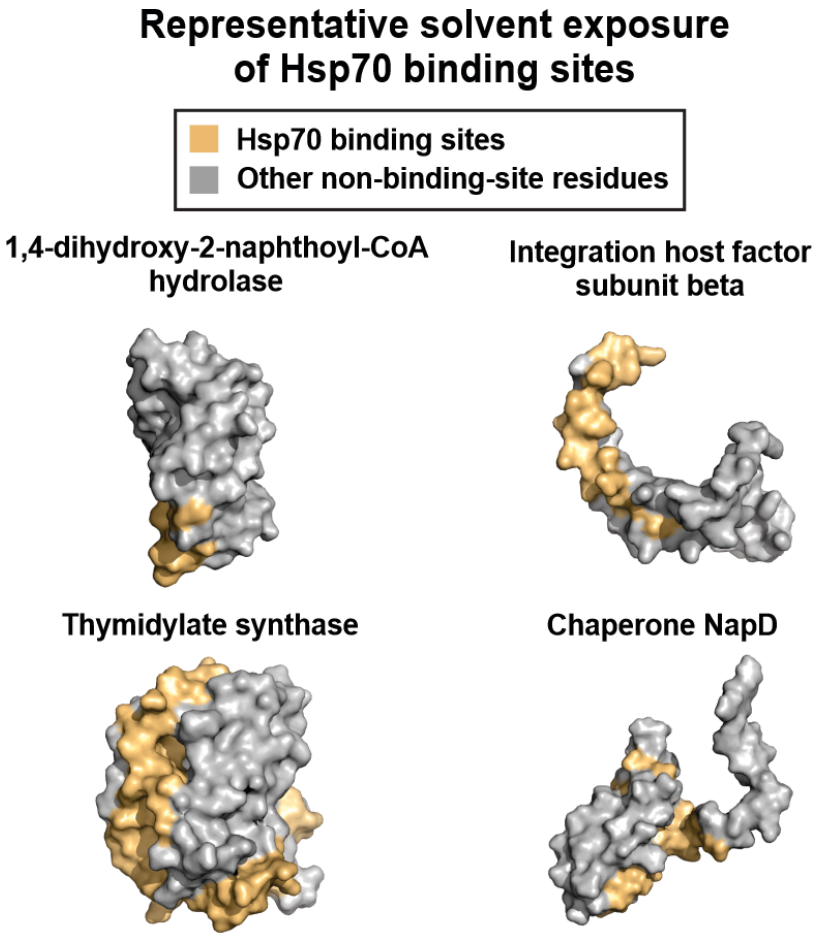


Figure 5-6. Representative solvent exposure of Hsp70 binding sites. Four representative proteins with various molecular weight, hydrophobicity and net charge are chosen. Top row: 1,4-dihydroxy-2-naphthoyl-CoA hydrolase (PDB: 4K4B) and integration host factor subunit beta (PDB: 1OWG) are examples of proteins with one Hsp70 binding site; bottom row: thymidylate synthase (PDB: 1AIQ) and chaperone NapD (PDB: 2JSX) are examples of proteins with multiple Hsp70 binding sites. Hsp70 binding sites are highlighted in yellow.

multiple binding sites, however, show a large range of solubilities, and only 35% show greater than 70% solubility. In addition, only a small portion of all nonpolar sites are Hsp70 binding sites, suggesting that some nonpolar patches do not require Hsp70 to become buried from solvent. Interestingly, Hsp70 binding sites are more solvent-exposed than other nonpolar sites in protein native states. In all, our findings suggest that the number of Hsp70 binding sites within client proteins is optimized to minimize unnecessary chaperone-protein interactions and keep chaperones available for proteins that require them. Finally, Hsp70 binding sites tend to be largely buried, demonstrating that proteins cannot fold completely while they are chaperone-bound. Importantly, Hsp70 binding sites are more solvent-exposed than other nonpolar regions of the native state. Greater solvent accessibility of the binding sites may allow non-chaperone-bound regions of the protein to fold and start burying nonpolar residues before the chaperone releases the protein. Overall, our work provides valuable insights into how the Hsp70 binding site distribution enables the *E. coli* bacterium, and potentially other organisms, to make the most effective use of this remarkable molecular chaperone.

5.5 Materials and Methods

Identification of Hsp70 binding sites. Using the sequence-based algorithm developed by Rüdiger *et al.*, we identified Hsp70 binding sites for 2,258 *E. coli* cytoplasmic proteins.[31] These proteins were selected due to their readily available solubility information as reported by Niwa *et al.*[13] The Hsp70 binding site prediction algorithm is based on the Hsp70 binding motif which consists of a five-residue hydrophobic core flanked on either side by four positively charged residues. The algorithm determines a score for each amino acid's energy contribution to binding to Hsp70. The combined energy value of a specific sequence predicts the likelihood that Hsp70

binds to that region. We used a sequence window size of 13 amino acids and an energy cutoff of ≤-5 to identify binding sites. Using this same window size and energy cutoff, Rüdiger *et al.* correctly predicted 95% of experimentally observed Hsp70 binding sites.[31]

Identification of hydrophobic sites. Mean fractional area loss, denoted as f by Rose *et al.*, was used as a measure of hydrophobicity.[76] The mean fractional area loss of each amino acid residue is determined according to

$$f = (A^0 - < A >)/A^0 \quad . \tag{2}$$

where A^0 is the solvent accessible surface area of a given residue in the stochastic standard state, defined as the mean surface area of an ensemble of extended tripeptides Gly-X-Gly, where X is the residue of interest; $< A >$ is the average solvent accessible surface area of that residue in natively folded proteins; $A^0 - < A >$ is the mean area buried on transfer from the standard state to the folded protein.

The mean fractional area loss is the average area a residue buries upon folding, normalized by its standard state area. This parameter can be used to classify amino acids as hydrophobic, moderately polar, and very polar. We computed the combined mean fractional area loss using ProtScale Analysis Tool with a window size of five amino acids. Residues with f > 0.785 were classified as nonpolar sites. We identified the nonpolar sites in all 86 proteins with one Hsp70 binding site and in 200 randomly selected proteins with multiple binding sites.

Determination of solvent accessibility of Hsp70 binding sites and nonpolar regions. We determined the solvent exposure of the nonpolar and Hsp70 binding sites of a subset of the previous set of 286 proteins. To choose these proteins, we identified all the proteins with solved crystal structures (34 proteins with one binding site, 75 proteins with multiple binding sites). We generated unfolded structures for these proteins with the Bax group NMR Utility Server, with

dihedral angles ψ = 120° and ϕ= -120°.We then determined the solvent accessible surface area (SASA) of Hsp70 binding sites and other nonpolar sites in the native conformation for these proteins using the Surface Racer computational program.[77] We calculated the SASA ratio for Hsp70 binding sites (SASA $ratio_{Hsp70}$) and for nonpolar regions (SASA $ratio_{nonpolar}$) by dividing the SASA in the unfolded structure by the SASA in the folded structure as shown in eqs. 3-4.

$$SASA\ ratio_{Hsp70} = \frac{SASA_{Hsp70,\ folded}}{SASA_{Hsp70,\ unfolded}}\ , \tag{3}$$

$$SASA\ ratio_{nonpolar} = \frac{SASA_{nonpolar,\ folded}}{SASA_{nonpolar,\ unfolded}}\ . \tag{4}$$

Generation electrostatic surface-potential maps of Hsp70 chaperone. PDB files were converted to PQR format using the PDB2PQR software and used as input for electrostatic surface potential calculations solving the nonlinear Poisson-Boltzmann equation with the Adaptive Poisson-Boltzmann Solver (APBS) software package. Hsp70 was treated as a dielectric continuum with dielectric constant 2.0 embedded in a solvent of dielectric constant 78.0. In Figure 5-5C, the molecular surface is colored according to the electrostatic charge of the solvent-excluded surface area. Blue denotes regions of positive potential (>+1 k_BT/e) and red denotes regions of negative potential (<-1k_BT/e). k_BT is 4.11 X 10^{-21} J at room temperature where k_B is the Boltzmann constant in J/Kelvin (K), T is the temperature in K, and e is the charge of a single electron in Coulomb. Electrostatic surface potential maps were generated directly within the PyMOL software package 2.4.0 (PyMOL Molecular Graphics System, Schrödinger, LLC) via the APBS plug-in.

Multiple linear regression analysis. Multiple linear regression analysis was performed using the regression tool in the Excel Data Analysis Tool Pack (Microsoft). The number of Hsp70 binding sites was set as the dependent variable and protein size and solubility were set as the independent variables. The variance inflation factor was determined from the R^2 value of the

linear fit of protein size as a function of protein solubility:

$$\frac{1}{1\text{-}R^2} \quad . \qquad (5)$$

5.6 Acknowledgments

We are grateful to Frank Delaglio and Ad Bax for advice and assistance with the surface-area calculations. We thank Lucas Bartel for help with the variable regression analysis. We are grateful to the National Science Foundation (NSF) for funding (grants MCB-2124672 and CBET 1912259 to S.C.). R.B.H. also thanks the National Institute of General Medical Sciences of the National Institute of Health for a TEAM-Science Fellowship (award number R25GM083252).

5.7 References

(1) Deuerling, E.; Schulze-Specking, A.; Tomoyasu, T.; Mogk, A.; Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **1999,** *400*, 693-6.

(2) Teter, S. A.; Houry, W. A.; Ang, D.; Tradler, T.; Rockabrand, D.; Fischer, G.; Blum, P.; Georgopoulos, C.; Hartl, F. U. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **1999,** *97*, 755-765.

(3) Genevaux, P.; Keppel, F.; Schwager, F.; Langendijk-Genevaux, P. S.; Hartl, F. U.; Georgopoulos, C. In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep* **2004,** *5*, 195-200.

(4) Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H. Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 8937-8942.

(5) Frydman, J. Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu. Rev. Biochem* **2001,** *70*, 603-647.

(6) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. Recent advances in understanding catalysis of protein folding by molecular chaperones. *FEBS Lett.* **2020,** *594*, 2770-2781.

(7) Clerico, E. M.; Meng, W.; Pozhidaeva, A.; Bhasne, K.; Petridis, C.; Gierasch, L. M. Hsp70 molecular chaperones: Multifunctional allosteric holding and unfolding machines. *Biochem. J.* **2019,** *476*, 1653-1677.

(8) Fatima, K.; Naqvi, F.; Younas, H. A review: Molecular chaperone-mediated folding, unfolding and disaggregation of expressed recombinant proteins. *Cell Biochem. Biophys.* **2021,** *79*, 153-174.

(9) Dahiya, V.; Buchner, J. Functional principles and regulation of molecular chaperones. *Adv. Protein Chem. Struct. Biol.* **2019,** *114*, 1-60.

(10) Nillegoda, N. B.; Wentink, A. S.; Bukau, B. Protein disaggregation in multicellular organisms. *Trends Biochem. Sci.* **2018,** *43*, 285-300.

(11) Tans, S. J. Picturing protein disaggregation. *Nat. Chem. Biol.* **2022,** *18*, 240-241.

(12) Wyszkowski, H.; Janta, A.; Sztangierska, W.; Obuchowski, I.; Chamera, T.; Klosowska, A.; Liberek, K. Class-specific interactions between Sis1 J-domain protein and Hsp70 chaperone potentiate disaggregation of misfolded proteins. *Proc. Natl. Acad. Sci. USA* **2021,** *118*, e2108163118.

(13) Niwa, T.; Ying, B. W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 4201-4206.

(14) Nishikawa, S. I.; Fewell, S. W.; Kato, Y.; Brodsky, J. L.; Endo, T. Molecular chaperones in the yeast endoplasmic reticulum maintain the solubility of proteins for retrotranslocation and degradation. *J. Cell Biol.* **2001,** *153*, 1061-1070.

(15) Baneyx, F.; Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **2004,** *22*, 1399-1408.

(16) Rosenzweig, R.; Moradi, S.; Zarrine-Afsar, A.; Glover, J. R.; Kay, L. E. Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. *Science* **2013,** *339*, 1080-1083.

(17) Witt, S. N. Hsp70 molecular chaperones and Parkinson's disease. *Biopolymers* **2010,** *93*, 218-228.

(18) Repalli, J.; Meruelo, D. Screening strategies to identify Hsp70 modulators to treat Alzheimer's disease. *Drug Des. Devel. Ther.* **2015,** *9*, 321-331.

(19) Chiti, F.; Dobson, C. M. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* **2017,** *86*, 27-68.

(20) Ciccocioppo, F.; Bologna, G.; Ercolino, E.; Pierdomenico, L.; Simeone, P.; Lanuti, P.; Pieragostino, D.; Del Boccio, P.; Marchisio, M.; Miscia, S. Neurodegenerative diseases as proteinopathies-driven immune disorders. *Neural Regen. Res.* **2020,** *15*, 850-856.

(21) Kim, D. K.; Kim, T. H.; Lee, S. J. Mechanisms of aging-related proteinopathies in caenorhabditis elegans. *Exp. Mol. Med* **2016,** *48*, 1-9.

(22) Lashuel, H. A. Rethinking protein aggregation and drug discovery in neurodegenerative diseases: Why we need to embrace complexity? *Curr. Opin. Chem. Biol.* **2021,** *64*, 67-75.

(23) Levy, E.; El Banna, N.; Baille, D.; Heneman-Masurel, A.; Truchet, S.; Rezaei, H.; Huang, M. E.; Beringue, V.; Martin, D.; Vernis, L. Causative links between protein aggregation and oxidative stress: A review. *Int. J. Mol. Sci* **2019,** *20*, 3896.

(24) Monaco, A.; Fraldi, A. Protein aggregation and dysfunction of autophagy-lysosomal pathway: A vicious cycle in lysosomal storage diseases. *Front. Mol. Neurosci.* **2020,** *13*, 37.

(25) Hunt, C.; Morimoto, R. I. Conserved features of eukaryotic Hsp70 genes revealed by comparison with the nucleotide sequence of human Hsp70. *Proc. Natl. Acad. Sci. USA* **1985,** *82*, 6455-6459.

(26) Daugaard, M.; Rohde, M.; Jaattela, M. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Lett.* **2007,** *581*, 3702-3710.

(27) Rebeaud, M. E.; Mallik, S.; Goloubinoff, P.; Tawfik, D. S. On the evolution of chaperones and cochaperones and the expansion of proteomes across the tree of life. *Proc. Natl. Acad. Sci. USA* **2021,** *118*, e2020885118.

(28) Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U. DnaK functions as a central hub in the e. coli chaperone network. *Cell Rep.* **2012,** *1*, 251-264.

(29) Bertelsen, E. B.; Chang, L.; Gestwicki, J. E.; Zuiderweg, E. R. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 8471-8476.

(30) Zhu, X.; Zhao, X.; Burkholder, W. F.; Gragerov, A.; Ogata, C. M.; Gottesman, M. E.; Hendrickson, W. A. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **1996,** *272*, 1606-1614.

(31) Rüdiger, S.; Germeroth, L.; Schneider-Mergener, J.; Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **1997,** *16*, 1501-1507.

(32) Mayer, M. P.; Schroder, H.; Rudiger, S.; Paal, K.; Laufen, T.; Bukau, B. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **2000,** *7*, 586-593.

(33) Mayer, M. P.; Rudiger, S.; Bukau, B. Molecular basis for interactions of the DnaK chaperone with substrates. *Biol. Chem.* **2000,** *381*, 877-885.

(34) Schmid, D.; Baici, A.; Gehring, H.; Christen, P. Kinetics of molecular chaperone action. *Science* **1994,** *263*, 971-973.

(35) Mecha, M. F.; Hutchinson, R. B.; Lee, J. H.; Cavagnero, S. Protein folding in vitro and in the cell: From a solitary journey to a team effort. *Biophys. Chem.* **2022,** *287*, 106821.

(36) Sekhar, A.; Santiago, M.; Lam, H. N.; Lee, J. H.; Cavagnero, S. Transient interactions of a slow-folding protein with the Hsp70 chaperone machinery. *Protein Sci.* **2012,** *21*, 1042-1055.

(37) Imamoglu, R.; Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein. *Nat. Commun.* **2020,** *11*, 1-13.

(38) Sharma, S. K.; De los Rios, P.; Christen, P.; Lustig, A.; Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **2010,** *6*, 914-920.

(39) Kellner, R.; Hofmann, H.; Barducci, A.; Wunderlich, B.; Nettels, D.; Schuler, B. Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proc. Natl. Acad. Sci. USA* **2014,** *111*, 13355-13360.

(40) Sekhar, A.; Nagesh, J.; Rosenzweig, R.; Kay, L. E. Conformational heterogeneity in the Hsp70 chaperone-substrate ensemble identified from analysis of NMR-detected titration data. *Protein Sci.* **2017,** *26*, 2207-2220.

(41) Rosenzweig, R.; Sekhar, A.; Nagesh, J.; Kay, L. E. Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *Elife* **2017,** *6*, e28030.

(42) Nordquist, E. B.; Clerico, E. M.; Chen, J.; Gierasch, L. M. Computationally-aided modeling of Hsp70–client interactions: Past, present, and future. *J. Phys. Chem. B* **2022,** *126*, 6780-6791.

(43) Nordquist, E. B.; English, C. A.; Clerico, E. M.; Sherman, W.; Gierasch, L. M.; Chen, J. Physics-based modeling provides predictive understanding of selectively promiscuous substrate binding by Hsp70 chaperones. *PLoS Comput. Biol.* **2021,** *17*, e1009567.

(44) Van Durme, J.; Maurer-Stroh, S.; Gallardo, R.; Wilkinson, H.; Rousseau, F.; Schymkowitz, J. Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput. Biol.* **2009,** *5*, e1000475.

(45) Gutierres, M. B. B.; Bonorino, C. B. C.; Rigo, M. M. Chaperism: Improved chaperone binding prediction using position-independent scoring matrices. *Bioinformatics* **2020,** *36*, 735-741.

(46) Schneider, M.; Rosam, M.; Glaser, M.; Patronov, A.; Shah, H.; Back, K. C.; Daake, M. A.; Buchner, J.; Antes, I. BiPPred: Combined sequence- and structure-based prediction of peptide binding to the Hsp70 chaperone BiP. *Proteins* **2016,** *84*, 1390-1407.

(47) Niwa, T.; Ying, B. W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H., Esol solubility database of all *E. coli* proteins. 2009.

(48) Mogk, A.; Tomoyasu, T.; Goloubinoff, P.; Rüdiger, S.; Röder, D.; Langen, H.; Bukau, B. Identification of thermolabile Escherichia coli proteins: Prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* **1999,** *18*, 6934-6949.

(49) Chung, H. J.; Bang, W.; Drake, M. A. Stress response of Escherichia coli. *Compr. Rev. Food Sci. Food Saf.* **2006,** *5*, 52-64.

(50) Clerico, E. M.; Tilitsky, J. M.; Meng, W.; Gierasch, L. M. How Hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.* **2015,** *427*, 1575-1588.

(51) Landry, S. J.; Jordan, R.; McMacken, R.; Gierasch, L. M. Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature* **1992,** *355*, 455-457.

(52) Chen, Z.; Kurt, N.; Rajagopalan, S.; Cavagnero, S. Secondary structure mapping of DnaK-bound protein fragments: Chain helicity and local helix unwinding at the binding site. *Biochemistry* **2006,** *45*, 12325-12333.

(53) Brodsky, J. L.; Chiosis, G. Hsp70 molecular chaperones: Emerging roles in human disease and identification of small molecule modulators. *Curr. Top. Med. Chem.* **2006,** *6*, 1215-1225.

(54) White, S. H.; Jacobs, R. E. Statistical distribution of hydrophobic residues in proteins - implications for protein folding. *Biophys. J.* **1990,** *57*, A431-A431.

(55) Schwartz, R.; King, J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. *Protein Sci.* **2006,** *15*, 102-112.

(56) Schwartz, R.; Istrail, S.; King, J. Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. *Protein Sci.* **2001,** *10*, 1023-1031.

(57) Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S. Complementary role of co- and post-translational events in de novo protein biogenesis. *J. Phys. Chem. B* **2020,** *124*, 6488-6507.

(58) Sekhar, A.; Lam, H. N.; Cavagnero, S. Protein folding rates and thermodynamic stability are key determinants for interaction with the Hsp70 chaperone system. *Protein Sci.* **2012,** *21*, 1489-502.

(59) Robertson, A. D.; Murphy, K. P. Protein structure and the energetics of protein stability. *Chem. Rev.* **1997,** *97*, 1251-1268.

(60) Ghosh, K.; Dill, K. Cellular proteomes have broad distributions of protein stability. *Biophys. J.* **2010,** *99*, 3996-4002.

(61) Ghosh, K.; Dill, K. A. Computing protein stabilities from their chain lengths. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 10649-10654.

(62) Varela, A. E.; England, K. A.; Cavagnero, S. Kinetic trapping in protein folding. *Protein Eng. Des. Sel.* **2019,** *32*, 103-108.

(63) Bukau, B.; Walker, G. C. Mutations altering heat shock specific subunit of rna polymerase suppress major cellular defects of E. coli mutants lacking the DnaK chaperone. *EMBO J.* **1990,** *9*, 4027-4036.

(64) Hesterkamp, T.; Bukau, B. Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in e.Coli. *EMBO J.* **1998,** *17*, 4818-28.

(65) Richards, F. M. Structure of proteins. *Annu. Rev. Biochem.* **1963,** *32*, 269-300.

(66) Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **1974,** *248*, 338-339.

(67) Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **1976,** *105*, 1-12.

(68) Grantcharova, V. P.; Baker, D. Folding dynamics of the src SH3 domain. *Biochemistry* **1997,** *36*, 15685-15692.

(69) Mazouchi, A.; Zhang, Z.; Bahram, A.; Gomes, G. N.; Lin, H.; Song, J.; Chan, H. S.; Forman-Kay, J. D.; Gradinaru, C. C. Conformations of a metastable SH3 domain characterized by smFRET and an excluded-volume polymer model. *Biophys. J.* **2016,** *110*, 1510-1522.

(70) 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.

(71) Kurt, N.; Rajagopalan, S.; Cavagnero, S. Effect of Hsp70 chaperone on the folding and misfolding of polypeptides modeling an elongating protein chain. *J. Mol. Biol.* **2006,** *355*, 809-820.

(72) Sekhar, A.; Velyvis, A.; Zoltsman, G.; Rosenzweig, R.; Bouvignies, G.; Kay, L. E. Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. *Elife* **2018,** *7*, e32764.

(73) Sekhar, A.; Rosenzweig, R.; Bouvignies, G.; Kay, L. E. Hsp70 biases the folding pathways of client proteins. *Proc. Natl. Acad. Sci. USA* **2016,** *113*, E2794-E2801.

(74) Palleros, D. R.; Shi, L.; Reid, K. L.; Fink, A. L. Hsp70-protein complexes. Complex stability and conformation of bound substrate protein. *J. Biol. Chem.* **1994,** *269*, 13107-13114.

(75) Lee, J. H.; Zhang, D.; Hughes, C.; Okuno, Y.; Sekhar, A.; Cavagnero, S. Heterogeneous binding of the SH3 client protein to the DnaK molecular chaperone. *Proc. Natl. Acad. Sci. USA* **2015,** *112*, E4206-E4215.

(76) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. *Science* **1985,** *229*, 834-838.

(77) Tsodikov, O. V.; Record, M. T., Jr.; Sergeev, Y. V. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* **2002,** *23*, 600-609.

5.8 Appendix

Supporting figures

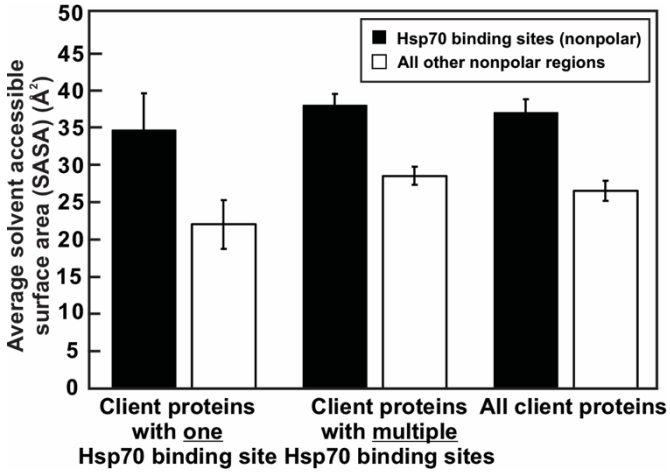


Figure 5-S1. Average solvent accessible surface area (SASA) of Hsp70 binding sites and all other nonpolar regions. These data show that the average SASA of Hsp70 binding sites is much greater than that of all other nonpolar regions across cytoplasmic *E. coli* proteins, suggesting that Hsp70 binding sites are more solvent exposed within the native state of proteins. SASA per site was calculated as the average SASA values of all amino acids belonging to a binding site or nonpolar site. Error bars denote standard error.