Insights into some key aspects of protein folding

I. Relation between protein folding and aggregation under physiologically relevant conditions

II. Probing the influence of newly synthesized proteins on ribosomal components

by Angela E. Varela

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biophysics)

at the University of Wisconsin-Madison

2018

Date of final oral examination: November 16, 2018, 11:15 am, 9341 Chemistry
This dissertation is approved by the following members of The Final Oral Committee
Silvia Cavagnero, Professor, Chemistry
David Brow, Professor, Biomolecular Chemistry
Judith Burstyn, Professor Chemistry
Jeffrey Hardin, Professor, Integrative Biology
Gail Robertson, Professor, Neuroscience

To my parents, who pray for me daily.

Philippians 4:13

First and foremost I would like to thank God for the grace He has bestowed upon me throughout my life, especially during the last 6 years. I would not be here if it weren't for the unrelenting, unwavering, and undeserving favor of God and for that, I am forever grateful.

I would like to thank Silvia Cavagnero for her constant support and her belief in me. Without her guidance, I could not have reached this point in my Ph.D. career. Thank you for all the time and love you put into your work and us graduate students and thank you for loving what you do day in and day out.

I would like to thank my committee members, Dr. Judith Burstyn, Dr. David Brow, Dr. Gail Robertson and Dr. Jeff Hardin, for their helpful advice and support throughout the last few years.

I would like to thank my undergraduate mentor, Dr. Emina Stojkovic for believing in me. You work incredibly hard for your students and it is because of your unrelenting spirit that we have been able to find a place for ourselves in the scientific world. Thank you for your passion towards creating a more diverse scientific community.

To the entire Cavagnero group: I am so grateful for each and every one of you.

Special thanks my previous collaborators, who I'd like to thank individually:

To Anders Knight, thank you for training me and being the first person I collaborated with in the lab. Your guidance has not been forgotten. (I still order my Eppendorf tubes blue, green, orange, yellow, and pink because of you).

To Jon Lang, your patience, kindness and mentorship has been much appreciated and I am so glad we were able to work together.

To Aniruddha Srivastava, you are one of the kindest, smartest, and hardworking people I have ever met. God was looking out for me when we were put on the same project. You have been

the best lab partner I could have asked for. I could never repay you for all the work you've put in to this project and all the help you have given me. Thank you.

To Andrew Stangl, your excitement for science both in and outside of the lab is awe-inspiring. You work incredibly hard and I have tried to emulate your gusto in my work. Thank you for always being there for me and letting me use your acrylamide for years. I am so glad I can call you 'friend'. Never change.

I would also personally like to thank current collaborators in the group:

To Meranda Masse, I could not have imagined a better graduate student to work with during my last year. You've been so helpful, understanding and kind. I've always said that I've worked with the best of the best in the lab and working with you has proven that statement time and time again. Thank you for always looking on the bright side of things. I am excited for you as you continue this journey and I can't wait to see where life takes you.

To Matthew Dalphin, words cannot express how grateful I am to know you. I love your excitement and enthusiasm for the things you love-it's magnetizing. Thank you for your invaluable advice in scientific discussions (and in life/food choices) and for always being there when I needed a friend to talk. You are one of the best teachers I know. Go get that burrito, you deserve it.

I would like to personally acknowledge current members of the Cavagnero group who were monumental in my support system during my Ph.D. career:

To Miranda Mecha, words cannot express how thankful I am that our paths crossed. We may have never worked together on a project but your presence in this lab has been one of the best things that could have happened to all of us. I don't know what will happen when you leave but I am so thankful that you got there right when I needed your friendship.

To Rachel Hutchinson, Hanming Yang and Andrew Fuchs, thank you for all the answers to the many questions I've asked you throughout the years, for all your helpful advice-I could always count on all three of you whenever I wasn't sure about something. Rachel, you are amazing. Hanming, you are wonderful. Andrew, you are awesome. Thank you for all the fun conversations about life and science over the years. I will miss you three very much.

I would also like to acknowledge a few people whose presence really impacted my time in graduate in the best way possible: Dongyu (Allison) Zhang, Jung Ho Lee, Yue Liu, and Valeria Guzman: your friendships during my time in Madison have meant the world to me.

Ken and Yusuke, thank you both for always being there when I needed a friend. Words could never express my gratitude for your presence in my life; you helped me through some of the most difficult times and you both are more like my brothers than friends.

To Heejun, thank you so much for always being there for me. You and Yusuke helped me so much academically and personally and I am so happy that our paths crossed when they did. I miss your smiles, your macarons and most of all, your hugs.

To Carlos and Jean, thank you for making my first teaching experience fly by with ease. Your support helped make that semester a very successful one.

To my family:

To my Madison family (Dave, Jean, Gloria, Frank, Abigail, Anneka, Jeremy, Charlotte, Bart, Matt, Katie, Seth, Gaby, Don and everyone else in CS@UW), thank you for your love and support throughout these years. I found you all at a very difficult time in my life and little did I know the impact you would have on me. I could never truly express my gratitude for your love, faith and prayers. I love you all very much.

To my amazing husband and best friend, Alexander, you have supported me no matter what and I am forever grateful for you. Thank you for being my biggest fan and believing in me when I could not believe in myself. You've been in my corner since day one and you've been there to talk me out of some of my worst moments. Thank you for always seeing the big picture when all I could see was the current situation. You are truly my better half and coming home to you (and our pets) will always be the highlight of my day. Home is where you are, my love.

To my parents, Christine and Ed, I love you both so much. Thank you for visiting me throughout my time here. You have no idea how much I needed your companionship after moving away from home. I am so glad you are mine and I am your favorite. I could never repay you for all you've done for me.

To my brother Joseph, thank you for being one of my favorite people. I love you.

To my Turton family, Aunt Gloria, Josh, Pam, Aunt Myrtle, and everyone else: thank you all for your constant support and presence in my life. I love you all.

To my four-legged furry friends:

To Jack, thank you for being a constant reminder to not take life too seriously. You are the brightest star in the night sky. I love you.

To Andy, you have always been yourself and I love you more for it. You have taught me to go after what I want and not to be ashamed of it. I will do my best to make you proud.

To Scrappy, you are my biggest cheerleader and coming home to Chicago to see you has been one of the best highlights of the last six years. Thank you for being the most loyal, vigilant, supportive and smartest dog I have ever known. I strive to be worthy of your love.

To Confetti, thank you for being my constant companion during all the late, late nights I spent working on this thesis. You are my favorite reading buddy.

To Jewel, I am so grateful you came into the lives of my family. We are so lucky to have you and I can't wait to get to know you more, little one.

Last but not least, I'd like to acknowledge a few important four-legged companions who could not be with me today, but their impact on my life (especially during these last few years) has been significant.

To my Toffee, I love you very much and I am so grateful you chose us.

January 11, 2006 – April 13, 2018

To my Myles, adopting you during graduate school was hands down the second best thing I've ever done. You brought so much joy and love into my life the moment I picked you up from the shelter. You made our apartment a home. Seeing you each and every evening will always be one my favorite graduate school memories. No day was a bad day when I got to come home to your wagging tail. You exuded love and you taught me to love more fiercely. I love you with all my heart. Thank you for getting me through some of my hardest days.

September 17, 2014 – June 2, 2018

To my Franky, everything I've done in life, I've done with you in mind. You've been a constant companion of mine for most of my life and you always inspired me to do my best to make you proud. Graduate school was extremely difficult, but visiting you always made it better. Thank you for being the best friend I could have ever dreamed up. I think of you every day, my little love, and I will love you until my heart stops. I am so thankful you were mine.

August 2, 1996 – March 30, 2015

Table of Contents

Abstract

We still know very little about protein folding in the cell and what parameters and molecular machines cause soluble proteins to achieve their native conformation. Although Anfinsen's thermodynamic hypothesis explains the origin of protein stability, it does not account for the presence of protein aggregates and where they lie in an energy landscape relative to the native state, under physiologically relevant conditions. Protein folding starts within the ribosome during biosynthesis. While the ribosome and cotranslationally active molecular chaperones help nascent chains achieve their native state, little is known about the influence of nascent polypeptides and proteins on the ribosome and its apparent stability. In this thesis I show that Anfinsen's thermodynamic hypothesis needs to be modified to include soluble and insoluble protein aggregates because most proteins are kinetically trapped relative to their aggregated states. In addition, I explore some of the determinants of the apparent stability of the 70S ribosome in *E. coli*.

This thesis is divided into four chapters. Chapter 1 explains Anfinsen's thermodynamic hypothesis together with its major implications, and outlines the major aspects of the ribosome structure and function. This chapter serves as the conceptual background for the remaining portions of this document.

Chapters 2 and 3 explore the concept of protein kinetic trapping under physiologically relevant conditions and demonstrates the need to extend Anfinsen's hypothesis to include protein aggregates. More specifically, Chapter 2 introduces an experimental strategy denoted as the "cyclic perturbation approach" to show that two structurally distinct, non-amyloid aggregate states of sperm whale apomyoglobin are kinetically trapped relative to each other and relative to the native state. The experiments presented in Chapter 3 show that kinetic trapping is a more general

phenomenon than previously thought, and that it can be extended to the large majority of soluble proteins in *E. coli*. A series of kinetic simulations confirms this phenomenon.

Chapter 4 showcases the influence of the nascent chain on the apparent stability of the 70S *E. coli* ribosome. The presence of the newly synthesized polypeptide was found to stabilize the ribosome. Surprisingly, characteristic properties of the nascent chain, including its N- and C-terminal residues, net charge, and length (beyond 32 residues) do not have any influence on the structural stabilization of the 70S complex by the nascent polypeptide. This chapter also introduces a working model for denaturant-induced ribosomal disassembly.

Overall, this dissertation shows that the kinetic trapping of native, unfolded and intermediate protein states relative to aggregates under physiologically relevant conditions is a general phenomenon in Nature. In addition, it shows that nascent proteins contribute to increasing the apparent stability of the bacterial ribosome.

Chapter 1

Introduction

CHAPTER 1

1.1 General aspects of protein folding

Understanding how proteins fold in the cellular context is one of the most compelling unsolved problems in modern biology. While researchers studied the balance between protein folding and aggregation for years, we still know very little about how proteins achieve their native state.

About five decades ago, Cyrus Levinthal theorized that unfolded proteins have to refold via specific pathways (as opposed to via random routes), given that the available time for folding limits the number of conformations that a protein is able to sample (Figure 1-1 A).[1].

Within the ensemble of states that can be populated at any given temperature, Anfinsen's thermodynamic hypothesis states that, upon folding, a protein reaches its most thermodynamically stable conformation (Figure 1-1 B). This state is only determined by the amino acid sequence, in a given environment.[2] Therefore, according to Anfinsen and given the above constraints, the protein folding process is currently believed to be mostly under thermodynamic control (Figure 1-2).

For instance, Anfinsen showed that ribonuclease A folds under thermodynamic control since it can refold and regain its activity after treatment with urea and 2-mercaptoethanol. This important experiment showed that proteins can adopt their native conformation based on their primary structure alone.[2] In addition, most small single-domain and monomeric proteins are able to fold and unfold fast and reversibly, suggesting that the folding of these proteins is also governed by thermodynamic principles.[3-4] Moreover, the fact that proteins can achieve the same native state both *in vivo* and *in vitro* also supports thermodynamic control of protein folding.

Although most proteins are known to follow Anfinsen's thermodynamic hypothesis, there are a few that fold under kinetic control. An interesting example is α-lytic protease (Figure 1-3 A), which possesses an intermediate state that is kinetically trapped as a result of high barriers unless its sequence contains a covalently bound pro region, which serves as a folding catalyst.[5] A representative chemical-potential landscape for this protein is provided in Figure 1-3 B. The pro region also serves the purpose of stabilizing the folded state (corresponding to the N' species in Figure 1-3 B), thus thermodynamically driving its formation. The pro region is then cleaved, to give rise to the metastable and kinetically trapped, yet biologically active, native state (corresponding to the N species in Figure 1-3 B).[6] Another example is that of proteins from the serpin family (Figure 1-3 C), which initially fold into an active protein form but slowly convert into a stable, inactive conformation over time (Figure 1-3 D).[5, 7]

In summary, proteins have, in most cases, evolved to allow the unfolded state to convert into the bioactive native state under thermodynamic control. However, in contrast with kinetic trapping relative to unfolded or intermediates, a protein may also be trapped relative to its aggregated states. This thesis demonstrates that this concept applies to single, purified proteins. In addition, it illustrates this idea in the case of most of *E. coli* soluble proteins. In all, a broader view of Anfinsen's hypothesis is necessary to properly take into account the multiple states of proteins (native, unfolded, intermediates, and aggregates).

1.2 Expanding Anfinsen's thermodynamic hypothesis to include protein aggregates

As a result of Anfinsen's pioneering work, many researchers developed various protein folding and aggregation pathway models. These models typically apply to non-physiologically relevant concentrations and do not emphasize the kinetic barriers between the native and aggregated protein

states.[8] While it has been suggested that kinetic barriers between bioactive protein states and aggregates must be large[9], little has been done to characterize them. To date, one experimental study briefly alluded to the idea that high kinetic barriers exist between native proteins and their aggregates.[10]

Research by Baldwin and coworkers showed that small, amyloid-prone proteins (< 150 amino acids) often exceed their critical concentrations for aggregation *in vivo* so that amyloid fibril formation should, in principle, occur. However, no aggregation is, in practice, observed.[10-11] This works shows that the cellular machinery ensures that peptides and small proteins remain kinetically trapped relative to amyloid aggregates.

Chapters 2 and 3 of this thesis significantly expand upon the latter concept by demonstrating that a model eukaryotic protein (apomyoglobin) as well as most soluble *E. coli* proteins are kinetically trapped relative to nonamyloid aggregates, under physiologically relevant conditions (Figure 1-4).

Chapter 2 specifically characterizes apomyoglobin after it has undergone a cyclic pathway that systematically changes its state from unfolded to native to aggregated under physiologically relevant conditions without introducing any covalent modifications. Long-term incubation steps were included to gain explicit evidence about kinetic stability. In order to demonstrate the generality of our original findings, Chapter 3 further extends the concept of kinetic trapping relative to aggregates by including the entire water-soluble proteome of *E. coli*.

Overall, our work differs from protein studies by others[10] on several counts. First, my experiments showed that native protein states can be kinetically trapped relative to nonamyloid aggregates. In other words, using both purified apomyoglobin and the water-soluble *E. coli*

proteome, I was able to determine that the aggregates generated from the experiments outlined in Chapters 2 and 3 lack amyloid character.

Second, I proved that the relevant aggregates are often soluble, unlike the majority of the aggregated species related to the pathology of several neurodegenerative diseases.

Third, the findings presented here relate to proteins of variable length. Regardless of size, most proteins are kinetically trapped on timescales relevant to the *E. coli* lifetime.

Last, the kinetic simulations in Chapter 3 show that kinetic trapping is not necessarily due to the presence of high kinetic barriers. Indeed, proteins at very low concentrations with small kinetic barriers still can be kinetically trapped in their native states relative to aggregates.

1.3 Kinetically trapped proteins in the cell

The work presented in Chapters 2 and 3 proves that folded proteins are kinetically trapped relative to their aggregated states under physiologically relevant conditions. In the bacterial cellular environment and upon completion of translation and release from the ribosome, newly synthesized proteins kinetically partition between fully folded, native states and aggregated conformations (Figure 1-5). Recent protein folding studies demonstrated this phenomenon for both prokaryotic and eukaryotic proteins (Figure 1-6). This kinetic partitioning is extremely sensitive to amino acid sequence and single-point mutations (that have no effect on native-protein stability and tertiary structure) can easily tilt the balance in favor of aggregate formation. Co- and post-translational interactions between the nascent polypeptide chain and the ribosome are critical to drive successful kinetic partitioning and generate the native state instead of aggregates. Due to the typical success of this immediately post-translational kinetic partitioning, cellular proteins (Figure 1-5) are consistently able to attain their native state in the cell with much higher yields than upon

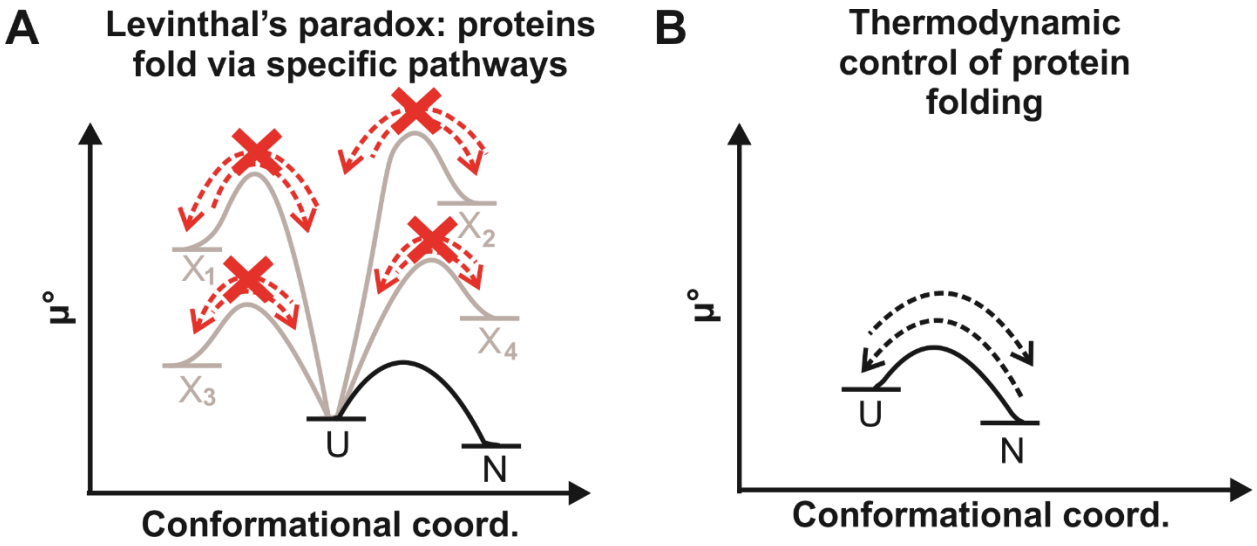


Figure 1-1. Protein folding is governed by fundamental principles. A) Time-constraints limit the number of available folding pathways. Therefore, consistent with Levinthal's paradox[1], proteins attain their native state by traveling through folding routes that have the lowest free energy barriers. Red dashed lines with a red cross denote kinetically inaccessible routes. Kinetically inaccessible routes are shown in gray. U and N denote unfolded and native states, respectively. X_n denotes alternate inaccessible protein conformational states. **B)** The native state is the most stable conformation of a protein under physiologically relevant conditions. Native, unfolded and intermediate states interconvert freely, under these conditions. U and N denote unfolded and native states, respectively.[3]

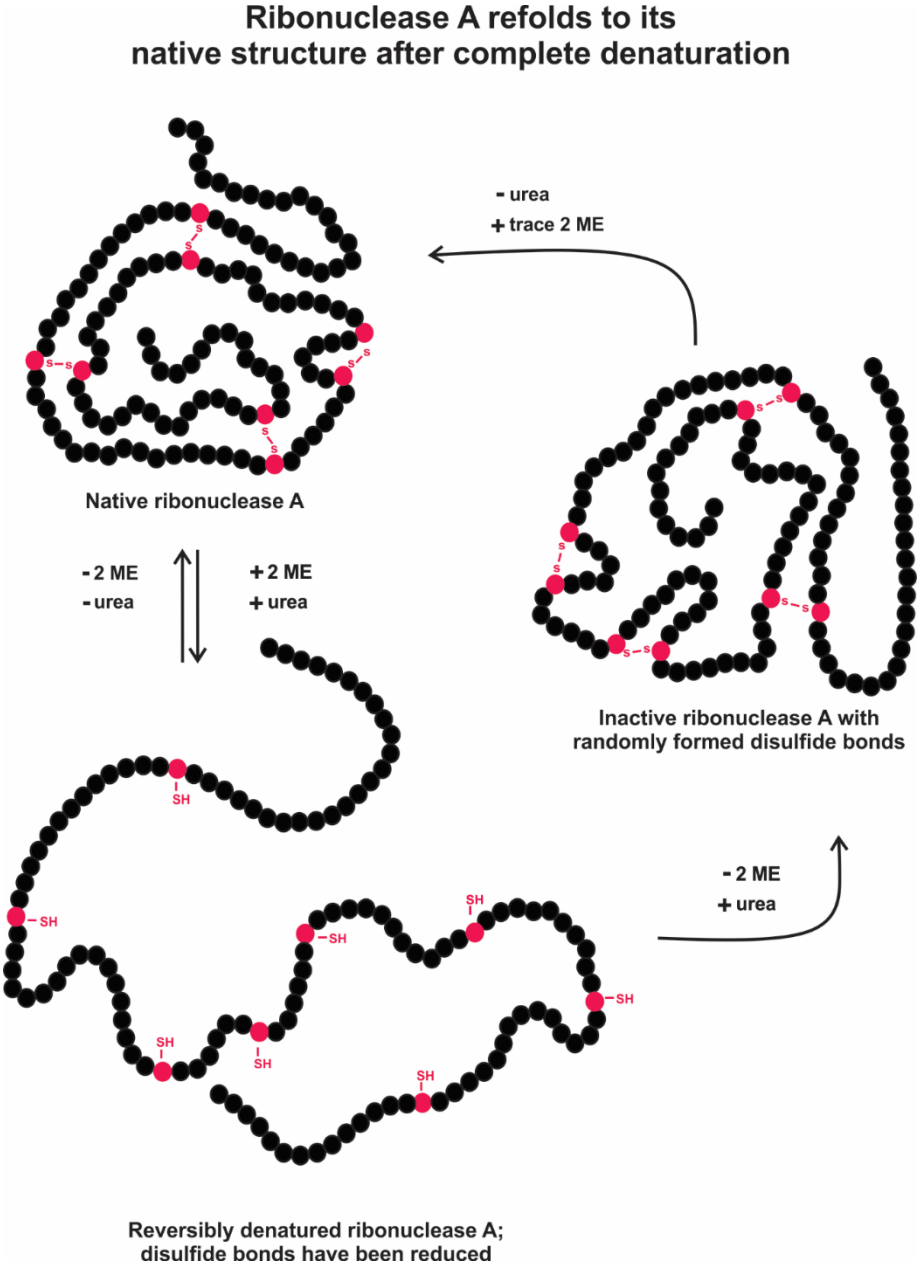


Figure 1-2. Anfinsen's ribonuclease A experiments showed that the native state of a protein is determined solely by its sequence, within a given environment and at a specific temperature and pressure.[2, 12] 2ME denotes 2-mercaptoethanol. Figure adapted from Horton, 2006.[13]

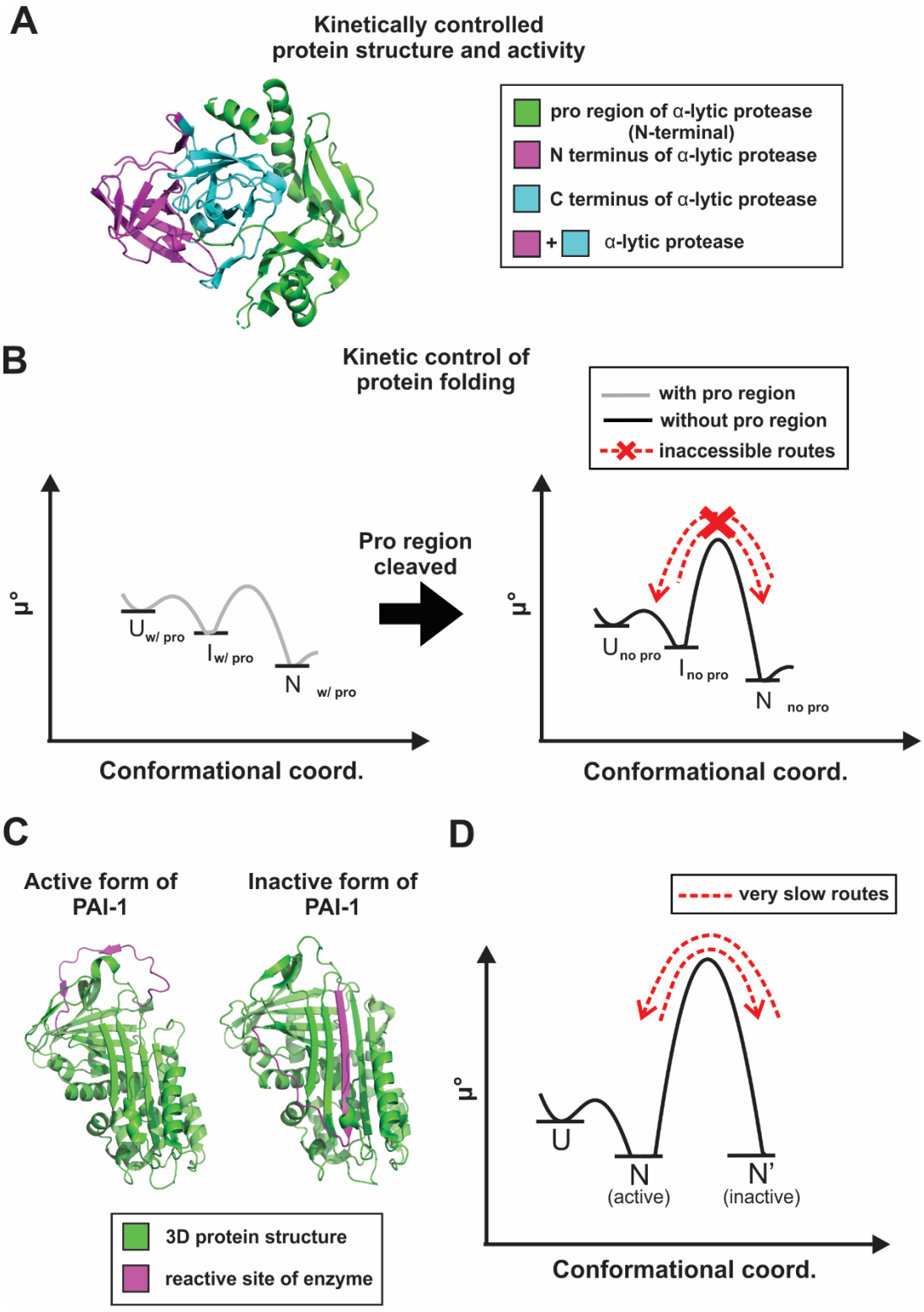


Figure 1-3 Some proteins fold under kinetic control. A) Crystal structure of α-lytic protease (magenta and cyan) complexed with a noncovalently bound pro region (green). PDB code: 4PRO.

B) The unfolded ($U_{w/pro}$) state of α-lytic protease can fold into an intermediate ($I_{w/pro}$) that is able to reach its native state ($N_{w/pro}$) when it is bound to its N-terminal pro-region. When the pro region is cleaved, the unfolded ($U_{no\ pro}$) and intermediate state ($I_{no\ pro}$) are unable to access the native state ($N_{no\ pro}$) due to high kinetic barriers. **C)** Crystal structure of the active (PDB: 1C5G) and latent (PDB: 1DB2) forms of plasminogen activator inhibitor (PAI-1) PDB. **D)** PAI-1 initially folds into its native state (N) and contains a reactive loop (magenta). Over time it slowly converts into a stable, inactive form (N') as the loop rearranges to form a β-sheet structure.

in vitro refolding starting from chemically denatured states. As shown in my work, after the native state has been formed in the cell (with assistance from the ribosome and molecular chaperones), it is kinetically trapped relative to aggregates (Figure 1-5). According to the above mechanism, proteins are biologically active without experiencing any aggregation, on typical cellular lifetime timescales.

1.4 Similarities between this work and the familiar process of egg cooking

The research presented in Chapters 2 and 3 presents a number of similarities with the well-known process of egg boiling followed by cooling (before consumption). For simplicity, we focus only on comparisons with the albumen of eggs[14-19], as the yolk is more complex.

The most abundant protein in egg white is ovalbumin. This protein contains one disulfide bridge and four sulfhydryl groups in its native state at room temperature [20-21]. Ovalbumin is known to undergo extensive disulfide bridge rearrangements upon heating followed by cooling in an oxygen-containing atmosphere.[15-16] This process facilitates formation of large soluble and insoluble aggregates at pH ca. 7.6.[14, 22]

However, formation of new disulfide bridges upon heating occurs concurrently or immediately after nonpolar surfaces have been exposed upon heating.[16] Aggregation is therefore believed to be driven by solvent exposure of nonpolar groups at high temperature, rather than by disulfide exchange.[14] On the other hand, disulfide exchange, which occurs shortly after hydrophobicity-induced aggregation[16], reinforces the gel-like properties of heated-and-cooled egg albumen, and likely enhances the thermodynamic stability of the aggregates.[14]

In the case of the complex S100 *E. coli* protein mixture analyzed in Chapter 3, insoluble protein aggregates are also generated, and their population is slightly decreased upon performing the

heating-and cooling process under reducing conditions. This result is fully consistent with the data gathered for egg white.[14]

Interestingly, the heated-and-cooled ovalbumin aggregates display an enhanced β-sheet secondary-structure content, similarly to the results that we achieved upon heating-and-cooling the apomyoglobin protein, as discussed in Chapter 2.

The work presented in Chapter 3 shows that regardless of the presence of reducing agents, insoluble aggregates are formed. This result highlights the fact that formation of insoluble aggregates upon heating-and-cooling is not merely a consequence of a nonreducing environment.

The experiments on the pure model protein discussed in Chapter 2 employ pure apomyoglobin, a cysteine-free protein. Experimental analysis of this model system is of key importance in this thesis, because it enables unequivocally ruling out any effects on protein kinetic trapping relative to aggregates) due to disulfide bond formation or disulfide exchange. In addition, we also ruled out other heating-induced covalent modifications or degradative events. The latter three processes complicate clear-cut interpretations of egg heating-and-cooling phenomena.

In summary, as briefly outlined above, there are clear analogies between the work in Chapters 2 and 3 of this thesis and the familiar egg-cooking process. Yet, the experiments reported here enable drawing clear-cut conclusions as disulfide-bond formation and/or reshuffling is either ruled out or explicitly taken into account.

1.5 The cellular machinery helps prevent aggregate formation

In the bacterial cell, the ribosome and several cotranslationally active chaperones help proteins achieve their bioactive conformation. Proteins are known to be able to begin folding during translation, even before becoming bioactive.[23-30] In the case of cotranslational protein

folding, the ribosome is believed to modulate peptide folding and play an active role in translation, though the exact mechanism underlying this phenomenon is still under investigation.[31-33]

Once soluble proteins traverse the ribosomal exit tunnel and emerge into the cytosol, they utilize assistance of chaperones other than the ribosome, such as trigger factor (TF) and DnaK in *E. coli* (Figure 1-7 A). Here I limit the discussion to cotranslationally active chaperones in *E. coli*. TF is the first chaperone that interacts with translating peptide chains as they are produced within the cellular environment.[34] TF is the only chaperone known to bind to the 50S subunit of the ribosome in *E. coli* and uses ribosomal protein L23 as a docking site.[35-37] The binding motif of TF consists of a sequence of approximately eight amino acids that are aromatic and/or basic (Figure 1-7 B).[38] TF can bind both unfolded and partially folded proteins[34, 39], and slow their folding so that proteins are more likely to achieve their final native state rather than aggregate.[40-43]

In addition to TF, DnaK is an intracellular chaperone that might interact co- and/or post-translationally with newly synthesized polypeptides to provide additional assistance in folding.[39] In fact, it is estimated that DnaK interacts with ~25% of all newly synthesized cytosolic proteins.[44] DnaK is known to bind protein sequences that contain two regions of basic residues, which are separated by a core of approximately five hydrophobic or aromatic residues (Figure 1-7 B).[45] In general, DnaK facilitates the folding of non-native protein intermediates by cycling between substrate bound and unbound states until the final protein structure has been attained (Figure 1-7 C).

Previous research showed that the ribosome destabilizes the folding of translating proteins that are in close proximity to its surface[27, 46-47], likely preventing unwanted associations to facilitate the formation of correctly folded protein structures.[48] Other research mentioned in subsequent

sections of this chapter discuss interactions between the nascent chain and various ribosomal components.

While both TF and DnaK can compete in assisting proteins to fold, neither chaperone is essential for cell viability. Deletion of either the *tig* or *dnak* gene increases the binding of the remaining chaperone to newly synthesized proteins.[39, 76] As a result, at least one of these chaperones is critical for cell viability since both TF and DnaK have somewhat overlapping functions.[77]

How the cell machinery is able to promote formation of folded proteins and disfavor aggregation is still unclear. For instance, do all proteins have a tendency to aggregate co- and post-translationally, and how does the ribosome help prevent aggregation and possibly promote folding? Given that proteins originate on the ribosome in the cell, Chapter 4 of this thesis work examines the potential role of the ribosome-associated nascent protein chain in ribosome stabilization. More specifically, I investigated whether the nascent chain has any effect on the apparent stability of the peptidyl transferase center (PTC) of the ribosome and the ribosomal proteins in *E. coli*.

1.6 Overview of the ribosome structure and function

The ribosome is a macromolecular structure composed of both RNA and protein (Figure 1-8). In *E. coli*, the intact ribosome (70S) comprises a large subunit and a small subunit, known as the 50S and 30S, respectively.[78]. The 50S subunit consists of 23S and 5S RNA and contains 33 proteins while the 30S subunit is comprised of 16S RNA and contains 21 proteins.[79-80] The intact 70S complex weighs approximately 2.7 million Daltons and is ca. 200 Å in diameter.[81]

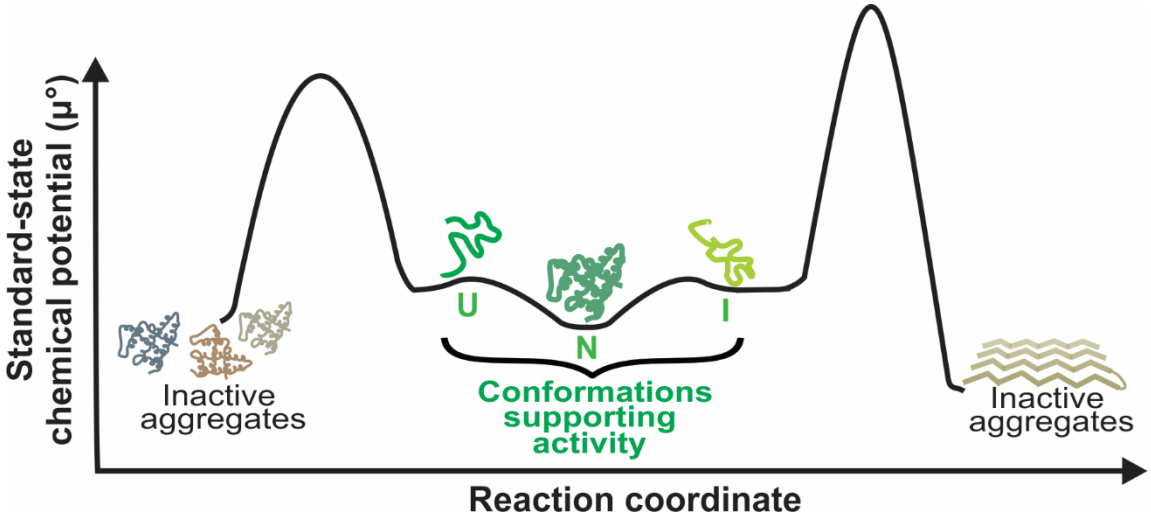


Figure 1-4. As shown in Chapters 2 and 3 of this thesis, the bioactive states, unfolded (U), native (N) and intermediates (I) of many proteins are kinetically trapped relative to stable aggregates. The latter are believed to be often detrimental to cell viability.

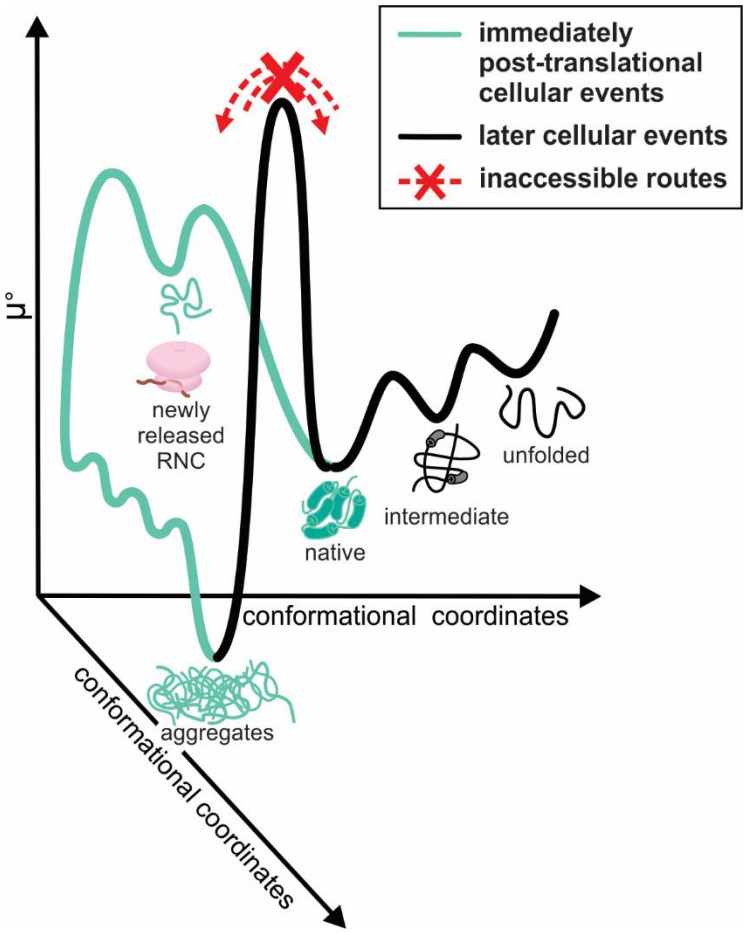


Figure 1-5. Chemical-potential landscape of cellular proteins upon release form the ribosome. Protein folding upon release from the ribosome is governed by kinetic partitioning between native and aggregated states. This process is usually successful for evolutionarily optimized proteins, leading to formation of the native state. Once the native state is generated, it is kinetically trapped relative to aggregates. Figure adapted from Addabbo et al. (*in preparation*).

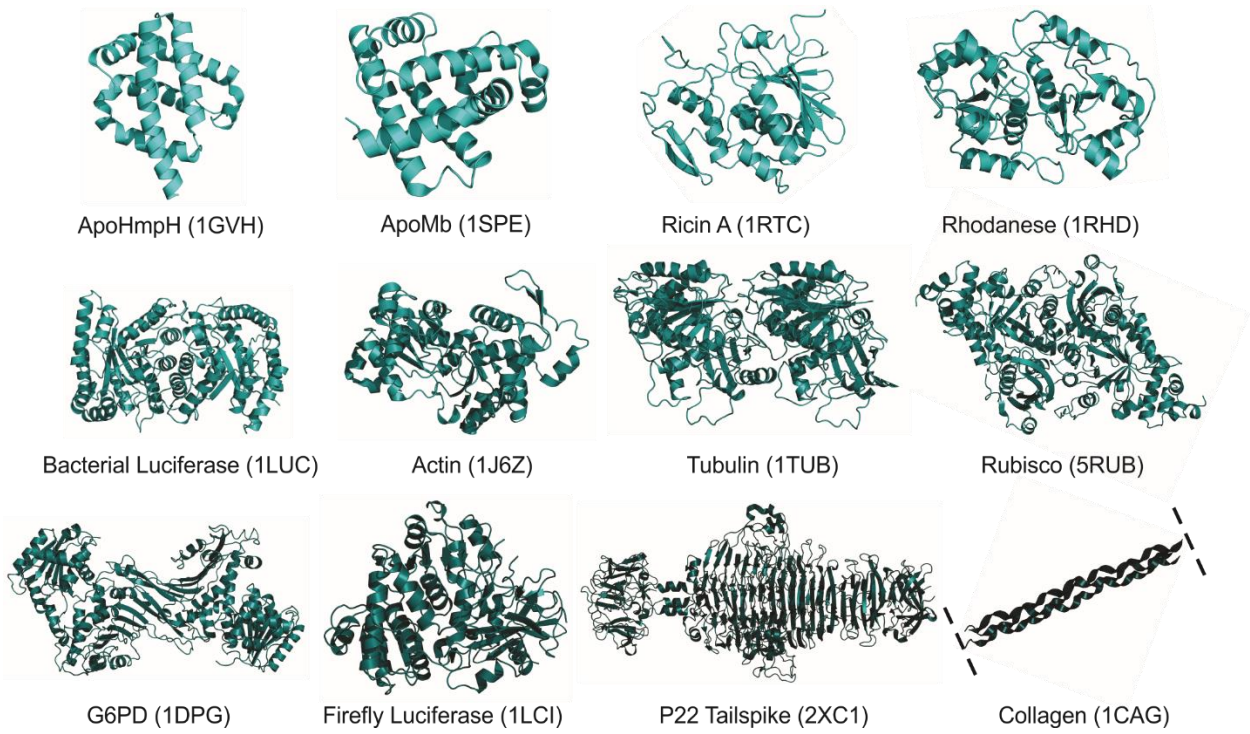


Figure 1-6. Prokaryotic and eukaryotic proteins that have been featured in protein folding studies in the cellular context. Protein structures are ordered from left to right in increasing size. PDB codes are shown in parentheses. Citations for the protein folding studies of the above proteins are as follows: ApoHmpH[49], ApoMb[49-50], ricin A[51-52], rhodanese[53-55], bacterial luciferase[56-58], actin[59-60], tubulin[59, 61], rubisco[62-63], G6PD[64-67], firefly luciferase[40, 61, 68-69], P22 tailspike[70-72], and collagen[73-75]. Figure adapted from Mecha et al. (*in preparation*).

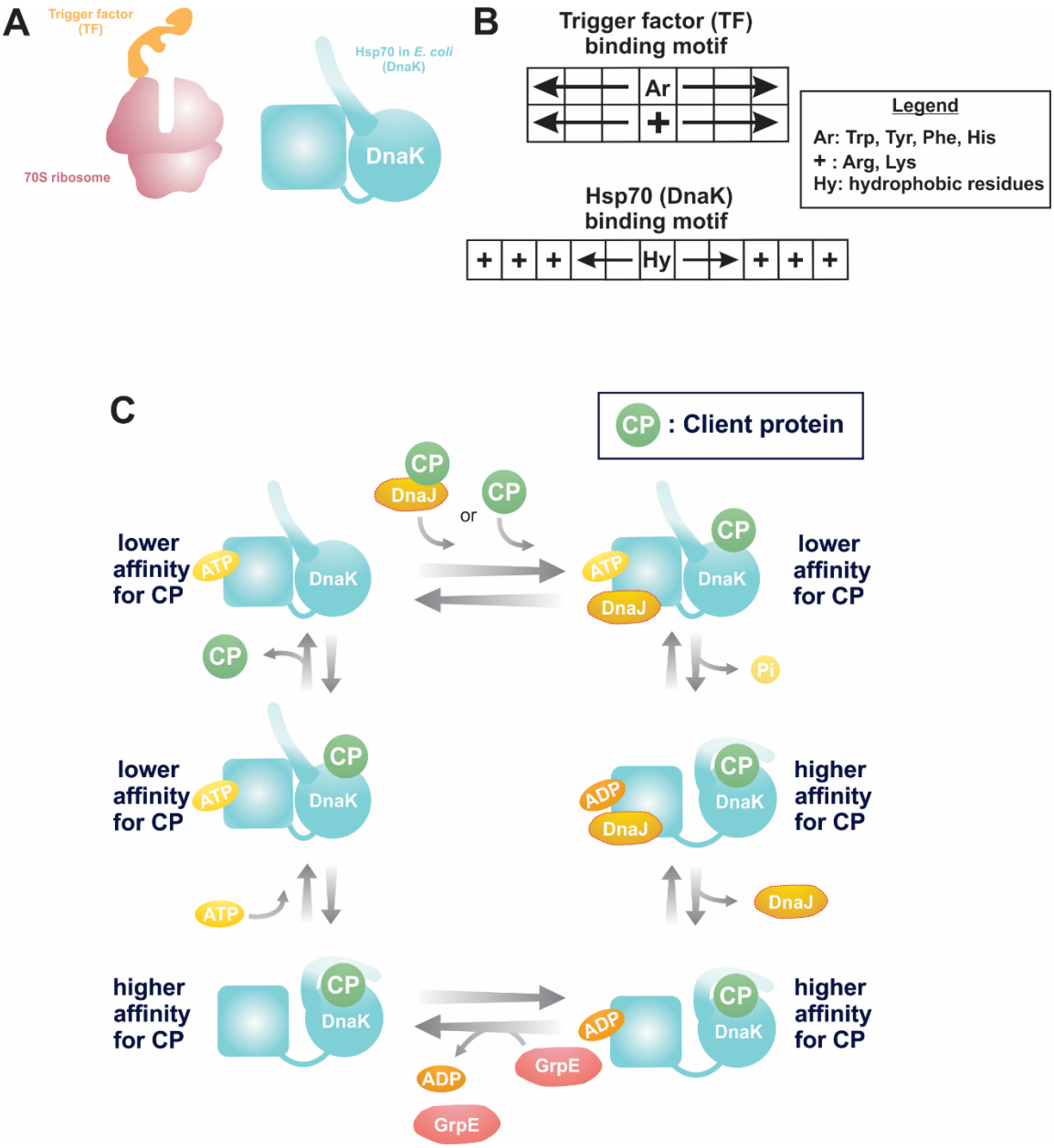


Figure 1-7. Selected co- and post-translationally active chaperones in E. *coli*. **A)** Cartoon representation of the *E. coli* chaperones trigger factor (TF) and DnaK. DnaK is also known as the *E. coli* heat shock factor HsP70. **B)** Client-protein binding motifs of TF and DnaK adapted from Patzelt et al. 2001[38] and Rüdiger et al. 1997[45], respectively. **C)** DnaK chaperone cycle. Figure adapted from Mecha et al. (*in preparation*).

In the cell, the ribosome synthesizes proteins with high accuracy.[82] It is estimated that the bacterial ribosome adds approximately 10-22 amino acids per second to the growing nascent chain during translation. [83-89] Together, the 30S and 50S work in concert to decode mRNA message and catalyze peptide bond formation during protein biosynthesis.

Protein synthesis can be divided into four steps: initiation, elongation, termination and recycling (Figure 1-9). During initiation, the 30S subunit binds to initiation factors IF-1 and IF-3.[90-91] This complex then binds to IF1, IF2, initiator tRNA and an mRNA strand at the Shine Dalgarno sequence[92] upstream of the start codon in order to correctly position the 5' AUG sequence in the peptidyl site.[93-97] Guanosine triphosphate (GTP)-bound IF-2 recruits formyl methionyl-tRNA (fMet-tRNA^{fMet}) to form a codon-anticodon pair with the 5' AUG sequence.[98-99] As the 50S subunit binds the 30S to form the 70S, GTP is hydrolyzed and simultaneously released with IF-1, IF-2, and IF-3.[80, 99-100]

Elongation starts when an aminoacyl-tRNA (aa-tRNA) binds to a complex between elongation factor (EF)-Tu and GTP.[99] When the aa-tRNA-EF-Tu-GTP complex binds to the aminoacyl (A) site of the ribosome, GTP is hydrolyzed to form guanosine-diphosphate (GDP) and released with EF-Tu.[101-103] A peptide bond is formed between the amino acids on the tRNA. A dipeptidyl-tRNA is formed in the A-site, and the tRNAs move into a hybrid state spanning two peptidyl transferase center sites.[99] Translocation occurs at the end of the elongation step when the ribosome moves along the mRNA strand to the next codon.[80, 103] This movement shifts the tRNA in the P site to the E site and the tRNA in the A site to the P site. After tRNAs reach the E site, they are released into the cytosol and recycled. The elongation cycle repeats until the ribosome encounters a stop codon on the mRNA.

Termination begins once the elongation cycle is halted. When a termination codon (UAA, UAG or UGA) occupies the A site, one of three release factors (RF-1, RF-2 or RF-3) binds at the A-site to hydrolyze the terminal peptidyl tRNA bond and release the newly synthesized polypeptide. RF-1 and RF-2 both recognize UAA, but UAG and UGA are only recognized by RF-1 and RF-2, respectively.[80, 104-105] After the protein is released, the 70S complex dissociates into the 50S and 30S subunits, which are recycled to begin a new cycle of protein synthesis.[80, 96-97, 99, 106-107]

1.7 Ribosome stability

The apparent thermodynamic stability of the ribosome has been investigated via a combination of techniques involving thermal or chemical denaturation. Several studies established that the thermal stability increases as a function of ribosomal component size, with the 30S being less stable than the 50S, and the 50S being less stable than the 70S.[108-109] Protein-RNA interactions within the ribosome are believed to contribute to the overall stability of the ribosome.[110] However, the complex, multi-component structure of the ribosome has made it difficult to assess the contributions of each ribosomal component to the overall stability of the intact 70S. [109]

Previous chemical denaturation studies focused on structural alterations of the ribosomal proteins (r-proteins) in solutions containing high concentrations of urea. For example, 6 M urea reduces the helical contact of the ribosomal proteins, which alters the ribosome's overall structure.[111] Although some of the r-proteins still remain noncovalently bound to the ribosome at high concentrations of urea, many are easily detached when the ionic strength increases.[111] Similar effects have also been observed for the 50S ribosomal subunit.[112]

One limitation of chemical denaturation studies is that they do not yet address how ribosome-bound nascent chains (RNCs) contribute to the stability of the 70S subunit. To date, one research group has compared the stability of the *E. coli* 70S complex and a stalled RNC by using urea titrations and sucrose gradients to monitor when the RNA falls apart (at 260 nm).[46] While 70S complexes were stable in up to 1 M urea, SecM-stalled RNCs were stable at 3 M urea.[46] However, the increased stability of ribosomes with RNCs they reported may not be a direct result of the presence of the nascent chain, because SecM stalling reinforces contact between the ribosomal subunits and may artificially stabilize the complex.[115] Therefore, additional studies are needed to show the effect of RNCs on the ribosomal stability.

Chapter 4 compares the apparent stability of the 70S ribosomal complex in the absence and presence of RNCs. In addition, the overall stability of the peptidyl transferase center and the ribosomal proteins in the context of the ribosome are also addressed in this chapter.

1.8 The peptidyl-transferase center

The peptidyl transferase center (PTC) is arguably the most important component in protein synthesis as it is responsible for catalyzing peptide bond formation. Protein synthesis and the function of the PTC is explained in Section 1.6 of this chapter. Briefly, the PTC is located in a cavity at the interface between the 50S and 30S ribosomal subunits (Figure 1-10).[116] Domain V of the 23S RNA is directly responsible for the catalytic activity of the PTC[116]. Since the PTC is composed of RNA, which is characteristically unstable in the absence of divalent cations, it is not surprising that Mg^{2+} ions are present in the PTC, as they probably stabilize its structure.[117-118] High resolution structural data[116] and 3D comparisons among the large subunits of the ribosome from

different organisms[119] revealed that ribosomal proteins are absent in the PTC and its nearby surrounding area, rendering it one of the universally conserved ribozymes in Nature.

Due to its importance and highly conserved nature, a great deal of research has been done on the PTC, including the inhibition of protein synthesis using antibiotics. Chapter 4 utilizes puromycin, which is an antibiotic derived from *Streptomyces alboniger* [120] and mimics the 3' end of an amino-acylated-tRNA (aa-tRNA).[121] It binds to the A-site in the 50S ribosomal subunit where it deacylates the peptidyl-tRNA in the P-site and prematurely releases the nascent chain.[122-125] More details about the use of puromycin will be discussed in Chapter 4.

1.9 The ribosomal exit tunnel

The ribosomal tunnel is located in the 50S subunit and is estimated to be ~80-100 Å[126] long and 10-20 Å wide (Figure 1-11 A). [116, 127-129] It extends from the PTC site through the 50S to the surface of the ribosome opposite the PTC (Figure 1-11 B). The tunnel wall consists of mostly domains I to V of 23S RNA.[116] During translation, the ribosomal exit tunnel guides the growing nascent chain into the cytosol of the cell.

Elongating nascent polypeptides are capable of forming some secondary structures while still buried within the tunnel. [23-26, 28, 129-136] In fact, previous studies probing the electrostatic potential of the tunnel demonstrated that specific folding zones within the ribosomal exit tunnel can either promote or inhibit polypeptide compaction (Figure 1-11 C and D).[23, 25, 28, 137] Formation of secondary structure may also be assisted by interactions between the 23S RNA and ribosomal proteins facing the interior wall of the ribosomal exit tunnel. For instance, at the beginning of the exit tunnel, near the PTC, the nascent chain can interact with 23S RNA.[138-139] As the polypeptide continues to extend through the tunnel, it can also interact with several ribosomal proteins.

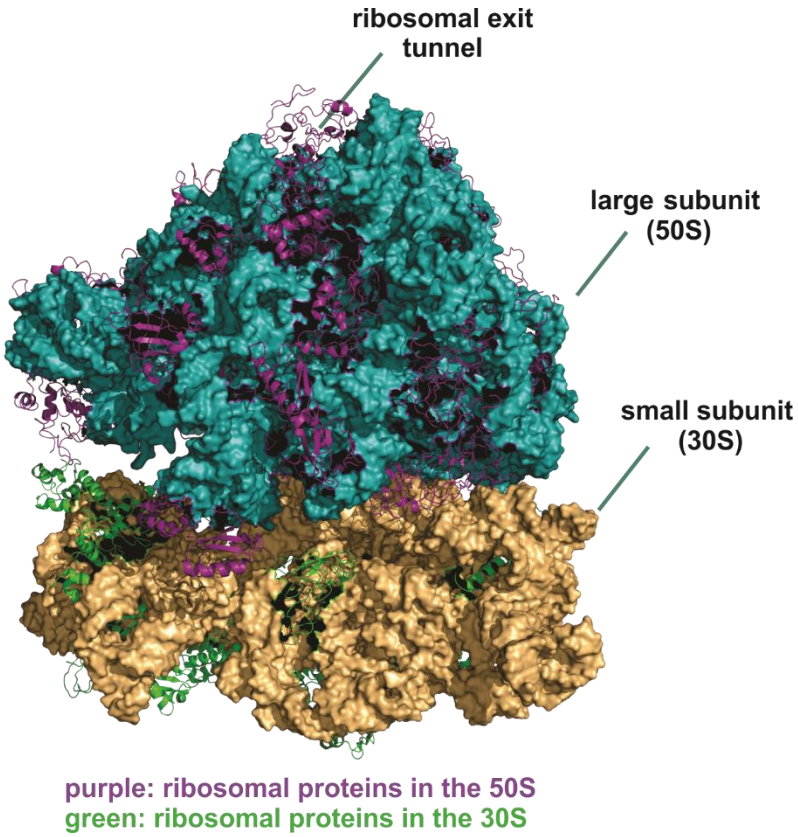


Figure 1-8. Structure of the 70S *E. coli* ribosome at 3.5 Å resolution. The RNA structure of the 50S subunit is shown in turquoise and the 30S subunit is shown in yellow (PDB code: 2AVY and 2AW4).[113] Figure adapted from Fedyukina and Cavagnero, 2011.[114]

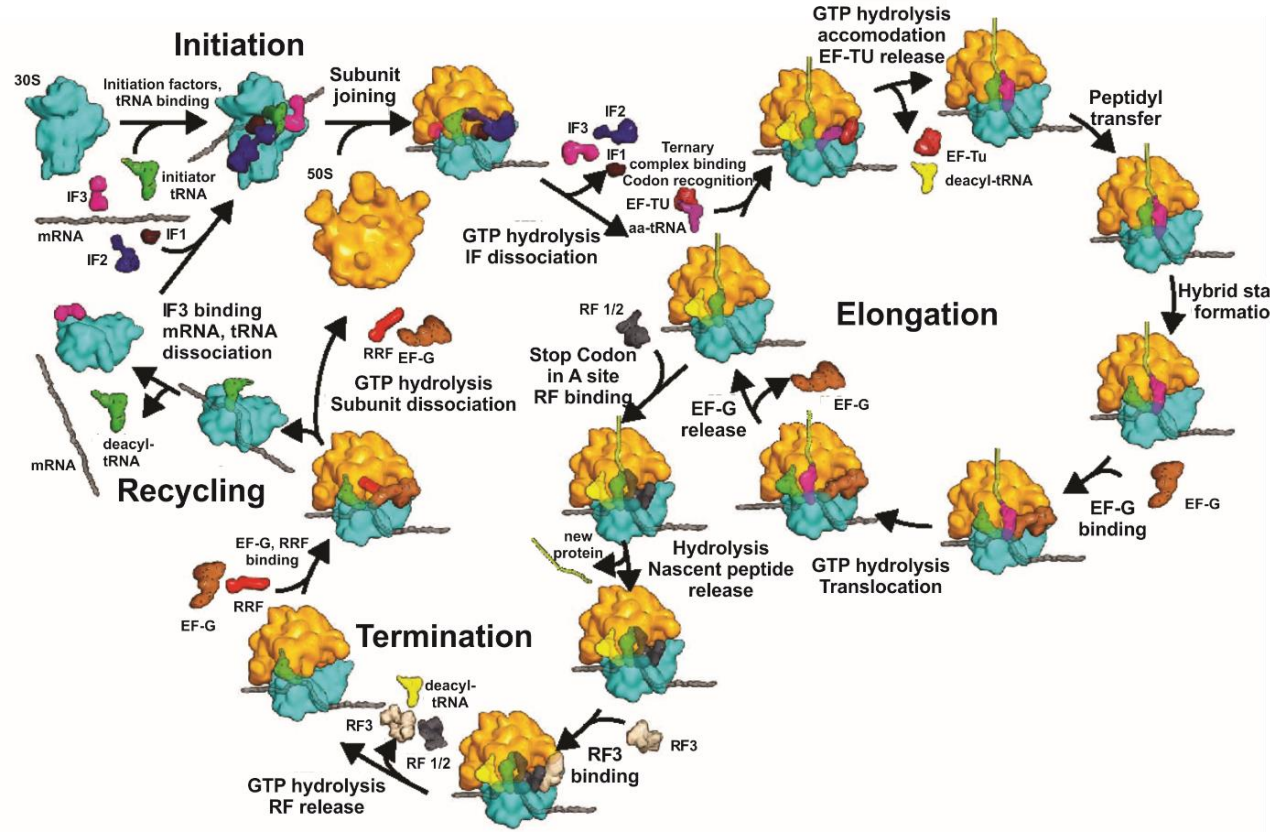


Figure 1-9. Protein biosynthesis in *E. coli*. Bacterial translation takes place in four stages: initiation, elongation, termination and recycling. Figure adapted from Schmeing and Ramakrishnan, 2009.[80]

In *E. coli*, three ribosomal proteins, L4, L22 and L23, face the interior of the ribosomal exit tunnel wall in two distinct regions. L4 and L22 are located in portions of the tunnel wall at the narrowest part of the tunnel[118, 126-127, 140] whereas L23 is located near the ribosome exit tunnel.[127] This constriction site is roughly 20-35 Å away from the PTC.[127] Multiple crosslinking studies demonstrated that nascent chains are capable of interacting with both L4 (also known as L4 in eukarya) and L22 (denoted as L17 in eukarya).[23-24, 140-145] While neither protein is required for the PTC to undergo catalytic activity, researchers suggested that these proteins may allosterically relay information about the nascent chain to the ribosomal surface, although the mechanism for this process is currently unknown.[127] The last section of the exit tunnel farthest from the PTC widens into a vestibule[28] where L23 is embedded within the tunnel wall and is able to interact with nascent chains bearing N-terminal signal sequences before nascent polypeptides enter the cytosol.[146-149] To date, the exact purpose for these interactions between nascent polypeptides and ribosomal proteins has eluded researchers.

1.10 Nascent chains bearing signal sequences interact with ribosomal proteins facing the outer surface of the ribosome

Once the nascent chain moves from the exit tunnel into the cytosol, it has the ability to interact with the surface of the ribosome. Due to the phosphodiester backbone of the ribosomal RNA (rRNA) and the charge separation within ribosomal proteins, the ribosome surface is highly negatively charged, and could establish interactions with the growing polypeptide chain (Figure 1-12 A).[116, 150-151] In fact, previous research showed that electrostatic interactions between the ribosomal surface and newly synthesized nascent chain may influence folding properties.[47, 152].

For instance, systematic mutations to alter the overall net charge of the phosphorylated insulin receptor interaction region of Grb14 (PIR), an intrinsically disordered protein (IDP), showed that positively charged mutants significantly interacted with the surface of the ribosome.[152]

Negatively charged variants exhibited repulsive effects with the surface of the ribosome and were more likely to be highly dynamic in solution.[152] Increasing the ionic strength did not alter the properties of either species in complex with the ribosome.[152] However, increasing the ionic strength of a system with foldable protein mitigated the destabilizing effect of the ribosomal surface.[47] T4 lysozyme, a protein with two globular regions, exhibited a faster apparent folding rate under high salt conditions, suggesting that folding is somewhat mediated by electrostatic interactions with the ribosome.[47] Ongoing research in the Cavagnero group shows that the nascent chain preferentially interacts with ribosomal proteins on a specific side of the ribosomal surface (Figure 1-12 B, Fuchs et al. *in preparation*).

Additional experiments probed interactions of nascent protein with the ribosomal surface by crosslinking and cryo-electron microscopy (cryo-EM). Cross-linking studies showed that nascent chains are able to interact with ribosomal proteins on the surface.[146-149, 153] These data also depict nascent chain interactions that favor ribosomal proteins that are found on one side of the ribosomal exit tunnel; however, more work is currently underway to determine the basis of these interactions (Fuchs et al. *in preparation*).

It is important to note that the published data on these interactions involve proteins containing N-terminal signal sequences, which are present on proteins targeted for transport across membranes or secretory pathways; cystolic proteins do not contain targeting sequences. Fuchs et al. (*in preparation*) will be the first study, to our knowledge, that shows explicit cross-linking evidence between a protein lacking a signal sequence and r-proteins on the ribosomal surface. This further

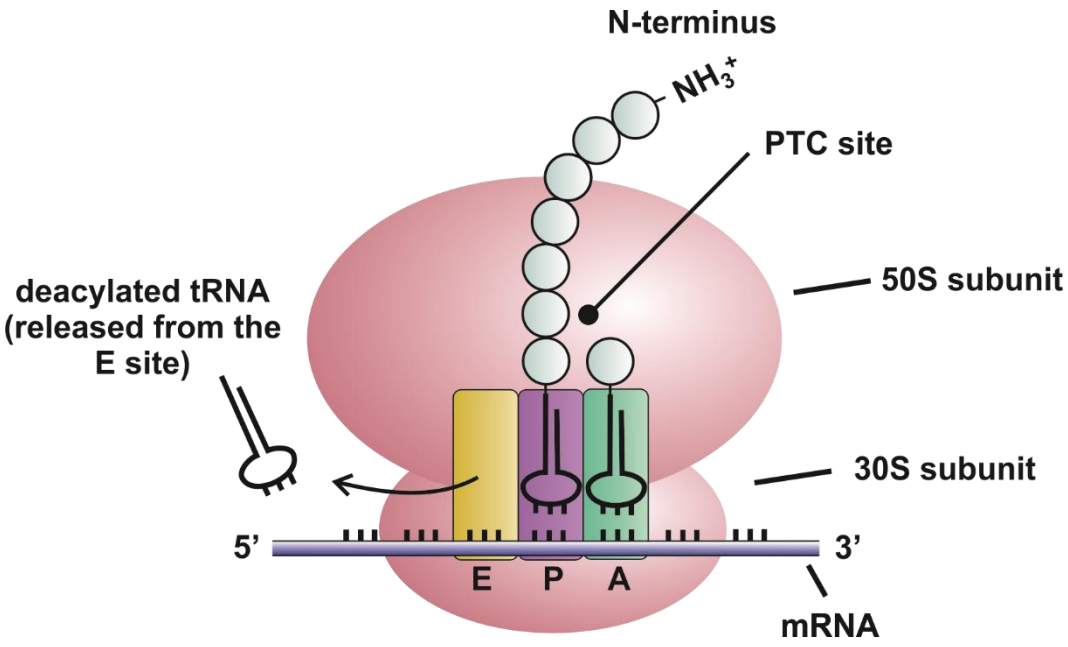


Figure 1-10. Schematic structure of the ribosome illustrating the A, P and E sites. The peptidyl transferase center of the ribosome catalyzes peptide bond formation, as the mRNA strand is translated.

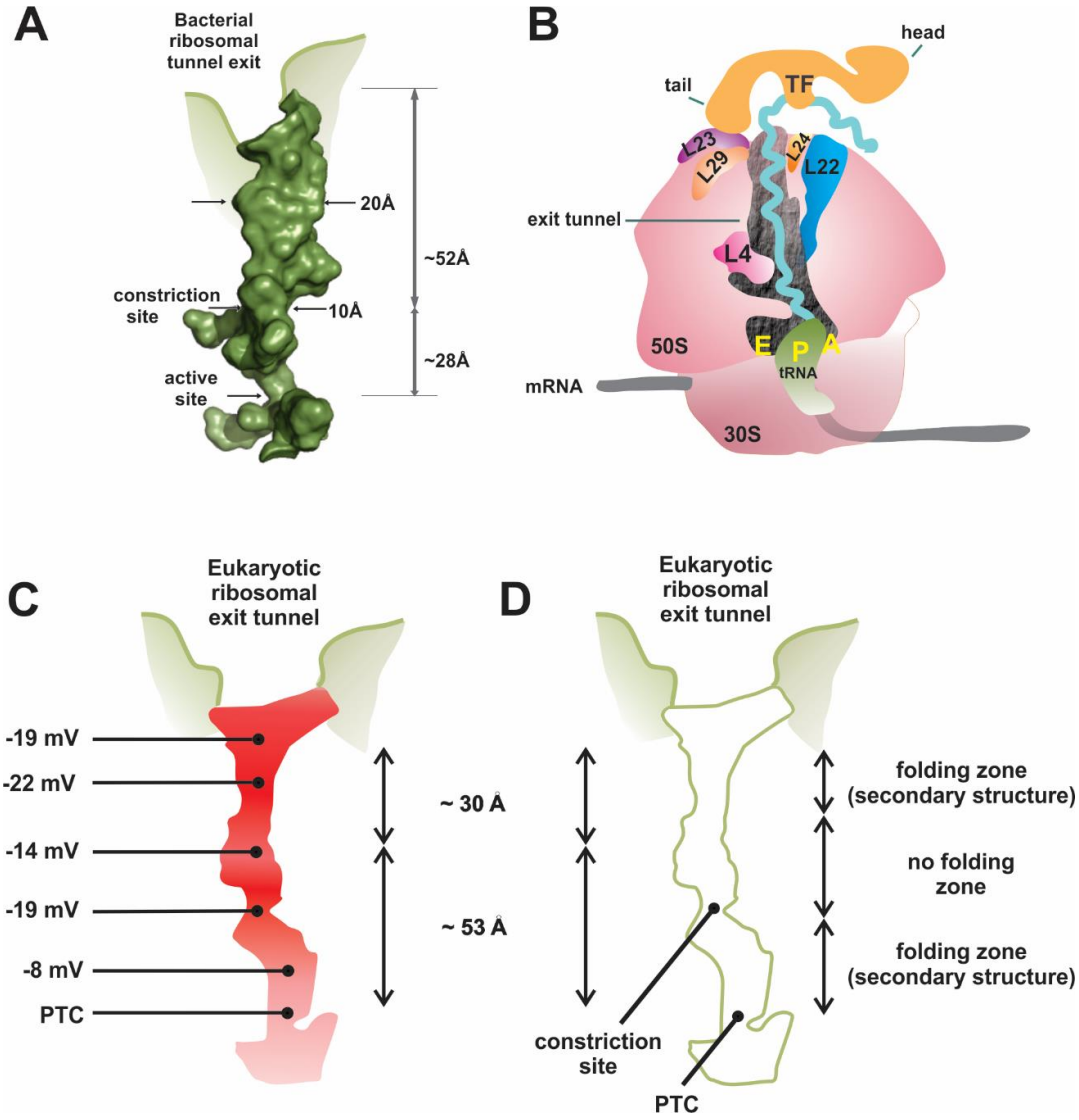


Figure 1-11. The ribosomal exit tunnel. A) Structure of the ribosomal exit tunnel.[126] **B)** Schematic cartoon representation of the 70S complex illustrating the nascent protein chain within the ribosomal exit tunnel. This figure is adapted from Fedyukina and Cavagnero, 2011. [114] **C)** Electrostatic potential of the ribosomal tunnel as described by Lu *et al.* 2007. [137] **D)** Folding zones of the ribosome exit tunnel, according to Bhushan *et al.* 2010 and Lu and Deutsch 2005.[23, 25]

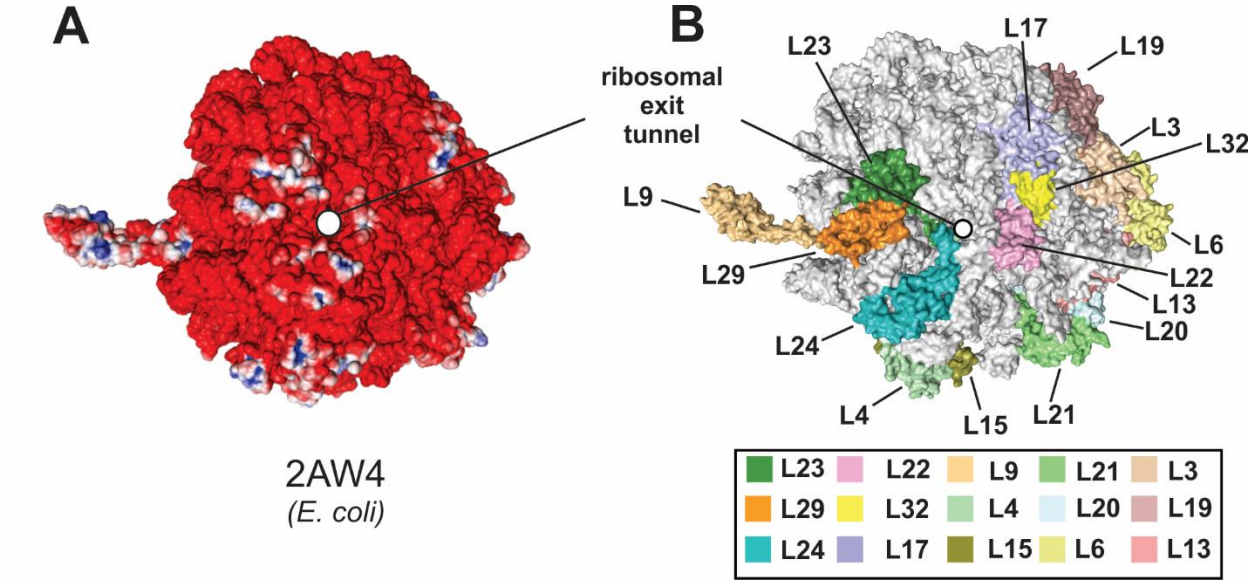


Figure 1-12. Top view of the *E. coli* 50S ribosomal subunit. A) Surface electrostatic potential map of the *E. coli* 50S subunit. Positive potential regions are shown in blue (> +3 k_bT/e) and negative potential regions are shown in red (> -3 k_bT/e). Figure adapted from Fedyukina et al., 2014.[151] **B)** Structural representation of the ribosomal proteins that are on the surface of the 50S near the exit tunnel. Figure adapted from Fuchs et al., *in preparation*.

confirms that interactions between the nascent chain and proteins on ribosomal surface are not a consequence of an N-terminal signal sequence.

Within the past decade, advances in high-resolution NMR enabled characterization of the dynamic and structural features of RNCs.[154-160] Constructs closer to the ribosomal surface are more unstructured whereas constructs that are farther from the surface adopt more compact states. However, one major drawback to NMR studies of RNCs is that researchers use resonance line-broadening as a probe for the absence/presence of RNC-protein interactions. In NMR, line-broadening can be attributed to a number of causes, including intra- and inter-molecular interactions. The two cannot be discriminated. As a result, NMR studies based on line width assessment do not unequivocally prove that interactions are present between the RNCs and the ribosomal surface.

1.11 This thesis work

In summary, the work presented in this thesis highlights a number of findings in the field of protein folding. First, Chapter 2 shows that apomyoglobin's native state is kinetically trapped relative to two distinct, nonamyloid aggregates under physiologically relevant conditions.

Chapter 3 expands this concept to the entire *E. coli* soluble proteome, and shows that kinetic trapping relative to non-amyloid aggregates is a phenomenon of general significance.

Chapter 4 explores the effect of nascent chains on the 70S ribosome. This work demonstrates that the ribosome is stabilized by the presence of non-SecM nascent chains in the ribosomal tunnel, regardless of their sequence.

Overall, this thesis highlights the need to expand Anfinsen's thermodynamic hypothesis to include nonamyloid aggregates, and it advances our understanding of the effect of the nascent chain on the apparent stability of the bacterial ribosome.

1.12 References

(1) Levinthal, C., *Mossbauer spectroscopy in biological systems*. Urbana:Univ. of Illinois Press: 1969.

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

(3) Jackson, S. E. How do small single-domain proteins fold? *Fold Des.* **1998**, *3*, R81-R91.

(4) Epstein, C. J.; Goldberger, R. F.; Anfinsen, C. B. In *The genetic control of tertiary protein structure: studies with model systems*, Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press: 1963; pp 439-449.

(5) Baker, D.; Agard, D. A. Kinetics versus thermodynamics in protein folding. *Biochemistry* **1994**, *33*, 7505-7509.

(6) Kelch, B. A.; Salimi, N. L.; Agard, D. A. Functional modulation of a protein folding landscape via side-chain distortion. *Proc. Natl. Acad. Sci.* **2012**, *109*, 9414-9419.

(7) Hekman, C.; Loskutoff, D. Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants. *J. Biol. Chem.* **1985**, *260*, 11581-11587.

(8) Jahn, T. R.; Radford, S. E. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **2008**, *469*, 100-117.

(9) Gazit, E. The "correctly folded" state of proteins: Is It a metastable state? *Angew. Chem.* **2002**, *41*, 257-259.

(10) 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.

(11) Thirumalai, D.; Reddy, G. Protein thermodynamics: Are native proteins metastable? *Nat.*

Chem. **2011**, *3*, 910-911.

(12) Haber, E.; Anfinsen, C. B. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **1962**, *237*, 1839-1844.

(13) Horton, H. R., *Principles of biochemistry*. 6th ed. ed.; Pearson Prentice Hall: Upper Saddle River, N.J., 2006.

(14) Mine, Y.; Noutomi, T.; Haga, N. Thermally induced changes in egg white proteins. *J. Agric. Food Chem.* **1990**, *38*, 2122-2125.

(15) Van der Plancken, I.; Van Loey, A.; Hendrickx, M. E. Changes in sulfhydryl content of egg white proteins due to heat and pressure treatment. *J. Agric. Food Chem.* **2005**, *53*, 5726-5733.

(16) Van der Plancken, I.; Van Loey, A.; Hendrickx, M. E. Effect of heat-treatment on the physico-chemical properties of egg white proteins: a kinetic study. *J. Food Eng.* **2006**, *75*, 316-326.

(17) Shigeru, H.; Shuryo, N. Contribution of hydrophobicity, net charge and sulfhydryl groups to thermal properties of ovalbumin. *Can. Inst. Food. Sci. Technol. J.* **1985**, *18*, 290-295.

(18) Croguennec, T.; Renault, A.; Beaufils, S.; Dubois, J.-J.; Pezennec, S. Interfacial properties of heat-treated ovalbumin. *J. Colloid Interface Sci.* **2007**, *315*, 627-636.

(19) Shimada, K.; Matsushita, S. Thermal coagulation of egg albumin. *J. Agric. Food Chem.* **1980**, *28*, 409-412.

(20) Nisbet, A. D.; Saundry, R. H.; Moir, A. J. G.; Fothergill, L. A.; Fothergill, J. E. The complete amino-acid sequence of hen ovalbumin. *Eur. J. Biochem.* **1981**, *115*, 335-345.

(21) Thompson, E.; Fisher, W. Amino acid sequences containing half-cystine residues in ovalbumin. *Aust. J. Biol. Sci.* **1978**, *31*, 433-442.

(22) Yang, M.; Dutta, C.; Tiwari, A. Disulfide-bond scrambling promotes amorphous aggregates in lysozyme and bovine serum albumin. *J. Phys. Chem. B* **2015**, *119*, 3969-3981.

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

(24) 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.

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

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

(27) Holtkamp, W.; Kokic, G.; Jäger, M.; Mittelstaet, J.; Komar, A. A.; Rodnina, M. V. Cotranslational protein folding on the ribosome monitored in real time. *Science* **2015**, *350*, 1104-1107.

(28) 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.

(29) 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.

(30) Khushoo, A.; Yang, Z.; Johnson, A. E.; Skach, W. R. Ligand-driven vectorial folding of

ribosome-bound human CFTR NBD1. *Mol. Cell* **2011**, *41*, 682-692.

(31) 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.

(32) 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.

(33) Pechmann, S.; Willmund, F.; Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **2013**, *49*, 411-421.

(34) Valent, Q. A.; Kendall, D.; High, S.; Kusters, R.; Oudega, B.; Luirink, J. Early events in preprotein recognition in E. coli: interaction of SRP and trigger factor with nascent polypeptides. *EMBO J.* **1995**, *14*, 5494-5505.

(35) 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.

(36) 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.

(37) Schlünzen, F.; Wilson, D. N.; Tian, P.; Harms, J. M.; McInnes, S. J.; Hansen, H. A.; 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.

(38) Patzelt, H.; Rüdiger, S.; Brehmer, D.; Kramer, G.; Vorderwülbecke, S.; Schaffitzel, E.;

Waitz, A.; Hesterkamp, T.; Dong, L.; Schneider-Mergener, J. Binding specificity of Escherichia coli trigger factor. *Proc. Natl. Acad. Sci.* **2001**, *98*, 14244-14249.

(39) 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.

(40) 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.

(41) Mashaghi, A.; Kramer, G.; Bechtluft, P.; Zachmann-Brand, B.; Driessen, A. J.; Bukau, B.; Tans, S. J. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **2013**, *500*, 98-101.

(42) Haldar, S.; Tapia-Rojo, R.; Eckels, E. C.; Valle-Orero, J.; Fernandez, J. M. Trigger factor chaperone acts as a mechanical foldase. *Nat. Commun.* **2017**, *8*, 668.

(43) 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.

(44) 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 Rep.* **2012**, *1*, 251-264.

(45) Rüdiger, S.; Buchberger, A.; Bukau, B. Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* **1997**, *4*, 342-349.

(46) 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.

(47) Kaiser, C. M.; Goldman, D. H.; Chodera, J. D.; Tinoco, I.; Bustamante, C. The ribosome modulates nascent protein folding. *Science* **2011**, *334*, 1723-1727.

(48) 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*, 444-448.

(49) Eun, Y. J.; Kurt, N.; Sekhar, A.; Cavagnero, S. Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* **2008**, *376*, 879-897.

(50) 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.

(51) Argent, R. H.; Parrott, A. M.; Day, P. J.; Roberts, L. M.; Stockley, P. G.; Lord, J. M.; Radford, S. E. Ribosome-mediated folding of partially unfolded ricin A-chain. *J. Biol. Chem.* **2000**, *275*, 9263-9269.

(52) Spooner, R. A.; Hart, P. J.; Cook, J. P.; Pietroni, P.; Rogon, C.; Höhfeld, J.; Roberts, L. M.; Lord, J. M. Cytosolic chaperones influence the fate of a toxin dislocated from the endoplasmic reticulum. *Proc. Natl. Acad. Sci.* **2008**, *105*, 17408-17413.

(53) 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*, 683-689.

(54) 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*, 36-42.

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

(56) Waddle, J. J.; Johnston, T. C.; Baldwin, T. O. Polypeptide folding and dimerization in bacterial luciferase occur by a concerted mechanism in vivo. *Biochemistry* **1987**, *26*, 4917-4921.

(57) Fedorov, A. N.; Baldwin, T. O. Process of biosynthetic protein folding determines the rapid formation of native structure1. *J. Mol. Biol.* **1999**, *294*, 579-586.

(58) Fedorov, A. N.; Baldwin, T. O. GroE modulates kinetic partitioning of folding intermediates between alternative states to maximize the yield of biologically active protein1. *J. Mol. Biol.* **1997**, *268*, 712-723.

(59) Vainberg, I. E.; Lewis, S. A.; Rommelaere, H.; Ampe, C.; Vandekerckhove, J.; Klein, H. L.; Cowan, N. J. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **1998**, *93*, 863-873.

(60) Gao, Y.; Thomas, J. O.; Chow, R. L.; Lee, G.-H.; Cowan, N. J. A cytoplasmic chaperonin that catalyzes β-actin folding. *Cell* **1992**, *69*, 1043-1050.

(61) Frydman, J.; Nimmesgern, E.; Erdjument-Bromage, H.; Wall, J.; Tempst, P.; Hartl, F. Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **1992**, *11*, 4767-4778.

(62) Goloubinoff, P.; Christeller, J. T.; Gatenby, A. A.; Lorimer, G. H. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **1989**, *342*, 884-889.

(63) Checa, S. K.; Viale, A. M. The 70-kDa heat-shock protein/DnaK chaperone system is

required for the productive folding of ribulose-bisphosphate carboxylase subunits in Escherichia Coli. *Eur. J. Biochem.* **1997**, *248*, 848-855.

(64) Das, B.; Chattopadhyay, S.; Gupta, C. D. 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.

(65) Chattopadhyay, S.; Das, B.; Dasgupta, C. Reactivation of denatured proteins by 23S ribosomal RNA: role of domain V. *Proc. Natl. Acad. Sci.* **1996**, *93*, 8284-8287.

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

(67) Nakamoto, H.; Fujita, K.; Ohtaki, A.; Watanabe, S.; Narumi, S.; Maruyama, T.; Suenaga, E.; Misono, T. S.; Kumar, P. K.; Goloubinoff, P. Physical interaction between bacterial heat shock protein 90 (Hsp90) and Hsp70 chaperones mediates their cooperative action to refold denatured proteins. *J. Biol. Chem.* **2014**, *289*, 6110-6119.

(68) 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. Mol. Biol.* **1999**, *6*, 697.

(69) Szabo, A.; Langer, T.; Schröder, H.; Flanagan, J.; Bukau, B.; Hartl, F. U. The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. *Proc. Natl. Acad. Sci.* **1994**, *91*, 10345-10349.

(70) 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*, 1516-1523.

(71) Gordon, C. L.; Sather, S. K.; Casjens, S.; King, J. Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. *J. Biol. Chem.* **1994**, *269*, 27941-27951.

(72) Sather, S. K.; King, J. Intracellular trapping of a cytoplasmic folding intermediate of the phage P22 tailspike using iodoacetamide. *J. Biol. Chem.* **1994**, *269*, 25268-25276.

(73) Leikina, E.; Mertts, M.; Kuznetsova, N.; Leikin, S. Type I collagen is thermally unstable at body temperature. *Proc. Natl. Acad. Sci.* **2002**, *99*, 1314-1318.

(74) Nagata, K. Hsp47: a collagen-specific molecular chaperone. *Trends Biochem. Sci.* **1996**, *21*, 23-26.

(75) Barnes, A. M.; Carter, E. M.; Cabral, W. A.; Weis, M.; Chang, W.; Makareeva, E.; Leikin, S.; Rotimi, C. N.; Eyre, D. R.; Raggio, C. L. Lack of cyclophilin B in osteogenesis imperfecta with normal collagen folding. *N. Engl. J. Med.* **2010**, *362*, 521-528.

(76) 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.

(77) 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.

(78) Tissières, A.; Watson, J. D. Ribonucleoprotein particles from Escherichia coli. *Nature* **1958**, *182*, 778-780.

(79) Stelzl, U.; Connell, S.; Nierhaus, K.; Wittmann-Liebold, B., *Ribosomal proteins: role in ribosomal functions*. John Wiley & Sons, Ltd.: 2001.

(80) Schmeing, T. M.; Ramakrishnan, V. What recent ribosome structures have revealed about

the mechanism of translation. *Nature* **2009**, *461*, 1234-1242.

(81) Berg, J.; Tymoczko, J.; Stryer, L., A ribosome is a ribonucleoprotein particle (70S) made of a small (30S) and a large (50S) subunit. In *Biochemistry*, 5th Edition ed.; WH Freeman: New York, 2002.

(82) Johansson, M.; Lovmar, M.; Ehrenberg, M. Rate and accuracy of bacterial protein synthesis revisited. *Curr. Opin. Microbiol.* **2008**, *11*, 141-147.

(83) Siller, E.; DeZwaan, D. C.; Anderson, J. F.; Freeman, B. C.; Barral, J. M. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* **2010**, *396*, 1310-1318.

(84) Pedersen, S. Escherichia coli ribosomes translate in vivo with variable rate. *EMBO J.* **1984**, *3*, 2895-2898.

(85) Liang, S.-T.; Xu, Y.-C.; Dennis, P.; Bremer, H. mRNA composition and control of bacterial gene expression. *J. Bacteriol.* **2000**, *182*, 3037-3044.

(86) Young, R.; Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli b/r as a function of growth rate. *Biochem. J.* **1976**, *160*, 185-194.

(87) Sørensen, M. A.; Pedersen, S. Absolute in vivo translation rates of individual codons in Escherichia coli: the two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **1991**, *222*, 265-280.

(88) Bremer, H.; Dennis, P. P. Modulation of chemical composition and other parameters of the cell by growth rate. *EcoSal Plus* **1996**, *2*, 1553-1569.

(89) Proshkin, S.; Rahmouni, A. R.; Mironov, A.; Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **2010**, *328*, 504-508.

(90) Dahlquist, K. D.; Puglisi, J. D. Interaction of translation initiation factor IF1 with the E. coli

ribosomal A site1. *J. Mol. Biol.* **2000**, *299*, 1-15.

(91) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Hartsch, T.; Wimberly, B. T.; Ramakrishnan, V. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **2001**, *291*, 498-501.

(92) 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.* **1974**, *71*, 1342-1346.

(93) Antoun, A.; Pavlov, M. Y.; Lovmar, M.; Ehrenberg, M. How initiation factors maximize the accuracy of tRNA selection in initiation of bacterial protein synthesis. *Mol. Cell* **2006**, *23*, 183-193.

(94) Grigoriadou, C.; Marzi, S.; Pan, D.; Gualerzi, C. O.; Cooperman, B. S. The translational fidelity function of IF3 during transition from the 30 S initiation complex to the 70 S initiation complex. *J. Mol. Biol.* **2007**, *373*, 551-561.

(95) Milon, P.; Konevega, A. L.; Gualerzi, C. O.; Rodnina, M. V. Kinetic checkpoint at a late step in translation initiation. *Mol. Cell* **2008**, *30*, 712-720.

(96) Karimi, R.; Pavlov, M. Y.; Buckingham, R. H.; Ehrenberg, M. Novel roles for classical factors at the interface between translation termination and initiation. *Mol. Cell* **1999**, *3*, 601-609.

(97) Peske, F.; Rodnina, M. V.; Wintermeyer, W. Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol. Cell* **2005**, *18*, 403-412.

(98) Myasnikov, A. G.; Marzi, S.; Simonetti, A.; Giuliodori, A. M.; Gualerzi, C. O.; Yusupova,

G.; Yusupov, M.; Klaholz, B. P. Conformational transition of initiation factor 2 from the GTP-to GDP-bound state visualized on the ribosome. *Nat. Struct. Mol. Biol.* **2005**, *12*, 1145-1149.

(99) Nelson, D. L.; Lehninger, A. L.; Cox, M. M., *Lehninger principles of biochemistry*. 5th ed.; Macmillan: 2008.

(100) Laursen, B. S.; Sørensen, H. P.; Mortensen, K. K.; Sperling-Petersen, H. U. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 101-123.

(101) Schmeing, T. M.; Voorhees, R. M.; Kelley, A. C.; Gao, Y.-G.; Murphy, F. V.; Weir, J. R.; Ramakrishnan, V. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **2009**, *326*, 688-694.

(102) Miller, D. L.; Weissbach, H., Factors involved in the transfer of aminoacyl-tRNA to the ribosome. In *Molecular Mechanisms of Protein Biosynthesis*, Weissbach, H.; Pestka, S., Eds. Academic Press: 1977; pp 323-373.

(103) Brot, N., Translocation. In *Molecular mechanisms of protein biosynthesis*, Weissbach, H.; Pestka, S., Eds. Academic Press: 1977; pp 375-411.

(104) Rydén, S.; Isaksson, L. A temperature-sensitive mutant of Escherichia coli that shows enhanced misreading of UAG/A and increased efficiency for tRNA nonsense suppressors. *Mol. Gen. Genet.* **1984**, *193*, 38-45.

(105) Kawakami, K.; Inada, T.; Nakamura, Y. Conditionally lethal and recessive UGA-suppressor mutations in the prfB gene encoding peptide chain release factor 2 of Escherichia coli. *J. Bacteriol.* **1988**, *170*, 5378-5381.

(106) Javed, A.; Christodoulou, J.; Cabrita, L. D.; Orlova, E. V. The ribosome and its role in protein folding: looking through a magnifying glass. *Acta Cryst. D* **2017**, *73*, 509-521.

(107) Ramakrishnan, V. What we have learned from ribosome structures. *Biochem. Soc. Trans.* **2008**, *36*, 567-574.

(108) Leon, S. A.; Brock, T. D. Effect of streptomycin and neomycin on physical properties of the ribosome. *J. Mol. Biol.* **1967**, *24*, 391-404.

(109) Mackey, B.; Miles, C.; Parsons, S.; Seymour, D. Thermal denaturation of whole cells and cell components of Escherichia coli examined by differential scanning calorimetry. *Microbiol.* **1991**, *137*, 2361-2374.

(110) 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.

(111) Spitnik-Elson, P.; Greenman, B.; Abramovitz, R. The influence of 6-M urea on 30-S ribosomes of Escherichia coli. *Eur. J. Biochem.* **1974**, *49*, 87-92.

(112) Langer, J.; Acharya, A.; Moore, P. Characterization of the particles produced by exposure of ribosomal subunits to urea. *BBA-Nucleic Acids and Protein Synthesis* **1975**, *407*, 320-324.

(113) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **2005**, *310*, 827-834.

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

(115) Mitra, K.; Schaffitzel, C.; Fabiola, F.; Chapman, M. S.; Ban, N.; Frank, J. Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol. Cell* **2006**, *22*, 533-543.

(116) 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.

(117) Klein, D. J.; Moore, P. B.; Steitz, T. A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **2004**, *10*, 1366-1379.

(118) 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.

(119) Hsiao, C.; Mohan, S.; Kalahar, B. K.; Williams, L. D. Peeling the onion: ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* **2009**, *26*, 2415-2425.

(120) Porter, J.; Hewitt, R.; Hesseltine, C.; Krupka, G.; Lowery, J.; Wallace, W.; Bohonos, N.; Williams, J. Achromycin: a new antibiotic having trypanocidal properties. *Antibiotics & Chemotherapy (Northfield, Ill.)* **1952**, *2*, 409-410.

(121) 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.

(122) Rabinovitz, M.; Fisher, J. A dissociative effect of puromycin on the pathway of protein synthesis by Ehrlich ascites tumor cells. *The Journal of biological chemistry* **1962**, *237*, 477-481.

(123) Morris, A. J.; Schweet, R. S. Release of soluble protein from reticulocyte ribosomes. *Biochim. Biophys. Acta* **1961**, *47*, 415.

(124) Allen, D. W.; Zamecnik, P. C. The effect of puromycin on rabbit reticulocyte ribosomes. *Biochim. Biophys. Acta* **1962**, *55*, 865-874.

(125) Nathans, D. Puromycin inhibition of protein synthesis: incorporation of puromycin into peptide chains. *Proc. Natl. Acad. Sci.* **1964**, *51*, 585-592.

(126) 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.

(127) 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.

(128) Beckmann, R.; Spahn, C. M.; Eswar, N.; Helmers, J.; Penczek, P. A.; Sali, A.; Frank, J.; Blobel, G. Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **2001**, *107*, 361-372.

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

(130) Woolhead, C. A.; Johnson, A. E.; Bernstein, H. D. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* **2006**, *22*, 587-598.

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

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

(133) Kosolapov, A.; Tu, L.; Wang, J.; Deutsch, C. Structure acquisition of the T1 domain of Kv1.3 during biogenesis. *Neuron* **2004**, *44*, 295-307.

(134) Tu, L.; Wang, J.; Deutsch, C. Biogenesis of the T1− S1 linker of voltage-gated K+ channels. *Biochemistry* **2007**, *46*, 8075-8084.

(135) 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.

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

(137) Lu, J.; Kobertz, W. R.; Deutsch, C. Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* **2007**, *371*, 1378-1391.

(138) Arenz, S.; Bock, L. V.; Graf, M.; Innis, C. A.; Beckmann, R.; Grubmüller, H.; Vaiana, A. C.; Wilson, D. N. A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. *Nat. Commun.* **2016**, *7*, 12026.

(139) Choi, K. M.; Brimacombe, R. The path of the growing peptide chain through the 23S rRNA in the 50S ribosomal subunit; a comparative cross-linking study with three different peptide families. *Nucleic Acids Res.* **1998**, *26*, 887-895.

(140) 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.

(141) 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*, 33697-33707.

(142) 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*, e1000581.

(143) 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.

(144) 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*, 138-146.

(145) Sohmen, D.; Chiba, S.; Shimokawa-Chiba, N.; Innis, C. A.; Berninghausen, O.; Beckmann, R.; Ito, K.; Wilson, D. N. Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. *Nat. Commun.* **2015**, *6*, 6941.

(146) 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.

(147) Eisner, G.; Koch, H.; Beck, K.; Brunner, J.; Muller, M. Ligand crowding at a nascent signal sequence. *J. Cell Biol.* **2003**, *163*, 35-44.

(148) 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.

(149) Ullers, R. S.; N.G. Houben, E.; Raine, A. H.-J., Corinne ; 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*, 679-84.

(150) Trylska, J.; Konecny, R.; Tama, F.; Brooks III, C. L.; McCammon, J. A. Ribosome motions modulate electrostatic properties. *Biopolymers* **2004**, *74*, 423-431.

(151) 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.

(152) 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.

(153) Choi, K. M.; Atkins, J. F.; Gesteland, R. F.; Brimacombe, R. Flexibility of the nascent polypeptide chain within the ribosome: Contacts from the peptide N-terminus to a specific region of the 30S subunit. *Eur. J. Biochem.* **1998**, *255*, 409-413.

(154) Hsu, S.-T. D.; Fucini, P.; Cabrita, L. D.; Launay, H.; Dobson, C. M.; Christodoulou, J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16516-16521.

(155) 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.

(156) Chan, S. H.; Waudby, C. A.; Cassaignau, A. M.; Cabrita, L. D.; Christodoulou, J. Increasing the sensitivity of NMR diffusion measurements by paramagnetic longitudinal relaxation enhancement, with application to ribosome–nascent chain complexes. *J. Biomol. NMR* **2015**, *63*, 151-163.

(157) Cabrita, L. D.; Cassaignau, A. M. E.; Launay, H. M. M.; Waudby, C. A.; Wlodarski, T.; Camilloni, C.; Karyadi, M.-E.; Robertson, A. L.; Wang, X.; Wentink, A. S., et al. A structural ensemble of a ribosome–nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* **2016**, *23*, 278.

(158) Waudby, C. A.; Launay, H.; Cabrita, L. D.; Christodoulou, J. Protein folding on the ribosome studied using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *74*, 57-75.

(159) Lange, S.; Franks, W. T.; Rajagopalan, N.; Döring, K.; Geiger, M. A.; Linden, A.; van Rossum, B.-J.; Kramer, G.; Bukau, B.; Oschkinat, H. Structural analysis of a signal peptide inside the ribosome tunnel by DNP MAS NMR. *Sci. Adv.* **2016**, *2*, e1600379.

(160) Eichmann, C.; Preissler, S.; Riek, R.; Deuerling, E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci.* **2010**, *107*, 9111-9116.

Chapter 2

Kinetic Trapping of Folded Proteins Relative to Aggregates under Physiologically Relevant Conditions

The text of this chapter overlaps with a published article (Angela E. Varela, Jonathan F. Lang, Yufan Wu, Matthew D. Dalphin, Andrew S. Stangl, Yusuke Okuno and Silvia Cavagnero. *Journal of Physical Chemistry B* **2018**, 122, 31, 7692-7698).

Author Contributions:
Angela E. Varela and Silvia Cavagnero designed the project. Angela E. Varela participated in sample preparation, UV-vis, CD, DLS, FTIR, fluorescence, and SDS-PAGE gel data collection, analysis and interpretation, and in discussions. Jonathan F. Lang participated in sample preparation, UV-vis, CD, and DLS data collection, analysis and interpretation, and in discussions. Yufan Wu participated in sample preparation, UV-Vis, CD, CD and DLS data collection, analysis and interpretation, and in discussions. Matthew D. Dalphin participated in computer simulations, discussions and manuscript writing. Yusuke Okuno participated in FTIR deconvolution analysis and discussions. Silvia Cavagnero participated in data analysis, discussions and manuscript writing.

2.1 Abstract

Anfinsen's thermodynamic hypothesis does not explicitly take into account the possibility of protein aggregation. Here, we introduce a cyclic-perturbation approach to prove that not only the native state but also soluble aggregates of most proteins can be highly populated under mild, physiologically relevant conditions, even at very low concentration. Surprisingly, these aggregates are not necessarily amyloid in nature, and are usually not observed in bioactive proteins due to the extremely low kinetic flux from the native state towards a region of the chemical-potential landscape encoding aggregates. We illustrate this concept for the representative model protein apomyoglobin – at room temperature and no denaturant – and demonstrate kinetic trapping of the native state relative to at least two different types of soluble, predominantly non-amyloid aggregates. The concentration and temperature dependence of aggregation confirm the above scenario. As a result, Anfinsen's thermodynamic hypothesis should be extended to encompass nonamyloid aggregated states.

2.2 Introduction

The ability of a protein to fold into an ordered, three-dimensional structure is the quintessential example of biological self-organization.[1-3] Given that protein folding is often a prerequisite for activity in the cell, it is not surprising that formation of aggregates, instead of the native state, is sometimes accompanied by severe pathological conditions.[4-6] The physiological need for functional proteins and the growing demand for large-scale protein production via *in vivo* overexpression in biotechnology and pharmaceutical industry[7] require a better understanding of the relation between protein folding and aggregation.[8]

In 1973, Christian Anfinsen proposed the thermodynamic hypothesis for protein folding, stating that the amino acid sequence and the environment entirely determine a protein's three-dimensional structure.[3] Since then, it has become widely assumed that the folding landscape of a protein has evolved with a bias towards the native state, which has the lowest standard-state Gibbs free energy, and freely exchanges with intermediates and unfolded states at equilibrium in a given environment (Figure 2-1A).[9]

However, folding/unfolding under thermodynamic control is not the only feasible scenario. For instance, native and unfolded states (or intermediates) may be kinetically inaccessible with respect to each other, as schematically illustrated in Figure 2-1B. The case of α-lytic protease falls under this category. For this protein, an intermediate is kinetically trapped due to high barriers to folding, unless an N-terminal pro region is included in the protein sequence.[10-13] Hence the folding of α-lytic protease is under kinetic control. On a related note, kinetic trapping of the native state (N) with respect to the unfolded ensemble has been observed in the cellular environment of several organisms by SDS-PAGE and capillary zone electrophoresis.[14-15]

Alternatively, as shown in Figure 2-1C, biologically relevant conformations (native, intermediate and unfolded states) may be kinetically trapped with respect to noncovalent aggregates. The representative landscape in this figure is convenient as it enables direct comparisons between species with a different number of conformational degrees of freedom. Hence, monomeric species and aggregates can be directly compared, upon adopting the standard-state-chemical potential landscapes shown in Figure 1 and in other sections of this study. Figure 1C illustrates the fact that aggregates could be kinetically stable due to large chemical-potential barriers separating them from unfolded (U) or intermediate (I) states under physiologically relevant conditions.[16-17]

Kinetic trapping of proteins may also occur in additional guises. For instance, the native, intermediate and unfolded states may be trapped relative to non-physiologically relevant aggregated states, as shown in Figure 2-1D. However, this concept has been extremely elusive and not conclusively proven under physiologically relevant conditions and in the case of non-amyloid aggregates.

In this work, we show that the N, I and U states of a large number of soluble proteins are kinetically trapped with respect to aggregates. We first demonstrate that the N state (together with any relevant I and U states) of a model protein can be experimentally converted to at least two types of aggregates populated under the same mild conditions that support the presence of the N, I and U states. We conclude that the mammalian protein apoMb is kinetically trapped in the N/I/U states relative to kinetically inaccessible aggregated states.

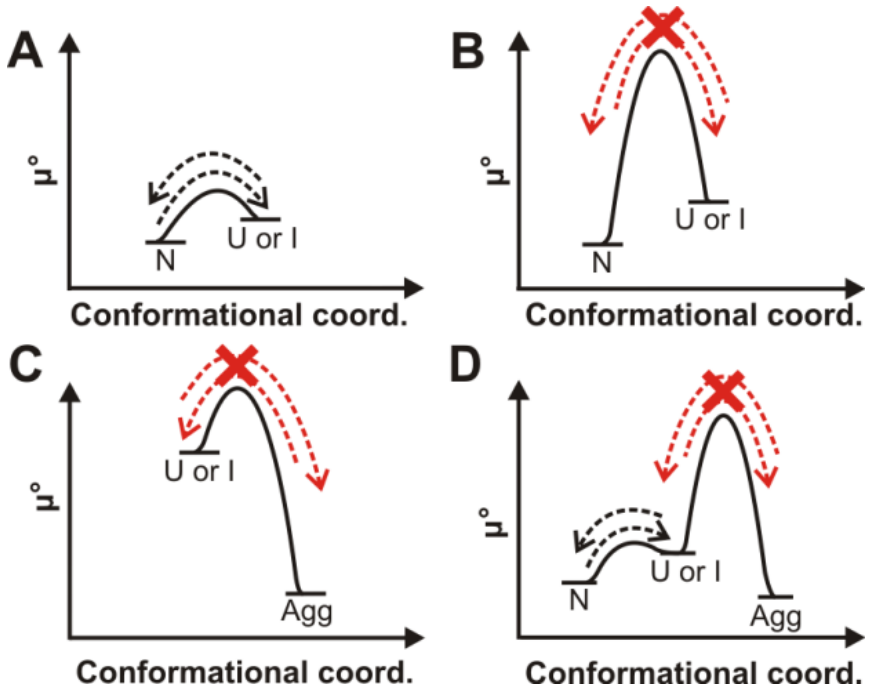


Figure 2-1. Minimalist chemical potential landscapes for protein folding, unfolding and aggregation. The standard-state chemical potential of the monomer (N, I or U) or aggregates (Agg) is shown as function of a generic conformational coordinate. These diagrams illustrate three limiting views that are represented in the current literature (A-C) and an alternative scenario (D). (A) The native state N, (and any relevant intermediate I), is more thermodynamically stable than the unfolded state U under physiologically relevant conditions. All states are kinetically accessible. (B) Kinetically trapped proteins have a native state (or a native state precursor) that does not freely exchange with the unfolded state. (C) Some aggregates, denoted as Agg, are kinetically trapped relative to the unfolded state. (D) Chemical potential landscape proposed in this work: the native, intermediate and unfolded states are kinetically trapped relative to the aggregates.

2.3 Results

Experimental design: the cyclic-perturbation approach

In order to explore the presence of kinetically trapped states, we developed a cyclic-perturbation approach whose main features are described below. This procedure is designed to generate a well-defined starting scenario (e.g., the monomeric protein comprising the N, I and U states in equilibrium with each other in solution), populated under physiologically relevant conditions, followed by generation of another scenario (e.g., a self-associated state) that does not typically interconvert with the former at the same temperature, pressure and solution conditions as the initial state. This goal is achieved upon switching between physiological, non-physiological and back to physiological conditions. This simple procedure enables probing whether the two classes of scenarios can exist, and whether the corresponding states can be populated under identical experimental conditions, i.e., within the same standard-state chemical potential landscape.

For instance, kinetically inaccessible states at room temperature can be generated upon transient treatment with denaturants or at high temperatures followed by slow denaturant dilution (upon dialysis) or by slow return to the original temperature (upon cooling). Excursions to non-physiologically relevant landscapes (at high temperature and in the presence of denaturants) enable the escape from any kinetically trapped scenarios, while returning to regular solution conditions ensures the applicability of our findings to the original physiologically relevant landscape. A final stage of the cyclic perturbation approach involves newly generating the initial state, to ensure that the protein has not undergone any covalent modifications during the process. It is worth noticing that, while individual portions of the cyclic perturbation approach have already been routinely

employed in the literature, the cyclic conversion of a native state to aggregate and then back to native is unique to the approach proposed here.

We applied the cyclic perturbation strategy to apomyoglobin (apoMb, Figure 2-2), a well-characterized representative model protein whose native structure is an example of the ubiquitous globin fold.[18-22]

A cartoon providing a pictorial illustration of the features of the cyclic perturbation approach is shown in Figure 2-3. Note that green and gray regions denote physiologically relevant and non-physiologically relevant conditions, respectively. A more detailed version of this approach, including further experimental details, is presented in Supplementary Figure S1.

As shown in detail in the next section, application of the cyclic perturbation approach to apoMb showed that the physiologically relevant N state (together with any relevant I and U states) can be experimentally converted to at least two types of aggregates that are usually not kinetically accessible. In addition, we subsequently converted these aggregates back to the native, intermediate and unfolded state to ensure that the same protein species is capable of again populating the original state.

The folded state of apoMb is kinetically trapped with respect to aggregates

The presence of apoMb monomer-(N/I/U) and soluble aggregated states was assessed by a combination of far-UV circular dichroism (CD) and dynamic light scattering (DLS). CD was employed to probe the protein secondary structure while DLS provided information on particle size in solution, respectively.

As shown in Figure 2-4, we started from the low-pH urea-denatured protein, which is known to be monomeric and devoid of significant secondary structure[23], as assessed by far-UV

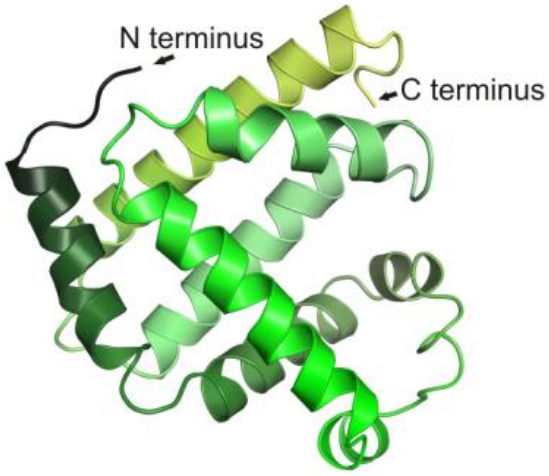


Figure 2-2. Three-dimensional structure of native sperm whale myoglobin. The image was created with PyMol 1.3 (Delano Sci.) employing known coordinates (PDB: 2mbw).

CD and DLS, respectively (Figure 2-4a, species a) at 17 ± 8 μM total protein concentration.

Dialysis into pH 6 buffer generated a polydisperse sample, denoted here as species b, with an overall α-helical content. DLS shows the presence of a species with a small hydrodynamic radius (R_g), which we attribute to the native state and two larger-size helical aggregates (Figure 2-4b, Supplementary Table S1). An estimate of the percent mass of the different peaks (Supporting Table S2) shows that the major species in this sample (ca. 94%) is the refolded monomer. Gel-filtration chromatography was then carried out to isolate the monomeric native state (Figure 2-4c, species c), presumably in rapid equilibrium with more poorly populated monomeric unfolded and intermediate states[24]. DLS analysis was carried out to confirm the exclusive presence of the monomer. If needed, the sample was subjected to filtration with 0.22 μm and/or 0.02 μm filters (Millipore, Supplementary Figure S1).

At this juncture, it is important to mention that DLS linewidths should not be interpreted as a measure of the size distribution of the different species, given that DLS linewidth values are affected by large errors inherent in the regularization analysis carried out with the DYNALS algorithm[25]. Accordingly, the linewidth of the DLS peaks varied somewhat from experiment to experiment (Supplementary Table S3) and cannot be meaningfully interpreted.

Incubation of species c at room temperature for about a week led to formation of no aggregate (Figure 2-5). This result shows that the pure native protein is kinetically trapped relative to any aggregates, on the experimental timescale of at least a few days.

Heating of species c followed by slow cooling produced species d (Figure 2-4d). The latter comprises a ca. equimolar mixture of the native state and a soluble aggregate that is rich in β-sheet content, according to CD and DLS (see also the percent mass estimates in Supplementary Table S2).

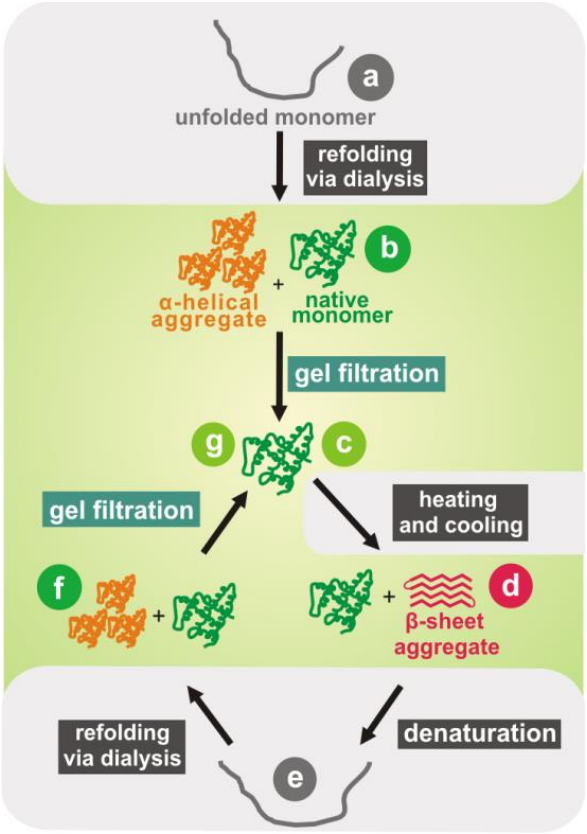


Figure 2-3. Schematic representation of the "cyclic perturbation" procedure employed in this study. Regions with a green and gray background denote physiologically and non-physiologically relevant conditions, respectively. (a) Pure apoMb was denatured in 6 M urea at pH 2.3, then (b) dialyzed against pH 6.0 buffer to generate the native monomer and an α-helical aggregate. Then, (c) gel filtration chromatography led to isolation of the pure monomeric native state. (d) The native monomer underwent three cycles of heating to 90°C followed by slow cooling to room temperature to obtain a mixture of α –helical native monomer and β-sheet aggregate at room temperature. (e) Unfolding in denaturant followed (in 6 M guanidinium hydrochloride at pH 2.3 for 72 hrs at room temperature). (f) The unfolded protein was dialyzed against pH 6.0 buffer to regenerate native monomer and α-helical aggregate. (g) The native monomer was then isolated via gel filtration. Species denoted as b, c, d, f and g were incubated at room temperature for one week before proceeding to the following step, to test kinetic stability.

The slow cooling (after the heating) enabled the system to return to the physiologically relevant chemical-potential landscape. Prolonged incubation of species d at room temperature did not lead to any significant changes in secondary structure. However, some very large aggregates were likely produced given that the quality of the DLS autocorrelation functions reproducibly degraded (Figure 2-5). Additional details are provided in the section focusing on slow escape from kinetic traps.

Denaturation of species d (Figure 2-4 e) followed by refolding upon dialysis into physiological buffer (Figure 2-4 f) and gel filtration generated, once again, the monomeric N/I/U states of the protein (Figure 2-4 g, species g). Mass spectrometry analysis of the final product showed no variations in mass relative to the starting protein, confirming that the above procedure introduced no covalent modifications.

In summary, as a result of the cyclic perturbation approach the monomeric α-helical native state, together with the α-helical and the β-strand aggregates of apoMb were shown to exist under the same physiologically relevant conditions, as pictorially shown by the structural cartoons in the green region of Figure 2-3.

Hence, we conclude that all of the above states of apoMb, monomers and aggregates, belong to the same chemical-potential landscape. The pure monomeric native state (species c of Figure 2-3) does not spontaneously interconvert into any aggregates upon incubation for several days at room temperature (species c of Figure 2-5). Hence we also deduce that pure native apoMb, in equilibrium with I and U states, is kinetically trapped relative to at least two classes of aggregated states, at room temperature.

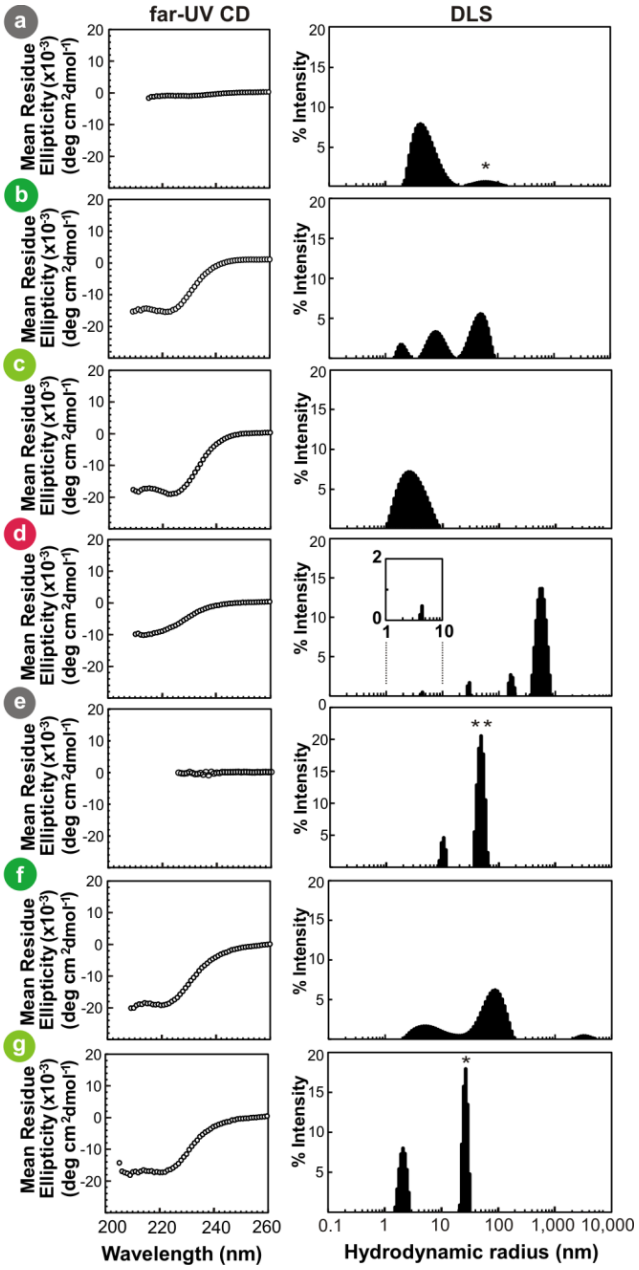


Figure 2-4. Dynamic light scattering and far-UV circular dichroism spectra of apoMb. Panel labels a-g correspond to the states described in Figure 2-3. (a) Unfolded apoMb (17 ± 8 μM) in 6 M urea, 10 mM sodium acetate (NaOAc), pH 2.3. (b) Native monomer and α-helical aggregates after dialysis in 10 mM sodium acetate at pH 6.0. (c) Native monomer after gel filtration in 10

mM NaOAc at pH 6.0. (d) Mixture of β-sheet aggregate and α-helical monomer in 10 mM NaOAc at pH 6.0 obtained after heating (3 cycles, each up to 90°C for 6 hrs) followed by slow cooling. (e) Unfolded protein in 6 M guanidinium chloride, pH 2.3. (f) Native monomer and α-helical aggregates in 10 mM NaOAc at pH 6.0. (g) Native monomer in 10 mM NaOAc at pH 6.0, obtained after denaturation, refolding, and gel-filtration of the β-sheet aggregate. The single asterisk in panels a and g denotes a poorly populated species with negligible ($\leq 0.1\%$) estimated percent mass (see Methods), while the double asterisk in panel e denotes a species with moderate (7.1%) estimated percent mass.

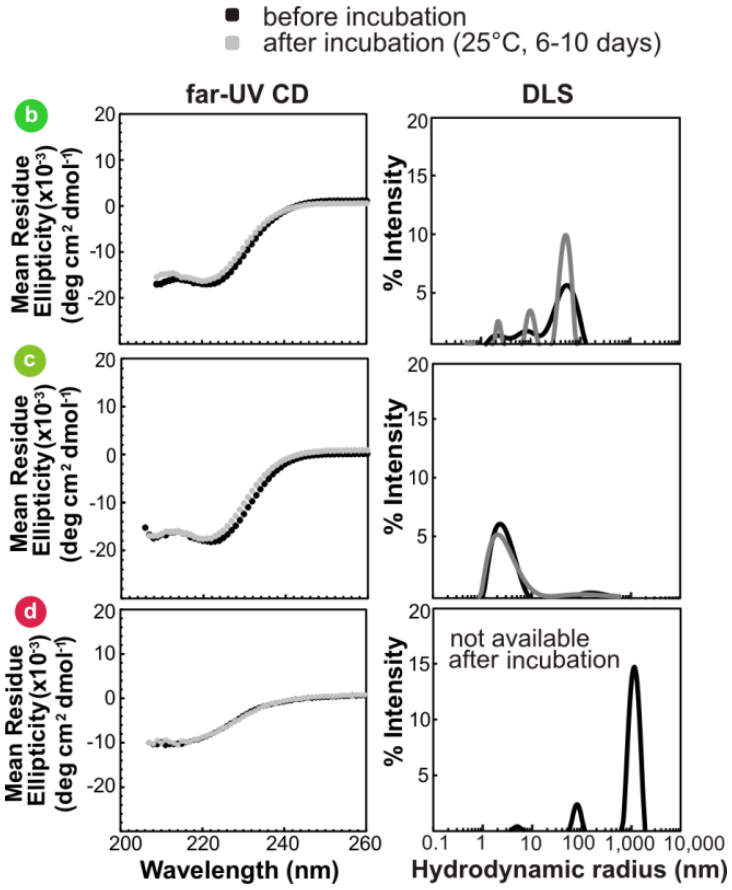


Figure 2-5. Effect of long-term incubation (6-10 days) at room temperature of apoMb species b-d (17±8 μM) of Figure 2-3. Secondary structure and aggregation state were monitored by far-UV circular dichroism (CD) and by dynamic light scattering (DLS). After DLS sample prep (centrifugation) for species d after incubation, the sample became too dilute to acquire reliable data, suggesting that the majority of the protein formed aggregates large enough to pellet during centrifugation.

The apoMb aggregates generated upon heating and cooling are large and have β-strand, mostly non-amyloid, character

To gain additional, more direct evidence on the size of the soluble apoMb aggregates that contribute to species d of Figure 2-3, generated upon heating and cooling, we performed static light scattering (SLS) analysis. We found that species d of Figure 2-3 encompasses very large aggregates with a weight-average mass of 12,550,000 ± 95,000 Da, corresponding to self-associated species encompassing, on the average, 724 ± 6 monomers (Supplementary Table S4). This result was obtained adopting a spherical shape model during the SLS data processing. Very similar values were obtained when models assuming other shapes were employed (Supplementary Table S4). Hence we conclude that soluble species d includes very large aggregates encompassing, on the average, hundreds of monomeric units of the protein.

The CD minimum centered at ca. 210 nm is compatible with both β-strands and local turns (see CONTINLL deconvolution analysis in Supplementary Information), hence we performed Fourier Transform infrared (FTIR) analysis to clearly discriminate between sheets and turns. Importantly, FTIR was also instrumental to tease apart regular β-sheets from β-sheets belonging to amyloids.[26]

Panels A and B of Figure 2-6 show the raw (black) Gaussian deconvoluted (gray) FTIR spectra apoMb, while panel C displays the corresponding second-derivative spectra before (dashed) and after (solid) the heating-and-cooling steps. Difference spectra are also shown in Supplementary Figure S2. The broad FTIR resonance centered at 1646 cm^{-1} dominated the spectrum of species c in Figure 2-6A (see also Table 2-1) and confirms the presence of a strong α-helical component.

The remaining FTIR spectral features of species c are consistent with the CD deconvolution data in Supplementary Table S5. After the heating and cooling step (Figure 2-3), the protein displays a much weaker FTIR feature close to 1646 cm^{-1} (Figure 2-6B, Table 2-1), suggesting a significant decrease in native-state α-helical content, consistent with some residual small population of α-helical native state, as detected by CD (Supplementary Table S5). The intense FTIR band of the heated-and-cooled sample at 1635 cm^{-1} is assigned to a non-amyloid β-sheet.[26] Interestingly, bands at 1635 and 1655 cm^{-1} similar to ours were also observed by Fink and coworkers in apoMb non-amyloid β-sheet inclusion bodies.[27] The FTIR peaks at 1617 and 1682 cm^{-1} are generally diagnostic of β-strand and (or) turn[28-29]. However, very intense bands in the 1610 - 1630 cm^{-1} region are known characteristics of amyloid-type β-sheets.[26] Given that the intensity of our 1617 cm^{-1} band is very weak in our case, we conclude that the heated-and-cooled sample is either non-amyloid or may include a small fraction of amyloid-type β-sheet.

This analysis is confirmed by the 2^{nd}-derivative spectra of Figure 2-6C. The 2^{nd}-derivative analysis is instrumental to identify complex spectral features in FTIR and is highly complementary to the spectral deconvolution analysis. Salient features of the parent spectra were identified via 2^{nd}-derivative local minima.

In summary, the far-UV CD data deconvolution in Supplementary Table S5 and the FTIR data of Figure 2-6 show that the apoMb aggregates formed upon heating-and-cooling have predominantly β-strand character.

In order to further explore whether the β-strand aggregates bear any relations with known amyloid aggregates previously identified for a related protein (horse myoglobin) at pH 9[30], we performed thioflavin T (ThT) binding assays on two classes of heated-and-cooled apoMb samples. The results are shown in Figure 2-7. First, we ran an experiment on a sample consisting of heated

and cooled apoMb at pH 9, prepared according to procedures by Fandrich *et al.*[30]. This reference sample displays a significant ThT response, consistent with amyloid character being acquired by heated-and-cooled sperm whale apoMb at the non-physiological pH 9. Then, we examined the β-stranded aggregated apoMb under the milder and more physiologically relevant pH 6. The latter sample displays no ThT response. Additional fluorescence emission spectra collected on the heated-and-cooled sample in the absence of ThT, testing for the presence of intrinsic amyloid fluorescence as described [31-33], yielded only a very weak band (data not shown).

Hence, the collective evidence gathered in this work shows that the β-sheet sperm whale apoMb aggregates generated upon heating and cooling at the physiologically relevant pH 6 have either no, or very little, amyloid character. This species is structurally very different from the control sperm whale aggregates generated at pH 9, which are amyloid in nature.

In summary, this study shows that the chemical-potential landscape of apoMb under physiologically relevant conditions includes a large variety of aggregated states, some of which have a non-amyloid nature (i.e., predominantly α-helical) and some others (i.e., primarily β-strand) that are not ThT-responsive and are therefore unlikely to be amyloid. Hence the concept of protein native state kinetic trapping with respect to aggregates is considerably more general than previously thought[34], as the aggregated states that we identified are not necessarily amyloid in nature.

Importantly, our heating/cooling procedure may not have exhaustively explored the chemical-potential landscape of apoMb. Hence the landscape of this protein may also include additional regions of conformational space not documented in this study.

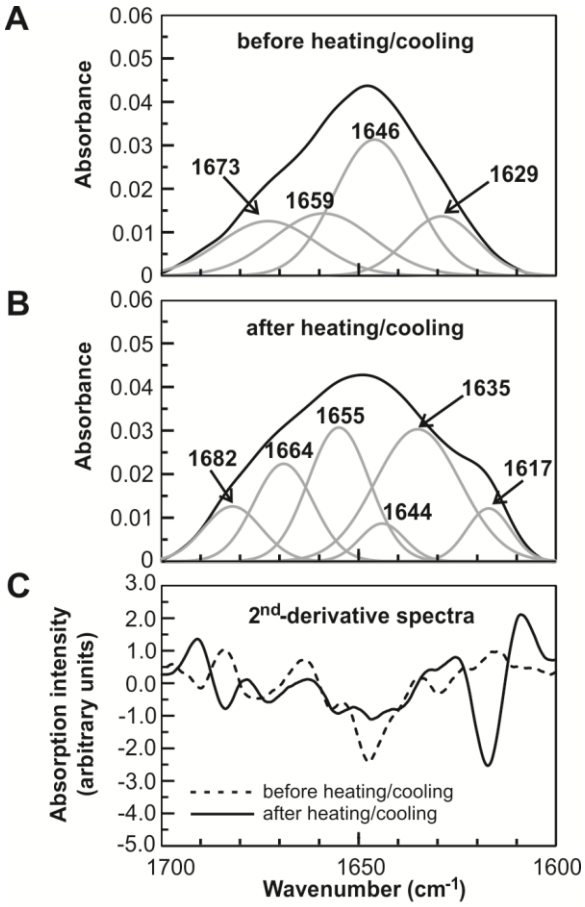


Figure 2-6. Fourier-transform infrared (FTIR) spectra of apoMb before and after heating and cooling. ApoMb (15-17 μM) in D_2O (A) before and (B) after the heating-and-cooling steps, corresponding to species c and d of Figure 2-3, respectively. Separate spectral features were assessed by fitting the data to Gaussian curves (grey). (C) Corresponding second-derivative FTIR spectra before and after the heating and cooling steps.

Table 2-1. FTIR-data corresponding to Figure 2-6, including secondary structure assignments [28-29, 35].

Before heating/cooling		
Wavelength at peak maximum (nm)	Normalized area[a]	Secondary structure[b]
1629	0.156	solvated[c]
1646	0.410	α
1659	0.234	α, un, β, t
1673	0.200	un, β, t
After heating/cooling		
Wavelength at peak maximum (nm)	Normalized area[a]	Secondary structure[b]
1617	0.072	β, t
1635	0.347	α, β, t
1644	0.0520	α, un, β, t
1655	0.253	α, un
1669	0.179	un, t
1682	0.099	β, t

[a] Areas were normalized relative to the total areas of all peaks.

[b] The symbols α, β, t, un denote α-helix, β-sheet, turn and unordered structure, respectively.

[c] Refs. [28, 35-36]

Slow escape of native apoMb from its kinetically trapped state can be promoted by the presence of aggregates with specific secondary structure. Long-term incubation of samples b and d led to additional insights. DLS analysis of sample b, which contains the native state and α-helical aggregates, shows that the identity of all the species does not vary with time over the course of several days (Figure 2-5). The DLS mass estimate confirms that the percent of native monomer in the sample stays unchanged (Supplementary Table S6). Hence the presence of helical aggregates does not seed the conversion of the monomer to aggregates.

In contrast, sample d, which contains the native state in the presence of β-sheet aggregates, gives rise to traces of insoluble species upon long-term incubation (Figure 2-5). This conclusion was reached because we could not collect meaningful DLS data after incubation: centrifugation before data collection renders the DLS data too noisy due to sample depletion via precipitation, and omission of the centrifugation step yields unusable data due to the presence of extremely large species suspended in solution.

In summary, the behavior of samples b and d shows that some types of aggregates (β-sheet-containing), but not others (α-helix-containing) are capable to slowly promote conversion of the apoMb native state to aggregates. Hence kinetic trapping of the native state can be slowly abrogated by the presence of some, but not all, types of aggregates. This finding has fundamental implications for our understanding of the interplay between protein folding and aggregation. For instance, this result suggests that production of newly synthesized protein in a β-sheet aggregated conformation must be avoided at all costs in the cellular environment. While the role of amyloid seeds in promoting aggregation is fairly well understood,[37] we still know very litle about how seed conformation (especially in the case of non-amyloids) relates to seeding efficiency.[37]

Higher temperatures enable more frequent barrier crossings

In order to probe the effect of incubation temperature on the degree of apoMb kinetic trapping, we incubated the protein and collected corresponding DLS data at temperature higher than room temperature. While pure folded apoMb was shown to be kinetically trapped in its native state with respect to aggregates at room-temperature (species c in Figure 2-5), incubation for a week at 37 °C led to three classes of outcomes as shown in Figure 2-8. In essence, either the protein stays unchanged in its monomeric state or it gives rise to some aggregates with hydrodynamic radius close to 100 nm. Hence, the pure folded protein exhibits stochastic behavior at 37 °C. At this temperature apoMb is still generally kinetically trapped in its native state but it occasionally hops to the aggregated side of the chemical potential landscape.

Therefore, as expected, higher temperature supplies extra thermal energy to enable crossing kinetic barriers that are otherwise inaccessible at 25 °C at low micromolar concentration.

In summary, our cyclic perturbation approach established that, at room temperature, atmospheric pressure and low micromolar concentration, the unfolded, major folding intermediate and folded states of sperm whale apomyoglobin (which are progressively more thermodynamically stable[24] and are sampled rapidly) are kinetically trapped relative to a variety of aggregated states.

2.4 Discussion

The native state of many proteins is kinetically trapped with respect to corresponding aggregated states

This study shows explicit evidence for the kinetic trapping of the native state of sperm whale apoMb relative to at least two different types (α-helix and β-strand-rich) of soluble and

insoluble aggregates under mild, physiologically relevant conditions, even at small (i.e., low μM) protein concentration.

A broader view of Anfinsen's thermodynamic hypothesis encompassing all protein states, including aggregates, is necessary to understand the behavior of proteins *in their natural environment.* Note that earlier studies have focused on nonphysiologically relevant conditions.[38] The implications of the above findings are significant as aggregated states are thermodynamically as viable as native states, for many proteins, even at moderate concentrations and mild pH values. Interconversion between monomers and aggregates does not usually take place under physiologically relevant conditions. Given that the ribosome, molecular chaperones and other relevant cellular components routinely ensure the exclusive production of the biologically relevant N/I/U states *in vivo* upon translation, non-physiologically relevant self-associated states are typically not sampled during the lifetime of a cell.

The above scenario may change dramatically in anomalous settings such as *in vivo* recombinant protein overexpression[39-40] or disease-related conditions[5] where, despite the preservation of some of the physiologically relevant variables (e.g., pH, temperature, ionic strength), other parameters vary drastically and lead to a high flux through the barrier (e.g., at high protein concentration) or to an entirely different landscape.

Relative thermodynamic stability of apoMb aggregates

We demonstrated that apoMb can populate the N/I/U states and two main types of soluble aggregates under physiologically relevant conditions, as illustrated in the representative chemical-potential landscape of Figure 2-9. Upon slow dialysis, the protein can presumably cross the barriers (of the progressively changing landscape during the refolding step) in either direction. Under these

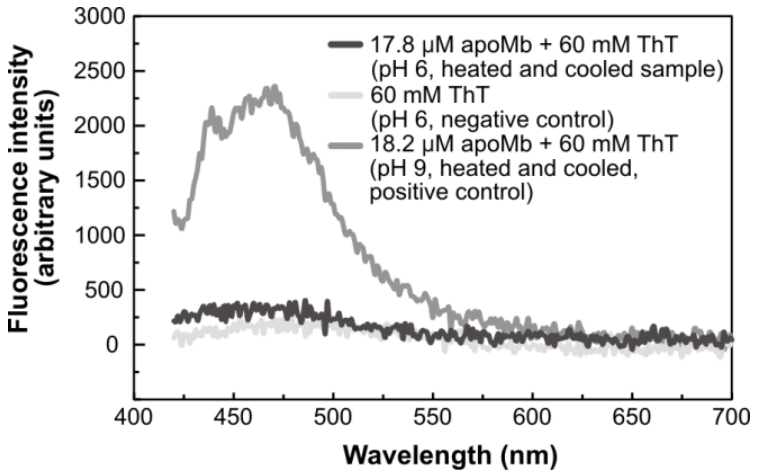


Figure 2-7. Heated and cooled apoMb aggregates do not contain amyloid aggregates. Thioflavin T binding assay of species d of Figure 2-3, showing no evidence for amyloid properties by the heated and cooled β-sheet apoMb soluble aggregate.

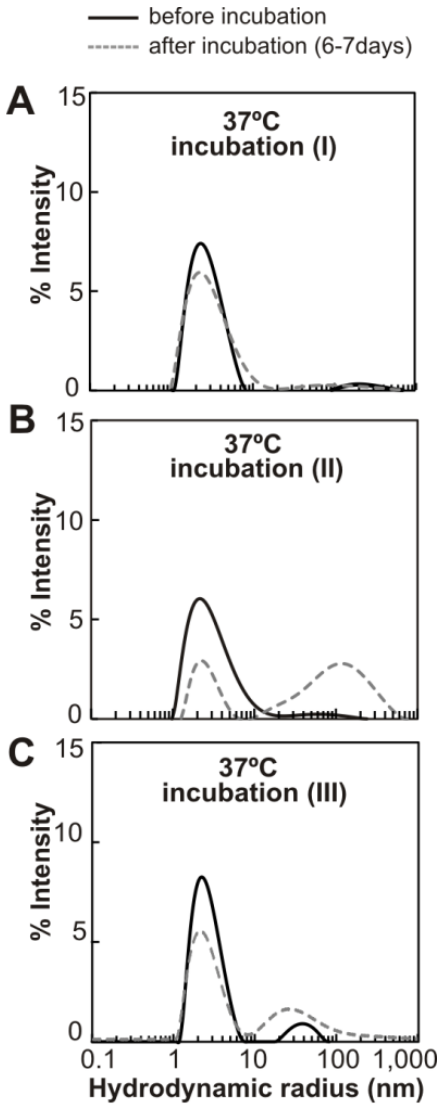


Figure 2- 8. High temperature incubation leads to stochastic behavior of apoMb. Effect of long-term incubation at 37 °C of gel-filtered monomeric native apoMb (i.e., species c of Figure 2-3, 13 µM). Panels A-C show the results of three representative independent experiments illustrating the heterogeneous behavior displayed by individual apoMb samples upon incubation for 7 days.

conditions, both native state and helical aggregate are produced.

This result suggests that the above states have similar standard-state chemical potentials. Similarly, the β-sheet aggregate generated upon heating and slow cooling can presumably reversibly sample both monomeric and aggregated states while at high temperatures. However, a gradual return to the original landscape by slow cooling leads to effective kinetic trapping. As a result of this process, comparable masses of native and aggregated conformations are produced. Hence we infer that the β-sheet and α-helical aggregates detected in this work likely have a thermodynamic stability similar to that of the native state.

Recurrence of kinetic trapping in Nature

A variety of fundamental processes in Nature are known to be under kinetic control, including the combustion of sucrose[41] and the hydrolysis of peptides, proteins and nucleic acids.[41-42]. Although the products are more thermodynamically stable than the reactants, the above processes do not occur under physiological conditions due to the extremely low flux across the relevant kinetic barriers. Further, in a non-biological context, glasses[43-44] are well-known materials formed under kinetic control. Hence the existence of kinetic trapping in proteins is yet another manifestation of a recurring concept in Nature.

Conclusions

In summary, we showed that the native state of the mammalian sperm whale apoMb is kinetically trapped under physiologically relevant conditions, relative to aggregated states, according to the representative landscape of Figure 9. We propose that this phenomenon is relevant

to other proteins from a variety of organisms, given that the sequence requirements for protein structure formation and aggregation are similar across species.[45-46]

2.5 Methods

ApoMb expression and purification

Recombinant sperm whale apoMb is a 154-residue, all-helical protein (including the N-terminal Met resulting from bacterial expression, Figure 2-2). ApoMb was expressed in *E. coli* and purified as described[23]. The purity of the final protein was assessed by reverse-phase HPLC and found to be ≥ 95%. The molecular mass of the final protein produced as a lyophilized powder, was verified by MALDI mass spectrometry (expected average molecular mass 17,331 Da).

General features of the "cyclic perturbation" approach

The main features of the cyclic perturbation approach introduced in this work are illustrated in Figure 2-3. The species generated at each step are labeled with letters a through g in Figure 2-3, and are referred to with the same letter in the sub-titles below, which describe additional experimental details. In addition, Supplementary Figure S1 provides a pictorial representation of many of the relevant experimental details. All buffers and stock solutions were filtered through a 0.22 μm syringe filter (Millipore) prior to use.

i. Preparation of unfolded apoMb (generation of species a)

Pure lyophilized apoMb was dissolved in 6.0 M urea and 10 mM sodium acetate adjusted to pH 2.3, subject to bath sonication (B1510MTH ultrasonic cleaner, Branson) for 7 minutes, and then filtered through 0.22 μm and 0.02μm (Whatman Anotop 10, Sigma-Aldrich) syringe filters. Unless otherwise stated, the total protein concentration was 17±8 μM. The solution was then incubated at room temperature for 24 hrs followed by spectroscopic analysis by electronic

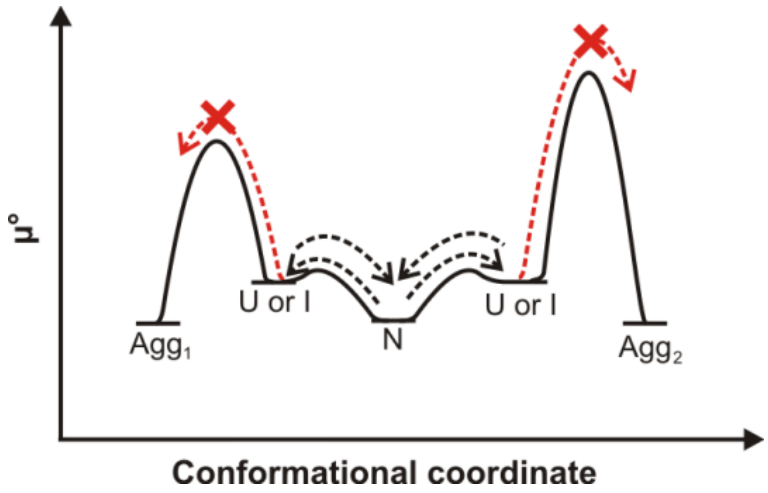


Figure 2-9. Minimalist representation of extended chemical-potential landscape encompassing monomeric and self-associated protein states, proposed in this work. This landscape is consistent with the experimental behavior of apoMb under physiologically relevant conditions. This landscape illustrates the fact that the native and one or more aggregated states (shown here as Agg_1 and Agg_2) are often kinetically trapped relative to one another. More complex landscapes were not considered due to the lack of supporting experimental evidence, at this stage.

absorption, far-UV circular dichroism (CD) and dynamic light scattering (DLS).

ii. **Preparation of native state and α-helical aggregate of apoMb (step a → b)**

Urea-unfolded apoMb was extensively dialyzed against 10 mM sodium acetate at pH 6.0 with a 3,500-MWCO 0.5-3 mL Slide-A-Lyzer dialysis cassette (Thermo Scientific) until urea concentration <6.0 μM. After dialysis, the solution pH was confirmed to be 6.0. If necessary, the refolded apoMb samples were concentrated with an Amicon Ultra 0.5-3 mL filter (3K MWCO, Millipore) to 17±8 μM concentration. Sample solubility was confirmed by electronic absorption at 280 nm before and after sample centrifugation (14,548 g for 10 min). Spectroscopic analysis by electronic absorption, CD and DLS was performed before and after 6-10-day incubation at room temperature.

iii. **Isolation of native apoMb monomer (step b → c)**

To separate the native apoMb monomer from the α-helical aggregate, apoMb samples were loaded onto a NAP-25 or PD-10 gel filtration column (GE Healthcare, fine grade) and eluted with aqueous 10 mM sodium acetate at pH 6.0. Protein concentration was then adjusted to 17±8 μM by ultrafiltration (see above for details). Electronic absorption, CD and DLS data were collected before and after 6-10-day incubation at room temperature. Acquisition of electronic absorption data at 280 nm before and after gel filtration showed that 29 ± 4 % of the protein was lost on the column, presumably as large soluble aggregates.

iv. **Generation of solution containing β-sheet aggregates of apoMb (step c → d)**

The gel-filtered solution containing pure native apoMb in 10 mM sodium acetate at pH 6.0 was passed through a 0.02 μm syringe filter to remove any potential dust or residual large aggregates. The solution was then transferred to PCR tubes and subject to three successive identical cycles of heating and cooling in a thermal cycler (Mastercycler PCR apparatus,

Eppendorf). Each cycle consisted of 6 hr heating at 90°C followed by overnight cooling to room temperature. Protein concentration was then verified again to test for any solvent losses during the heating steps, and lack of insoluble aggregates was verified by centrifugation (14,548 g for 10 min) followed by visual inspection for any residual pellets. No pellets were found in studies at 17±8 μM apoMb concentration. Electronic absorption, CD and DLS data were collected before and after 6-10-day incubation at room temperature.

***v.* Unfolding of solution containing β-sheet aggregates of apoMb (step d → e)**

Guanidinium hydrochloride (GnHCl) was added to the solution containing the β-sheet aggregates to a total GnHCl concentration of 6-10 M. The pH was adjusted to 2.3 and the solution was incubated for 72 hrs at room temperature. Electronic absorption, CD and DLS data were collected after the 72 hr incubation.

***vi.* Regeneration of native state and α-helical aggregate of apoMb (step e → f)**

The unfolded monomeric protein from the previous step was dialyzed extensively (see step *ii* for details) against 10 mM sodium acetate at pH 6.0. If necessary, the protein was concentrated (Amicon Ultra 0.5-3 mL, 3K MWCO Millipore) to preserve a protein concentration of 17±8 μM. Electronic absorption, CD and DLS data were collected before and after 6-10-day incubation at room temperature.

***vii.* Regeneration of native apoMb monomer and additional characterization (step f → g)**

The sample from the previous step was loaded onto a NAP-25 or PD-10 gel filtration column (GE Healthcare, fine grade) and eluted with aqueous 10 mM sodium acetate at pH 6.0. The protein was concentrated back to 17±8 μM. Electronic absorption, CD and DLS data were collected before and after 6-10-day incubation at room temperature. Electrospray mass spectrometry analysis was performed at this final stage of the cyclic perturbation procedure to

verify that apoMb had not undergone any major covalent modification. A single peak with an average mass of 17,338 Da was obtained, confirming the absence of proteolysis and Met oxidation.

Electronic Absorption

Protein concentration was determined by UV absorbance at 280 nm. The previously reported extinction coefficient (at 280 nm) of 14,134 $M^{-1}cm^{-1}$ was used for apoMb.[47]

Far-UV circular dichroism

Far-UV circular dichroism (CD) spectra of apoMb were recorded at room temperature on a MOS-450 spectropolarimeter (BioLogic Science Instruments) at 17±8 μM protein concentration with 0.1 cm path length, quartz cuvettes. Four scans were collected with 1-nm step size and 20 s averaging time per point. Data were adjusted against appropriate blanks. Raw ellipticities were converted to mean residue ellipticity (MRE, in deg cm^2 $dmol^{-1}$) according to

$$MRE = \frac{\theta}{10 \cdot C \cdot (N_A - 1)\ell} \ , \quad (1)$$

where θ is the experimentally determined ellipticity in millidegrees, C is the protein concentration (moles/liter), N_A is the number of amino acids, and ℓ is the cuvette path length in cm. The CD data were further analyzed by the data deconvolution program CONTINLL[48-49] to identify the relative amounts and types of secondary structure in each sample.

Dynamic light scattering

Dynamic light scattering (DLS) data were collected at 90° (angle between incident light and detector) at room temperature on a DynaPro Nanostar instrument (Wyatt Technology) equipped with an avalanche photodiode detector and a GaAs laser operating at 658 nm. Data for each sample were signal averaged (20 acquisitions, 5 s each) and analyzed with the DYNAMICS software (Wyatt Technology). Prior to the measurements, samples (250 μL aliquots) were centrifuged at 14,548 g for 10 min to remove dust or other spurious large particle. The top portion

(100 μL) of the supernatant was transferred to a quartz microcuvette (HE0171, Wyatt Technology, 1 mm) and immediately used for the measurements. Refractive index values of the relevant buffer systems, serving as reference solutions lacking the protein, were measured with an Abbe refractometer (Thermo Spectronic, Fisher Scientific), for each of the solution conditions analyzed in this work.

Translational diffusion gives rise to fluctuations in the intensity of the scattered light. Fluctuation frequency is a function of molecular size. Data were collected as autocorrelation functions, which are a measure of the fluctuating intensity of the scattered light due to the continuously changing distance between particles in solution. Autocorrelation functions were analyzed according to the regularization method[50], which explicitly accounts for the potential presence of multiple species of different size and fits the experimental autocorrelation function to a sum of exponentials according to the DYNALS algorithm.[25] DYNALS assumes that particle motions are exclusively due to translational diffusion.

Translational diffusion coefficients D_t for the different exponentially decaying components of the autocorrelation function were obtained via the DYNALS algorithm and converted to apparent hydrodynamic radii (R_h) via the Stokes Einstein equation

$$R_h = \frac{k_B T}{6\pi\eta D_t}, \tag{2}$$

where k_B is the Boltzmann constant (in $kg \cdot m^2 \cdot s^{-2} \cdot K^{-1}$), T is the temperature (in K), η is the viscosity of the solution in the absence of the protein (in $kg \cdot s^{-1} \cdot m^{-1}$), and R_h is the protein particle hydrodynamic radius (in m). R_h represents the average radius of a hypothetical hard sphere with a translational diffusion coefficient identical to the measured value. Viscosity values to be used in equation 2 for the relevant experimental conditions were gathered from the NanoStar software

library except for the solutions containing denaturing agents, in which case reference solution viscosities in the absence of the protein were deduced by manual extrapolation of published experimental data in the presence of the relevant denaturants.[51]. Approximate estimates of the percent mass corresponding to each peak were generated assuming spherical shape for each of the sample particles (Supplementary Table S2)

Static light scattering

ApoMb solutions at stages a and d of the cyclic perturbation procedure of Figures 2-3 and 2-4 were studied by static light scattering (SLS). SLS data were collected at room temperature on a DynaPro Nanostar instrument (Wyatt Technology, Santa Barbara, CA) at a 90° angle upon excitation at 658 nm. Samples were centrifuged as described in the DLS section and only the top portion of the supernatant was used. In addition, apoMb solutions were filtered (0.22 μm, Anotop 10 Whatman syringe filters) prior to data collection. The reference pure water and the pure buffer underwent the same filtration procedure. Scattering intensity of protein sample, buffer and pure water were measured independently. The refractive index of water and the buffer were measured with an Abbe refractometer (Thermo Spectronic, Fisher Scientific).

The Raleigh ratio is related to experimental scattering intensities according to

$$R_\theta = \left[\frac{I_{sample} - I_{buff}}{I_{ref}}\right] R_{ref,\theta} \left(\frac{n_{buff}}{n_{ref}}\right)^2, \quad (3)$$

where R_{ref} is the Raleigh ratio of the reference solvent (i.e., toluene), the I_{sample}, I_{buff} and I_{ref} are the scattering intensities of the sample, buffer and toluene, respectively; and n_{buff} and n_{ref} are the refractive index of the buffer and toluene. After experimental determination of the latter five parameters and knowing $R_{ref,\theta}$ (a constant of the instrument for $\theta = 90°$), $R_{\square}$ was obtained from relation 3. The Raleigh ratio R_θ is also equal to

$$R_\theta = K\,\langle M\rangle_w c_{tot} P\left(\theta, R_g\right), \quad (4)$$

where c_{tot} is the independently determined total concentration of protein monomeric particles (in mg/ml) and $P(\theta,R_g)$ is the shape factor whose value is known for given molecular shapes (e.g., sphere, coil, rod) and from experimental estimates of hydrodynamic radius carried out on the same sample via DLS. K is equal to

$$K = \frac{4\pi^2 n_{buff}^2}{\lambda^4 N_A}\left(\frac{dn}{dc}\right)^2, \quad (5)$$

where N_A is the Avogadro number, λ is the incident wavelength (658 nm) and dn/dc is the polarizability of the protein solution, which was set to 0.185 mL/g, as customary for proteins.

The weight-average mass of the protein $\langle M\rangle_w$, which was determined from equation 4 after experimentally assessing the remaining parameters, is equal to

$$\langle M\rangle_w = \Sigma c_i M_i / \Sigma c_i \ , \quad (6)$$

where c_i and M_i are the concentration and mass of each particle i. The reported weight-average mass is the average from two independent experiments.

Infrared Spectroscopy

Fourier transform infrared (FTIR) data were collected at room temperature on an iS10 FT-IR spectrometer (Nicolet-Thermo Scientific) equipped with a deuterated triglycine sulfate (DTGS) detector and nitrogen gas purging. A CaF_2 sample cell with a 56 μm Teflon spacer was used. 256 scans with a 1 cm^{-1} spectral resolution were collected for each sample. Samples were prepared as follows. Due to the presence of residual trifluoroacetic acid (TFA) an H_2O in the HPLC-purified lyophilized protein, which could significantly affect amide I band intensity[52], samples were treated as described[29] to remove any residual traces of TFA and H_2O. Briefly, the pure protein powder

was dissolved in 10 mM DCl and lyophilized overnight. The powder was then dissolved in an excess of D_2O, incubated at room temperature for one hr and lyophilized overnight. Prior to data collection, lyophilized samples were dissolved in D_2O and the pH* (uncorrected electrode reading) was adjusted to 5.6. After the heating/cooling step (step d of Fig. 3), the pH* was readjusted to 5.6 if necessary. FTIR spectra were baseline-corrected and smoothed according to a Savitsky-Golay algorithm[53], and 2nd derivative spectra were determined. The major local minima identified in the 2nd-derivative spectra were taken as evidence of individual transitions and were used as input for Gaussian deconvolution of the original baseline-corrected spectra.

2.6 Acknowledgments

We are grateful R. Burgess, K. Dill, D. Eisenberg and Z. Wang for useful discussion and to T. Zhang and M. Zanni for access to their FTIR instrumentation.

No competing financial interests have been declared. This work was funded by the National Science Foundation (NSF) grants MCB-0951209 and MCB-1616459 (to S.C.), including NSF REU supplements (to J.F.L.). A.E.V. was the recipient of an NSF Graduate Research Fellowship.

2.7 References

(1) Gething, M. J.; Sambrook, J. Protein folding in the cell. *Nature* **1992**, *355*, 33-45.

(2) Dill, K. A. Dominant forces in protein folding. *Biochem.* **1990**, *29*, 7133-7155.

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

(4) Ellis, R. J.; Pinheiro, T. Medicine: Danger — misfolding proteins. *Nature* **2002**, *416*, 483-484.

(5) Ross, C. A.; Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **2004**, *10*, S10-S17.

(6) Dobson, C. M. Protein aggregation and its consequences for human disease. *Protein Pept. Lett.* **2006**, *13*, 219-227.

(7) Walsh, G. Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **2014**, *32*, 992-1000.

(8) Mousseau, N.; Derreumaux, P. Exploring energy landscapes of protein folding and aggregation. *Front. Biosci.* **2008**, *13*, 4495-4516.

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

(10) Baker, D.; Sohl, J. L.; Agard, D. A. A protein-folding reaction under kinetic control. *Nature* **1992**, *356*, 263-265

(11) Baker, D.; Agard, D. A. Kinetics versus thermodynamics in protein-folding. *Biochem.* **1994**, *33*, 7505-7509.

(12) Jaswal, S. S.; Sohl, J. L.; Davis, J. H.; Agard, D. A. Energetic landscape of a-lytic protease optimizes longevity through kinetic stability. *Nature* **2002**, *415*, 343-346

(13) Jaswal, S. S.; Truhlar, S. M. E.; Dill, K. A.; Agard, D. A. Comprehensive analysis of

protein folding ctivation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J. Mol. Biol.* **2005**, *347*, 355-366.

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

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

(16) Kellermayer, M. S. Z.; Karsai, A.; Benke, M.; Soos, K.; Penke, B. Stepwise dynamics of epitaxially growing single amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 141-144.

(17) Raccosta, S.; Manno, M.; Bulone, D.; Giacomazza, D.; Militello, V.; Martorana, V.; San Biagio, P. L. Irreversible gelation of thermally unfolded proteins: structural and mechanical properties of lysozyme aggregates. *Eur. Biophys. J.* **2010**, *39*, 1007-1017.

(18) Dyson, H. J.; Wright, P. E., Elucidation of the protein folding landscape by NMR. In *Nuclear Magnetic Resonance of Biological Macromolecules, Part C*, James, T. L., Ed. 2005; Vol. 394, pp 299-+.

(19) Ballew, R. M.; Sabelko, J.; Gruebele, M. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759-5764.

(20) Gruebele, M. Downhill protein folding: evolution meets physics. *C. R. Biol.* **2005**, *328*, 701-712.

(21) 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.

(22) 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.

(23) 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.

(24) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993**, *262*, 892-896.

(25) Goldin, A. A. DYNALS: Software for particle size distribution analysis in photon correlation spectroscopy. http://www.softscientific.com/science/WhitePapers/dynals1/dynals100.htm (accessed July, 1 2018).

(26) Giorgia, Z.; R.H., K. M.; G., M. M.; Marcus, F. FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils. *Protein Sci.* **2004**, *13*, 3314-3321.

(27) Shivu, B.; Seshadri, S.; Li, J.; Oberg, K. A.; Uversky, V. N.; Fink, A. L. Distinct β-sheet structure in protein aggregates determined by ATR–FTIR spectroscopy. *Biochem.* **2013**, *52*, 5176-5183.

(28) Singh, B. R., *Infrared analysis of peptides and proteins: principles and applications*. ACS Publications: 1999.

(29) 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. *Biochem.* **2003**, *42*, 7090-7099.

(30) Fandrich, M.; Fletcher, M. A.; Dobson, C. M. Amyloid fibrils from muscle myoglobin -

Even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature* **2001**, *410*, 165-166.

(31) Shukla, A.; Mukherjee, S.; Sharma, S.; Agrawal, V.; Radha Kishan, K. V.; Guptasarma, P. A novel UV laser-induced visible blue radiation from protein crystals and aggregates: scattering artifacts or fluorescence transitions of peptide electrons delocalized through hydrogen bonding? *Arch. Biochem. Biophys.* **2004**, *428*, 144-153.

(32) Chan, F. T.; Schierle, G. S. K.; Kumita, J. R.; Bertoncini, C. W.; Dobson, C. M.; Kaminski, C. F. Protein amyloids develop an intrinsic fluorescence signature during aggregation. *Analyst* **2013**, *138*, 2156-2162.

(33) Pinotosi, D.; Buell, A. K.; Dobson, C. M.; Kaminski Schierle, G. S.; Kaminski, C. F. A Label-free, quantitative assay of amyloid fibril growth based on intrinsic fluorescence. *ChemBioChem* **2013**, *14*, 846-850.

(34) 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.

(35) Gulotta, M.; Gilmanshin, R.; Buscher, T. C.; Callender, R. H.; Dyer, R. B. Core formation in apomyoglobin: probing the upper reaches of the folding energy landscape. *Biochem.* **2001**, *40*, 5137-5143.

(36) Reisdorf Jr., W. C.; Krimm, S. Infrared amide I' band of the coiled coil. *Biochem.* **1996**, *35*, 1383-1386.

(37) Kim, H. J.; Chatani, E.; Goto, Y.; Paik, S. R. Seed-dependent accelerated fibrillation of α-

synuclein induced by periodic ultrasonication treatment. *J. Microbiol. Biotechnol.* **2007**, *17*, 2027-2032.

(38) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884-90.

(39) Burgess, R. R., Refolding solubilized inclusion body proteins. In *Guide to Protein Purification, Second Edition*, 2009; Vol. 463, pp 259-282.

(40) Baneyx, F.; Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **2004**, *22*, 1399-1408.

(41) Eisenberg, D.; Crothers, D. M., *Physical Chemistry with Applications to the Life Sciences*. Benjamin / Cummings: 1979.

(42) Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* **1998**, *45*, 351-353.

(43) Wong, J.; Angell, C. A., *Glass: Structure by Spectroscopy*. Marcel Dekker: New York, 1976.

(44) Angell, C. A. Thermodynamic aspects of the glass transition in liquids and plastic crystals. *Pure Appl. Chem.* **1991**, *63*, 1387-1392.

(45) Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riekel, C. The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc. Chem. Res.* **2006**, *39*, 568-575.

(46) Uversky, V. N. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell. Mol. Life Sci.* **2003**, *60*, 1852-1871.

(47) Kurt, N. R., 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.

(48) Sreerama, N.; Woody, R. W. A Self-Consistent Method for the Analysis of Protein

Secondary structure from circular dichroism. *Anal. Biochem.* **1993**, *209*, 167-176.

(49) Sreerama, N. W., R.W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **2000**, *287*, 252-260.

(50) Provencher, S. W. Inverse problems in polymer characterization - direct analysis of polydispersity with photo correlation spectroscopy. *Macromol. Chem. Phys.* **1979**, *180*, 201-209.

(51) Kawahara, K.; Tanford, C. Viscosity and density of aqueous solutions. *J. Biol. Chem.* **1966**, *241*, 3228-3232.

(52) Gaussier, H.; Morency, H.; Lavoie, M. C.; Subirade, M. Replacement of trifluoroacetic acid with HCl in the hydrophobic purification steps of pediocin PA-1: a structural effect. *Appl. Environ. Microbiol.* **2002**, *68*, 4803-4808.

(53) Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627-1639.

2.8 Appendix

Supplementary Results

CONTINLL deconvolution of far UV CD data

The CONTINLL algorithm[1-3] was used to deconvolute the CD spectral data and estimate the fraction of different types of apoMb secondary structure of species c and d of Figure 2-3, i.e., before and after the heating-and-cooling step. As shown in Supporting Table S5, the α-helical content of species c was 66%, consistent with the native structure of pure apoMb. After heating and cooling, the α-helical content decreased over three-fold to only 20% while the β-structure increased six-fold from 5% to 30%. This result is consistent with the generation of a new species bearing β-strand secondary structure, after heating and cooling.

Higher apoMb concentrations lead to formation of insoluble aggregates

Aggregation, which occurs primarily by a series of bimolecular encounters, is strongly concentration-dependent and is expected to be more severe at high concentration.

The experiments discussed so far were performed at low (i.e., 17 ± 8 μM) total apoMb concentration. To probe the concentration dependence of apoMb kinetic trapping, we also tested significantly higher (102 ± 4 μM) apoMb concentrations.

ApoMb species b (Figure 2-3) is fully soluble even at high protein concentration (~100 μM). However, we observed that ca. 20% of the total protein turns into insoluble aggregates upon incubation of concentrated (102 ± 4 μM) apoMb at room temperature for 7-9 days (Supplementary Table S7). This result shows that the chemical-potential landscape of apoMb also includes insoluble aggregates. However, these species can only be generated at high total protein concentration (i.e., ca. 100 μM in the case of apoMb), when the critical concentration for their

formation has been met. The slow formation of the insoluble species indicates that the native state is only partially kinetically trapped, at higher concentration.

Interestingly, Supplementary Table S8 also shows that the concentration of the soluble aggregates of species b stays virtually unchanged at low and high total protein concentration, after 7-9 days long-term incubation.

Additional evidence on the effect of higher concentration is provided by analysis of the gel-filtered native monomer (i.e., species c of Figure 2-3) after long-term incubation at the higher (102 ± 4 μM) protein concentration. This gel-filtered concentrated and purely monomeric species gives rise to a small amount (4 ± 2 %) of insoluble aggregate, but only when subject to long-term incubation (Supplementary Table S6). DLS analysis of the soluble fraction (Supplementary Figure S3) shows a large native-state peak, together with smaller peaks due to soluble aggregates. On the other hand, the DLS percent-mass analysis shows 100% mass for the native state and 0% mass for both the soluble aggregated species. Hence the population of soluble aggregates is in fact negligible.

Hence, we conclude that apoMb is for the most part kinetically trapped on the timescale of a few hours, even at higher protein concentrations up to ca. 100 μM, in the absence of other aggregated forms of the protein in solution. On the other hand, it is clear that, overall, higher protein concentrations favor formation of insoluble aggregates and render the kinetic trapping process less stringent.

Supplementary Materials and Methods

Thioflavin T fluorescence assays

Lyophilized sperm whale apoMb was dissolved in filtered (0.22 μm) 10 mM sodium acetate (pH 6.0). The protein (17±8 μM) was then heated for 3x 6 hr cycles followed by slow

cooling to room temperature after each heating step. A control sample was prepared as above except that apoMb was dissolved in 50 mM sodium borate (pH 9.0) and heated for 24 hrs at 90 °C to obtain amyloid fibrils as described.[4] After heating, both protein solutions were individually mixed with thioflavin T (ThT) in 10 mM sodium acetate (pH 6.0) to yield 10 μM apoMb and 60 uM ThT. Fluorescence emission spectra were recorded with a PC1 photon counting steady-state spectrofluorimeter (ISS, Champaign IL) upon excitation at 390 nm. Slit widths for both excitation and emission channels were 4 nm. No filters were used.

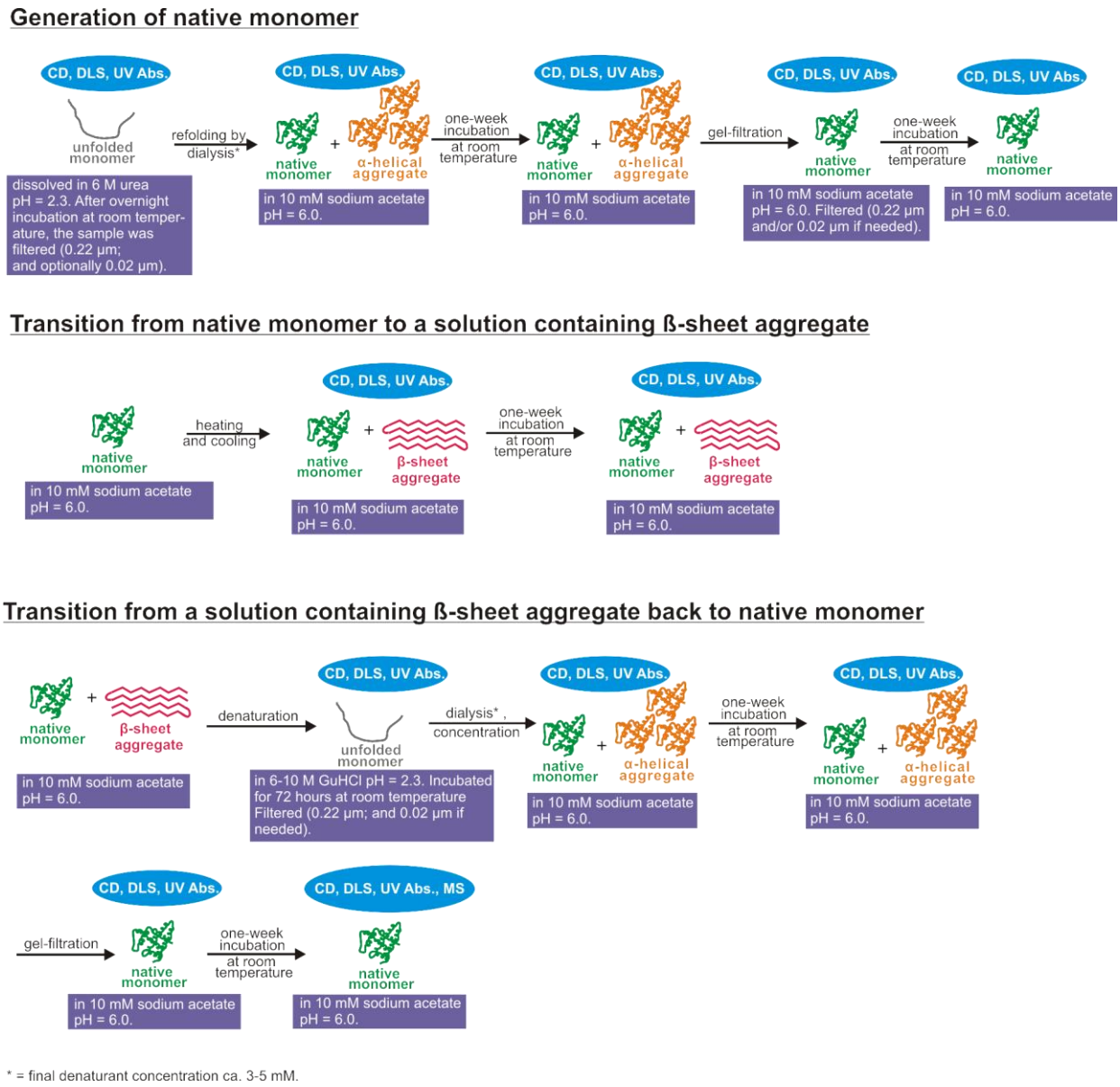
To assess the presence of fibrils in the soluble *E. coli* protein mixture in the reduced state, we heated the solution at 70 °C for 20 hrs at a concentration of 3.3 mg/mL in the presence of 1 mM DTT. Samples were then allowed to cool to room temperature. A 1:45 dye-to-protein (mg/mL) ratio was used to calculate the appropriate amount of protein used to prepare the sample for data collection. The final concentration of ThT was 20 μM in all samples. Lysozyme was used as a positive control to generate fibrils. Lysozyme solutions (3.3 mg/mL) were heated 48 hrs at 70 °C in PBS buffer at pH 2.0 and allowed to cool to room temperature. PBS buffer (66.8 mM K_2HPO_4, 10 mM NaCl, 0.8 mM $MgCl_2$, 0.1 μM $CaCl_2$, pH 7.4 or 2.0) was used to dilute the protein samples, if necessary. Fluorescence emission spectra for these experiments were recorded with a PC1 photon counting steady-state spectrofluorimeter (ISS, Champaign IL) upon excitation at 440 nm. Slit widths for both excitation and emission channels were 4 nm. No filters were used.

Assessment of fraction of insoluble aggregate of refolded apoMb

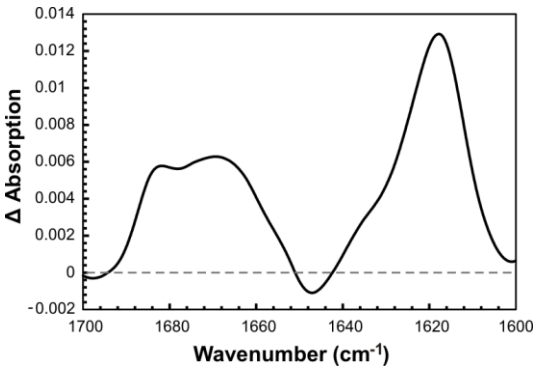
In order to estimate the percent of suspended insoluble aggregate of refolded apoMb (species b of Figure 2-3), samples were centrifuged, and electronic absorption spectra were collected before and after centrifugation. Briefly, after collection of an initial electronic absorption spectrum on species b of Figure 2-3, samples were centrifuged for 10 min at 13,000 rpm in an

Eppendorf 5402 microcentrifuge at 22 °C. The top 100 µL of the supernatant were then collected and UV absorbance was measured again. The difference between absorbance values at 280 nm before and after centrifugation was related to the protein concentration difference and then to the percent protein loss due to insoluble aggregate formation via the Lambert-Beer's law, using the known apoMb extinction coefficient of 14,134 $M^{-1}cm^{-1}$.[5] When needed, a λ^{-4} correction was applied to the electronic absorption spectra, to minimize artifacts due to elastic light scattering.

Supplementary Figures



Supplementary Figure S1. Detailed cartoon representation of the steps corresponding to the "cyclic perturbation" approach followed for apomyoglobin (apoMb) in this study. This cartoon corresponds to a more detailed view of the schematic procedure in Figure 2-3.



Supplementary Figure S2. Difference spectrum of the FTIR data of Figure 2-6 A and B (i.e., spectrum after heating – spectrum before heating).

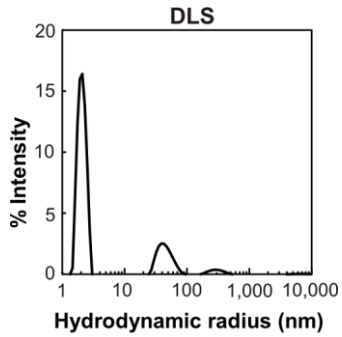


Figure S3. Dynamic light scattering (DLS) spectrum of concentrated (102.9 μM) refolded gel-filtered apoMb (species c of Figure 3) after long-term (17 days) incubation.

Supplementary Tables

Supplementary Table S1. Average number of peaks corresponding to aggregated apoMb species (see Figure 2-4) detected via dynamic light scattering (DLS). For each species, 2-3 independent experiments were carried out.

ApoMb species	Average number of peaks denoting apoMb aggregates
b	2.0 ± 0.1
d	3.0 ± 0.1
f	1.5 ± 0.3

Supplementary Table S2. Average hydrodynamic radii and estimated % masses of monomeric and aggregated apoMb species derived from DLS data.The different species are defined according to Figure 2-3.[(1)]

ApoMb species	Hydrodynamic radius R_h (nm)	Estimated % mass
a	5.2 ± 0.5	97 ± 1
b	2.5 ± 0.5	94 ± 2
	27 ± 15	6 ± 2
	94 ± 38	0.3 ± 0.1
c	3.2 ± 0.3	99 ± 1
d	2.5 ± 0.8	59 ± 17
	19 ± 5	1.4 ± 0.3
	124 ± 29	15 ± 13
	661 ± 160	24 ± 16
e	5 ± 1	93 ± 13
	45 ± 8	7 ± 4
f	5 ± 1	82 ± 9
	71 ± 20	16 ± 8
	1,720 ± 1,600	4 ± 3
g	2.0 ± 0.5	99 ± 1
	45 ± 17	0.1 ± 0.1

[(1)] The number of independent experiments ranged between 2 and 9.

Supplementary Table S3. Average linewidths and hydrodynamic radii of selected monomeric apoMb species (see Figure 2-4) detected via dynamic light scattering (DLS). For each species, 2-4 independent experiments were carried out.

ApoMb species	Average Hydrodynamic radius (nm)	Width at half height (nm)
a	6.0 ± 0.4	4.9 ± 0.6
c	3.5 ± 0.2	7.0 ± 1.2
e	3.6 ± 0.6	2.2 ± 0.6
g	1.6 ± 0.5	0.9 ± 0.2

Supplementary Table S4. Static light scattering (SLS) data for heated-and-cooled apoMb, i.e., species d according to the definition of Figure 2-3.

Shape	Weight-average mass (Da)	Number of monomers within aggregate
Sphere	12,551,000 ± 95,000	724 ± 6
Hollow sphere	12,938,000 ± 104,000	747 ± 6
Random coil	14,288,000 ± 142,000	824 ± 8
Star with 2 arms	14,304,000 ± 143,000	825 ± 8
Star with 3 arms	13,952,000 ± 132,000	805 ± 8
Star with 4 arms	13,747,000 ± 127,000	793 ± 7
Star with 5 arms	13,619,000 ± 123,000	786 ± 7

Supplementary Table S5. Distribution of secondary-structure types for species c and d of Figures 2-3 and 2-4 by deconvolution of the far-UV CD data via CONTINLL.

	Fraction of secondary structure[a]			
	α	β	t	un
Species c (gel filtered, native apoMb)	0.66	0.05	0.13	0.16
Species d (heated and cooled apoMb)	0.20	0.30	0.19	0.31

[a] The symbols α, β, t, un denote α-helix, β-sheet, turn and unordered structure, respectively.

Supplementary Table S6. Effect of total apoMb concentration on the aggregation of species c after room-temperature incubation for 7-9 days.

Total apoMb concentration (μM)	% of monomer and soluble apoMb aggregate	% of insoluble apoMb aggregate
17 ± 2	100	0
99 ± 2	96 ± 1	4 ± 2

Supplementary Table S7. Effect of initial apoMb concentration on the aggregation properties of species b of Figure 2-3 upon room-temperature incubation for 7-9 days.

Conc. [a] (μM)	M[b] (%)	SA[c] (%)	M + SA (%)	IA [d]	Total A[e]
20 ± 1	70 ± 6	30 ± 2	100 ± 0	0	30 ± 2
102 ± 4	54 ± 6	27 ± 4	81 ± 10	19 ± 10	46 ± 6

[a] Total-protein concentration.
[b] Monomer.
[c] Soluble aggregate.
[d] Insoluble aggregate.
[e] Total aggregate.

Supplementary Table S8. Average % mass of the non-aggregated species before and after incubation (6-9 days) at low concentration.

ApoMb species	Before incubation (% mass)	After incubation (% mass)
b	94 ± 2	95 ± 1
c	99 ± 1	100 ± 1
d	59 ± 7	-

Supplementary references

(1) Sreerama, N. W., R.W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **2000**, *287*, 252-260.

(2) Sreerama, N.; Woody, R. W. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **1993**, *209*, 167-176.

(3) Johnson, W. C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **1999**, *35*, 307-312.

(4) Fandrich, M.; Fletcher, M. A.; Dobson, C. M. Amyloid fibrils from muscle myoglobin - Even an ordinary globular protein can assume a rogue guise if conditions are right. *Nature* **2001**, *410*, 165-166.

(5) Kurt, N. R., 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.

Chapter 3

Kinetic Trapping of Folded Proteins Relative to Aggregates under Physiologically Relevant Conditions: Extension to the *E. coli* Soluble Proteome

The text of this chapter overlaps with a published article (Angela E. Varela, Jonathan F. Lang, Yufan Wu, Matthew D. Dalphin, Andrew S. Stangl, Yusuke Okuno and Silvia Cavagnero. *Journal of Physical Chemistry B* **2018**, 122, 31, 7692-7698).

Author Contributions:
Angela E. Varela and Silvia Cavagnero designed the project. Angela E. Varela participated in sample preparation, ThT data collection, SDS-PAGE analysis and interpretation, discussions and manuscript writing. Matthew D. Dalphin carried out the computer simulations, discussions and manuscript writing. Andrew J. Stangl participated in SDS-PAGE gel analysis, discussions and manuscript writing. Silvia Cavagnero participated in data interpretation, discussions and manuscript writing.

3.1 Abstract

As illustrated in Chapter 2, an extension of Anfinsen's thermodynamic hypothesis is needed to include protein aggregates, in addition to the native state, unfolded ensemble and folding/unfolding intermediates. Here we show that the majority of the soluble *E. coli* proteome is kinetically trapped in the non-aggregated state. Hence, the existence and low kinetic accessibility of large aggregates-that are not necessarily amyloid at room temperature and pH 6-7 is a general phenomenon. We also show that the average critical protein concentration for aggregation of most of the bacterial proteome is extremely small, much lower than the typical cellular protein concentration. As a result, the thermodynamic driving force for protein aggregation is large even if aggregation does not usually occur in healthy cells due to kinetic trapping. A broader view of Anfinsen's thermodynamic hypothesis encompassing all protein states, including aggregates, is necessary to understand the behavior of proteins in their natural environment.

3.2 Introduction

Anfinsen's Nobel prize-winning work demonstrated that a protein's most thermodynamically favored state is its native state. Although this concept has been assumed throughout protein folding, not all protein folding pathways are governed by thermodynamic control. Several well-known proteins fold via kinetically-controlled pathways (e. g., α-lytic protease, serpins, etc.).[1-2]

Because Anfinsen's work assumes that the native state is the most thermodynamically favored state, the potential presence and role of aggregates in the thermodynamic landscape of a protein remain largely unexplored.

In Chapter 2 we explicitly demonstrated that two non-amyloid aggregated states of apomyoglobin are kinetically inaccessible relative to its native state under physiologically relevant conditions. Interestingly, once populated, these aggregates do not covert back to the native state, hence they are kinetically trapped. Likewise, the native state is also kinetically trapped relative to the aggregates under physiologically stable conditions.

In this work, we extend our analysis to the entire soluble portion of the proteome of the *E. coli* bacterium. We show that aggregated states exist under physiologically relevant conditions, yet the biologically relevant conformations (N, I and U) are separated from the aggregates by potential-energy barriers with negligible flux in either direction. This concept is schematically illustrated in Figure 2-1 D of Chapter 2. Our experiments show that the average critical protein concentration (CPC) for aggregation of most of the bacterial proteome is much smaller than the typical protein concentration.

We conclude that, at pH 7, the large majority of the soluble proteins in *E. coli* are kinetically trapped in the N/I/U states relative to kinetically inaccessible aggregated states.

3.3 Results

Most of the soluble bacterial proteome is kinetically trapped relative to aggregates

To determine if whether the finding presented in Chapter 2 is of general significance, i.e., whether native-state kinetic trapping also applies to other known proteins, we examined the entire soluble bacterial proteome of *E. coli.* We prepared an S100-like collection of *E. coli* soluble proteins[3] and analyzed all of them in parallel. We estimate that this collection encompasses over 1,000 folded proteins. Given that the cyclic perturbation approach used in Chapter 2 is too complex to be applied to the S100 protein mixture, we resorted to a simplified version of the method comprising an initial heating and cooling step. The heating and cooling was carried out upon incubating the solution to 70 °C for 20 hrs followed by cooling back to room temperature, followed by SDS-PAGE analysis and mass spectrometry characterization. We focused on insoluble aggregates only, which are a subset of all possible aggregates and could be readily identified via a solubility assay. The procedure was repeated starting from different initial S100 total-protein concentration, to assess the dependence of our findings on macromolecular concentration.

In order to ensure that the results would not be biased by the generation of potential intermolecular disulfide crosslinks arising from Cys-containing proteins during the heating and cooling steps, we performed heating/cooling both under non-reducing and reducing conditions. The results shown in Figure 3-1 (see also Figs. S1 and S2) indicate that the majority of the S100 protein collection, i.e., ca. 85% and 75% of the S100 proteins under non-reducing and reducing conditions, respectively, is kinetically trapped in the native state relative to insoluble aggregates, at total protein concentration greater than ca. 0.5 mg/ml (Fig. 3-1, panels B and D).

The latter value corresponds to the apparent critical protein concentration (CPC) beyond which insoluble aggregates should form, assuming a simple nucleated polymerization model[4-5], if

the system were under thermodynamic control. Remarkably, this experimentally detected concentration (Fig. 3-1, panels B and D) is orders of magnitude smaller than the typical protein concentration of live *E. coli* cells[6-7] suggesting that the majority of the soluble *E. coli* proteome is kinetically trapped with respect to insoluble aggregates.

One may note that it would be best to consider CPC values of individual proteins within the *E. coli* proteome, rather than globally assessing the entire S100 mixture, to support of the above argument. This task is technically challenging due to insufficient gel resolution. On the other hand, an estimate of individual CPC values of the major *E. coli* proteins can be obtained upon direct visual inspection of the concentration-dependence profiles of soluble and insoluble fractions of individual proteins in the gels of Figures S1 and S2. This analysis shows that most individual proteins have CPC values remarkably close to those of the entire mixture.

The soluble bacterial proteome undergoes only a moderate degree of degradation upon heating

The gels in Figure 3-1 show that the protein bands before and after the heating step have similar molecular-weight patterns, suggesting the absence of major degradative phenomena. To address at a higher degree of resolution whether the heating-and-cooling process may have caused covalent modifications, we sought for the presence of oxidation and/or deamidation products in the S100 soluble proteome. We performed this test after the heating-and-cooling step, by nano-liquid chromatography electrospray ionization tandem mass spectrometry (nano-LC-ESI - MS/MS). We analyzed the 10 most abundant proteins of the S100 mixture after the heating-and-cooling step, as representative samples, identified according to their exponentially modified protein abundance index (emPAI, Tables S1-S3).

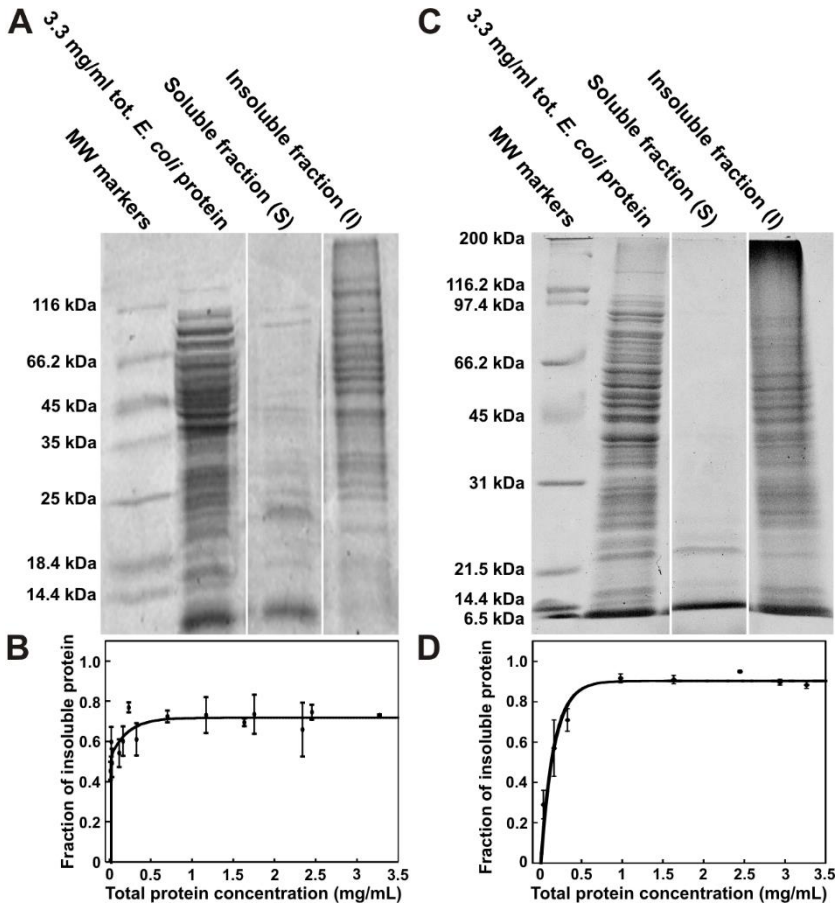


Figure 3-1. Evidence for kinetic trapping of the native state of the soluble *E. coli* proteome relative to the corresponding aggregated states. (A) Representative lanes of SDS-PAGE analysis of soluble *E. coli* proteome upon heating for 20 hrs at 70 °C followed by slow cooling to room temperature. Sample centrifugation led to the generation of a supernatant (S) and a pellet (P), shown separately in the gel. Samples contained 1mM DTT. (B) Plot of the fraction of insoluble, aggregated *E. coli* proteome generated by the above procedure as a function of total-protein concentration. The solid line is meant to guide the eye. Error bars represent the standard error of 3 independent experiments. Panels (C) and (D) display data similar to panels A and B, respectively, except that DTT was omitted (non-reducing conditions).

These proteins are regarded as a reliable sample of the population due to the large emPAI scores. We targeted the heating-and-cooling performed under reducing conditions. Two parameters were assessed. First, we evaluated the decrease in unmodified protein sequence coverage upon heating. The heated and-cooled insoluble aggregate (see pellet sample of Fig. 3-1) underwent a 1.7% ± 1.8% decrease in unmodified sequence coverage after heating-and-cooling, suggesting that this step did not cause any significant changes in the covalent structure of the insoluble protein aggregate. Second, we evaluated the frequency of occurrence of intact and modified (via oxidation and/or deamidation) individual secondary-fragmentation peptides. This type of analysis revealed that only a moderate fraction (18% ± 4%) of the proteins in the insoluble fraction generated after heating and cooling shows any signs of modification.

In summary, the two above complementary data-analysis procedures provided an overall estimate of the covalent quality of the pellet after the heating-and-cooling step, and showed that only a moderate fraction of the S100 pellet undergoes some modifications as a result of heating. Further details are provided in the Supplementary Methods and Tables S1-S3.

Does both homo- and hetero-aggregation occur in the bacterial proteome?

The observed formation of insoluble aggregates, which led us to propose that the *E. coli* soluble proteome is kinetically trapped, may in principle involve aggregate formation among chains of either the same (homoaggregates) or different (heteroaggregates) proteins.

It is plausible to postulate that the propensity to undergo heteroaggregate formation may be facilitated by sequence homology among different proteins. In *E. coli,* 1,345 proteins have ≥ 30% sequence identity (over more than 60% of their chains), when compared to other proteins in the genome of this bacterium.[8] However, previous studies showed that all the detectable

intracellular self-associated states in live cells are homoaggregates [9], suggesting that formation of heteroaggregates is generally rare. In addition, studies based on experiments and simulations showed that phase transitions of polypeptides and proteins, including crystallization and aggregation, are highly favored in solutions containing chains of homogeneous composition, and disfavored when the sequences display either conformational or compositional (e.g., a single-point mutation) heterogeneity. [10-11]

As a result of the above, we conclude that it is somewhat unlikely that a large fraction of the aggregates in the *E. coli* protein collection is heterologous in nature. In any case, homo-and hetero-aggregation remains an active field of research, and future work aimed at assessing the identity of the aggregates in our *E. coli* protein mixture is desirable.

Regardless, even if some heteroaggregates were present in the pellet fraction of Figure 3-1, their presence does not undermine the concept of native-state kinetic trapping. Heteroaggregates would simply populate an additional region of the chemical-potential landscape. Hence native states of many *E. coli* proteins can be regarded as kinetically trapped relative to both homo- and any potential hetero-aggregates.

Most of the *E. coli* proteins that do not form insoluble aggregates are involved in regulation and stress response

Next, we examined the biological function of the individual proteins of the S100-like mixture that remain soluble after the heating-and-cooling step. Liquid-chromatography coupled to electrospray ionization LC-MS/MS mass spectrometry (Tables S1-S3) shows that the majority of the *E. coli* protein mixture that remains soluble after the heating-and-cooling step (total

concentration > 1 mg/ml) has either a regulatory function or is dedicated to protecting the cell from stress conditions.

Intriguingly, the function of these proteins includes different kinds of stress, many of which are completely unrelated to temperature perturbations (Table S2). This result suggests that proteins with regulatory and broad anti-stress functions may have evolved the ability to withstand the harshest environmental stresses without losing their native fold upon return to the original conditions, to most effectively aid preserving cell viability.

The kinetically trapped aggregates detected in the *E. coli* proteome are not amyloid

To test for the presence of amyloid fibrils upon heating and cooling, the ThT assay was carried out on the entire *E. coli* S100 protein collection (soluble E. coli proteome) before and after the heating-and-cooling step (Fig. 3-2 and Fig. S3). Somewhat surprisingly, the *E. coli* soluble protein mixture exhibits some ThT fluorescence before the heating-and-cooling step, possibly due to interactions of ThT with nucleic acids[12], or due to other unknown reasons. After the heating-and-cooling step, however, no significant increase in ThT fluorescence was observed, consistent with the fact that the heating process does not generate any ThT-sensitive amyloid fibril formation within the bacterial proteome.

Given that lack of ThT response does not strictly guarantee the absence of amyloid fibrils[13], we also performed transmission electron microscopy (TEM) experiments on the *E. coli* protein collection (under reducing conditions) before and after the heating-and-cooling process. After extensive TEM analysis of several regions of the specimen of insoluble heated-and-cooled *E. coli* protein mixture, we could not identify any fibrils (Fig. 3-2C).

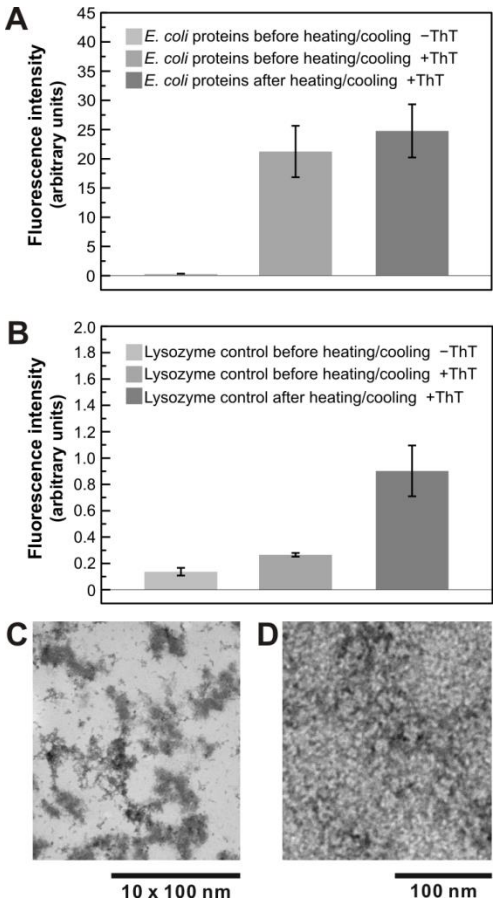


Figure 3-2. Thioflavin T (ThT) binding assay of cellular proteins. (A) ThT binding of soluble *E. coli* proteome before and after heating at 70 °C for 20 hrs at pH 7.0. (B) Reference experiment showing the ThT binding of pure lysozyme after heating at 70 °C for 48 hrs, pH 2.0. The data in panel A show no evidence for the formation of any amyloid-like aggregates upon heating/cooling. (C) Transmission electron micrograph of the heated pellet sample from the *E. coli* protein mixture. (D) Transmission electron micrograph of the unheated *E. coli* protein mixture for reference. The sample is densely packed as expected however, there are no fibrils.

Likewise, no fibrils could be detected by TEM in the grid sample derived from soluble protein mixture before the heating process (Fig. 3-2D). We conclude that the insoluble aggregates generated within the *E. coli* protein mixture upon heating and cooling under reducing conditions do not contain any significant amount of amyloid fibrils.

Native-state kinetic trapping relative to aggregates does not require high energy barriers

Some key features of the kinetic trapping of folded proteins with respect to aggregates were probed computationally, so that a few underlying principles could be deduced. We carried out simulations with a simple kinetic model including reversible steps for the folding, unfolding and aggregation of a single protein, using known folding/unfolding thermodynamic and kinetic parameters of sperm whale apomyoglobin [14 15 16 17] and aggregation related rate constants whose values were adjusted to yield consistent kinetics with the experiments discussed in this work (Table S4). While more complex kinetic models for aggregation could be adopted, we refrained from doing so given the lack of supporting experimental evidence.

The key features of our simple model are illustrated in Figure 3-3A. The self-association of a putative 20-mer (A_{20}) was modeled as a nucleated polymerization process [18] comprising an aggregation-prone intermediate (M*) and a putative dimeric nucleus (A_2). All the rate constants for elongation of the nucleus were set to be the same. Additional potentially relevant processes (e.g., secondary nucleation [19] and fragmentation [20]) were omitted, for simplicity.

First, the nucleation and elongation rate constants were adjusted so that native state (N) and endpoint aggregates (A_{20}) remain kinetically trapped with respect to each other on the timescale of ca. a year (Fig. 3-3B). This setup ensured that no interconversion between pure N and pure A_{20} took place upon 10-day incubation at our experimental initial total-protein concentrations.

While the simulated kinetic trapping of species b and c (Fig. 2-3C of Chapter 2) is consistent with the experimental findings of Figure 2-5 (Chapter 2), our simulations (Fig. 3-3C) did not manage to capture the additional aggregation of species d that was experimentally detected upon long-term incubation.

To address this discrepancy, we explored whether monomers may circumvent their kinetic trapping in the presence of some preformed aggregates serving as seeds (see section below on the role of seeding). The latter scenario is consistent with the experimental nature of species d in Figure 2-3 of Chapter 2 before long-term incubation.

Next, we temporarily modified the model to verify that it properly captures expected key features of nucleated polymerization. Namely, longer aggregation times when the aggregation-prone intermediate M* is rendered less thermodynamically stable (Figure S4) and slower aggregation timescales (and longer lag times) when the aggregation nucleus is simulated to be larger in size[21] (Figure S5).

We then reverted to the original version of the model (Fig. 3-3A and Table S4) and explored the role of nucleation barriers. As shown in Figure 3-4, a surprising result is that large barriers to nucleation are not needed to kinetically trap proteins in their monomeric state, including the native state N. Instead, modest barriers of approximately 6 kcal/mol are sufficient for kinetic trapping. The latter was observed to persist over the ca. 10-day incubation time employed in the apoMb experiments conducted in Chapter 2.

The experimental results on the *E. coli* proteome were simulated in Figure 3-5. First, we showed that, across a large and physiologically-relevant concentration range of monomeric protein (1 – 1000 μM), aggregation occurs very slowly, over a timescale exceeding by far the average

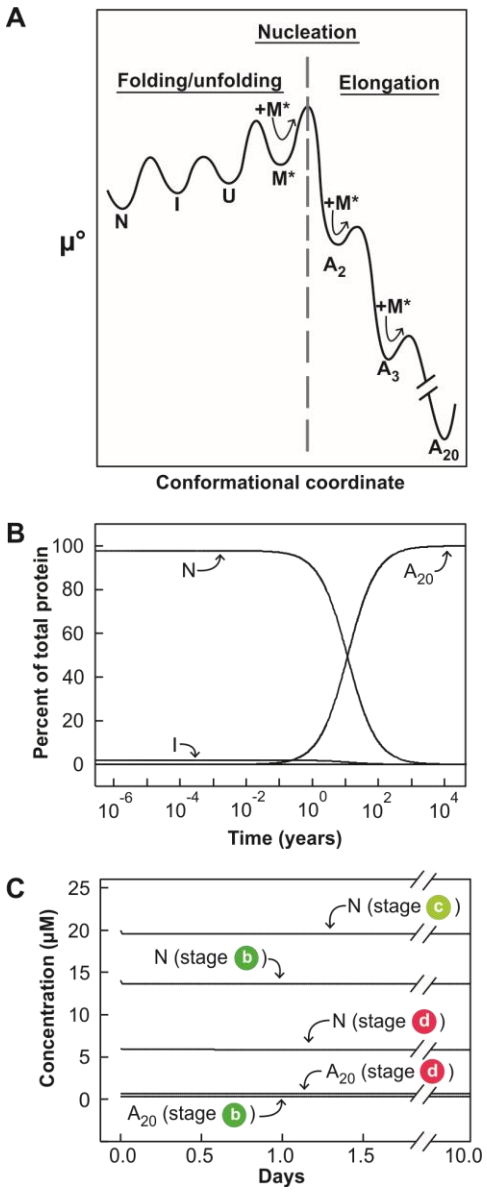


Figure 3-3. Kinetic simulation of protein folding and aggregation under physiologically relevant conditions consistent with experimental results. (A) Pictorial representation of the simulated standard-state chemical potential landscape. (B) Computed timecourse for the accumulation and/or depletion of species N, I, U, M*, A_2, A_5, A_{10}, A_{15} and A_{20} during long-term incubation. No appreciable buildup of U, M_2, M_5, M_{10} and M_{15} was observed. (C) Simulated 10-day incubation of species b, c and d (defined as in Fig. 2-3 of Chapter 2). The initial concentrations

of N and A_{20} were estimated from the relative amounts of monomer and soluble aggregate listed in Supplementary Table S7 of Chapter 2 for species b and from the relative loss in alpha helical secondary structure listed in Supplementary Table S5 in Chapter 2 for species d.

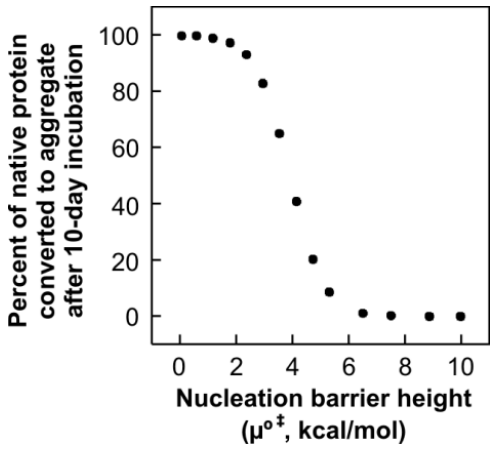


Figure 3-4. Simulated effect of nucleation-barrier height (the barrier from M* to A_2) for the conversion of native protein to aggregated states (A_2 to A_{20} in Fig. 2-3 of Chapter 2). The initial simulated native protein concentration was 20 μM. ApoMb was used as the protein in this simulation to be consistent with Chapter 2.

lifetime of *E. coli* cells of c.a. 30 minutes[22-23] (Figure 3-5A). Then, we simulated the extent of aggregation after 7-and 17-day incubations of the monomeric states (including N) at 10 and 100 μM concentrations, respectively (Fig. 3-5B). The results are fully consistent with the concentration dependence of apoMb aggregation observed in Supplementary Figure S3 of Chapter 2.

Role of seeding

The computational model corresponding to the chemical-potential diagram of Figure 3-3A was also instrumental to estimate whether preformed aggregates could serve as seeds. As shown in Figure S6, seeding the system with aggregates of variable size enables circumventing the slow barrier to nucleation so that monomeric proteins are rapidly converted to aggregates (Fig. S6A). Hence seeding was generally found to accelerate aggregation. Interestingly, upon close inspection of the role of seed size, we found that smaller seeds are more effective at converting monomeric proteins to aggregates, and seeding with the final large aggregate (A_{20}) has actually no effect (Figure S6A). In addition, smaller seeds promote aggregation at lower concentration than larger seeds (Fig. S6B) while the final aggregate A_{20} is ineffective at promoting aggregation even at large concentrations. Seeding with smaller aggregates is also compatible with a wider distribution of final aggregate sizes (Fig. S6C) and faster formation of the final aggregate (Fig. S6D), under the conditions of our simulation.

As highlighted in Figure S7, the seeds are promptly incorporated into the aggregates at early times, until fully used up. After all the seeds have been consumed, the aggregation-prone intermediate M* present in solution regardless of the seeds gets utilized, for aggregation to continue. The latter scenario is particularly clear in the central panel of Figure S7, which pertains to A_{15}, i.e., a seed of 15-residue nominal size.

In summary, we showed that even the simple model presented here, which does not encompass secondary nucleation nor postulates mechanistic changes upon addition of seeds, explains how seeding with small aggregates may capture aggregation-prone M* monomers in solution and convert them to large aggregates in the context of a fast, nucleation-free process.

We did not deliberately add any seeds in our experiments. On the other hand, the simulations suggest that the selective conversion of some of the monomer to aggregates only in the presence of aggregates of a given type, upon incubation, (e.g., see β-sheet-containing aggregates in sample d and Fig. 2-5 and Table S2 of Chapter 2) may be due to the presence of low-molecular weight seeds in solution consisting of lower-order aggregates. More detailed investigations are needed to shed further light on this topic and will be the subject of future investigations.

3.4 Discussion

Relations between this work and previous findings

The concept of protein kinetic trapping with respect to aggregates is not entirely new in the literature. For instance, Baskakov and coworkers showed that the monomeric recombinantly expressed prion protein from mouse is kinetically trapped with respect to a β-sheet aggregated isoform at low pH (pH 3.6).[24] This study was based on unfolding equilibrium thermodynamic measurements on a β-sheet self-associated version of the prion protein from mouse. However, the pH employed in this study was far from physiological. Later on, other investigators proposed that a few amyloidogenic globular proteins analyzed at different pH, including two proteins at pH 7 (human lysozyme and αβ-crystallin) are metastable, hence kinetically trapped in their native state with respect to the more thermodynamically stable amyloid fibrillar aggregates.[25] On a related line

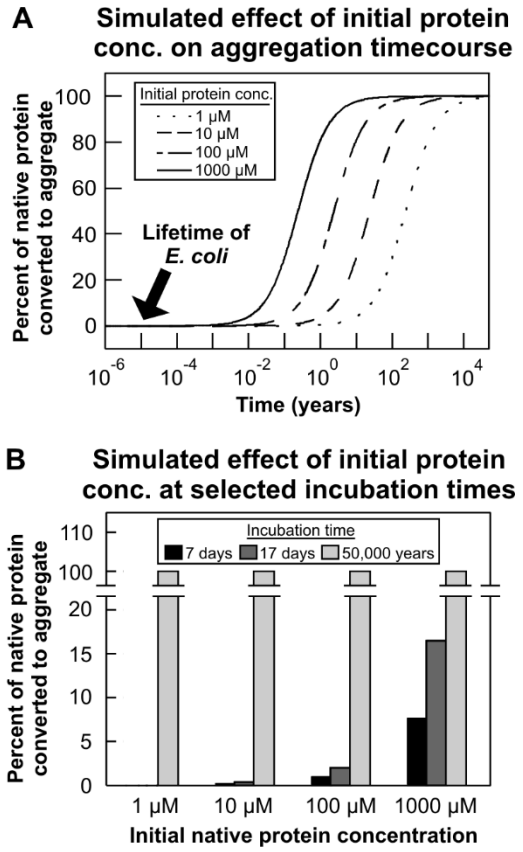


Figure 3-5. Protein concentration effects the rate of aggregate formation, but not during the lifetime of *E. coli.* Simulated effect of initial native protein concentration on the kinetics of aggregate conversion during (A) long-term incubation and (B) at selected incubation times.

of thought, supersaturation of a native protein, deemed metastable relative to fibrils, was linked to amyloid formation[26]. Finally, computational studies on whole proteomes[27] and experimental work on *C. elegans* and mammalian brains[28-30] showed connections between insoluble amyloid formation and upregulation of metastable protein expression upon aging.

Our study, however, differs from the above reports in a number of fundamental ways.

First, we found that kinetic trapping extends to non-amyloid aggregates. Hence, protein native states are often kinetically trapped not only relative to amyloids, but also relative to other types of aggregates, which contain either native-like or non-native-like secondary structure. In essence, the kinetically trapped aggregates may be amyloid in nature but they do not have to be so. In fact, none of the aggregates detected in this work (which may not be a collectively exhaustive set of all possible aggregates) display amyloid character.

Second, we determined that the aggregates do not have to be insoluble. Soluble aggregates are in fact more common at low concentrations.

Third, the critical length (< L*~150 amino acids) for fibrillation proposed by Baldwin et al.[25] does not apply to the aggregates identified here, which are not fibrillar. In fact, proteins of all lengths fit the kinetic trapping concept, as shown in Figure 3-1. The few exceptions are proteins found to be at true thermodynamic equilibrium at room temperature, as testified by the *E. coli* proteins that remain soluble even after heating and cooling (Figure 3-1). Interestingly, these proteins are actually fairly small in size (< 20 kDa).

Finally, our kinetic simulations based on a simple nucleated polymerization model show that the nucleation barriers do not have to be high at all. For protein kinetic trapping under physiologically relevant conditions it is sufficient that the kinetic flux, i.e., the transition rate, through the relevant barriers is slow. For aggregation nuclei that encompass two or more

monomers, rate-determining nucleation can be slow simply due to the high reaction order (> 1), implying extremely slow reactions at physiologically relevant monomer concentrations.

3.5 Methods

Preparation of *E. coli* protein mixture

Single colonies of the A19 *E. coli* strain were inoculated into Luria broth (LB) medium at 37°C and grown overnight at 250 rpm. Afterwards, a small portion of the culture was transferred to an excess of fresh LB medium so that the starting optical density at 600 nm (OD_{600}) was 0.1before further incubation at 37°C at 250 rpm. Once the large growth reached OD_{600} 0.87, it was centrifuged at 30,590 g for 15 minutes. The pellet was homogenized in a buffer containing 50 mM tris base, 2 mM EDTA, 7 mM 2-Mercaptoethanol (BME), 5 mM benzamidine HCl, and 2 mM phenylmethylsulfonyl fluoride (PMSF), while trying to avoid air bubbles. The solution was then sonicated (Fisher Scientific, Model 550 Sonic Dismembrator) with 1 s bursts followed by 1 s quiescent periods for a total sonication time of 1 min per cycle. This process was repeated three times with a one-minute break between each run. 2 mM PMSF was then added immediately before sonication was initiated to minimize proteolysis. The sample was then subject to a series of centrifugations. The first two were at 30,590 g for 20 min, to remove genomic DNA.[31] The supernatant was removed each time and centrifuged. After the second centrifugation, the supernatant was centrifuged once more at 106,000 g at 4°C for 2 hrs to remove ribosomes.[32] The supernatant was then dialyzed in PBS buffer (66.8 mM K_2HPO_4, 10 mM NaCl, 0.8 mM $MgCl_2$, 0.1 µM $CaCl_2$, pH 7.4) for 2 hrs using a 3,500 MWCO Slida-A-Lyzer Dialysis Cassette (Thermo Scientific, Rockford, IL). A single buffer exchange was done after 2 hrs and the sample was

dialyzed overnight. The resulting sample was aliquoted and then flash frozen and stored -80 °C for future use. All buffers were filtered through a 0.22 μm filter (Millipore) prior to use.

Kinetic trapping of proteins comprising the soluble *E. coli* proteome

Samples of variable concentrations of the *E. coli* soluble protein mixture (ranging from 0.03 mg/mL up to 3.3 mg/mL total protein concentration) were prepared in PBS buffer starting form a 3.3 mg/mL stock solution. The samples were heated at 70 °C for 20 hrs in a glass vial capped with heat-resistant stoppers to prevent solvent evaporation. After heating, the samples were centrifuged at 15,493 g to separate soluble and insoluble portions. The pellets were resuspended in a sodium dodecyl sulfate buffer (SDS) buffer (comprising 0.14 M SDS, 2.7 M glycerol, and 0.1 M tris base at pH 6.8) for gel analysis. A 1:2 dilution of supernatant with SDS buffer in the presence of 0.7 M β-mercaptoethanol was loaded on the gels after boiling for 10 min. Pellets and supernatants were run on 15% SDS-PAGE gels. Volumes of soluble and insoluble fractions were tailored so that, if summed, their intensities would display the same intensity as the reference lane (showing the sample mixture before heating).

The above experiments were also carried out under reducing conditions, i.e., in the presence of 1 mM dithiothreitol (DTT) added to the initial PBS buffer, to mimic the reducing environment of the *E. coli* cytoplasm. Fresh DTT (1 mM) was added every 24 hrs to maintain a consistent reducing environment.

ImageJ software analysis

SDS-PAGE gels were analyzed with the ImageJ Software. In order to determine the amount of protein loaded in each lane, the protein concentration of the cell lysate was calculated. The mass of protein loaded in the marker lane (Bio-Rad, California) was calculated and utilized as a quantitative comparison to the unheated reference lane. The intensity ratio between the reference

lane and the marker lane was then used to determine the mass of protein of protein in the reference lane. This was used to obtain the total protein concentration in the protein mixture (3.3 mg/mL). The total intensity of pellet was divided by the sum of the pellet and supernatant intensities for a given concentration. Repeat experiments of each concentration were averaged to get a more accurate value. Standard error values were the calculated using three independent repeat experiments.

Kinetic simulations

Computational modeling of apoMb aggregation was carried out with the Gepasi Biokinetic Simulator software package (version 3.30[33-35]), running on Microsoft Windows. All simulated unimolecular or bimolecular elementary steps were assumed to be reversible according to a nucleated polymerization diagram corresponding to the process

$$\boldsymbol{N \rightleftarrows I \rightleftarrows U \rightleftarrows M^{*} \rightleftarrows M_2 \rightleftarrows M_3 \rightleftarrows \cdots \rightleftarrows M_n \rightleftarrows M_{n+1} \rightleftarrows \cdots \rightleftarrows M_{20}} \quad , \qquad (1)$$

where N, I and U are the folded, intermediately-folded and unfolded states of apoMb, M* is a hypothetical high-energy, aggregation-prone species. The M_n species are aggregates resulting from the stepwise addition of monomer M* to the growing aggregate. The dimer M_2 is the aggregation nucleus, whose aggregation rate-determining formation is required for the subsequent elongation steps. The elongation steps were designed so that the aggregation is strongly downhill, hence the system bears an ultra-low CPC (< 1 molecule/mL) according to

$$\boldsymbol{\Delta\mu^{o}_{elongation} \approx RTln(CPC)} \qquad (2)$$

where R is the ideal gas constant and T is the temperature, set to 298.15 K. All the rate constants used in the simulations are listed in Tables S4-6 . The rate constants for apoMb folding and unfolding were chosen to match known experimental values for this protein[14, 16, 36], and the rate

constants for the nucleation, elongation and dissociation steps were designed to produce an overall downhill aggregation process, with chemical-potential barrier heights estimated according to

$$\boldsymbol{k = k_{max} e^{\frac{-\Delta\mu^{\circ\ddagger}}{RT}}} \quad , \qquad (3)$$

where k is the microscopic rate constant, k_{max} is the maximum expected rate constant for a barrierless process, and $\Delta\mu^{\circ\ddagger}$ is the standard-state activation chemical potential. The k_{max} parameter was estimated to correspond to diffusion-limited encounters, i.e., 10^{10} $M^{-1}s^{-1}$ for bimolecular nucleation and elongation events, and was set to 10^{10} s^{-1} for unimolecular dissociation events. The apoMb chemical potential barrier heights for the formation and unfolding of the burst-phase folding intermediate[15-16] were roughly estimated adopting typical kinetic parameters for highly cooperative helix-coil transitions (1.25×10^6 s^{-1} and 3.87×10^6 s^{-1} respectively).[37] The system of differential equations defining the rate laws corresponding to scheme 1 was solved with the Livermore Solver of Ordinary Differential Equations Algorithm (LSODA) via the Gepasi software package.[38]

3.6 Acknowledgments

We are grateful to G. Sabat, G. Barrett-Wilt, R. Burgess, K. Dill, D. Eisenberg and Z. Wang for useful discussion.

No competing financial interests have been declared. This work was funded by the National Science Foundation (NSF) grants MCB-0951209 and MCB-1616459 (to S.C.), including NSF REU supplements (to J.F.L.). A.E.V. was the recipient of an NSF Graduate Research Fellowship.

3.7 References

(1) Baker, D.; Agard, D. A. Kinetics versus thermodynamics in protein-folding. *Biochem.* **1994**, *33*, 7505-7509.

(2) Hekman, C.; Loskutoff, D. Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants. *J. Biol. Chem.* **1985**, *260*, 11581-11587.

(3) Menninger, J. R.; Mulholland, M. C.; Stirewalt, W. S. Peptidyl-tRNA hydrolase and protein chain termination. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1970**, *217*, 496-511.

(4) Cantor, C. R. S., P. R., *Biophysical chemistry part I the conformation of biological macromolecules*. W.H. Freeman and Co, New York: 1980.

(5) Oosawa, F.; Kasai, M. A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol.* **1962**, *4*, 10-21.

(6) Cayley, S.; Lewis, B. A.; Guttman, H. J.; Record, M. T. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions in vivo. *J. Mol. Biol.* **1991**, *222*, 281-300.

(7) Posey, A. E.; Ruff, K. M.; Harmon, T. S.; Crick, S. L.; Li, A.; Diamond, M. I.; Pappu, R. V. Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers. *J. Biol. Chem.* **2018**, *293*, 3734-3746.

(8) Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F., et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453-1462.

(9) Rajan, R. S.; Illing, M. E.; Bence, N. F.; Kopito, R. R. Specificity in intracellular protein

aggregation and inclusion body formation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13060-13065.

(10) McManus, J. J.; Lomakin, A.; Ogun, O.; Pande, A.; Basan, M.; Pande, J.; Benedek, G. B. Altered phase diagram due to a single point mutation in human gamma D-crystallin. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16856-16861.

(11) Thurston, G. M. Protein anisotropy turns solubility on its head. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18877-18878.

(12) Ilanchelian, M.; Ramaraj, R. Emission of thioflavin T and its control in the presence of DNA. *J. Photochem. Photobiol. A: Chem.* **2004**, *162*, 129-137.

(13) Cloe, A. L.; Orgel, J. P. R. O.; Sachleben, J. R.; Tycko, R.; Meredith, S. C. The japanese mutant aβ (δe22-aβ1−39) forms fibrils instantaneously, with low-thioflavin t fluorescence: Seeding of wild-type aβ1−40 into atypical fibrils by δe22-aβ1−39. *Biochem.* **2011**, *50*, 2026-2039.

(14) Eun, Y. J.; Kurt, N.; Sekhar, A.; Cavagnero, S. Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* **2008**, *376*, 879-97.

(15) 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*, 1171-1176.

(16) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993**, *262*, 892-896.

(17) 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.

(18) Oosawa, F.; Asakura, S., *Thermodynamics of the polymerization of protein*. Academic Press: 1975.

(19) Ferrone, F. A.; Hofrichter, J.; Eaton, W. A. Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. *J. Mol. Biol.* **1985**, *183*, 611-631.

(20) Tanaka, M.; Collins, S. R.; Toyama, B. H.; Weissman, J. S. The physical basis of how prion conformations determine strain phenotypes. *Nature* **2006**, *442*, 585-589.

(21) Kar, K.; Jayaraman, M.; Sahoo, B.; Kodali, R.; Wetzel, R. Critical nucleus size for disease-related polyglutamine aggregation is repeat length dependent. *Nat. Struct. Mol. Biol.* **2011**, *18*, 328-336.

(22) Reshes, G.; Vanounou, S.; Fishov, I.; Feingold, M. Timing the start of division in *e.Coli*: A single-cell study. *Phys. Biol.* **2008**, *5*, 046001.

(23) Sezonov, G.; Joseleau-Petit, D.; D'Ari, R. *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* **2007**, *189*, 8746-8749.

(24) Baskakov, I. V.; Lagname, G.; Prusiner, S. B.; Cohen, F. E. Folding of prion protein to its native a-helical conformation is under kinetic control. *J. Biol. Chem.* **2001**, *276*, 19687-19690.

(25) 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.

(26) Yoshimura, Y.; Lin, Y.; 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.

(27) Ciryam, P.; Tartaglia, G. G.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M. Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.* **2013**, *5*, 781-790.

(28) Freer, R.; Sormanni, P.; Vecchi, G.; Ciryam, P.; Dobson, C. M.; Vendruscolo, M. A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease. *Sci. Adv.* **2016**, *2*, e1600947.

(29) Ciryam, P.; Kundra, R.; Freer, R.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M. A transcriptional signature of Alzheimer's disease is associated with a metastable subproteome at risk for aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 4753-4758.

(30) Walther, D. M.; Kasturi, P.; Zheng, M.; Pinkert, S.; Vecchi, G.; Ciryam, P.; Morimoto, R. I.; Dobson, C. M.; Vendruscolo, M.; Mann, M., et al. Widespread proteome remodeling and aggregation in aging C. elegans. *Cell* **2015**, *161*, 919-932.

(31) Zubay, G.; Watson, M. The absence of histone in the bacterium Escherichia coli: I. Preparation and analysis of nucleoprotein extract. *The Journal of Cell Biology* **1959**, *5*, 51-54.

(32) Felske, A.; Engelen, B.; Nubel, U.; Backhaus, H. Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl. Environ. Microbiol.* **1996**, *62*, 4162-4167.

(33) Mendes, P. Gepasi: A software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.* **1993**, *9*, 563-571.

(34) Mendes, P. Biochemistry by numbers: Simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.* **1997**, *22*, 361-363.

(35) Mendes, P.; Kell, D. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* **1998**, *14*, 869-883.

(36) Xu, M.; Beresneva, O.; Rosario, R.; Roder, H. Microsecond folding dynamics of apomyoglobin at acidic pH. *J. Phys. Chem. B* **2012**, *116*, 7014-7025.

(37) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76-88.

(38) Petzold, L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci. Comput.* **1983**, *4*, 136-148.

(39) Bozzola, J. J.; Russell, L. D., *Electron microscopy*. Jones & Bartlett Learning, LLC: 1991.

(40) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646-4658.

(41) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, Y.; Rappsilber, J.; Mann, M. Exponentially modified proetin abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of seqeunced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265-1272.

(42) Silhavy, T. J.; Boos, W.; Kalckar, H. M., The role of the Escherichia coli galactose-binding protein in galactose transport and chemotaxis. In *Biochemistry of sensory functions*, Jaenicke, L., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 1974; pp 165-205.

(43) Ueguchi, C.; Seto, C.; Suzuki, T.; Mizuno, T. Clarification of the dimerization domain and its functional significance for the Escherichia coli nucleoid protein H-NS. *J. Mol. Biol.* **1997**, *274*, 145-151.

(44) Grainger, D. C.; Hurd, D.; Goldberg, M. D.; Busby, S. J. W. Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res.*

2006, *34*, 4642-4652.

(45) Wang, W.; Li, G.-W.; Chen, C.; Xie, X. S.; Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **2011**, *333*, 1445-1449.

(46) Yim, H. H.; Villarejo, M. osmY, a new hyperosmotically inducible gene, encodes a periplasmic protein in Escherichia coli. *J. Bacteriol.* **1992**, *174*, 3637-3644.

(47) Weber, A.; Kögl, S. A.; Jung, K. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in Escherichia coli. *J. Bacteriol.* **2006**, *188*, 7165-7175.

(48) Keskin, O.; Bahar, I.; Flatow, D.; Covell, D.; Jernigan, R. Molecular mechanisms of chaperonin groel-groes function. *Biochem.* **2002**, *41*, 491-501.

(49) Numoto, N.; Kita, A.; Miki, K. Crystal structure of the co-chaperonin Cpn10 from Thermus thermophilus HB8. *Proteins: Struct. Funct. Bioinform.* **2005**, *58*, 498-500.

(50) Gajiwala, K. S.; Burley, S. K. Hdea, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *J. Mol. Biol.* **2000**, *295*, 605-612.

(51) Surette, M. G.; Miller, M. B.; Bassler, B. L. Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: A new family of genes responsible for autoinducer production. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1639-1644.

(52) Gentry, D. R.; Burgess, R. R. rpoZ, encoding the omega subunit of Escherichia coli RNA polymerase, is in the same operon as spot. *J. Bacteriol.* **1989**, *171*, 1271-1277.

(53) Wissenbach, U.; Six, S.; Bongaerts, J.; Ternes, D.; Steinwachs, S.; Unden, G. A third periplasmic transport system for l-arginine in Escherichia coli: Molecular characterization of the artPIQMJ genes, arginine binding and transport. *Mol. Microbiol.* **1995**, *17*, 675-686.

3.8 Appendix

Supplementary Results

Analysis of covalent modifications of *E. coli* proteins after heating and cooling by nano-LC-ESI MS/MS

The extent of modification (oxidation/deamidation) of the soluble proteins of the *E. coli* proteome (present in the S100 bacterial extract, see main manuscript) after the heating and cooling process was assessed via nano-liquid chromatography electrospray ionization (ESI) MS/MS. As described in the main manuscript, the heating-and-cooling procedure consisted of a single heating period for 20 h at 70°C followed by slow cooling to room temperature

We found that only a moderate fraction of the proteins in the supernatant and pellet are covalently modified, after heating and cooling. Namely, we detected a 24% ± 5% and 18% ± 4% decrease in the number of unmodified amino acids in the supernatant and pellet after heating and cooling, respectively. While these values are non-negligible, they support the conclusion that the large majority of the proteins were not covalently modified by the heating and cooling process.

To further support the above analysis, we also examined the sequence coverage of the unheated and heated sample fractions. The supernatant and pellet showed a 3.0% ± 1.0 % and 1.7% ± 1.8% decrease in sequence coverage after heating and cooling, respectively. Both variations are small, suggesting that the above analysis of percent of unmodified amino acids refers to a comparable pool of amino acid sequences.

Interestingly, as discussed in the main manuscript, a significant fraction of the top ten most abundant proteins within the soluble fraction of the *E. coli* S100 mixture (after heating and cooling) has a regulatory function or is involved in stress response (Table S11). Hence, this class of proteins must have evolved to be particularly robust towards temperature perturbations beyond the typical

heat-shock temperatures of ca. 42 °C. These proteins are able to remain soluble despite the large temperature perturbations.

Thioflavin T fluorescence assays

To assess the presence of fibrils in the soluble *E. coli* protein mixture in the reduced state, we heated the solution at 70 °C for 20 hrs at a concentration of 3.3 mg/mL in the presence of 1 mM DTT. Samples were then allowed to cool to room temperature. A 1:45 dye-to-protein (mg/mL) ratio was used to calculate the appropriate amount of protein used to prepare the sample for data collection. The final concentration of ThT was 20 µM in all samples. Lysozyme was used as a positive control to generate fibrils. Lysozyme solutions (3.3 mg/mL) were heated 48 hrs at 70 °C in PBS buffer at pH 2.0 and allowed to cool to room temperature. PBS buffer (66.8 mM K_2HPO_4, 10 mM NaCl, 0.8 mM $MgCl_2$, 0.1 µM $CaCl_2$, pH 7.4 or 2.0) was used to dilute the protein samples, if necessary. Fluorescence emission spectra for these experiments were recorded with a PC1 photon counting steady-state spectrofluorimeter (ISS, Champaign IL) upon excitation at 440 nm. Slit widths for both excitation and emission channels were 4 nm. No filters were used.

Transmission electron microscopy

After heating and cooling as described in the previous sections, the insoluble fraction of the *E. coli* protein mixture was immersion fixed in a 2.5% glutaraldehyde, 2.0% paraformaldehyde solution in 0.1 M sodium phosphate buffer (PB) and incubated at 4 °C overnight. The fixed pellet was then rinsed 5 times for 5 min in PB and post-fixed in 1% osmium tetroxide with 1% potassium ferrocyanide in 0.1 M PB for 2 hr at room temperature. The pellet was then rinsed in PB 5 times for 5 min. The sample was dehydrated with aqueous ethanol (35%, 50%, 70%, 80%, 90% for 15 min at each step, followed by treatment with 95% ethanol for 30 min and 100% for 3 x 15 min) at room temperature and then in propylene oxide (PO) 2 x 7 min at room temperature. The fully

dehydrated sample was infiltrated in increasing concentrations of PolyBed 812 (Polysciences Inc., Warrington, PA) and PO mixtures according to the following steps: 1) 25% PolyBed 812, 75% PO for 60 min at room temperature. 2) 50% Polybed 812, 50% PO for 60 min at room temperature. 3) 75% PolyBed 812, 25% PO o/n at room temperature. 4) 100% PolyBed 812, 0% PO for 4 x 45 min cycles at 60 °C. Afterwards, embedding and polymerization took place in fresh PolyBed 812 for 24 hours at 60 °C. Samples were then sectioned on a Leica EM UC6 Ultramicrotome (Leica Mircosystems Inc., Buffalo Grove, IL) at 100 nm. The sections were collected on Cu, Pioloform/carbon coated 2x1 slot grids (EMS, Hatfield, PA), and post-stained in uranyl acetate and lead citrate. Samples were viewed at 80 kV on a Phillips CM120 transmission electron microscope, equipped with an AMT BioSprint12 (Advanced Microscopy Techniques, Corp., Woburn, MA).

The unheated *E. coli* protein mixture and the soluble fraction after the heating and cooling steps were treated as follows. The solutions containing the soluble proteins were negative stained with Nano-W (Nanoprobes, Yaphank, NY) with a known two-step method.[1] A 2 μL droplet of the sample was placed on a Pioloform (Ted Pella Inc., Redding, CA) coated 2x1 Cu slot grid (EMS, Hatfield, PA) with the coating side down. Any excess sample was removed with filter paper. Before complete drying, a 2 μL of Nano-W was applied, wicked again with new filter paper and allowed to dry. Samples were viewed on a Philips CM120 transmission electron microscope at 80kV and documented with an AMT BioSprint12 (Advanced Microscopy Techniques, Corp., Woburn, MA).

Enzymatic "in liquid" digestion of the *E. coli* protein mixture for nano-LC-ESI MS/MS

An unheated S100-like mixture (~100 μg protein) was TCA/acetone precipitated (10% TCA, 28% acetone final concentration) and the pellets were re-solubilized and denatured in 30 μl of a

solution containing 8 M urea, 50 mM NH_4HCO_3 (pH 8.5), and 1 mM Tris-HCl for 10 min. Samples were diluted to 120 µl and reduced using a solution containing 5 µl of 25 mM DTT, 10 µl MeOH, and 75 µl 25 mM NH_4HCO_3 (pH 8.5). They were incubated at 52 °C for 15 min, cooled on ice to room temperature before 6 µl of 55 mM indole-3-acetic acid (IAA) was added for alkylation and incubated in the dark at room temperature for 15 minutes. The reaction was quenched by adding 16 µl of 25 mM DTT. Afterwards, 16 µl of a Trypsin/Lys-C solution (100 ng/µl *Trypsin/Lys-C* mix in 25 mM NH_4HCO_3) and 42 µl of 25 mM NH_4HCO_3 (pH 8.5) was added to reach a 200 µl final volume. The heat precipitated sample (protein amount estimated at approx. ~300 µg) was directly solubilized in 90 µl of a solution consisting of 8 M urea, 50 mM NH_4HCO_3 (pH 8.5), and 1 mM Tris-HCl for 10 minutes. The sample was then diluted to 360µl and reduced using a solution containing 15 µl of 25 mM DTT, 30 µl MeOH, and 225 µl of 25 mM NH_4HCO_3 (pH 8.5). Afterwards, it was incubated at 52 °C for 15 min, cooled on ice to room temperature and 18 µl of 55 mM IAA was added for alkylation. The sample was incubated in the dark at room temperature for 15 min and then the reaction was quenched by adding 48 µl of 25 mM DTT. Next, 48 µl of the aforementioned Trypsin/Lys-C solution and 126 µl of 25 mM NH_4HCO_3 (pH 8.5) was added to reach a 600 µl final volume. The digestions were conducted for 2 hrs at 42 °C. An additional 8 µl of Trypsin/Lys-C solution for the unheated S100-like mixture and 24 ul for pellet was added and the digestion proceeded o/n at 37 °C. The reaction was terminated by acidification with 2.5% TFA to reach a 0.3% final concentration.

Nano-LC-ESI MS/MS of *E. coli* protein mixture

An aliquot of each digest (50 µg of protein) was cleaned up using OMIX C18 SPE cartridges (Agilent, Palo Alto, CA) per manufacturer protocol and eluted in 20 µl of 60/40/0.1% ACN/H_2O/TFA, dried to completion and finally reconstituted in 40 µl of 0.1% formic acid. The

peptides were analyzed by nano-LC-ESI MS/MS using the Agilent 1100 nanoflow system (Agilent) connected to a new generation hybrid linear ion trap-orbitrap mass spectrometer (LTQ-Orbitrap Elite™, Thermo Fisher Scientific) equipped with an EASY-Spray™ electrospray source. Chromatography of the peptides prior to mass spectral analysis was accomplished using a capillary emitter column (PepMap® C18, 3µM, 100Å, 150x0.075mm, Thermo Fisher Scientific) onto which 2 µl of extracted peptides were automatically loaded. The NanoHPLC system delivered solvents A: 0.1% (v/v) formic acid, and B: 99.9% (v/v) acetonitrile-0.1% (v/v) formic acid was loaded at a rate of 0.50 µL/min to load the peptides (over a 30 minute period) and 0.2 µl/min to elute the peptides directly into the nano-electrospray with a gradual gradient from 3% (v/v) B to 20% (v/v) B over 154 minutes and concluded with a 12 min fast gradient from 20% (v/v) B to 50% (v/v) B at which time a 5 min flash-out from 50-95% (v/v) B took place. As the peptides were eluted from the HPLC-column/electrospray source, survey MS scans were acquired in the Orbitrap with a resolution of 120,000 followed by MS2 fragmentation of the 20 most intense peptides detected in the MS1 scan from 380 to 1800 m/z; redundancy was limited by dynamic exclusion.

Analysis of nano-LC-ESI MS/MS data

Raw MS/MS data were converted to an mgf file format using MSConvert (ProteoWizard: Open Source Software for Rapid Proteomics Tools Development) for downstream analysis. The resulting mgf files were used to search against an *Escherichia coli* amino acid sequence database containing decoy reverse entries and a list of common contaminants (8,484 total entries) using an in-house *Mascot* search engine 2.2.07 (Matrix Science) containing variable methionine oxidation, asparagine and glutamine deamidation plus a fixed cysteine carbamidomethylation. Peptide mass tolerance was set at 15 ppm and the fragment mass was set at 0.6 Da. Protein annotations, significance of identification and spectral-based quantification was done with help of Scaffold

software (version 4.3.2, Proteome Software Inc., Portland, OR). Protein identifications were accepted if they could be established at greater than 80.0% probability within 1% False Discovery Rate and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm.[2] Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

In order to approximate the abundance of proteins that were unmodified after heating, we created the following set of criteria. First, we examined the top ten most abundant proteins in the pellet and supernatant samples as identified by the highest emPAI (Exponentially Modified Protein Abundance Index) values.[3] For each of these proteins, we then identified the MS2 peptides that contained at least one M, Q, and/or N residue and were present in both the unheated and heated (supernatant or pellet) MS data. Three of these peptides were randomly chosen and used as representatives to measure the degree of degradation of the protein as long as the majority of their Mascot ion scores were at or above 35. This criterion was followed for more than 90% of peptides examined in the pellet and 79% of peptides in the supernatant. Exemptions occurred because either the peptides found between the unheated and heated samples did not match and thus could not be compared and/or the majority of their individual ion scores were not above 35. Next we counted how many times this peptide was unmodified vs modified by determining the frequency by which the unmodified peptide occurs. This was also done for the same peptide in the unheated fraction so that we could subtract the heat-treated sample from the unheated sample to get a difference between them. The differences were then averaged and the standard error was determined to yield an estimate of the extent of modification of the protein in either the supernatant or pellet. We used the unheated % as a baseline for both the supernatant and pellet proteins since sample preparation for MS data collection introduces modifications to the protein.

Supplementary Figures

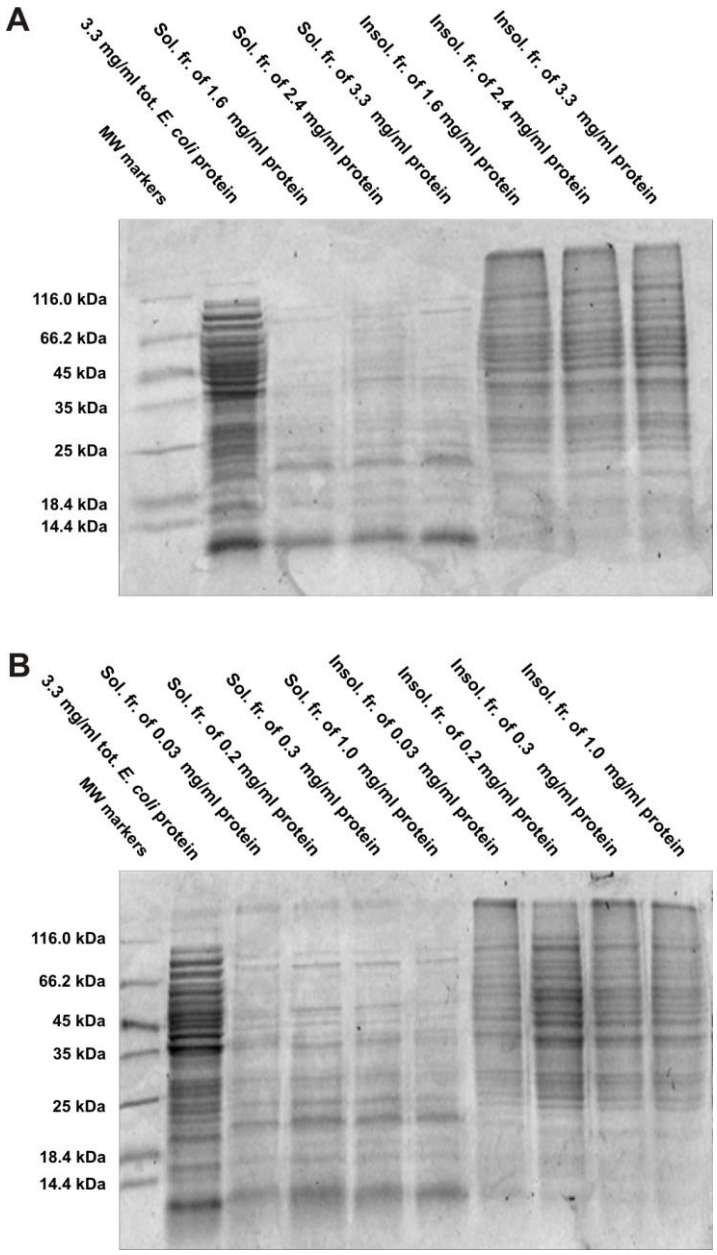


Figure S1. Representative SDS-PAGE gels showing the soluble and insoluble fractions of the proteins of the *E. coli* bacterium under reducing conditions, i.e., in the presence of 1 mM dithiothreitol (DTT). A) Gel showing representative raw data used to generate the plot in Fig. 3-1B. B) Same as panel A except that the data correspond to a lower total-protein concentration regime.

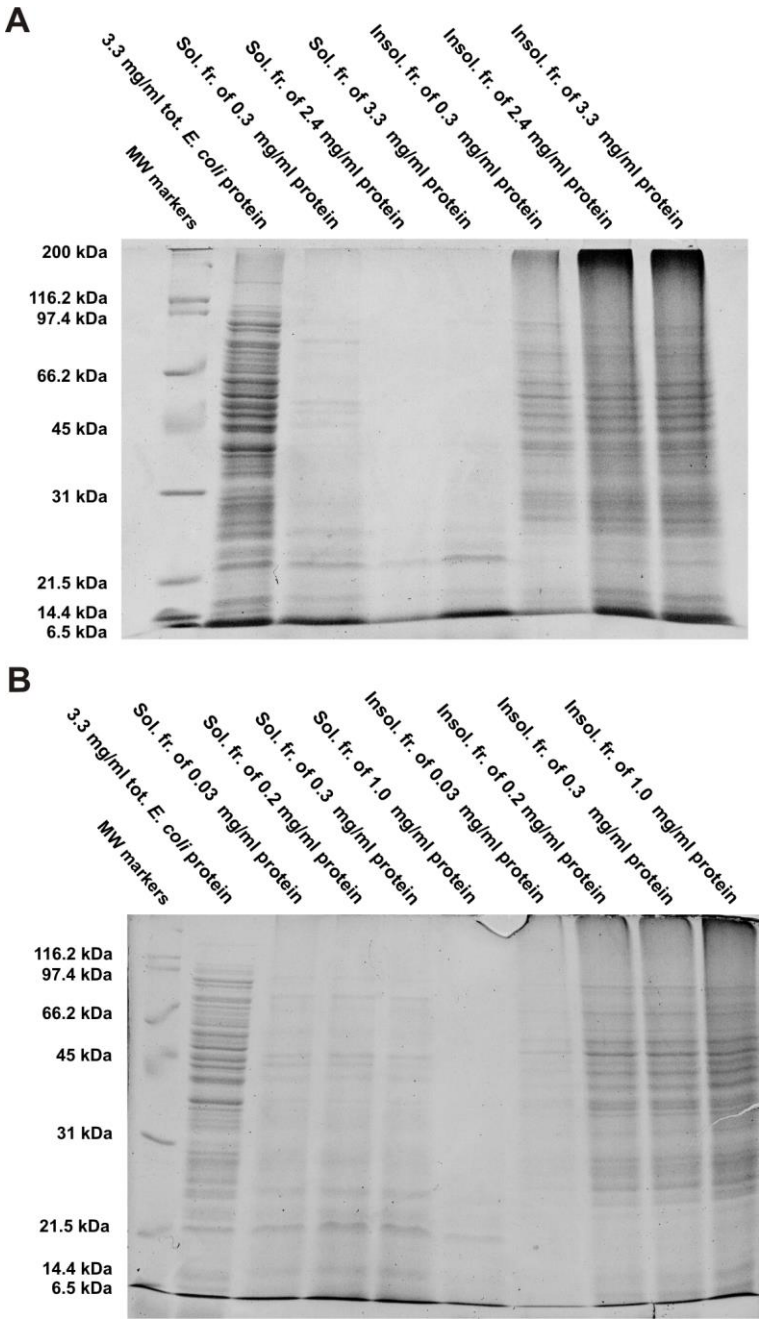


Figure S2. Representative SDS-PAGE gels showing the soluble and insoluble fractions of the S100 proteins of the *E. coli* bacterium under non-reducing conditions. A) Gel showing representative raw data used to generate the plot in Fig. 3-1D. B) Same as panel A except that the data correspond to a lower total-protein concentration regime.

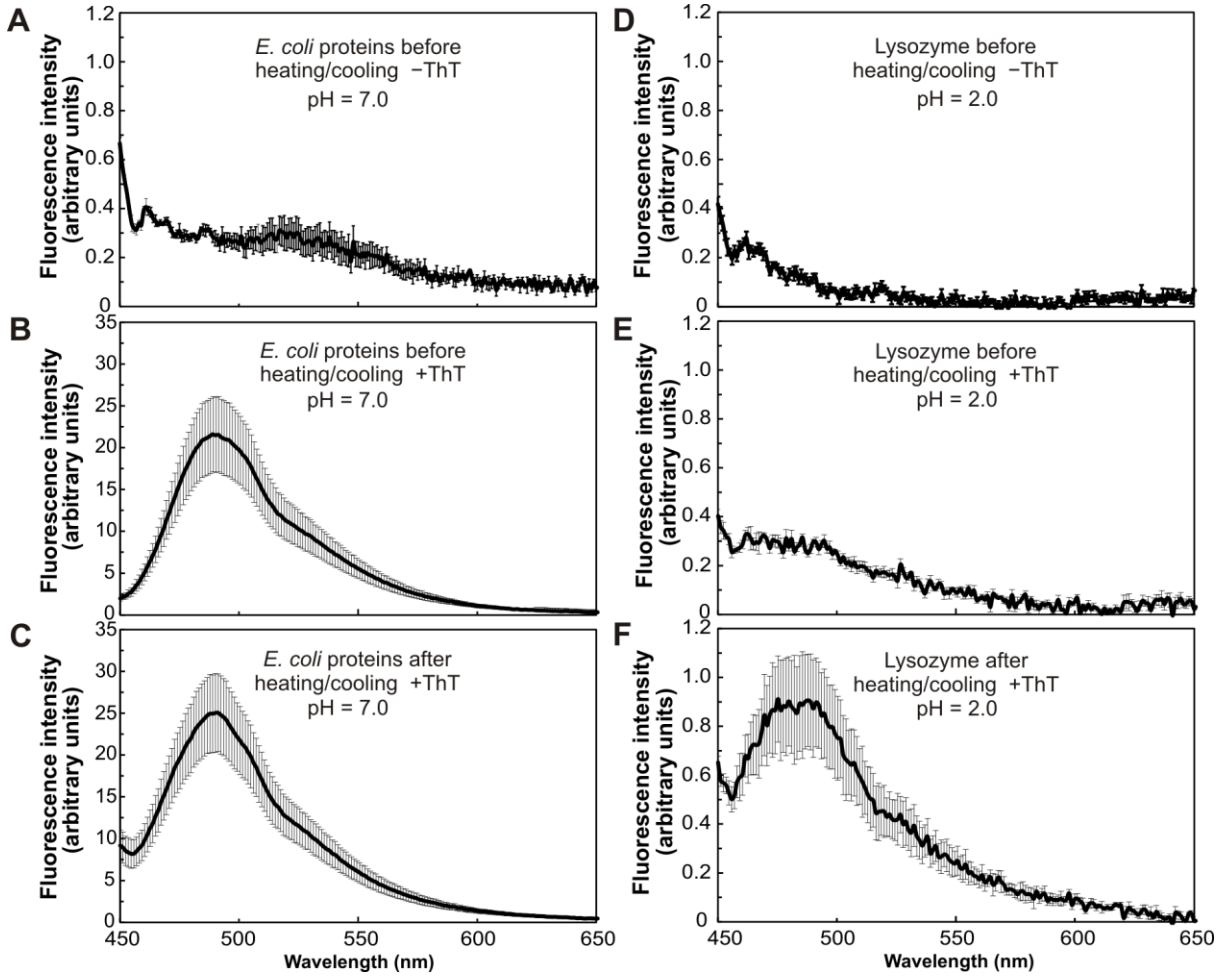


Figure S3. Fluorescence emission spectra corresponding to the ThT assay data in Figure 3-2A) Fluorescence emission spectrum of the soluble *E. coli* proteome before heating (at 70 °C) and cooling (to room temperature) in the absence of ThT. B) Fluorescence emission spectrum of the soluble *E. coli* proteome before heating and cooling in the presence of ThT. C) Fluorescence emission spectrum of the soluble *E. coli* proteome after heating and cooling in the presence of ThT. D) Fluorescence emission spectrum of lysozyme before heating (at 70 °C for 48 hrs) and cooling (to room temperature) in the absence of ThT. E) Fluorescence emission spectrum of lysozyme before heating and cooling in the presence of ThT. F) Fluorescence emission spectrum of lysozyme after heating and cooling in the presence of ThT. (see Supplementary Methods for additional experimental details).

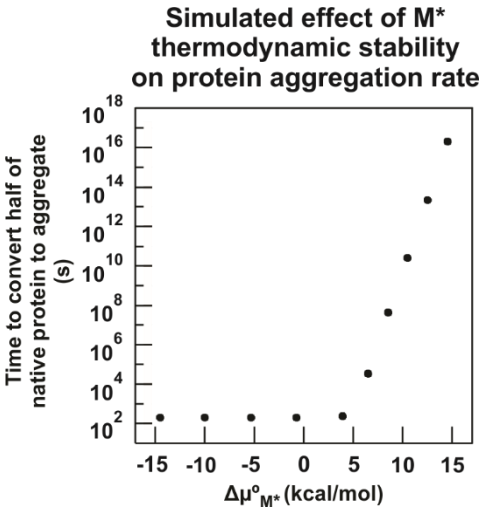


Figure S4. Result of computer simulations showing the effect of the thermodynamic stability of the M* aggregation-prone intermediate relative to the native state N (defined as in Fig. 3-3A) on the aggregation timecourse of the native protein. The initial simulated concentration of native protein is 20 μM.

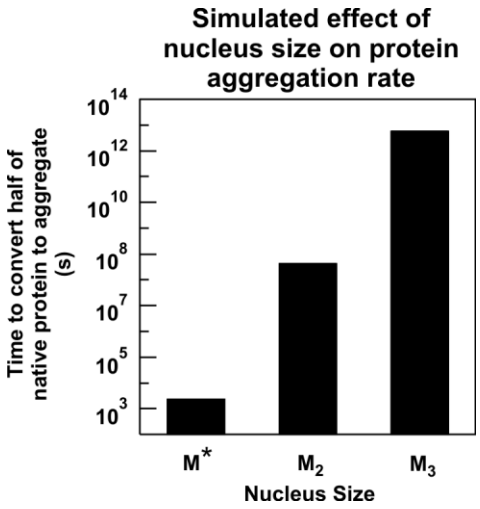


Figure S5. Simulated effect of nucleus size on protein aggregation half-life. The initial concentration of native protein was set to 20 μM. To generate the models bearing the M* and M_3 nuclei, the rate constants for the rate-determining nucleation step of the original model (Fig. 3-3A and Table S4), which features the M_2 nucleus, were varied as shown in Tables S5 and S6.

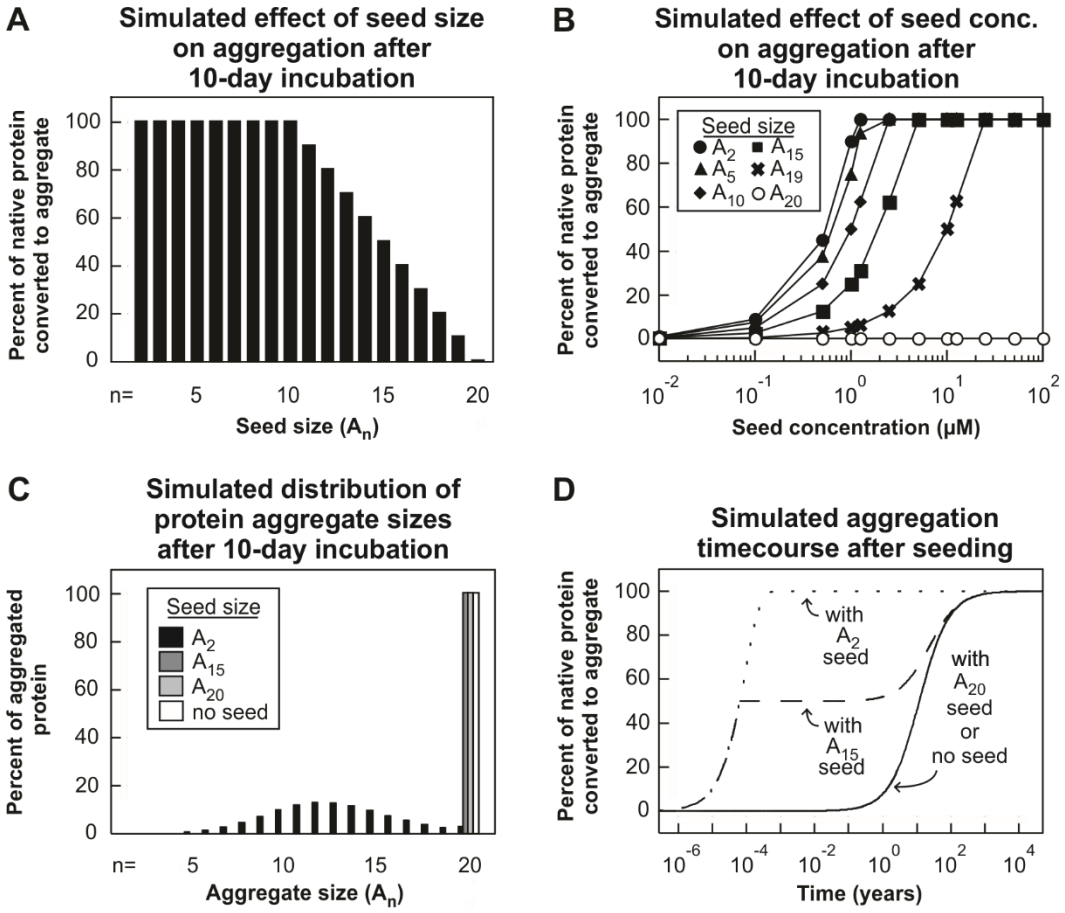


Figure S6. Simulated effect of seeding on the conversion of native protein to aggregates. Effect of seed (a) size and (b) concentration on aggregation kinetics. Effect of seeding on the (c) distribution of aggregated protein among differently-sized aggregates after ten-day incubation and (d) aggregation time course during long-term incubation. In all panels, the starting concentration of native protein is 20 μM. Except for (b), the concentration of the seed was 2 μM. The lines in (b) are meant to guide the eye. Aggregation in (b) remained independent of A_{20} seed concentration even at very high seed concentrations (tested up to 100 M; data not shown). Considering the starting concentrations of native protein and seeds, the total monomer concentrations in (a) range from 24 μM to 60 μM depending on which seed size was added. The total monomer concentrations in (c) and (d) are either 24 μM, 50 μM, 60 μM or 20 μM depending on whether a seed size of A_2, A_{15}, A_{20} or no seed was added.

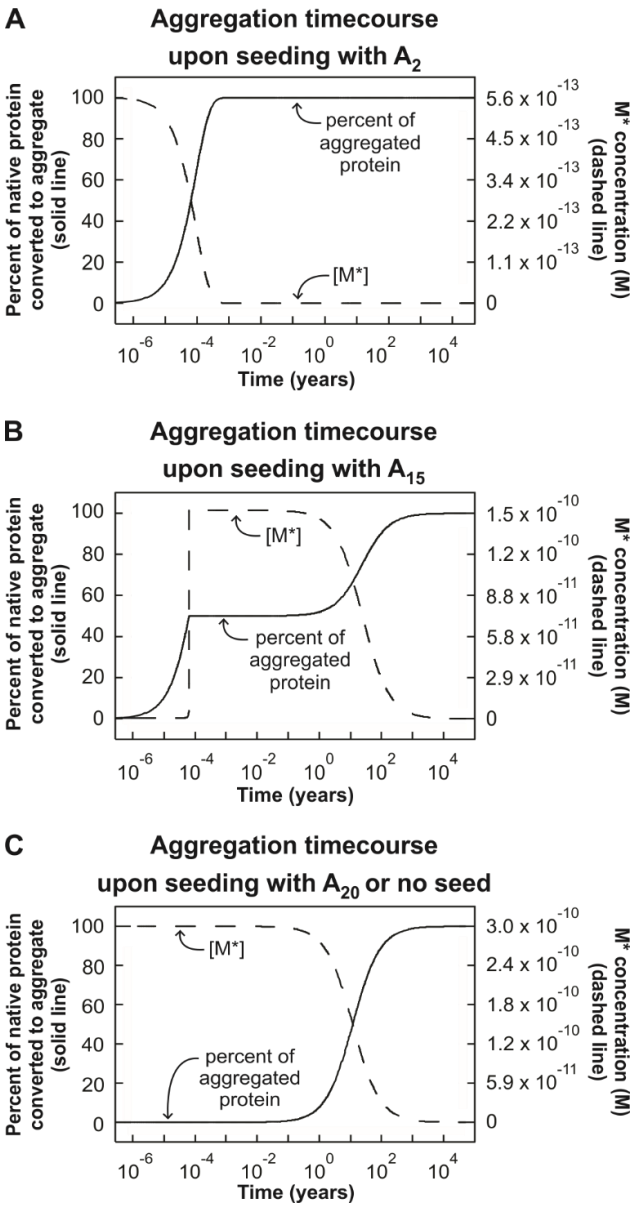


Figure S7. Simulated timecourse of high-energy protein monomer ([M*]) and total-aggregate concentration during long-term incubation, upon seeding with 2 μM (a) A_2, (b) A_{15} or (c) A_{20} aggregates. In all panels, the starting concentration of native protein was 20 μM. Taking into account the simulated starting concentrations of both native protein and seeds, the total monomer concentration in each panel is (a) 24 μM, (b) 50 μM, (c) 60 μM. Panel c also illustrates the aggregation time course in the absence of seeding, corresponding to a total monomer concentration of 20 μM.

Supplementary Tables

Table S1. Top ten most abundant proteins identified via nano-LC-ESI MS/MS in the pellet fraction of the *E. coli* S100 protein mixture after heating and cooling.

Protein name	UniProt ID	MW (kDa)	emPAI score	% Sequence coverage	Total ion score	Spectral counts
Glyceraldehyde-3-phosphate dehydrogenase A	P0A9B2	35,681	942	87	10,873	544
Malate dehydrogenase	P61889	53,221	280	95	7,448	258
Tryptophanase	P0A853	53,204	267	88	9,399	543
Enolase	P0A6P9	69,130	185	81	6,440	303
Glutamate decarboxylase alpha	P69908	20,862	122	81	8,141	351
Glutamate decarboxylase beta	P69910	10,381	108	81	8,093	343
2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	P62707	45,683	102	79	2,549	153
Alkyl hydroperoxide reductase subunit C	P0AE08	19,575	101	91	2,123	112
60 kDa chaperonin	P0A6F5	27,126	85	86	7,781	365
Transaldolase B	P0A870	32,488	79	86	2,698	145

Table S2. Top ten most abundant proteins identified via nano-LC-ESI MS/MS in the soluble fraction of the *E. coli* S100 protein mixture after heating and cooling.

Protein name	UniProt ID	MW (kDa)	emPAI score	% Sequence	Total ion score	Total spectral counts	Function
D-galactose-binding periplasmic protein	P0AEE5	35,690	20,591	90	10,745	407	Transport of galactose and glucose; plays a role in galactose chemotaxis[4]
DNA-binding protein H-NS	P0ACF8	15,587	13,138	94	5,639	306	Transcription[5] and translation regulator; binding non-coding DNA[6]; chromosome organization [7]
Osmotically-inducible	P0AFH8	21,061	6,138	96	18,646	508	Stress-induced with unknown function[8-9]
10 kDa chaperonin	P0A6F9	10,381	2,285	68	3,537	209	Supresses ATPase activity[10,11]
Acid stress chaperone HdeA	P0AES9	11,964	1,917	74	7,564	383	Exhibits chaperone activity at low pH[12]
UPF0337 protein YibI	P68206	8,320	1,617	89	2,455	150	Stress induced with unknown function[9]
S-ribosylhomocysteine lyase	P45578	19,575	793	84	1,536	121	Synthesis of an autoinducer that plays a role in quorum sensing[13]
DNA-directed RNA polymerase subunit	P0A800	10,230	618	69	826	42	Promotes RNA polymerase assembly[14]
Acid stress chaperone	P0AET2	12,149	515	98	2,741	170	Exhibits chaperone activity at low pH[14]
Putative ABC transporter arginine-	P30859	27,027	482	72	3,390	182	Involved in arginine transport[15]

Table S3. Top ten most abundant proteins identified via nano-LC-ESI MS/MS in the unheated *E. coli* S100 protein mixture.

Protein name	UniProt ID	MW (kDa)	emPAI score	% Sequence coverage	Total ion score	Spectral counts
Glyceraldehyde-3-phosphate dehydrogenase A	P0A9B2	35,681	552	89	12690	558
Glutamate decarboxylase alpha	P69908	53221	340	82	10641	480
Glutamate decarboxylase beta	P69910	53204	283	82	10572	469
Chaperone protein DnaK	P0A6Y8	69130	283	87	11329	476
Alkyl hydroperoxide reductase subunit C	P0AE08	20862	159	89	1989	101
10 kDa chaperonin	P0A6F9	10381	129	86	1195	82
Enolase	P0A6P9	45683	106	86	6621	323
S-ribosylhomocysteine lyase	P45578	19575	99	88	1117	85
Triosephosphate isomerase	P0A858	27126	92	93	2538	184
Malate dehydrogenase	P61889	32488	87	97	7980	261

Table S4. Kinetic steps and corresponding rate constants used in the kinetic simulations with a dimeric nucleus presented in this work.

Reaction	Rate constants
N ⇄ I	$k_{NI} = 0.47\ s^{-1}$ (1) $k_{IN} = 22.4\ s^{-1}$ (2)
I ⇄ U	$k_{IU} = 61.5\ s^{-1}$ (3) $k_{UI} = 3.47 \times 10^{3}\ s^{-1}$ (4)
U ⇄ M*	$k_{UM^*} = 0.94\ s^{-1}$ $k_{M^*U} = 23.1\ s^{-1}$
M* + M* ⇄ A_2	$k_{M^*A_2} = 3.04 \times 10^{4}\ M^{-1}s^{-1}$ $k_{A_2M^*} = 1.92 \times 10^{-12}\ s^{-1}$
A_n + M* ⇄ A_{n+1}	$k_{A_nA_{n+1}} = 6.06 \times 10^{9}\ M^{-1}s^{-1}$ $k_{A_{n+1}A_n} = 7.82 \times 10^{-12}\ s^{-1}$

(1) Rate constant calculated from stability data[16,17] along with kinetics data.[18]

(2) Rate constant estimated from kinetics data.[18]

(3) Rate constant calculated from stability data[17] and kinetics data.[19]

(4) Rate constant estimated from kinetics data.[19]

Table S5. Kinetic steps and corresponding rate constants used in the kinetic simulations with a monomeric nucleus presented in this work.

Reaction	Rate constants
N ⇄ I	$k_{NI} = 0.47\ s^{-1}$ (1) $k_{IN} = 22.4\ s^{-1}$ (2)
I ⇄ U	$k_{IU} = 61.5\ s^{-1}$ (3) $k_{UI} = 3.47 \times 10^{3}\ s^{-1}$ (4)
U ⇄ M*	$k_{UM^*} = 0.94\ s^{-1}$ $k_{M^*U} = 2.41 \times 10^{-16}\ s^{-1}$
M* + M* ⇄ A_2	$k_{M^*A_2} = 3.04 \times 10^{4}\ M^{-1}s^{-1}$ $k_{A_2M^*} = 1.92 \times 10^{-12}\ s^{-1}$
A_n + M* ⇄ A_{n+1}	$k_{A_nA_{n+1}} = 6.06 \times 10^{9}\ M^{-1}s^{-1}$ $k_{A_{n+1}A_n} = 7.82 \times 10^{-12}\ s^{-1}$

(1) Rate constant calculated from stability data[16,17] along with kinetics data.[18]

(2) Rate constant estimated from kinetics data.[18]

(3) Rate constant calculated from stability data[17] and kinetics data.[19]

(4) Rate constant estimated from kinetics data.[19]

Table S6. Kinetic steps and corresponding rate constants used in the kinetic simulations with a trimeric nucleus presented in this work.

Reaction	Rate constants
N ⇄ I	$k_{NI} = 0.47\ s^{-1}$ (1) $k_{IN} = 22.4\ s^{-1}$ (2)
I ⇄ U	$k_{IU} = 61.5\ s^{-1}$ (3) $k_{UI} = 3.47 \times 10^{3}\ s^{-1}$ (4)
U ⇄ M*	$k_{UM^*} = 0.94\ s^{-1}$ $k_{M^*U} = 23.1\ s^{-1}$
M* + M* ⇄ A_2	$k_{M^*A_2} = 3.04 \times 10^{4}\ M^{-1}s^{-1}$ $k_{A_2M^*} = 1.64 \times 10^{5}\ s^{-1}$
A_n + M* ⇄ A_{n+1}	$k_{A_nA_{n+1}} = 6.06 \times 10^{9}\ M^{-1}s^{-1}$ $k_{A_{n+1}A_n} = 7.82 \times 10^{-12}\ s^{-1}$

(1) Rate constant calculated from stability data[16,17] and kinetics data.[18]

(2) Rate constant estimated from kinetics data.[18]

(3) Rate constant calculated from stability data[17] and kinetics data.[19]

(4) Rate constant estimated from kinetics data.[19]

Supporting references

(1) Bozzola, J. J.; Russell, L. D., *Electron microscopy*. Jones & Bartlett Learning, LLC: 1991.

(2) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646-4658.

(3) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, Y.; Rappsilber, J.; Mann, M. Exponentially modified proetin abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of seqeunced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265-1272.

(4) Silhavy, T. J.; Boos, W.; Kalckar, H. M., The role of the Escherichia coli galactose-binding protein in galactose transport and chemotaxis. In *Biochemistry of sensory functions*, Jaenicke, L., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 1974; pp 165-205.

(5) Ueguchi, C.; Seto, C.; Suzuki, T.; Mizuno, T. Clarification of the dimerization domain and its functional significance for the Escherichia coli nucleoid protein H-NS. *J. Mol. Biol.* **1997**, *274*, 145-151.

(6) Grainger, D. C.; Hurd, D.; Goldberg, M. D.; Busby, S. J. W. Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res.* **2006**, *34*, 4642-4652.

(7) Wang, W.; Li, G.-W.; Chen, C.; Xie, X. S.; Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **2011**, *333*, 1445-1449.

(8) Yim, H. H.; Villarejo, M. osmY, a new hyperosmotically inducible gene, encodes a periplasmic protein in Escherichia coli. *J. Bacteriol.* **1992**, *174*, 3637-3644.

(9) Weber, A.; Kögl, S. A.; Jung, K. Time-dependent proteome alterations under osmotic stress

during aerobic and anaerobic growth in Escherichia coli. *J. Bacteriol.* **2006**, *188*, 7165-7175.

(10) Keskin, O.; Bahar, I.; Flatow, D.; Covell, D.; Jernigan, R. Molecular mechanisms of chaperonin groel-groes function. *Biochem.* **2002**, *41*, 491-501.

(11) Numoto, N.; Kita, A.; Miki, K. Crystal structure of the co-chaperonin Cpn10 from Thermus thermophilus HB8. *Proteins: Struct. Funct. Bioinform.* **2005**, *58*, 498-500.

(12) Gajiwala, K. S.; Burley, S. K. Hdea, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *J. Mol. Biol.* **2000**, *295*, 605-612.

(13) Surette, M. G.; Miller, M. B.; Bassler, B. L. Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: A new family of genes responsible for autoinducer production. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1639-1644.

(14) Gentry, D. R.; Burgess, R. R. rpoZ, encoding the omega subunit of Escherichia coli RNA polymerase, is in the same operon as spot. *J. Bacteriol.* **1989**, *171*, 1271-1277.

(15) Wissenbach, U.; Six, S.; Bongaerts, J.; Ternes, D.; Steinwachs, S.; Unden, G. A third periplasmic transport system for l-arginine in Escherichia coli: Molecular characterization of the artPIQMJ genes, arginine binding and transport. *Mol. Microbiol.* **1995**, *17*, 675-686.

(16) 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.

(17) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993**, *262*, 892-896.

(18) Eun, Y. J.; Kurt, N.; Sekhar, A.; Cavagnero, S. Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* **2008**, *376*, 879-97.

(19) 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*, 1171-1176.

Chapter 4

Nascent Polypeptide Chains Have a Stabilizing Effect on Ribosomal Components

The text of this chapter is part of a manuscript in preparation by Angela E. Varela, Meranda M. Masse, Aniruddha Srivastava and Silvia Cavagnero, titled “Nascent Polypeptide Chains Have a Stabilizing Effect on Ribosomal Components”.

4.1 Abstract

The role of the ribosome in the cell extends well beyond catalysis of peptide-bond formation. For instance, the ribosome is emerging as a key player in cotranslational protein folding. A number of studies showed that the ribosome interacts with newly synthesized nascent protein chains. Little is known, however, on how these interactions may affect the ribosome. Here, we investigate whether the presence of the nascent chain affects the urea sensitivity of the bacterial ribosome, and how various ribosome-bound nascent chains affect the apparent stability of the entire ribosome and two of its components: the peptidyl-transferase center (PTC) and ribosomal proteins. Several constructs derived from the *E. coli* model protein flavohemoglobin (Hmp) were employed to test whether nascent-chain length or amino-acid sequence influence the apparent stability of ribosomal components. Sucrose-gradient data showed that the empty ribosome is overall more sensitive to urea denaturation than the ribosome in complex with the nascent chain (RNC). The stabilizing effect imparted by the presence of the nascent protein is not chain-length nor amino-acid-sequence dependent. Further studies on RNCs showed that the PTC and the ribosomal proteins display a similar sensitivity to urea. This effect is, in turn, less pronounced than the global urea dependence of ribosomal-subunit separation (i.e., 70S into 30S and 50S). To determine if this effect is due to the presence of a nascent chain or the tRNA in complex with the ribosome, control experiments on significantly truncated RNCs (containing 2-3 amino acids) will be carried out in the near future. Thus far, our results support a hierarchical model according to which the bacterial ribosome is stabilized by the presence of the nascent chain. The latter helps holding ribosomal subunits together as well as non-specifically stabilizing the PTC and ribosomal proteins.

4.2 Introduction

While recent evidence shows that the ribosome plays an active role in cotranslational protein folding, little is currently known about the effect of nascent protein chains on the ribosome [1-4]. During translation, the nascent chain traverses the ribosomal exit tunnel, where it starts acquiring some secondary structure.[5-7] The ribosomal exit tunnel is approximately 80-100 Å long, ranges in width between 10 and 20 Å.[8-10], and fits approximately 30 to 40 amino acids [11-13]. The nascent chain may interact with ribosomal components within the tunnel. The inner wall of the bacterial ribosomal exit tunnel consists primarily of 23S rRNA and ribosomal proteins L4, L22 and L23 [8, 10, 14]. Different regions of the tunnel are known to have a different sensitivity to nascent-protein side chains.[15-16]; In all, however, it is not known whether the overall stability of the ribosome is influenced by the presence of the native protein.

As summarized in Table 1, nascent-chain-ribosome interactions can be divided into four categories, depending on whether these interactions involve nascent proteins bearing an N-terminal signal sequence, or whether the interacting portion of the ribosome is located inside or outside the ribosomal exit tunnel. Single-particle cryo-electron microscopy (cryo-EM) and chemically or photochemically-induced crosslinking provide the most direct evidence. In short, interactions between nascent chains and ribosomal components have been established except for RNCs lacking N-terminal signal sequences. Nascent protein chains may experience weak noncovalent bonding with either the ribosomal exit tunnel or the ribosomal surface. The currently available evidence on nascent-chain-ribosome interactions is summarized in Table 1.

Photo-crosslinking and cryo-electron microscopy (cryo-EM) in prokaryotic and eukaryotic *in vitro* systems showed that nascent chains bearing signal sequences interact with the 23S RNA and with the three ribosomal proteins located within the tunnel. [14, 17-19] In contrast, nascent chains

lacking a signal sequence were only found to interact with the ribosomal proteins in the exit tunnel, to date.[20-22]

On the other hand, once the nascent protein chain is sufficiently long to protrude outside of the exit tunnel, it may interact with the ribosomal surface. This putative mode of action is supported by recent findings.[23] Ribosome-bound nascent chains (RNCs) containing N-terminal signal sequences were found to crosslink with *E.coli* ribosomal proteins facing the outer surface of the ribosome L24, L29, 23S RNA and, surprisingly, even with ribosomal proteins facing the surface of the 30S subunit [24-28]. However, N-terminal signal sequences are only relevant to a specific class of proteins that need to be post-translationally targeted to the cell membrane, to the periplasm, or to secretion pathways.[29]

Although high resolution NMR and molecular dynamics simulations suggest that some RNCs lacking a signal sequence may also interact with the ribosomal surface, clear evidence for this phenomenon is lacking. Hence, to date it has not been unequivocally proven that these interactions exist, due to lack of explicit evidence derived from cross linking and/or cryo-EM. [30-32] Ongoing cross-linking studies in the Cavagnero group (not part of this thesis) are presently addressing this lack of knowledge.

While, overall, the presence of some interactions between the ribosome and RNCs has been established, virtually nothing is known about how the nascent protein chain affects the stability of the ribosome, possibly due to the above interactions. Here, we compare the urea sensitivity of RNCs lacking a signal sequence to that of empty 70S ribosomes. In addition, we address how nascent protein chains affect the stability of different ribosomal components, namely the peptidyl transferase center (PTC) and ribosomal proteins. Briefly, we found that RNCs render the ribosome less sensitive to urea denaturation. On the other hand, the peptide sequence and length (beyond ca

32 residues) do not play any significant role in this effect. Finally, based on our data we propose a working model for urea-induced ribosome disassembly. According to this model, the 30 and 50S subunits of 70S RNCs disassemble at a urea concentration between 1 and 2 molar, followed by concerted disassembly of the PTC and unfolding of ribosomal proteins at higher urea concentration (ca. 3 – 4 molar).

4.3 Results

Design of ribosome-bound nascent proteins and protein fragments

The *E. coli* protein flavohemoglobin (Hmp) was used throughout this work. Hmp plays an important role in O_2, NO and CO transport in the *E. coli* bacterium and is involved in a variety of signaling pathways. The crystal structure of Hmp reveals three domains, an N-terminal heme-binding domain (D1), a flavin adenine dinucleotide-binding domain and a C-terminal nicotinamide adenine dinucleotide-binding domain (Figure 4-1A and Figure 4-1B).[33]

Eight RNC constructs derived from Hmp were generated, as shown in Figure 4-1C. The D1N, D2N and D3N constructs comprise the first 32 N-terminal residues of Hmp domains 1, 2 and 3, respectively. In contrast, the D1C, D2C and D3C constructs contain the last 32 C-terminal residues of domains 1, 2 and 3, respectively. The above RNCs were generated to specifically test for any trends that may discriminate the behavior of N- and C- terminal residues exposed to either the ribosomal exit tunnel or to the outer ribosomal surface. The construct D1 contains 140 amino acids and comprises the entire first domain of Hmp. As a result, D1 contains the same 32 N- and C-terminal residues as D1N and D1C, respectively. In isolation, this domain shares the same all α-helical globin fold of the parent protein and has been extensively characterized.[34-35] The purpose of the D1 construct is to determine the influence of a partially folded domain in close proximity

Table 1. Summary of published experimental evidence on interactions between nascent protein chains (RNCs) and the ribosome, based on single-particle cryo-EM or covalent cross-linking.

RNC – ribosome interactions		
	Inside tunnel	Outside tunnel
+ N-terminal signal sequence	✓	✓
- N-terminal signal sequence	✓	n.a.

n.a. denotes not available.

✓ denotes the established presence of interactions between nascent protein chains and the ribosome.

to the ribosome since this RNC construct is known to be compact and partially folded.[36] The last construct, D1D2N, comprises 189 residues and includes D1and D2N in tandem. The purpose of this construct is to probe whether the ribosome is stabilized when the entire sequence of the D1 domain lays outside the ribosomal exit tunnel.

Sensitivity of the empty 70S ribosome and RNCs to urea

In order to explore the urea sensitivity of the ribosome, we chose to examine both empty 70S ribosomes and a representative RNC construct derived from the D1 domain of Hmp (see Figure 4-1). We then carried out sucrose-gradient experiments to examine ribosomal-subunit behavior as a function of urea concentration. Elution profiles were monitored at both 260 and 280 nm in separate experiments, to probe for any differences in the response of ribosomal RNA and ribosomal proteins. The data obtained in both sets of experiments (with detection at 260 and 280 nm) show comparable results (Figure 4-2). This result suggests that the sensitivity of the ribosomal RNA and proteins is similar at comparable urea concentrations.

The sucrose gradient data of Figure 4-2 show that the 70S peak (purple region) due to the empty 70S ribosomes decreases in intensity at around 1.0 M urea. On the other hand, the sucrose-gradient profiles of the RNCs have comparable 70S intensities up to 2.0 M urea and only become less intense at > 2 M urea. In all, it is clear that the empty 70S ribosomes are more sensitive to urea denaturation than the RNCs.

At > 2 M urea, both the empty-ribosome and RNC peaks undergo severe line broadening. This phenomenon is likely to due to ribosomal-subunit unfolding, possibly followed by RNA unfolding and/or aggregation. At the highest urea concentrations, the line broadening is so extensive that it is difficult to deconvolute the contribution of individual components.

Agarose gels were run on the empty-70S ribosome peaks of the sucrose gradients (see 70S traces with detection at 260 nm in Figure 4-2A). Detection was carried out with ethidium bromide, to selectively visualize the RNA component of the ribosome. The results (Figure 4-2C) show that the band intensities generally decrease at urea concentrations beyond ca 1 M. We attribute this intensity loss to RNA unfolding followed by covalent degradation (via backbone hydrolysis or other complex processes). At 0.0 M urea, the 70S lane contains two bands, which are diagnostic of the presence of 23S RNA and 16S RNA, belonging to the 50S and 30S ribosomal subunits, respectively. Interestingly, at 4.0 M urea, the 23S RNA band intensity increases compared to the 3.0 M urea peak. Although somewhat puzzling, we attribute this peak to the aggregation of the ribosomal RNA likely caused by self-associated states of degraded RNA fragments.

A proposed scheme for the urea unfolding of the empty ribosome is illustrated in Figure 4-3, consistent with the results of the agarose gel in Figure 4-2C and with the sucrose gradient data. Stage 1 (Figure 4-3) represents intact empty 70S ribosomes at 0 M urea. When small amounts of urea are added, the empty 70S dissociates (stage 2, Figure 4-3). At higher concentrations of urea, the ribosomal RNA unfolds (stage 3, Figure 4-3), likely followed by ribonucleotide hydrolysis (stage 4, Figure 4-3).

To better visualize how the 70S peak decreased as a function of urea concentration, the 70S peaks in Figure 4-6A and B were plotted according to their absorbance between 0 and 2.0 M urea (Figure 4-4). Due to complex unfolding processes and line broadening, the 70S peak intensities were not plotted after 2.0 M urea. Overall, the empty 70S ribosome data show significant decreases in the 70S peak between 0.0 M and 1.0 M urea (Figure 4-4A). The 70S peak in 2.0 M urea was not reliable at 2.0 M due to significant line broadening of the 50S peak (Figure

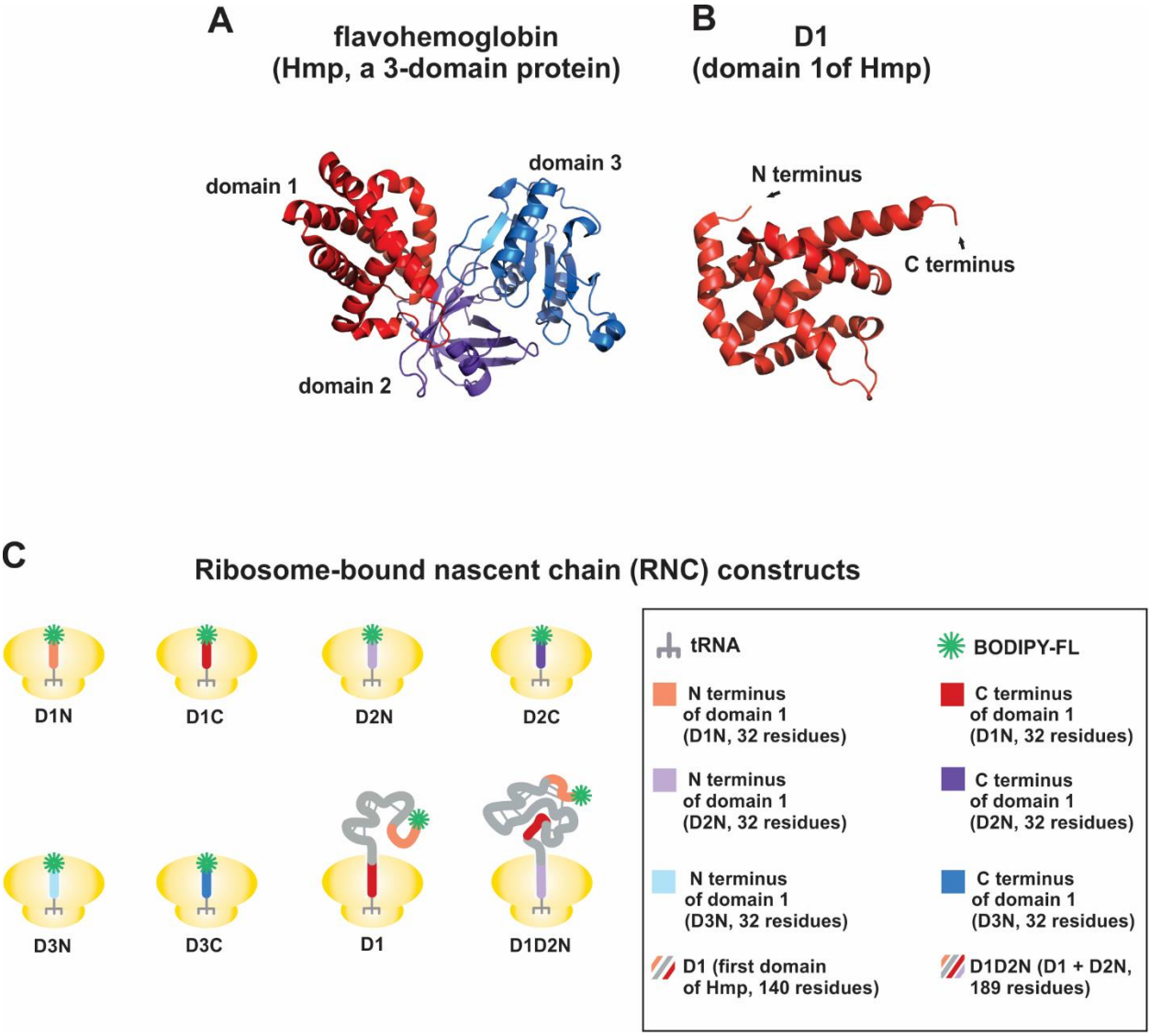


Figure 4-1. Proteins of interest and constructs utilized in this work. A) *E. coli* flavohemoglobin (Hmp) is the model protein used in this study. Hmp has three domains (shown in red, green, and blue). PDB code: 1GVH. **B)** Structure of the N-terminal domain of Hmp, denoted as D1N. **C)** Cartoon illustration of the eight RNC constructs derived from Hmp employed in this work.

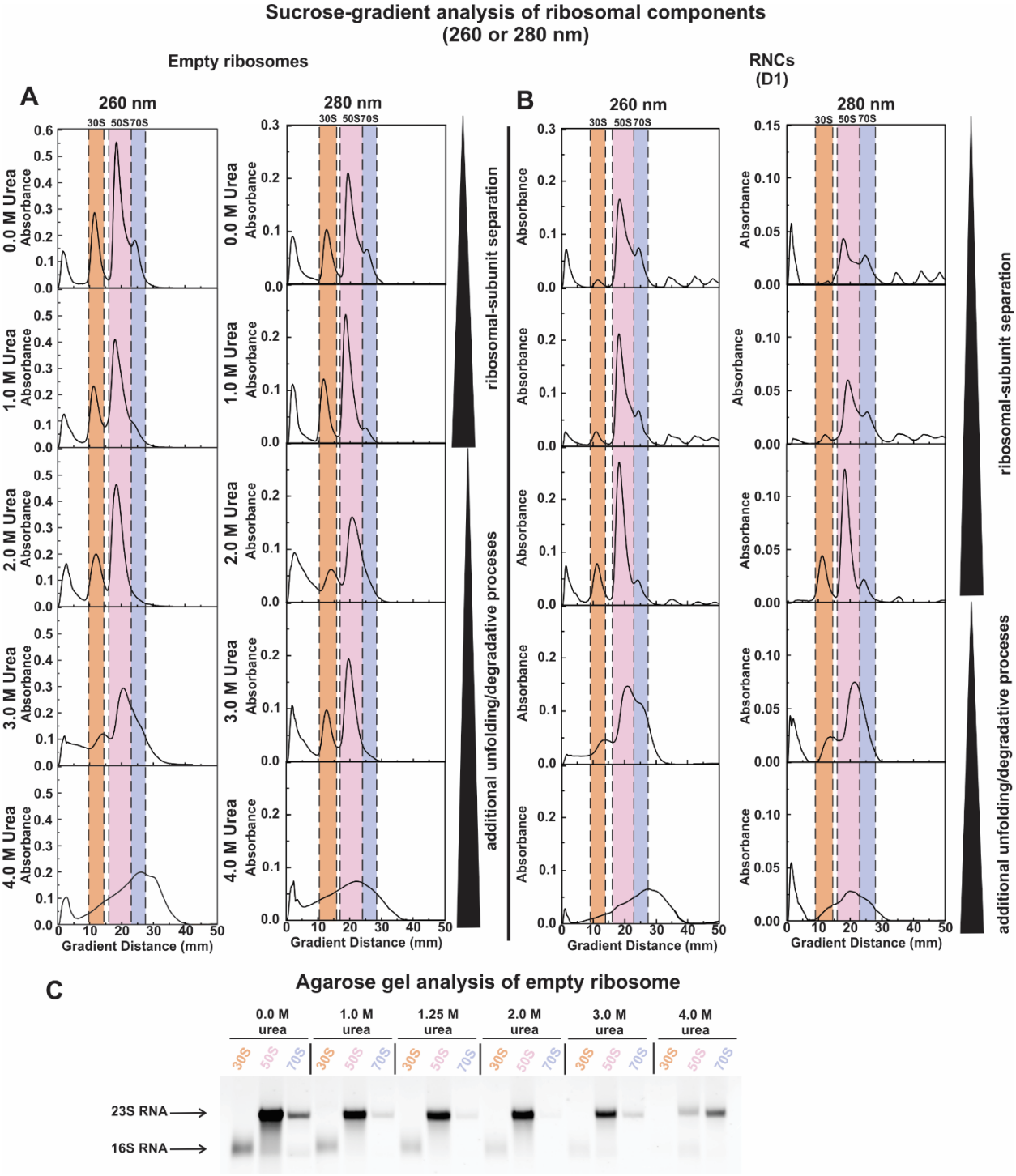


Figure 4-2. Sucrose gradient profiles of empty ribosomes and RNC complexes. A) Side-by-side comparison of sucrose gradient profiles of empty 70S ribosomes, detected by electronic

absorption spectroscopy at 260 nm and 280 nm. Orange-, pink- and purple-colored regions denote the 30S, 50S, and 70S subunits, respectively. **B)** Comparison of the D1 RNC gradient profile at 260 nm and 280 nm. Orange, pink and purple denote 30S, 50S, and 70S peaks, respectively. **C)** RNA agarose gel for collected sucrose-gradient peaks of empty 70S ribosomes (260 nm, see panel A). The 23S 16S RNA bands are indicative of the presence of the 50S and 30 ribosomal subunits, respectively. Together, both bands testify the presence of the 70S ribosomal complex.

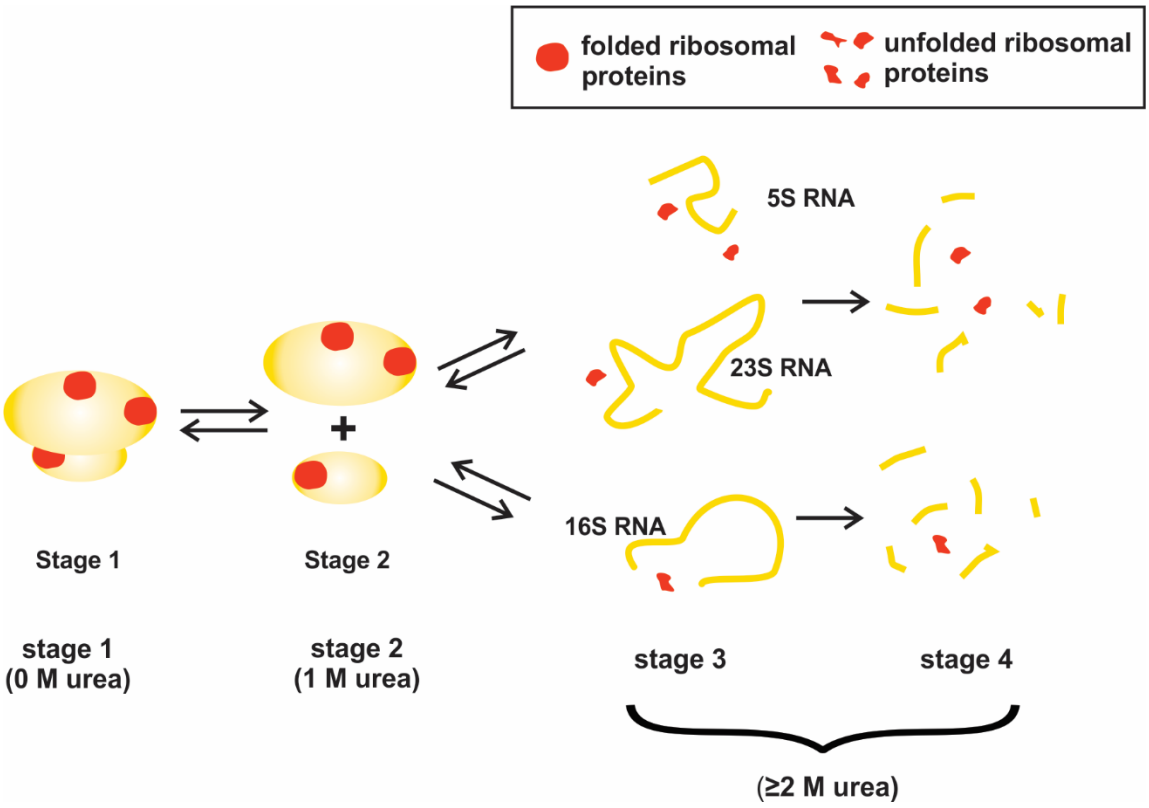


Figure 4-3. Proposed scheme for the unfolding of empty ribosomes upon addition of urea. At 0 M urea the 70S ribosome remains intact (stage 1). Concentrations of urea ca. 1-2 M cause the 70S complex to dissociate into the 30S and 50S subunits (stage 2). At concentrations > 2 M, the subunits unfold and undergo some RNA degradative processes (stages 3 and 4).

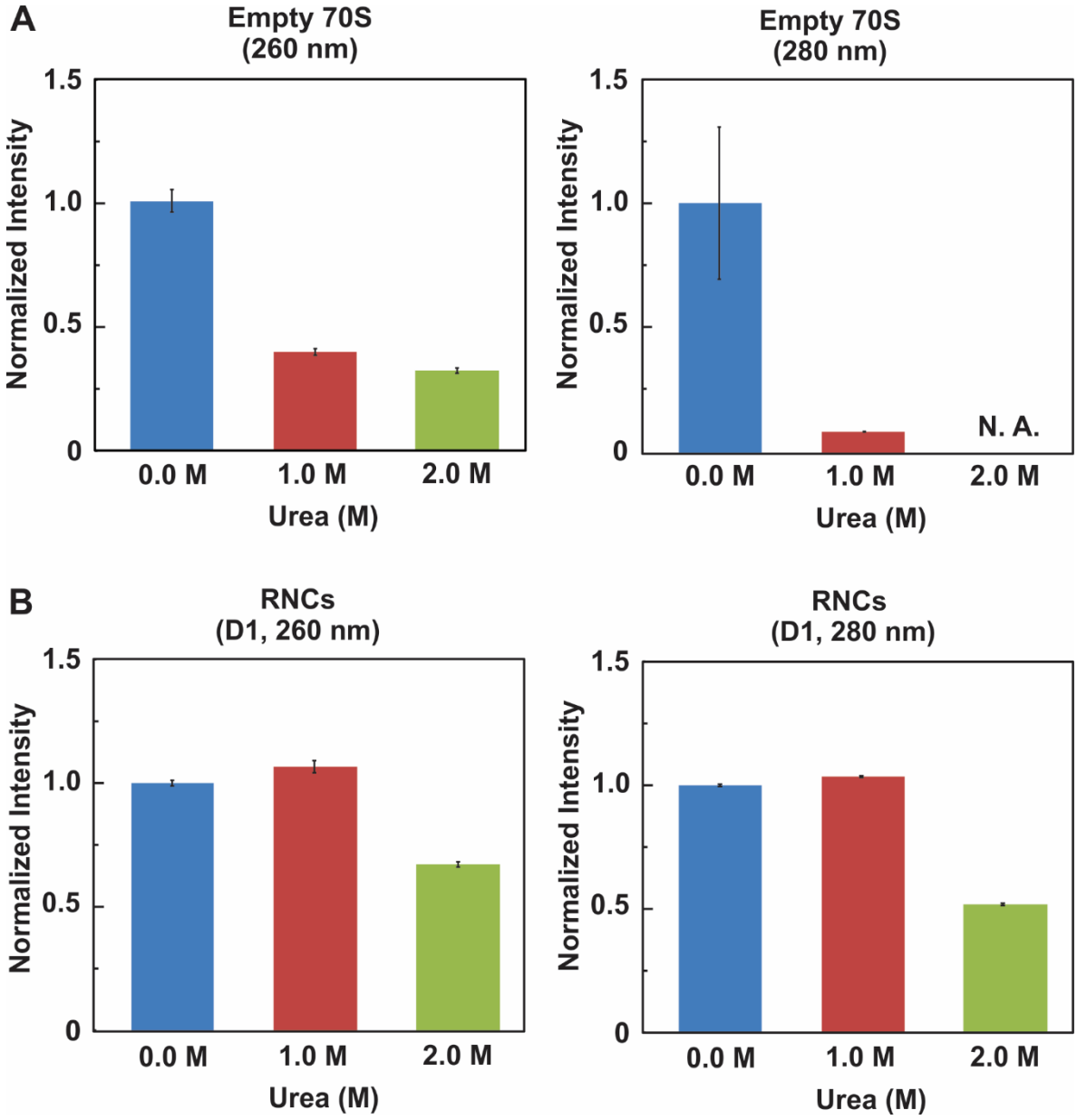


Figure 4-4. The ribosome-bound nascent chain complex is less sensitive to low concentrations of urea than the empty 70S ribosome. A) The 70S peak of empty 70S ribosomes measured at 260 nm and 280 nm from Figure 4-2A show that empty ribosomes are more sensitive to small changes in urea concentrations. N.A. denotes not applicable due to significant line broadening of the 50S peak that distorts the 70S peak in the sucrose gradient data from Figure 4-2A. **B)** The 70s peak of D1 RNCs measured at 260 nm and 280 nm from Figure 4-2B.

4-2A) and was not considered here. In contrast, RNC complexes exhibited less sensitivity to 1.0 M urea and remained mostly intact at 2.0 M urea (Figure 4-4A).

In summary, data show that empty 70S ribosomes are more sensitive to lower concentrations of urea than RNCs, which fall apart at 1.0 and > 2.0 M urea, respectively.

The apparent stability of ribosomal proteins is unaffected by nascent-chain length (beyond 32 residues) and amino-acid sequence

In order to discern why RNCs can withstand higher concentrations of urea than the empty 70S ribosome, the ribosomal protein stability was probed using the eight constructs generated from Hmp. A tryptophan fluorescence assay was utilized to track ribosomal protein unfolding in RNCs as a function of urea concentration since tryptophan is a well-established fluorescent reporter of its environment (Figure 4-5). Because the ribosomal proteins collectively contain 32 tryptophan residues (Figure 4-5A), and each one contributes to the overall fluorescence emission, the resulting spectra should represent and average depiction of the entire structure. Small environmental changes of the tryptophan residues (i.e., non-polar hydrophobic core versus polar, solvent exposed settings) due to ribosomal component unfolding and exposure to polar microenvironments will result in red-shifted emission spectra (Figure 4-5B and 4-5C).

It is important to note that D1 contains a single intrinsic tryptophan residue at position 120 and D1D2N contains an additional tryptophan as position 149; however, we do not believe they significantly influence the average fluorescence spectra since there are striking similarities between the stabilities of the data collected for D1 and D1D2N and the smaller 32-residue constructs, which do not contain intrinsic tryptophan residues.

The spectral center of mass for each emission spectra was calculated as a function of increasing urea concentrations to track red-shifted peaks as a result of unfolded ribosomal proteins in the presence of different Hmp constructs. Figure 4-6 A-H show representative titration curves, which were fit to a curve using a linear extrapolation method[37-38].

Because the unfolding of the ribosome is an irreversible process, a true $\Delta G°_{unfold}$ value cannot be obtained from the extrapolation method; however, an apparent stability value can be inferred and compared between constructs. We refer to the apparent stability of unfolding as $\Delta G°_{app,\ unfold}$ throughout this work. The data in Figure 4-6I show average $\Delta G°_{app,\ unfold}$ for each construct. Overall, all eight constructs show statistically similar results (Figure 4-6J).

The difference between the effects of D1 and D1C, which contain the same 32 C-terminal residues in the ribosomal exit tunnel (Figure 4-1C) was statistically insignificant ($p = 0.5$) , which suggested that the residues outside of the tunnel in D1 did not contribute to added stability by possible interactions with ribosomal proteins near the exit tunnel surface. Similar results can be drawn from D1D2N and D2N ($p = 0.75$), which both possess the same 32 C-terminal residues within the tunnel but differ in the presence polypeptide extensions outside of the tunnel. We also did not notice any striking differences between N- and C-terminal constructs in the tunnel.

The PTC is largely unaffected by nascent-chain sequence and length (beyond 32 residues)

To probe whether nascent chain length and sequence affect the apparent stability of the PTC, a puromycin-release gel assay was developed to measure the functionality of puromycin as urea concentration is increased. Puromycin is an antibiotic that induces premature release of the nascent polypeptide during translation because it mimics the adenosine on the 3’ end of an aminoacylated

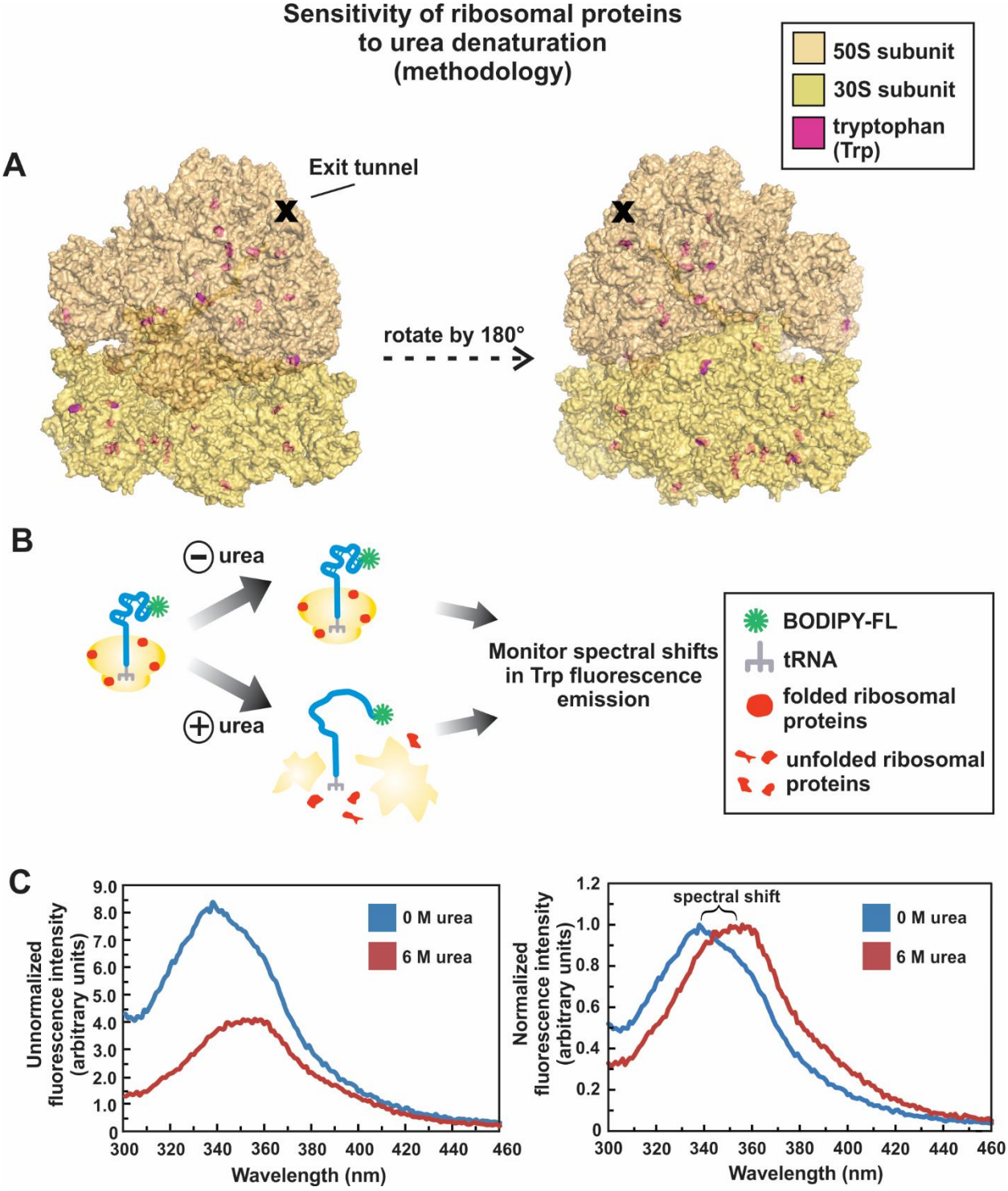
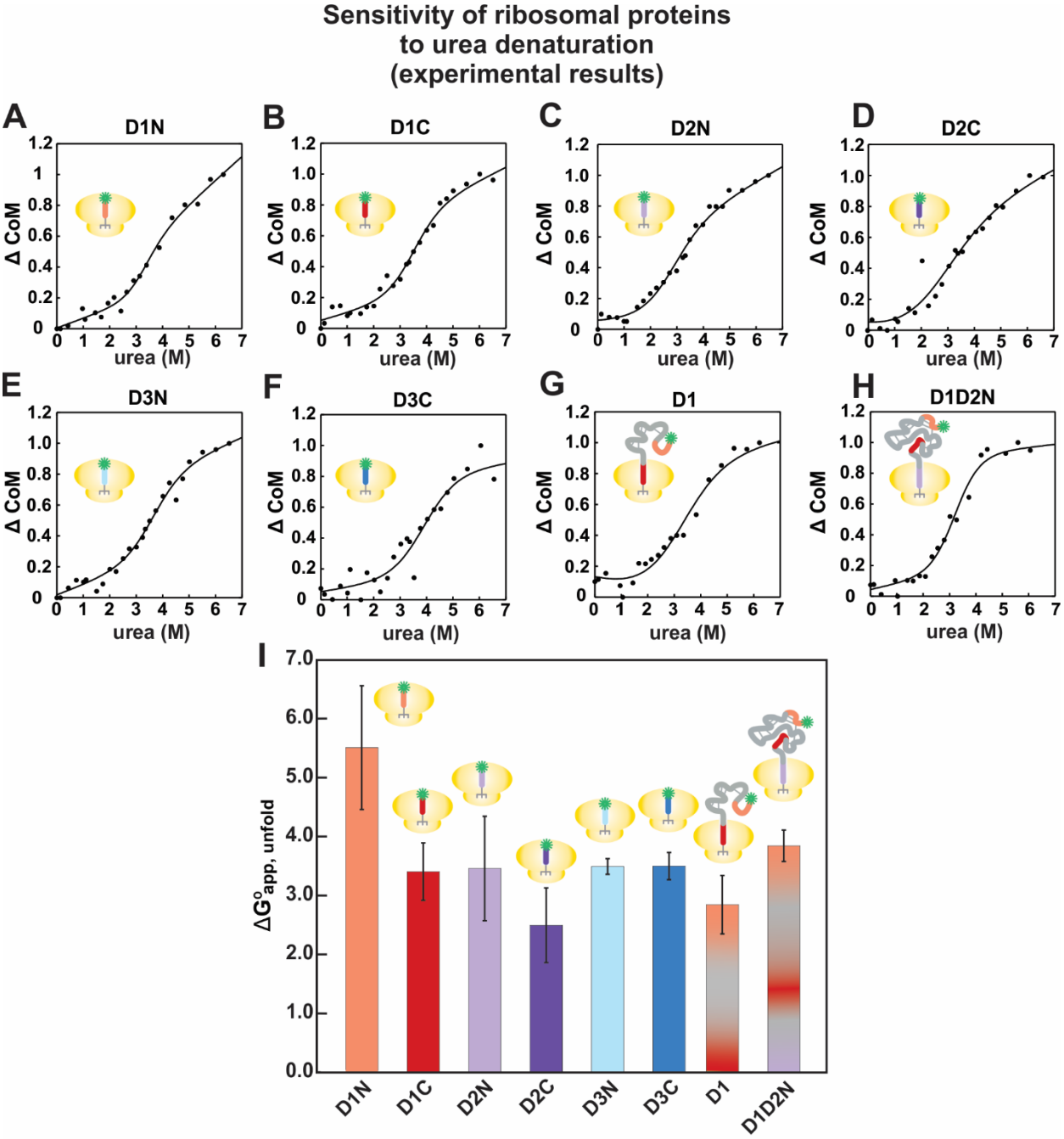


Figure 4-5. **Tryptophan fluorescence of the ribosome. A)** The ribosomal proteins in the *E. coli* 70S ribosome collectively contain 32 tryptophans (Trp, magenta). Sixteen Trps are located within

the proteins of the 50S subunit and sixteen are found in the ribosomal proteins of the 30S subunit. PDB codes: 2WWL and 2WWQ. **B)** Schematic illustration of the basic methodology followed in the ribosomal-protein stability experiments. **C)** Trp fluorescence becomes red-shifted at increasing urea concentrations due to environmental changes by the Trp fluorophore towards a more polar milieu.



J p-values of Ribosomal-Protein Stability

	D1	D1D2N	D1N	D1C	D2N	D2C	D3N	D3C
D1	—							
D1D2N	0.22	—						
D1N	0.26	0.37	—					
D1C	0.50	0.51	0.32	—				
D2N	0.61	0.75	0.27	0.96	—			
D2C	0.69	0.14	0.13	0.34	0.47	—		
D3N	0.43	0.36	0.31	0.89	0.98	0.26	—	
D3C	0.34	0.48	0.85	0.42	0.44	0.31	0.42	—

LEGEND
Same
Different

Figure 4-6. Summary of tryptophan fluorescence experiments illustrating the unfolding of ribosomal proteins. A-H) Representative urea titration curves obtained by steady-state fluorescence spectroscopy of D1N, D1C, D2N, D2C, D3N, and D3C, D1 and D1D2N, respectively. ΔCoM denotes the change in the spectral center of mass as shown in equation 4. **I)** $\Delta G^{\circ}_{app,unfold}$ values of each construct. Error bars represent the standard error based on 2-3 experimental repeats. **J)** A *p*-value table was constructed from the results of a two-tailed Student's T-test (Welch's test) comparing the apparent free energy of unfolding values for each construct. Tan boxes highlight statistically similar results according to a >95% confidence interval and blue boxes denote statistically different results..

tRNA and becomes incorporated into the released peptide.[39] However, puromycin can only function properly to release the nascent polypeptide if the A and P sites in the 50S subunit are intact (Figure 4-7A).[40] At high concentrations of urea, the PTC site denatures and puromycin is unable to effectively release the nascent protein chain from the tRNA molecule (Figure 4-7B).

Low-pH SDS-PAGE analyses were performed on selected Hmp RNC constructs (D1N, D1C, D1 and D1D2N) to specifically probe differences between N- and C- terminal fragments as well as partially and fully folded domains outside of the ribosomal tunnel. Low-pH SDS-PAGE analyses were used to generate peptidyl-tRNA (PT) bands that were quantified based on their intensity as a function of increasing urea concentration (Figure 4-7C and Figure 4-8A). The peptidyl-puromycin band (PP) was not used to track release because of its variability in fluorescence, which is noticeable in D1 in Figure 4-8A. Titration curves were generated based on the relative band intensities and the data was fit to a curve with the same linear extrapolation method used for the tryptophan work to generate the $\Delta G°_{app,\ unfold}$ for each construct [37-38] (Figure 4-8B and 4-8C, see Methods for further details). The $\Delta G°_{app,\ unfold}$ for D1N was the lowest at 3.3 (± 0.3 SEM) kcal/mol. D1C showed a higher $\Delta G°_{app,\ unfold}$ at 4.1 (± 0.8 SEM) kcal/mol. Interestingly, D1, which contains the same 32 C-terminal residues within the ribosomal exit tunnel as D1C, had the highest $\Delta G°_{app,\ unfold}$ at 4.7 (± 0.6 SEM) kcal/mol. D1D2N had a $\Delta G_{app,\ unfold}$ of 3.4 (± 0.6 SEM) kcal/mol.

A *p*-value table with the results from a one-tailed Student's test assuming unequal variances (Welch's test) was constructed in Figure 4-8D to assess any significant differences between the properties of the different constructs. The Student's test shows that all four constructs display an equivalent behavior. D1C and D1 shared a *p*-value of 0.28, which suggests that the N-terminal region of D1 that lies outside of the tunnel does not affect the PTC. Similarly, D1N and Hmp share

a *p*-value of 0.44, confirming the important idea that the peptide-chain portions outside of the tunnel do not significantly contribute to PTC stability. Surprisingly and somewhat unexpectedly, D1 and D1N were the only constructs to show statistically significant slight differences ($p = 0.04$) in $\Delta G°_{app,\ unfold}$. This result might imply some weak transient interactions between partially folded nascent chains and the proteins on the outer surface of the ribosome. The potential presence of these interactions will be further explored in the near future.

4.4 Discussion

RNCs have a higher apparent stability than the empty 70S ribosome

Over the years, researchers have attempted to characterize the stability of the ribosome by thermal denaturation. It was previously confirmed through thermal studies that the unfolding of the *E. coli* ribosome is an irreversible process,[41] and thus, stability experiments have yielded little information other than that the 30S is more thermodynamically unstable than the 50S.

Surprisingly, however, very little work has been done to characterize ribosomal stability using chemical denaturants. Urea unfolding experiments were previously conducted on the *E. coli* 50S subunit and showed that its unfolding as a function of urea is a reversible process.[42]

Only one study has attempted to characterize the stability of the empty 70S ribosome in comparison to a RNC and found that the RNC is stable > 3.0 M urea whereas the 70S is stable > 2.0 M urea.[43] However, the increased stability of their experiments in comparison to ours is likely influenced by Sec-M stalling methods that we did not employ in this work. Sec-M is believed to increase contacts between the 30S and 50S subunits, which would increase the overall stability of the 70S complex.[44] Overall, however, the trends remain the same: the presence of a nascent chain increases the stability of the 70S complex.

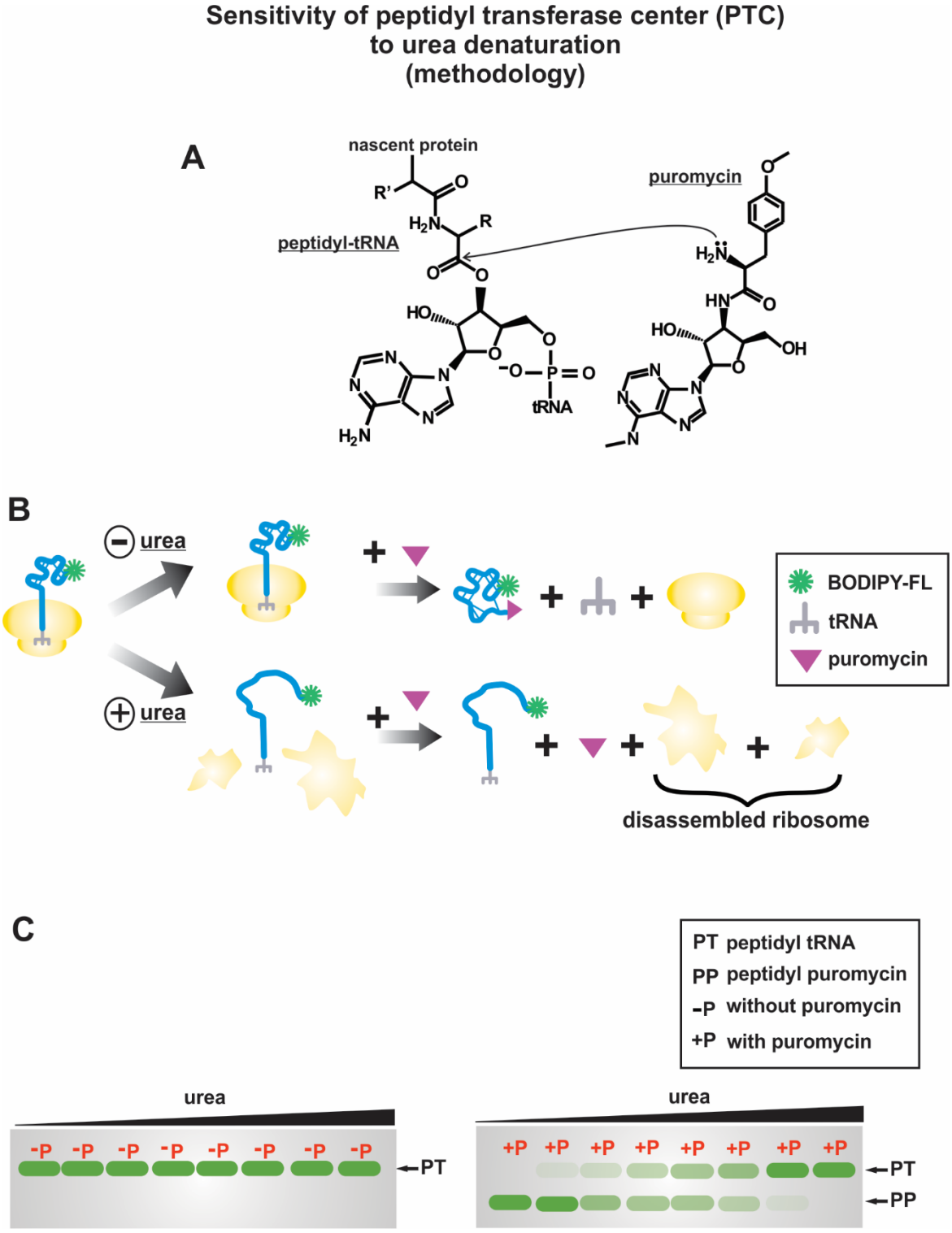


Figure 4-7. Puromycin releases nascent polypeptides from the tRNA at low denaturant concentration. A) Puromycin mimics the 3' end of the aminoacyl-tRNA and covalently attaches

to the nascent chain to release it from the peptidyl-tRNA. **B)** Puromycin is unable to perform its function at high urea concentration because of significant PTC structural modifications. **C)** Low-pH SDS-PAGE analysis is used to quantify the band intensities of peptidyl-tRNA (PT) as a function of urea concentration.

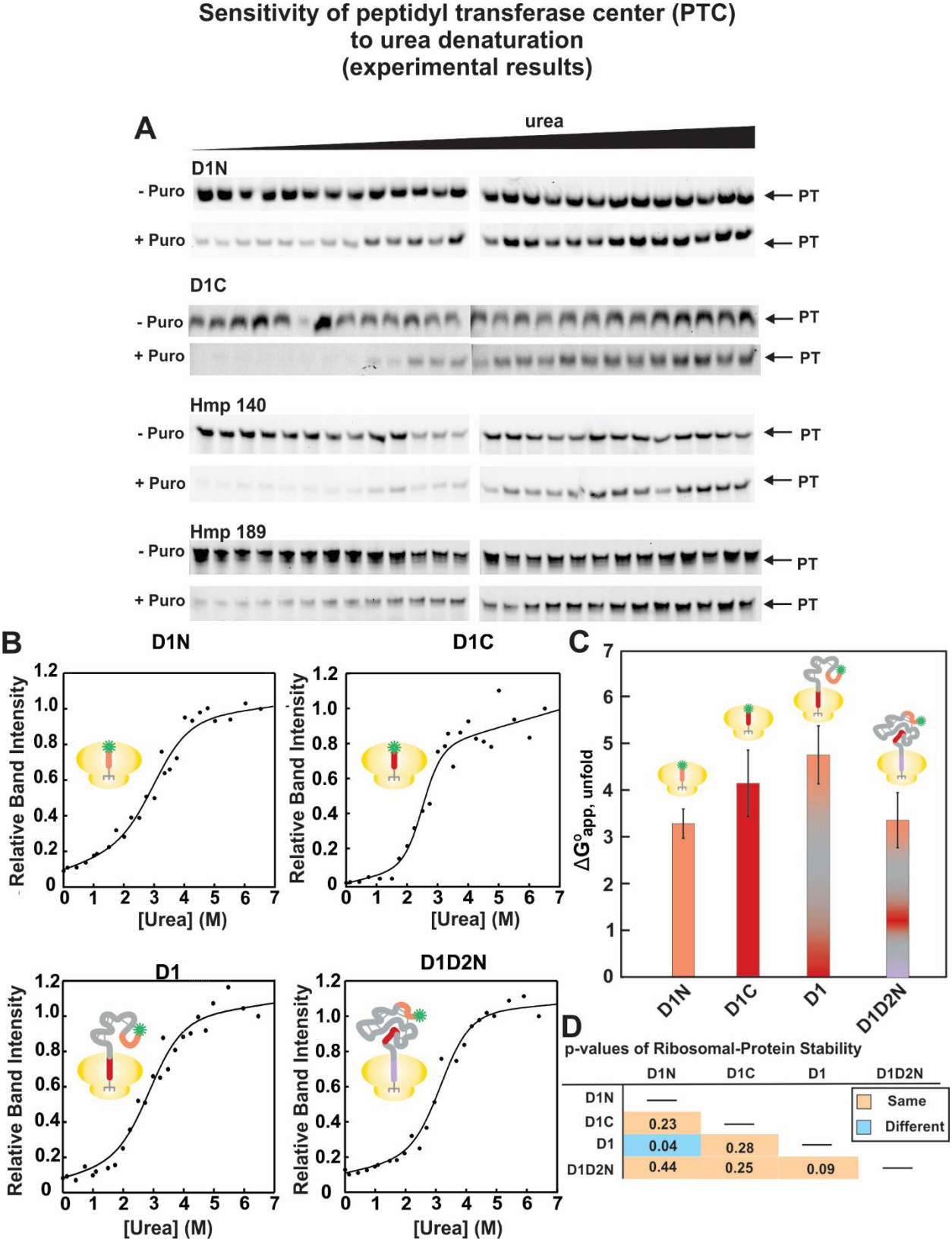

Sensitivity of peptidyl transferase center (PTC)
to urea denaturation
(experimental results)
A
urea
D1N
- Puro
+ Puro
PT
D1C
Hmp 140
Hmp 189
B
D1N
D1C
D1
D1D2N
Relative Band Intensity
[Urea] (M)
C
$\Delta G^{\circ}_{app,\ unfold}$
D1N
D1C
D1
D1D2N
D
p-values of Ribosomal-Protein Stability
D1N
D1C
0.23
D1
0.04
0.28
D1D2N
0.44
0.25
0.09
Same
Different

Figure 4-8. The peptidyl-transferase center is largely unaffected by a nascent chain length (beyond 32 residues) and amino acid sequence. **A)** Representative low-pH gels for each construct show that the peptidyl-tRNA bands are unaffected by increasing concentrations of urea; however, the addition of puromycin results in a decrease in peptidyl-tRNA (PT) band intensities when reacted with 1 mM puromycin at low concentrations of urea. **B)** Representative urea titration curves generated for each construct and fit to a curve using a linear extrapolation method (see Methods for additional details). **C)** $\Delta G^{\circ}_{app,\ unfold}$ values of each construct. Error bars represent standard error based on 2-7 experimental repeats. **D)** A *p*-value table containing the results of a one-tailed Student's test assuming unequal variances (Welch's t-test). Blue highlighted values are statistically significant according to a $\geq$ 95% confidence interval whereas tan highlighted values are statistically similar.

Ribosomal protein stability is unaffected by nascent chain length (beyond 32 residues) and sequence

During elongation, a nascent chain can likely interact with multiple ribosomal proteins as it traverses the tunnel into the cellular milieu. In *E. coli*, ribosomal proteins L4 and L22 are embedded within the ribosomal exit tunnel wall at the constriction site and growing evidence shows them interacting with nascent chains as they travel through the tunnel.[10, 21, 45-46] L23 is located within the tunnel wall in the vestibule[47] of the exit tunnel and can also interact with the nascent polypeptide.[10] In addition, several ribosomal proteins are found on the ribosomal surface in close proximity to the tunnel exit.

The experiments conducted here probe whether interactions with the ribosomal proteins mentioned above result in an increased overall stability of those proteins. In essence, the data show that ribosomal proteins are unaffected by nascent protein length and sequence dependence based on N- or C- terminal residues in the exit tunnel. In addition, the constructs D1 and D1D2N probe for interactions with ribosomal proteins outside of the tunnel and show no significant differences when compared to their shorter C-terminal constructs (D1C and D2N, respectively).

One area unexplored in this work would be to determine specific regions in the ribosomal proteins that interact with the growing nascent chain and mutate those residues to inhibit interaction and compare those results to those shown here. Similar results to those presented here would suggest that the ribosomal proteins are entirely unaffected by the nascent chain. Lower stability values would indicate that interactions with the nascent chain stabilize the ribosomal proteins and higher stability values would hint that interactions with the nascent chain somehow destabilize ribosomal proteins.

The apparent stability of the PTC is not affected by nascent chain length (beyond 32 residues) and amino-acid sequence

The development of the puromycin-release gel assay provides a tool to experimentally determine the structural integrity of the PTC as a function of increasing denaturant concentration. Although our results show statistically insignificant results, we were able to demonstrate that the PTC is unaffected by nascent chain length or the presence of N and C termini within the ribosomal exit tunnel.

The D1 RNC contains a partially folded domain outside of the ribosomal exit tunnel that may transiently interact with the surface of the ribosome. This interaction may, in principle allosterically affect other parts of the ribosome. D1 and D1C share the same 32 C-terminal residues within the ribosomal exit tunnel and comparisons between these constructs yields statistically insignificant results (p = 0.28). This result suggests that, in case D1 RNCs were to interact with the ribosomal surface, allosteric effects between the ribosomal surface and the PTC do not play a role in stabilizing the PTC.

Interestingly, D1 is the only tested construct containing a partially folded structure outside of the ribosome.[36] Note that the D1D2N construct is expected to be fully folded outside of the ribosomal tunnel because the entire sequence of this construct is available for conformational sampling out of the ribosomal exit tunnel. Now, the D1 construct shows a slightly higher apparent ribosomal-protein stability ($\Delta G°_{app,unfold}$ = 4.7 ± 0.6 kcal/mol) relative to the D1N RNC construct ($\Delta G°_{app,unfold}$ = 3.3 ± 0.3 kcal/mol, Fig. 4-8C and D). Hence we propose that constructs with a chain length between D1N (32 residues) and D1 (140 residues) may cause an even higher apparent stability of ribosomal proteins due to an even more limited degree of independent folding, leading to increased interactions with proteins on the outer surface of the ribosome. Previous work in the

Cavagnero group[48] suggests that nascent chains bearing ca. 57 amino acids toggle between a compact and an unstructured state, which likely interacts with the ribosomal surface. Ongoing studies in the Cavagnero group based on chemical crosslinking are currently testing for the presence of actual interactions with the ribosomal surface. Hence, future experiments related to this chapter's research include testing additional constructs with chain length comprised between D1N and D1, to probe for any additional stabilization of ribosomal proteins due to interactions between nascent chain and ribosomal proteins.

A working model for *E. coli* ribosome-nascent-chain complex disassembly in the presence of urea

Based on the experimental results presented here, we propose a working model for *E. coli* RNC disassembly as a function of urea concentration, as illustrated in Figure 4-9. Briefly, starting from intact RNCs at stage 1 (see Figure 4-2B and 4-2C at 0 M urea), the two ribosomal subunits disassemble into separate 30S and 50S components as the urea concentration reaches ca. 1-2 M, in stage 2 (Figure 4-2B at 1 and 2 M urea). As urea concentration increases up to 3-4 M, the PTC and the ribosomal proteins unfold and nucleotide hydrolysis takes place (see stages 3 and 4 of Figure 4-9, respectively, and Figure 4-10 for a comparison of RNC disassembly with empty 70S ribosome disassembly). Trp fluorescence and PTC experiments show that the ribosomal proteins and the ribosomal components of the PTC region have a similar apparent thermodynamic stability of 3-4 kcal·mol^{-1}, suggesting concurrent unfolding of RNA and ribosomal proteins (Figure 4-6I and 4-8C) after the 30S and 50S subunits have separated at ca. 1-2 M urea, as shown by the RNC sucrose gradients (Figure 4-2B). Consistent with the data in Figures 4-6 and 4-8, sucrose gradient

peak broadening at 3-4 M urea (Figure 4-2) testifies the presence of RNA unfolding likely followed by covalent degradation into smaller fragments.

Future Directions

The research pertaining to this work is still undergoing and will be completed in collaboration with Meranda Masse, another graduate student in the Cavagnero group.

Additional pending experiments include probing the urea sensitivity of RNCs that contain very short (1-2 amino acids) peptides to determine whether the increased stability we observe is dominated by the interaction of either the nascent polypeptide chain or the tRNA with the ribosome.

In addition, we will examine the $\Delta G°_{app,unfold}$ of a relatively short, partially folded construct that contains 55 amino acids. The purpose of this experiment is to investigate whether partially folded constructs interact with the ribosomal surface and increase the overall apparent stability of the RNC.

4.5 Methods

Preparation of RNCs

RNCs were generated using an in-house prepared A19 *E.* coli transcription-translation coupled cell-free system as previously described.[48-49] Transcription-translation proceeded for 30 min at 37 °C in the presence of BODIPY-FL-Met-tRNA^{fmet} to N-terminally label the RNCs. BODIPY-FL-Met-tRNAfmet was prepared as described.[48] RNCs were stalled at various lengths to generate each apoHmmpH construct using oligodeoxynucleotide-directed mRNA cleavage. An anti-SsrA oligonucleotide[48] was also added to the system to prevent premature release of the stalled RNC prior to synthesis.

RNC pellets were isolated using 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)[48] and subjected to ultracentrifugation at 160, 000 g for 1 hr at 4 °C. The purified pellet was resuspended 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 at 250 rpm on ice for 1 hr.

RNC denaturation

A stock 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 obtain its starting concentration. RNCs were incubated in varying concentrations of urea for 1 hr at ambient temperature in the dark.

Stability of 70S ribosome and RNCs

An in-house prepared A19 *E. coli* mixture (with 70S ribosomes) was subjected to a 1 hr incubation in varying concentrations of urea 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 0.3 mm/sec. The absorbance was measured at 260 nm and 280 nm to check for intactness of the ribosomal RNA and ribosomal 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.

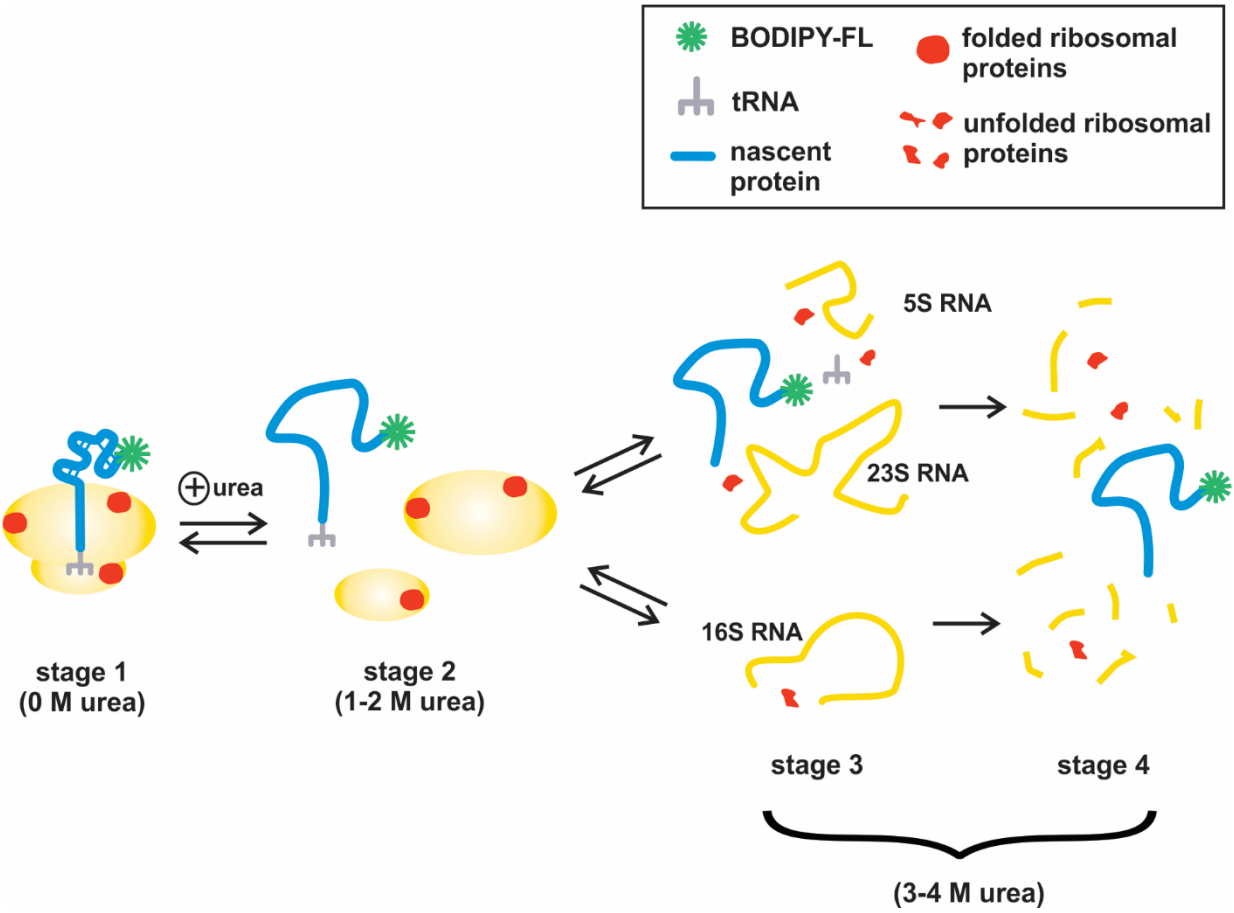


Figure 4-9. A working model for *E. coli* RNC disassembly in the presence of urea. The RNC progressively disassembles into 30S and 50S subunits at low urea concentrations (1-2 M). At ca. 3-4 M urea, the 70S is completely disassembled. Within the latter urea-concentration range, concerted disassembly and unfolding of the PTC and ribosomal proteins takes place, followed by RNA degradative processes, likely including ribonucleotide hydrolysis.

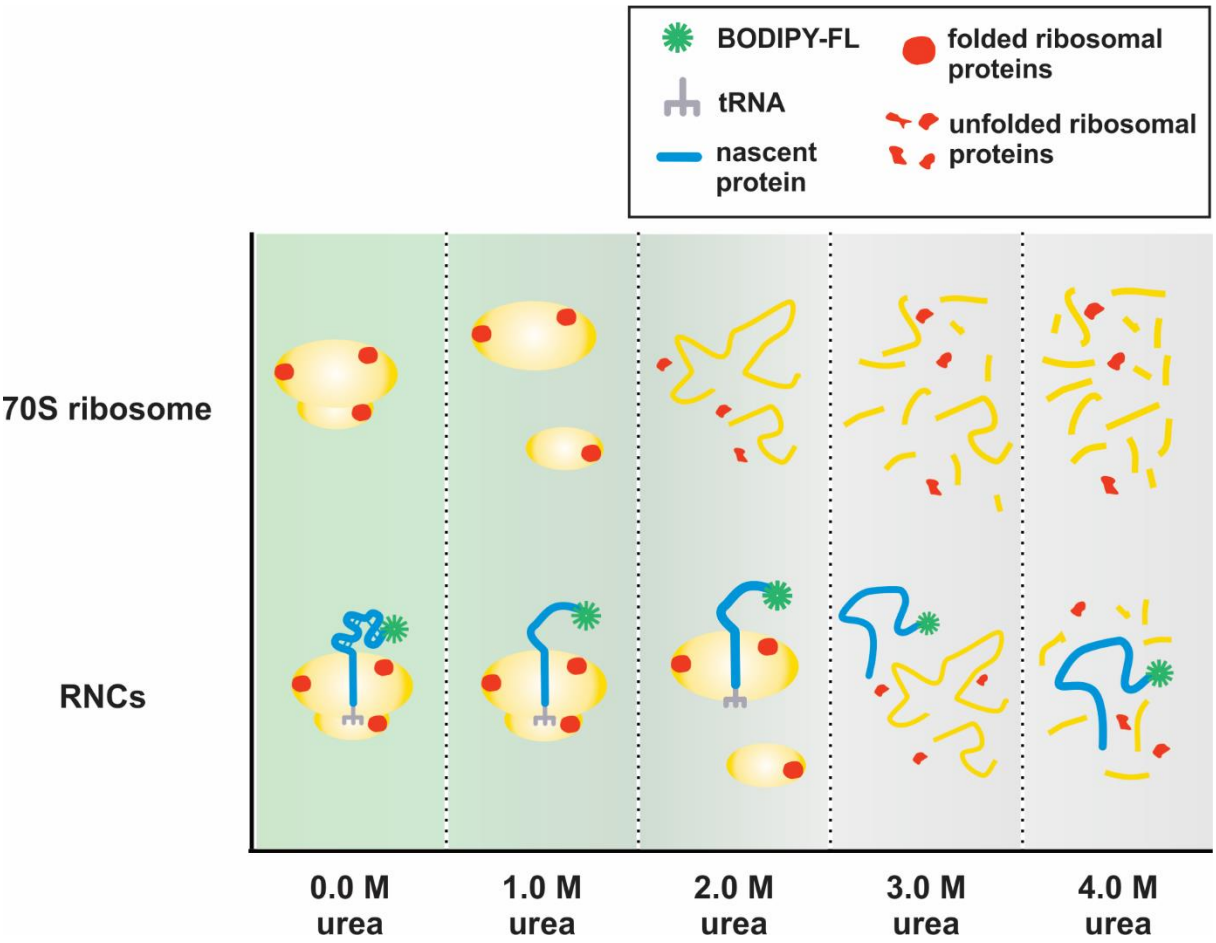


Figure 4-10. Summary of *E. coli* ribosome and RNC disassembly in the presence of urea. The 70S ribosome dissociates into the 50S and 30S subunits ca. 1.0 M urea. At 2.0 M urea, the ribosomal proteins and RNA unfold, which is followed by RNA degradation at urea concentrations ≥ 3.0 molar. In contrast, RNCs remain intact at 1.0 M urea, although the nascent chain begins to unfold. At 2.0 M urea the subunits dissociate. The ribosomal proteins and RNA unfold ca. 3.0 M urea followed be degradation at 4.0 M urea. The green to gray background represents the transition from 0.0 to 4.0 M urea, respectively.

PTC stability assessment by puromycin-assisted nascent-chain release assay

To confirm that the polypeptide was attached to the ribosome, low-pH SDS-PAGE[50] 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 ambient temperature. 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 70 °C for two 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 using Equation 1:

$$I_{\text{intergel}} = \frac{I_{\text{2.23 M urea,gel 1}}}{I_{\text{2.23 M urea,gel 2}}} * I_{\text{x,gel 2}} \quad (1)$$

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 an compare band intensities between the bound and released sample gels, Equation 2 was used:

$$I_{\text{relative}} = \frac{I_{+\text{puro}}}{I_{-\text{puro}}} \quad (2)$$

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 [37], the transition region slope and y-intercept were deduced to obtain the m-value and $\Delta G°_{H2O}$ with the Kaleidagraph software.[38]

Ribosomal protein stability by a tryptophan fluorescence emission assay

The *E. coli* ribosomal proteins collectively contain 32 tryptophan residues. Tryptophan 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.[51-53] 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.[54]

To generate a titration curve for the RNC complex, the spectral center of mass was obtained using the emission spectra between 300-385 nm and Equation 3

$$\text{Spectral center of mass (nm)} = \frac{\Sigma(\lambda) \times I_{\lambda}}{\Sigma(I)} \quad (3)$$

where λ is the wavelength and I_{λ} is the intensity at a specific wavelength.

The relative change in the spectral center of mass (ΔCoM) using Equation 4 was calculated using the wavelength values derived from Equation 3.

$$\Delta\text{CoM} = \frac{\lambda - \lambda_{fold}}{\lambda_{unfold} - \lambda_{fold}} \quad (4)$$

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. 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 using a known extrapolation method. [37-38]

4.6 Acknowledgments

We are thankful to A. Fuchs and M. Dalphin for useful discussions on sucrose gradient data collection and analysis, and for discussions on methods to assess apparent thermodynamic stability of RNCs. This work was funded by the National Science Foundation (NSF) grant MCB-1616459 (to S.C). A. E. V. received a National Science Foundation Graduate Research Fellowship Program and a Science and Medicine Graduate Research Scholars Fellowship from the University of Wisconsin-Madison.

4.7 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*, 589-597.

(3) Fedyukina, D. V.; Cavagnero, S., Protein folding at the exit tunnel. *Annu. Rev. Biophys.*, *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) Lu, J.; Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **2005**, *44*, 8230-8243.

(6) 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.

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

(8) 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.

(9) 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.

(10) 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.

(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) 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.

(15) Lu, J.; Hua, Z.; Kobertz, W. R.; Deutsch, C. Nascent peptide side chains induce rearrangements in distinct locations of the ribosomal tunnel. *J. Mol. Biol.* **2011**, *411*, 499-510.

(16) Lu, J.; Deutsch, C. Regional discrimination and propagation of local rearrangements along the ribosomal exit tunnel. *J. Mol. Biol.* **2014**, *426*, 4061-4073.

(17) 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*, 33697-33707.

(18) 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.

(19) Sohmen, D.; Chiba, S.; Shimokawa-Chiba, N.; Innis, C. A.; Berninghausen, O.; Beckmann, R.; Ito, K.; Wilson, D. N. Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. *Nat. Comm.* **2015**, *6*, 6941.

(20) 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*, 138-146.

(21) 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.

(22) 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.

(23) 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.

(24) Choi, K. M.; Atkins, J. F.; Gesteland, R. F.; Brimacombe, R. Flexibility of the nascent polypeptide chain within the ribosome: Contacts from the peptide N-terminus to a specific region of the 30S subunit. *Eur. J. Biochem.* **1998**, *255*, 409-413.

(25) 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.

(26) Eisner, G.; Koch, H.; Beck, K.; Brunner, J.; Muller, M., Ligand crowding at a nascent signal sequence. *J. Cell Biol.* **2003**; *163*, 35-44.

(27) Ullers, R. S.; N.G. Houben, E.; Raine, A.; Hagen-Jongman, C.; 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) Owji, H.; Nezafat, N.; Negahdaripour, M.; Hajiebrahimi, A.; Ghasemi, Y. A comprehensive review of signal peptides: structure, roles, and applications. *Eur. J Cell Biol.* **2018**, *97*, 422-441.

(30) Deckert, A.; Waudby, C. A.; Wlodarski, T.; Wentink, A. S.; Wang, X.; Kirkpatrick, J. P.; Paton, J. F.; Camilloni, C.; Kukic, P.; Dobson, C. M. 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.

(31) 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.

(32) Cabrita, L. D.; Cassaignau, A. M. E.; Launay, H. M. M.; Waudby, C. A.; Wlodarski, T.;

Camilloni, C.; Karyadi, M.-E.; Robertson, A. L.; Wang, X.; Wentink, A. S., et al. A structural ensemble of a ribosome–nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* **2016**, *23*, 278-285.

(33) 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.

(34) Eun, Y. J.; Kurt, N.; Sekhar, A.; Cavagnero, S. Thermodynamic and kinetic characterization of apoHmpH, a fast-folding bacterial globin. *J. Mol. Biol.* **2008**, *376*, 879-897.

(35) Zhu, L.; Kurt, N.; Choi, J.; Lapidus, L. J.; Cavagnero, S. Sub-millisecond chain collapse of the Escherichia coli globin ApoHmpH. *J. Phys. Chem. B* **2013**, *117*, 7868-7877.

(36) 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.

(37) Santoro, M. M.; Bolen, D. W. Unfolding free-energy changes determined by the linear extrapolation method. 1. Unfolding of pheynlmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **1988**, *27*, 8063-8068.

(38) Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **1990**, *8*, 93-98.

(39) 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.

(40) Wohlgemuth, I.; Beringer, M.; Rodnina, M. V. Rapid peptide bond formation on isolated 50S ribosomal subunits. *EMBO Rep.* **2006**, *7*, 699-703.

(41) 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.

(42) Roberts, M. E.; Walker, I. Structural studies on Escherichia coli ribosomes: III. Denaturation and sedimentation of ribosomal subunits unfolded in urea. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **1970**, *199*, 184-193.

(43) 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.

(44) Mitra, K.; Schaffitzel, C.; Fabiola, F.; Chapman, M. S.; Ban, N.; Frank, J. Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol. Cell* **2006**, *22*, 533-543.

(45) 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.

(46) Voss, N.; Gerstein, M.; Steitz, T.; Moore, P. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **2006**, *360*, 893-906.

(47) 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.

(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) 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.

(50) 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.

(51) Beechem, J. M.; Brand, L. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem* **1985**, *54*, 43-71.

(52) Demchenko, A. P., *Ultraviolet spectroscopy of proteins*. Springer Science & Business Media: 2013.

(53) Weber, G. Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds. *Biochem. J.* **1960**, *75*, 335-345.

(54) Teale, F.; Weber, G. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.* **1957**, *65*, 476-482.