THE IMPACT OF THE SMALL HEAT SHOCK PROTEINS ON PROTEIN AGGREGATE STRUCTURE AND AGGREGATE DISASSEMBLY

A Dissertation

by

ANDREW GLENN ROTH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hays Rye
Committee Members,	Steve Lockless
	Thomas Meek
	Gregory Reinhart
Head of Department,	Josh Wand

May 2020

Major Subject: Biochemistry

ABSTRACT

Cells have evolved a set of highly conserved proteins known as chaperones to assist in cellular function and homeostasis by preventing protein misfolding and aggregation. However, under extended stress, some proteins still misfold and aggregate, resulting in diseases such as cardiomyopathies and numerous neuropathies. The small heat shock proteins (sHsps) are a class of chaperones capable of inhibiting aggregation and assisting with protein disaggregation. However, the mechanisms by which the sHsps carry out these functions are poorly understood. In order to address this problem, we have employed the *E. coli* sHsp, IbpA and IbpB (IbpAB) as a model system, along with an established aggregation-prone substrate RuBisCO (from *R. Rubrum)*. We have found that RuBisCO forms either amorphous or fibril-like aggregates under slightly different conditions. Because standard bulk assays cannot be used to characterize the state distribution of protein aggregates in detail, we have employed a single particle fluorescence technique known as Burst Analysis Spectroscopy (BAS), which permits the minimally perturbative, free-solution observation of aggregate nano-particle size distributions. Using BAS, we observe that IbpAB can limit aggregate particle growth to a strikingly similar but limited size range, which is independent of the aggregation pathway. Using a two-color variant of BAS (MC-BAS), we show that IbpAB displays distinctive binding patterns when interacting with structurally different RuBisCO aggregate particles, suggesting IbpAB can sense and react to specific aggregate characteristics. Using ensemble inter- and intra-molecular FRET assays, we show that the binding of IbpAB to RuBisCO aggregates alters both the average relative proximity of different RuBisCO monomers within an aggregate, as well as the average conformation of the RuBisCO monomer itself. Time-resolved BAS measurements of aggregate disassembly in the presence of the *E. coli* bi-chaperone disaggregase, consisting of DnaK, DnaJ, GrpE, and ClpB (KJE/B), demonstrate that (1) IbpAB dramatically enhances RuBisCO aggregate disassembly and (2) that IbpAB release from aggregate particles is co-incident with disassembly of the aggregate itself. Additionally, IbpAB can inhibit aggregate growth of an aggregating RuBisCO sample depleted in the non-native monomer pool. However, unlike IbpAB inhibition at the beginning of an aggregating sample, late inhibition does not lead to stimulated disaggregation by the KJE/B disaggregase. Overall, these observations are most consistent with a model in which IbpAB bind to early aggregating states of RuBisCO, incorporate into the nascent aggregate particle and directly alter the conformational properties of the aggregate so that both ongoing particle growth is blocked and subsequent aggregate disassembly is enhanced.

DEDICATION

To my parents, sisters, and partner.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Hays Rye, and my committee members, Dr. Reinhart, Dr. Meek, and Dr. Lockless, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University manageable.

Finally, thanks to my mother and father for their encouragement and to my partner for her patience.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Hays Rye, Professor Gregory Reinhart, and Professor Thomas Meek of the Department of Biochemistry and Biophysics and Professor Steve Lockless of the Department of Biology.

The data and analysis for Chapter II was generated in part by Dr. Daniel Shoup. The data for Appendix A was generated in part by Dr. Jeremy Weaver and Mingqui Jiang. The data for Appendix B was generated in part by Dr. Mohsin M. Naqvi, Mario J. Avellaneda, Eline J. Koers, and Vanda Sunderlikova, and Günter Kramer

All other work conducted for the dissertation was completed by the student independently.

This work was made possible by N.I.H. under R01-GM114405

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

History of Protein Folding

Over evolutionary time, thousands of biological functions have developed requiring a wide variety of protein structures. Even with an extensive range of different protein structures, every protein begins as an unstructured amino acid chain before adopting the native fold. The process by which a protein adopts its functional structure from an unstructured polypeptide chain is called protein folding.

Understanding how proteins fold begins with identifying their elementary unit. The idea that proteins were made from elementary units of amino acids, which are linked in a chain called a peptide, was first proposed by Franz Hofmeister and Emil Fischer in 1902 and was referred to as the peptide theory (1). However, at that time peptide theory had little experimental evidence demonstrating how these assembled molecules could form a native protein. In the absence of evidence, peptide theory remained contested until the 1950's when Frederick Sanger developed a technique to determine the sequence of amino acids in the protein insulin (2, 3). With this technique, it was discovered that each protein has a unique amino acid sequence which did not match other proteins. With the ability to identify the amino acid sequence of an individual protein, the next unknown was how protein sequences were able to adopt stable, functional structures. Sanger highlighted this in his work by stating,

> “It would thus seem that no general conclusions can be drawn from these results concerning the general principles which govern the arrangement of the amino-acid residues in protein chains. In fact, it would seem more probable that there are no such principles, but that each protein has its own unique arrangement; an arrangement which endows it with its particular properties and specificities and fits it for the function that it performs in nature” (3).

Before Sanger’s breakthrough, it was known that proteins lost their function when denatured, either chemically or through heat, and that protein function could be restored upon gradual return to a native environment (1, 4, 5). These observations suggested that loss of protein function under specific experimental conditions altered protein stability and affected protein structure. Even so, there was little agreement on how the active protein is structured because peptide theory was not experimentally supported.

After Sanger’s breakthrough, the peptide theory was used as a starting point for understanding how proteins reached their functional or native structure. Even before the first protein structure was solved, Linus Pauling and Robert Corey predicted that polypeptides fold into α-helices and β-sheets (6, 7). They made these predictions based on X-ray crystallography structures of amino acids along with predicted constraints from the amide bond between amino acids in a polypeptide chain (6). Eventually, their theoretical models were supported by experimental evidence when the first protein structure, myoglobin, was solved in 1959 (8). The myoglobin structure displayed amide bonds between amino acids forming a polypeptide chain which then adopted an α-helix stabilized by hydrogen bonding. Following the myoglobin structure, Pauling and Corey's prediction of a β-sheet fold was supported when the structure of lysozyme was solved, giving even more credibility to Pauling and Corey's theories (9). With the secondary structures of proteins now solidified as α-helices and β-sheets, a basic understanding was known for how hydrogen bonding leads to stabilized proteins.

After Pauling and Corey's work, many researchers believed hydrogen bonding was the answer to protein stability and the only force dictating protein folding. However, free energy measurements of protein folding and unfolding told a different story signifying the need for additional stabilization forces (10, 11). Some believed additional stabilization was generated through hydrogen bonding between amino acid R groups, along with the peptide backbone, but this was not the case. Eventually, free energy requirements were met through the discovery of tertiary stabilization forces such as disulfide bonds, van deer Waals interactions, and the hydrophobic effect (11-13).

Out of all the additional stabilization forces discovered, the hydrophobic effect became a main focus due to the influence on the protein folding pathway. The hydrophobic effect helps drive protein folding to a more energetically favorable conformation due to the increased presence of hydrophobic amino acids in the center of a folded globular protein. However, in the late 1900's, the hydrophobic effect was only known to assist with protein stability and contribute to the free energy of a native protein.

The first protein folding models were heavily influenced by the peptide theory because of unknown effects from free energy and other stabilizing forces. These early protein folding models hypothesized that a unique amino acid sequence coded for a specific native protein structure (1). Therefore, a change in the amino acid sequence caused a change in the protein fold. Christian Anfinsen was a pioneer in testing early protein folding models. He utilized the protein ribonuclease from *B. Taurus* to understand the relationship between the amino acid sequence, the protein folding pathway, and the native protein structure (14-19).

Contrary to the predictions of early proposed protein folding models, Anfinsen demonstrated that the amino acid sequence alone did not dictate the protein folding rate or final fold. Instead, protein folding appeared to be driven by the protein's search for the lowest Gibbs free energy state, an idea that came to be known as the thermodynamic hypothesis (12, 15). To reach this conclusion, Anfinsen mutated or removed certain amino acids within the N- and C- termini of ribonuclease and exhibited that the protein could still fold into a native state at the same folding rate (20). The ability to shuffle or remove amino acids at certain positions was evidence that each amino acid sequence did not code for a specific native protein structure.

Though mutating and removing certain amino acids in ribonuclease did not cause changes in the native fold, Anfinsen did discover that conserved amino acid sequences, or motifs, within each protein were vital for proper protein folding. These conserved motifs were essential for structural features, like disulfide bonds, that stabilized the folded state or maintained a needed geometric formation (14, 17, 19). Along with this, motifs were also conserved for amino acids used in the active site of a protein. Through his work with ribonuclease, Anfinsen revealed the delicate balance between amino acid sequence, structure, and free energy. All three factors playing a key role in a protein's search for the native state.

Before the thermodynamic hypothesis, mechanistic models describing a protein folding pathway were difficult to solidify. This difficulty stemmed from the vast difference in the protein folding rate measured under biological conditions versus the theoretical protein folding rate. Under biological conditions, proteins were able to fold into a native state in the range of seconds to sub-second. However, with models based on peptide theory, the theoretical rate for a 100 amino acid protein would take up to 10^{50} years to fold (21). This theoretical rate was based on a protein performing random sampling of all the possible folded states before finding the correct fold and came to be known as Levinthal's paradox (22).

Following the predictions of the thermodynamic hypothesis, two models were proposed as a way to solve Levinthal's paradox. Each model consisted of a rate limiting step in which a protein samples a small number of possibilities before a rapid collapse of the entire protein into a low energy native state. The rate limiting step was thought to be a non-native fold on the pathway to the native state, therefore referred to as an intermediate state (23-26). The first model to utilize an intermediate state was called the "random search nucleation and propagation" model. This model proposed that a small portion, or domain, of a protein would randomly searched for a stable free energy state. Once the stable state was found, it served as a starting point for subsequent polypeptide chain collapse into the native state (26, 27). The second model was called the "diffusion-collision" or framework model. This model proposed that a protein was divided into different parts, or micro-domains, which folded on their own. These micro-domains would then collide into the native conformation (26, 27). While both models required a specific domain, or domains, that needed to fold before the native fold was accomplished, the nucleation model followed a sequential order where one domain folded first, followed by collapse to the native state; while the diffusion-collision model had multiple domains folding in tandem, followed by the domains coming together to form the native protein.

Following proposal of the nucleation model and the diffusion-collision model, a main area of interest was characterizing possible intermediate states. Over time, many proteins were found to form partially stable structural intermediates (27-29). These stable intermediates were nonnative protein folds that, in some way, matched the final native fold.

Since intermediate structures seemed to be somewhat stable, it was probable a protein followed a unique and definite sequence of intermediate steps to reach the native state. Each step would be determined by a free energy minima and subsequent activation energy barrier that the protein must overcome in order to continue to the native fold (29). This came to be known as the protein folding pathway and suggested protein folding rates would be dictated by free energy barriers along the pathway before a protein reached the native protein fold (30).

While early protein folding pathway models consisted of defined sequential steps along a pathway, the hypothesis soon expanded to become a three dimensional free energy landscape where multiple intermediate states were possible for each native protein fold (21, 29-31). This energy landscape is represented by a funnel where numerous protein conformations are possible at the same free energy level (Figure I.1). The top of the funnel, where the opening is the largest, represented the higher energy states which contained a large array of folding possibilities; while the bottom of the funnel, where the opening is the smallest, represented the lower energy states that have fewer conformations available (30). Another vital part of the free energy landscape is the presence of traps and holes where stable and semi-stable intermediate states exist. This makes a rigid free energy landscape as a protein folds to the lowest-energy state (31).

A generally accepted, modern definition of a protein folding intermediate is, "a metastable state that is transiently populated along the protein folding energy landscape" (32). Protein intermediate stabilization differs depending on the local free energy

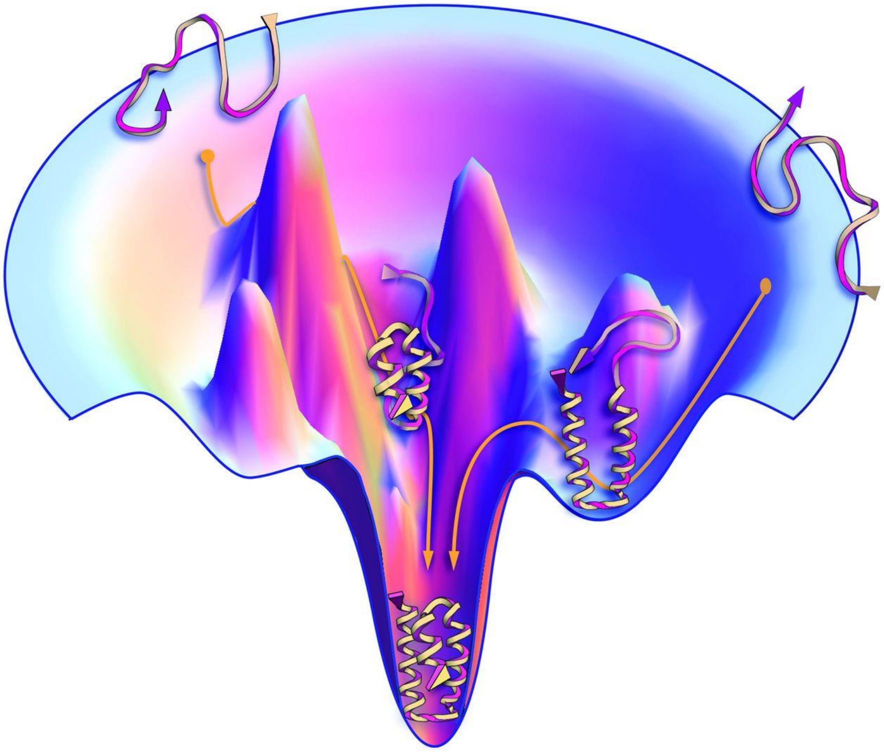


Figure I.1 Protein Folding Energy Landscape.
Proteins have a funnel shaped energy landscape with many high energy unfolded conformations and few low energy folded conformations. Local minima and activation barriers create folding intermediates throughout the energy landscape.
Reprinted with permission from "The Protein-Folding Problem, 50 Years On" by Ken A. Dill and Justin L. MacCallum, 2012, Science, 338, p. 1042-1046, Copyright [2012] by The American Association for the Advancement of Science

minima and transition state barriers. In turn, these energetic barriers dictate the time a folding intermediate state exists, or is populated. Some intermediates are undetectable, while others are populated for an extended period of time. As protein folding proceeds down the energy landscape, problems can arise when folding intermediates encounter a deep energetic local minima from which it cannot easily escape.

Protein Folding and Disease

Early in his career, Anfinsen observed how protein folding was influenced by the folding environment and mutations made in the amino acid sequence (14-18). Still, it was unclear exactly how the environment and mutations could positively or negatively impact protein folding. However, once the concepts of folding intermediates and the free energy landscape were solidified, it became clear that the folding environment and mutations could affect protein folding through altering intermediate formation or population time. If these conditions negatively impacted the intermediate state or population time, a highly stable intermediate state would form and the protein would not be able to continue to the native fold. These highly stable intermediate states are referred to as kinetically trapped intermediates (33-38) (Figure I.2).

A well-known example where mutations induce the formation of a kinetically trapped intermediate is in the cystic fibrosis transmembrane conductance regulator protein (CFTR). CFTR is a chloride channel that plays important roles in multiple tissues throughout the human body. Over 1000 mutations are known to induce CFTR to form non-native or misfolded states (39). The most common CFTR mutations cause a kinetically trapped intermediate to form that leads to the formation a highly stable misfolded state. The misfolded state is rapidly destroyed by the cellular protein quality control systems which causes the loss of a vital chloride channel and leads to severe forms of cystic fibrosis (40, 41).

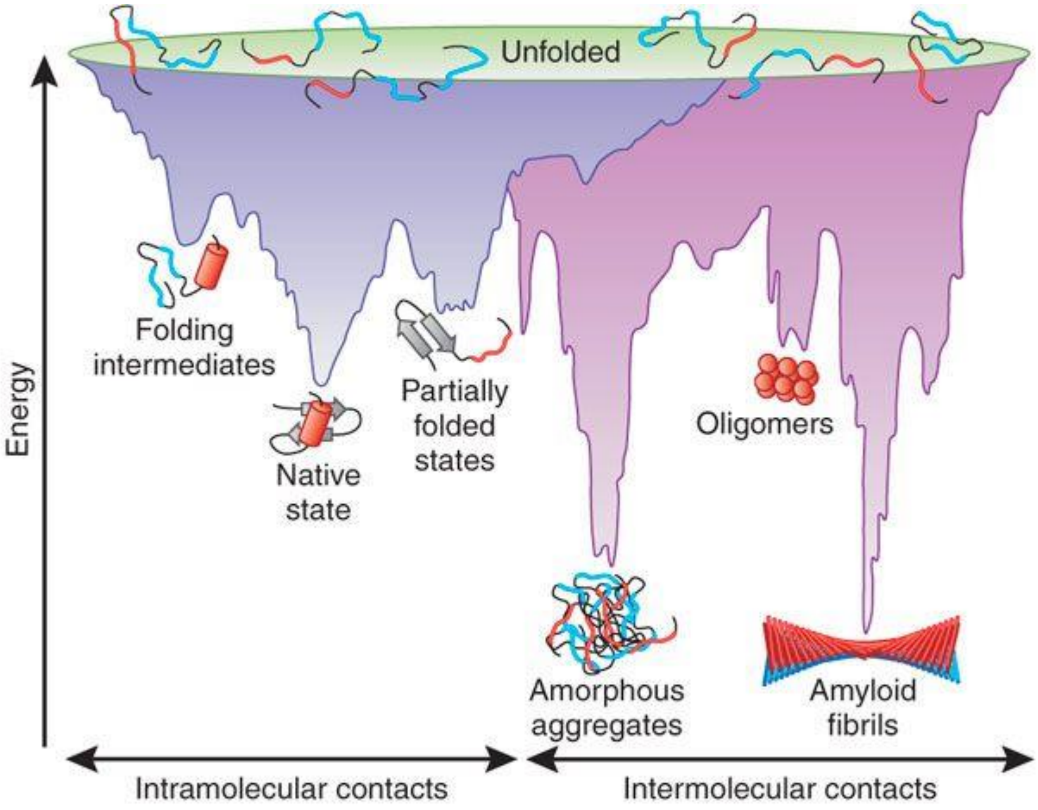


Figure I.2 Energy Landscape of Protein Folding and Aggregation.
High energy unfolded proteins have many folding outcomes with partially folded intermediates capable of folding into a native monomer or forming intermolecular contacts to form protein aggregates.
Reprinted with permission from “Converging Concepts of Protein Folding in vitro and in vivo” by F Ulrich Hartl and Manajit Hayer-Hartl, 2009, Nature Structural and Molecular Biology, 16, p. 574-81, Copyright [2009] by Springer Nature

While CFTR mutations are examples of kinetically trapped intermediates that lead to misfolded protein monomers, the majority of kinetically trapped intermediates do not remain monomers. Instead, these intermediates or misfolded proteins interact with each other to form protein aggregates. An accepted definition of protein aggregation is, “the association of two or more non-native protein molecules that are largely driven by hydrophobic forces and primarily result in the formation of either amorphous or fibril-like aggregates” (42-44). Without assistance, these aggregates are unable to disassemble or refold into a native state on a biologically relevant timescale (Figure I.2). In this work, we specifically distinguish between non-native aggregates and higher order states of native proteins when needed. An aggregated state does not refer to native proteins that form oligomers or become insoluble due to unnaturally high concentrations of the native protein.

It is important to distinguish between two general types of protein aggregates: amorphous and fibril-like. Amorphous aggregates are thought to primarily form through exposed hydrophobic residues from unstructured secondary and tertiary folds occurring within each protein monomer. Amorphous aggregates do not have a consistent structure or type of interaction making their overall aggregate structure irregular and heterogeneous (45). On the other hand, fibril-like aggregates have exposed hydrophobic patches originating from a highly ordered misfolded state. These misfolded states generally contain a β-strand fold very similar to a β-sheet. Through the formation of the β-strand fold, fibril-like aggregates can then interact with each other to form an ordered multimeric structure called a cross-β scaffold. From this scaffold, propagation, or addition of other subunits, occurs in a defined fashion as fibril-like aggregates grow. Fibril-like aggregates that continue to grow into large particles are often referred to as fibers or amyloids (46, 47) (Figure I.3).

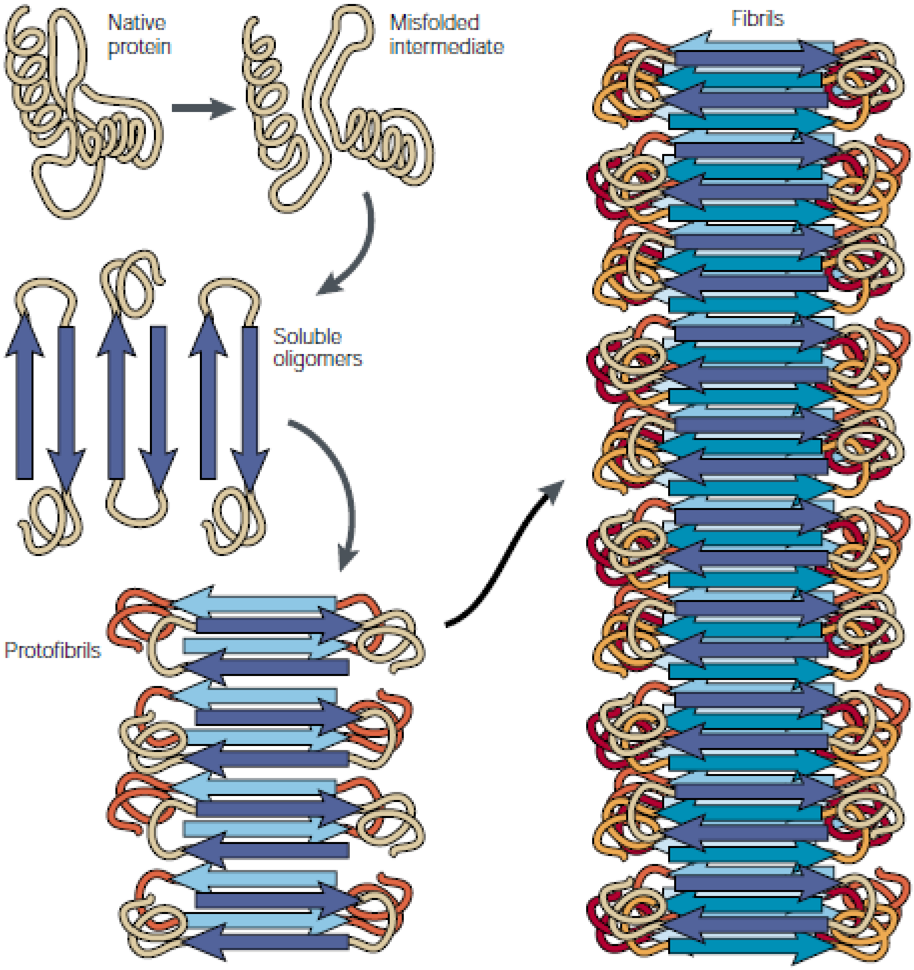


Figure I.3 Formation of Protein Fibrils.
Under certain conditions, native proteins can form misfolded intermediates. These intermediates can aggregate and form an oligomeric β-sheet structure. Additional β-sheet oligomers can interact with each other in an ordered manner to form fibils.
Reprinted with permission from “Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases” by Claudio Soto, 2003, Neuroscience, 4, p. 49-60, Copyright [2003] by Springer Nature

Protein aggregate toxicity depends on a multitude of factors ranging from environmental conditions to the exact misfolded state a protein adopts (48). For example, living *E. coli* can form large intracellular deposits of mostly aggregated proteins called inclusion bodies. When *E. coli* is exposed to a temporary elevation in temperature, the inclusion bodies typically increase in size as more proteins misfold (49). However, under these conditions the bacteria do not die, and once they are returned to an ideal temperature the inclusion bodies decrease in size with the bacteria returning to its normal physiologic state. This simple example shows that while protein aggregation is pervasive, cells have developed efficient responses that prevent aggregation from being immediately deadly.

By contrast, some protein aggregates can be extremely toxic, even under ideal environmental conditions. Aggregate toxicity occurs through two mechanisms. The first mechanism is through direct harm from the aggregates themselves, referred to as gain of function aggregates. The second mechanism is through the inability of the proteins, or molecules, within the aggregates to carry out their intended function, and is referred to as loss of function aggregates (47, 48, 50).

A highly studied gain of function protein aggregate is a misfolded version of the prion protein (PrP). PrP is capable of folding in two different ways, Prp^{C} or PrP^{sc} (51). PrP^{C} is the native conformation that associates with membranes, though the proposed function of this protein is still disputed (52-54). PrP^{sc}, on the other hand, is a misfolded state of PrP that contains the amyloid cross-β scaffold structure. PrP^{sc} is very stable and is resistant to degradation both *in vitro* and *in vivo*. This characteristic distinguishes PrP^{sc} from PrP^{C} as PrP^{C} is not resistant to degradation. PrP^{sc} can also catalyze the conversion of PrP^{C} into PrP^{sc}, causing an amplifying cycle of subunit formation for further amyloid growth; a process called seeding. Through seeding, PrP^{sc} forms a type of amyloid known as a prion that lead to the breakdown of the nervous system and causes Creutzfeldt-Jakob disease (51).

PrP is not the only protein with toxic traits stemming from a misfolded state. Many amyloid forming proteins associated with neurodegenerative diseases like amyloid-β, tau, and α-synuclein share similar behaviors with PrP, where misfolded states form amyloid structures (47, 52, 55-58). Amyloid fibers are also generally very stable and can resist degradation.

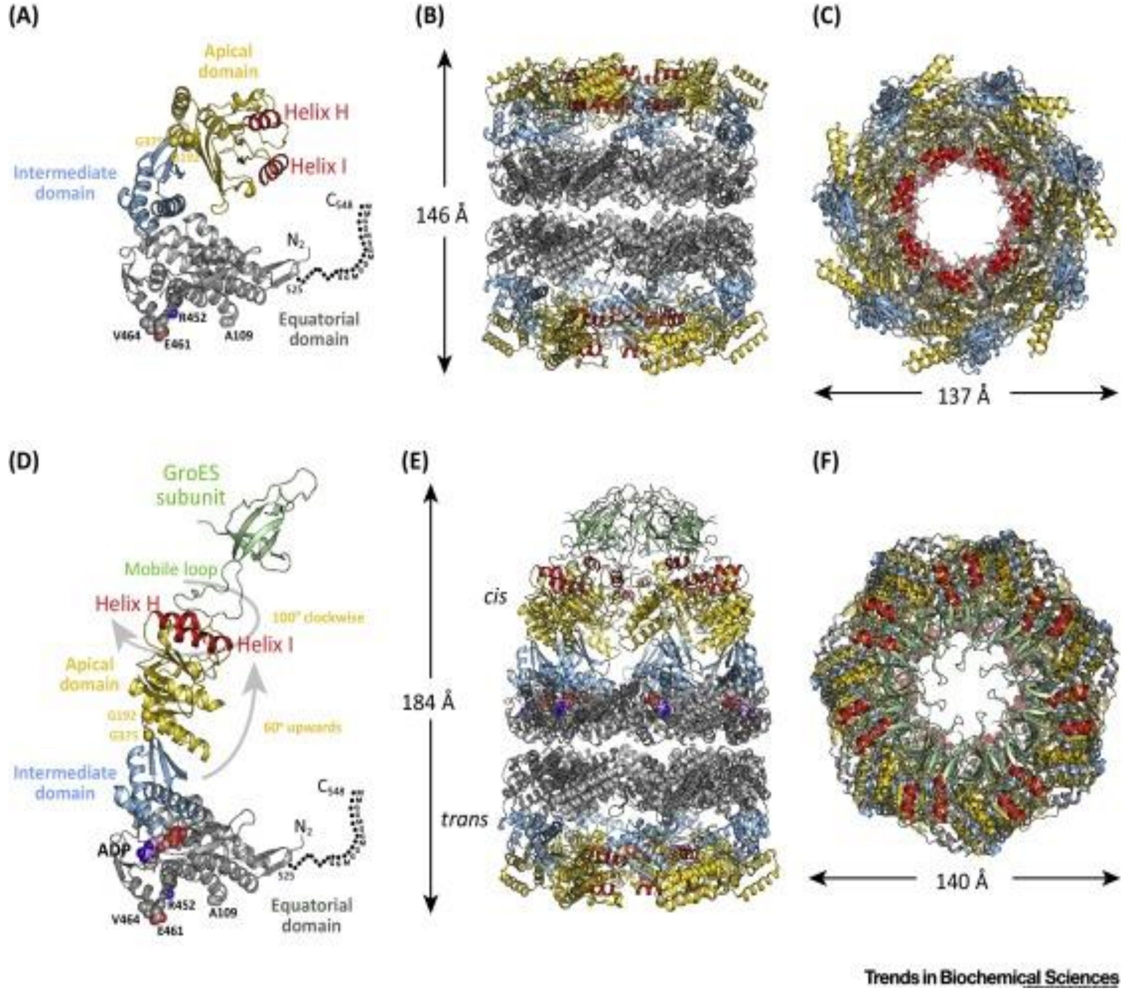
Even though protein misfolding and aggregation are the central cause for many debilitating diseases, most aggregates do not have a negative impact on cellular homeostasis under normal conditions. A cells' ability to remove protein aggregates derives in part from a group of regulatory protein families which assist in protein folding, aggregate recognition, and clearance. These families are known as the protein quality control (PQC) network. The action of the PQC network is essential for maintaining cellular proteostasis. There are two proteostasis networks involved with misfolded and aggregated protein removal; the first is the protein degradation network and the second is the aggregate disassembly and refolding network (59-63). The research in this dissertation focuses on the aggregate disassembly and refolding network, with a strong interest in the mechanisms involved with protein disassembly.

Protein Chaperones and Proteostasis

Protein folding *in vivo* is made complicated by a number of cellular conditions, including high protein concentrations, oxidative stress, post translational modifications, and changes in temperature. To address these issues, a set of specialized protein machines evolved to assist in the protein folding process. These machines are generally referred to as protein/molecular chaperones and were originally named for their ability to prevent “inappropriate interactions” between non-native proteins (64, 65).

Chaperones can also carry out additional roles including oxidative stress relief, protein transport across a membrane, and protein aggregate disassembly (65-68). Chaperones are ubiquitous, found in all organisms from single-celled bacteria to multi-celled organisms. Individual chaperones have also evolved to perform highly specific functions in different environments and organisms.

In this dissertation work, four different chaperone families are being utilized to understand the mechanisms of protein refolding, aggregate inhibition, and aggregate disassembly. These families are the Hsp60, Hsp70, Hsp100, and the small heat shock proteins.



Trends in Biochemical Sciences

Figure I.4 Crystal Structure of GroEL and GroEL-GroES Complex.
(A) The GroEL monomeric structure with the apical, intermediate, and equatorial domains in yellow, blue, and gray respectively. (B) Side view of the GroEL tetradecamer. (C) Top view of the GroEL tetradecamer. (D) GroEL and GroES subunit structures (E) Size view of the GroEL-GroES complex. (F) Top view of the GroEL-GroES complex.
Reprinted with permission from “The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding” by Manajit Hayer-Hartl, Andreas Bracher, and F. Ulrich Hartl, 2016, Trends in Biochemical Sciences, 41, p. 62-76, Copyright [2016] by Elsevier

The Hsp60 Family

The most well-known chaperone family is the Hsp60, or chaperonins, with the best understood family member being the GroEL-ES system of *E. coli* (69-76). The overall function of the chaperonins is to mediate proper folding of a monomeric protein substrate. Chaperonins are ATP driven protein folding machines, with the notable feature of two, seven subunit rings (77) (Figure I.4). Each ring contains a solvent-filled cavity, which can sequester up to a 60 kDa protein substrate.

Each GroEL subunit consists of three domains. The upper domain, known as the apical domain, is used to capture the substrate (72) (cite). The apical domain uses hydrophobic residues to interact with the non-native folding intermediate, and also binds to the dome shaped GroES heptamer to create an enclosed cavity for productive folding of the protein substrate. The domain below the apical domain is referred to as the intermediate domain which couples conformational changes between the apical and equatorial domain. The third GroEL domain below the intermediate domain is referred to as the equatorial domain (Figure I.4). The equatorial domain contains the ATP binding site for the GroEL machine. It is also responsible for important inter- and intra-EL ring contacts that hold the two ring complex together. These contacts between the two GroEL rings are vital for chaperonin function because the folding cycle depends on an ATP-driven sequence of protein capture, encapsulation, and release occurring on one ring at a time (75) (Figure I.5).

While many elements of the basic chaperonin functional cycle are now well established, precisely how these nanomachines productively facilitate protein folding has remained controversial. The GroEL functional cycle begins with a non-native folding

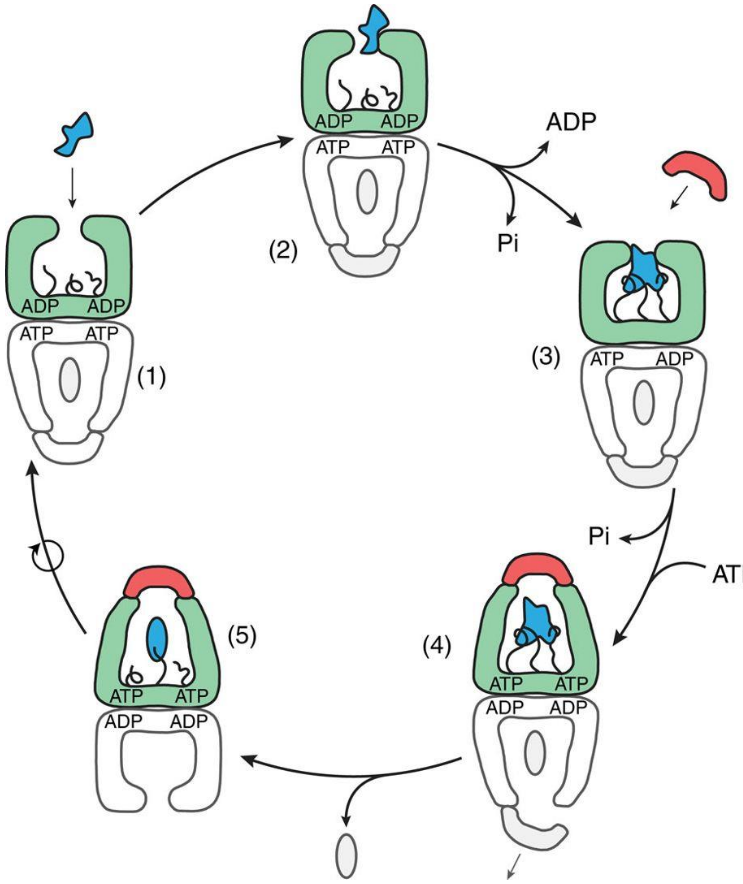


Figure I.5 Schematic of the Hypothesized GroEL/ES Cycle.
(1-2) Non-native substrate binds to the open apical domains of the GroEL ring and stimulates ADP release. (3) The C-terminal tail interact with the substrate protein and pull it deeper into the cavity. (4) GroES binds to the apical domain to create a closed cavity and cause forced unfolding and compaction of the substrate protein. This also causes the removal of the GroES on the opposite ring leading to substrate release. (5) The opposite ring is now primed for substrate binding and GroES capture leading to ATP hydrolysis.
Reprinted with permission from "GroEL actively stimulates folding of the endogenous substrate protein PepQ" by Jeremy Weaver, Mengqiu Jiang, Andrew Roth, Jason Puchalla, Junjie Zhang & Hays S. Rye, 2017, Nature Communications, 8, Copyright [2017] by Springer Nature

intermediate being captured by the apical domains of an open GroEL ring (78) (Figure I.5). Subsequent binding of ATP to the GroEL equatorial domains, followed by association of the GroEL co-chaperonin GroES to the apical domains, results in capture and enclosure of the substrate protein within the enclosed GroEL-GroES cavity. Encapsulation results in release of the substrate protein into the isolated GroEL-GroES chamber, which triggers the initiation of protein folding. Subsequent ATP hydrolysis within the GroEL-GroES cavity results in both a weakening of the cavity stability and communication of an allosteric signal across the tetradecamer ring-ring interface that permits ATP binding to the second ring. Binding of ATP to the second ring triggers decay of the spent folding cavity on the first ring, causing release of GroES, ADP and the substrate protein into free solution. If the open second ring has captured a folding intermediate, ATP binding on this ring, followed by GroES binding, results in assembly of a new folding cavity ready to initiate protein folding. The two GroEL rings alternate as folding cavities over the course of a full folding cycle. For protein molecules that have completed folding, or that have at least committed to the native state, their dependence on GroEL is compete. The folding intermediates of more uncooperative substrate proteins, generally referred to as “stringent proteins,” are recaptured for additional passages through the GroEL folding cycle (77, 79) (Figure I.5).

The literature suggests three possible mechanisms for how the GroEL-GroES cycle promotes proper protein folding: (1) isolation in the GroEL-ES folding cavity prevents aggregation and permits folding in the absence of second-order side reactions

(80), (2) unfolding of the substrate protein upon initial capture, driven both by multiple apical domain interactions in combination with C-terminal tail contacts helps disrupt kinetically stable misfolded states that would otherwise prevent folding, permitting the protein another chance at folding (81-83), and (3) confinement within the GroEL-GroES cavity alters the energetic and spatial constraints imposed on collapsed folding intermediates, resulting in folding dynamics that are distinct from and potentially faster than what is possible in free solution (77, 84).

Currently, a detailed study on the unstructured C-terminal tails of GroEL promotes the second model in which the misfolded substrate protein is actively unfolded upon initial capture and permitted another chance to fold within the GroEL-GroES cavity (82) (Appendix I). In this model, the C-terminal tails protrude upward into the cavity to interact with the misfolded substrate protein during initial substrate binding to the apical ring. During this initial binding process, the C-terminal tails pull the substrate further into the cavity causing the substrate protein to unfold. Once the ATP cycle begins, the substrate protein enters the folding cavity where it has the opportunity to fold in a secluded environment. Overall, this supports a mechanism where GroEL-GroES actively participate in the substrate folding pathway rather than simply providing an isolated environment.

The Hsp70 Network

While the Hsp60 chaperone family has specifically evolved to assist in productive monomeric protein folding, the Hsp70 chaperone family has evolved with a wider variety of functions (85). These functions include protein folding off of the ribosome, protein translocation across a membrane, and coordination of protein degradation machinery. However, for this work, the most important Hsp70 function is their ability to disassemble protein aggregates.

Hsp70 proteins are ATP driven machines that consist of two domains referred to as the ATPase domain and the substrate binding domain; each domain carries out the function their name implies (62) (Figure I.6). These domains are thought to be connected through a intrinsically disordered linker, through which the domains can communicate during substrate binding, ATP hydrolysis, and substrate release (85).

For the Hsp70 family, there are two important conformational states caused by ATP binding and hydrolysis (85). The first conformational state is cause by ATP binding to the ATP binding domain. This binding causes the substrate binding domain to have low substrate affinity and a high substrate sampling rate. The second conformational state is caused by ATP hydrolysis. During this process, the binding domain has high substrate affinity and a low substrate sampling rate. For the substrate binding affinity to change during each conformation, it is thought the binding domain acts like a clamp. In this model, ATP binding causes the clamp to open so the substrate protein can enter and

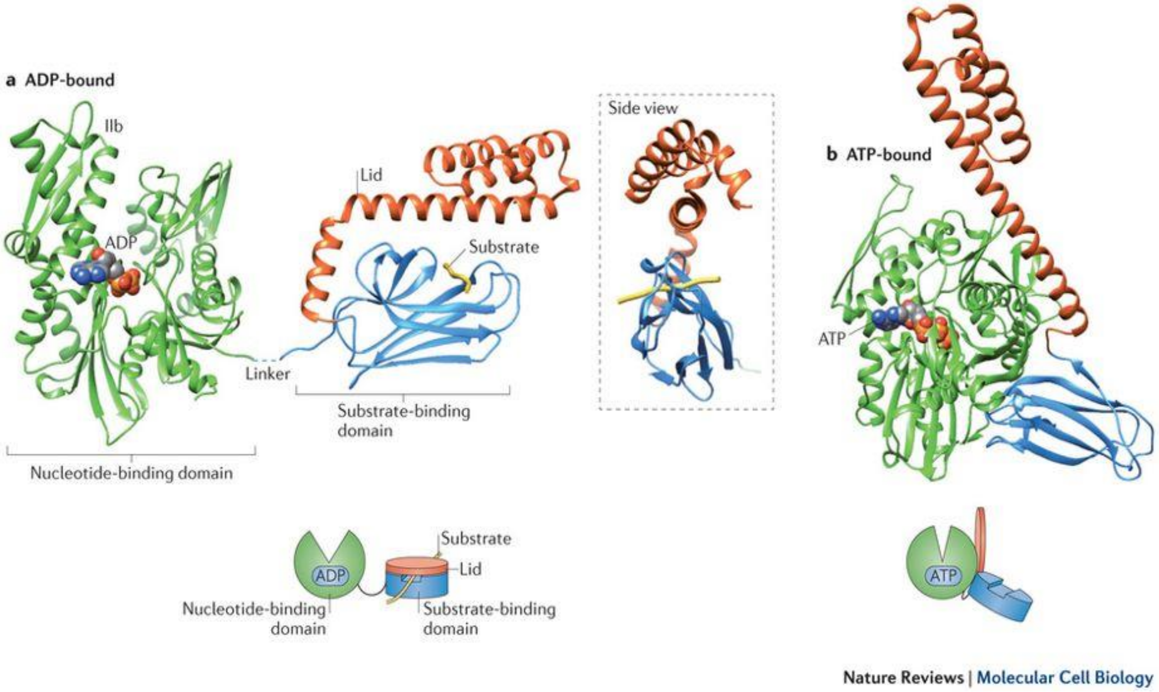


Figure I.6 Hsp70 Structure Bound to ATP and ADP.
Structure of the nucleotide binding domain, substrate binding domain, and lid in green, blue, and orange, respectively. (A) Hsp70 with ADP bound to the nucleotide binding domain and the substrate binding domain in the closed conformation. (B) Hsp70 with ATP bound to the nucleotide binding domain and the substrate binding domain in the open conformation.
Reprinted with permission from "Chaperone machines for protein folding, unfolding, and disaggregation" by Helen Saibil, 2013, Nature Reviews Molecular Cell Biology, 14, Copyright [2013] by Springer Nature

leave the domain at a fast rate (Figure I.6). Then, once ATP is hydrolyzed, the clamp closes trapping the substrate and creating the high substrate affinity (86-89). As the ATP cycle continues, the clamp opens and closes to assist with aggregate disassembly and protein substrate unfolding (Figure I.6).

ATP hydrolysis and substrate binding by Hsp70 alone are very inefficient (90). The lack of efficiency remained a mystery until the Hsp70 partner proteins, referred to as co-chaperones, were discovered. The co-chaperones increase the Hsp70 ATPase rate and substrate binding efficiency. The first co-chaperone is the Hsp40 family, also referred to as J-proteins due to the conserved family domain, the J-domain. Hsp40's are able to recognize, bind, and hold specific substrates at a high affinity until an Hsp70 is recruited (90). Through recruitment, Hsp70 associates with Hsp40 and brings the substrate into proximity of the substrate binding domain of Hsp70. By bringing the substrate and Hsp70 close, substrate transfer is stimulated and increases Hsp70 substrate binding efficiency (91). Once the substrate is secure in the binding domain, Hsp40 also stimulates Hsp70 ATP hydrolysis causing the clamp to close around the substrate (91).

Since Hsp40 stimulates Hsp70 substrate binding and ATP hydrolysis, the rate limiting step is removal of ADP from the Hsp70 ATP binding domain. To stimulate ADP removal, another co-chaperone is employed and is referred to as a nucleotide exchange factor (NEF). The NEF binds to the Hsp70 ATPase domain and causes a decrease in nucleotide affinity (85). This decrease in affinity stimulates the release of ADP from the ATPase domain and prepares the Hsp70 for another ATP cycle. With the NEF still bound to Hsp70, ATP binds to the empty Hsp70 ATPase domain causing the release of the NEF. This release opens the Hsp70 substrate binding domain to bind protein substrates or to be recruited by an Hsp40.

The Hsp100 Family

While the Hsp70 family is at the center of the proteostasis network, another molecular chaperone family, known as the Hsp100s, is needed to fully disassemble and unfold protein aggregates. Together, Hsp70 and Hsp100 form the primary prokaryotic disaggregation machinery referred to as the Hsp70/100 bi-chaperone system (63).

The Hsp100 proteins are ATP driven machines that completely unfold a misfolded or aggregated protein. They contain conserved AAAs domain which make them members of the AAA+ protein superfamily (62). AAA domains contain an ATP binding pocket and the ability to form homo-hexamers which, when assembled, contains a central pore through the complex. The most highly studied Hsp100 is ClpB from *E. coli*.

The central pore within the Hsp100 homo-hexameric complex is used to completely unfold a protein substrate and provides the opportunity for the substrate to refold correctly. The threading of a peptide chain is accomplished by conserved aromatic residues that line the central pore within each hexamer. Upon ATP hydrolysis, these aromatic residues complete a conformational change and pull the substrate through the central pore (63, 92). After ATP hydrolysis, ADP is released causing the residues to relax back to their original state and await the next cycle. While it was thought the AAA domains acted in unison, it is now hypothesized that the aromatic residues move in a sequential order around the hexameric ring (63). By using a sequential mechanism, the substrate protein cannot be released within the pore and travel in the wrong direction (92) (Figure I.7).

Hsp100 and Hsp70 appear to use a synergistic method of substrate transfer to maximize disaggregation efficiency (93). Similar to how Hsp70 and Hsp40 associate to

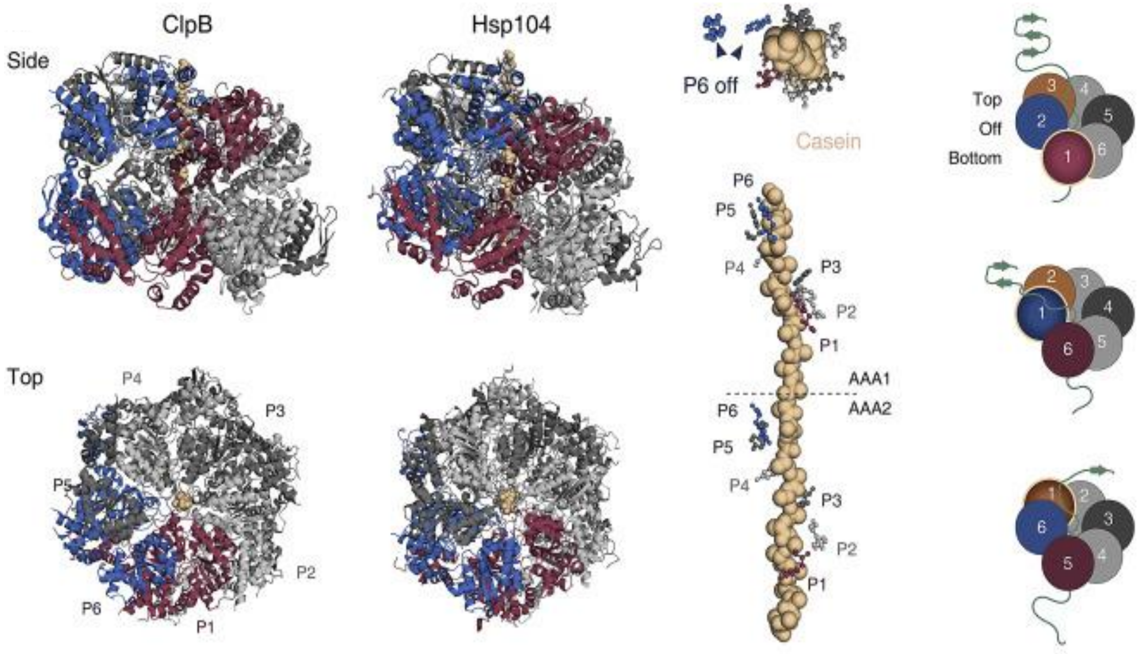


Figure I.7 Hsp100 with a Bound Substrate.
Left: side and top views of ClpB and Hsp104 bound to the substrate casein. Middle: Casein interacting with the conserved hydrophobic residues. Right: A model for substrate movement through the AAA hexamer pore as the subunits change their conformation around the ring.
Reprinted with permission from "Cellular Handling of Protein Aggregates by Disaggregation Machines" by Axel Mogk, Bernd Bukau, Harm Kampinga, 2018, Molecular Cell, 69, Copyright [2018] by Elsevier

transfer substrates for aggregate disassembly, Hsp100 associates with Hsp70 to receive a substrate protein for threading through the central pore. Because of the energy requirements of the threading process, substrate transfer from Hsp70 to an Hsp100 is highly regulated, using two mechanisms. The first mechanism utilizes the flexible N-termini of each Hsp100 subunit. Without ATP or Hsp70 interaction, all six Hsp100 N-termini cover the central pore to inhibit molecules from entering (85). The second regulatory mechanism is a two state conformational change occurring in the middle domain, a domain that wraps around the AAA domains (Figure I.8). The first conformational state promotes inactivity and is referred to as the "locked and ordered" conformation. The second conformational state is referred to as the "titled" conformation. The tilted conformation promotes activity and is activated by ATP binding and hydrolysis. Each conformation is also allosterically connected to the N-

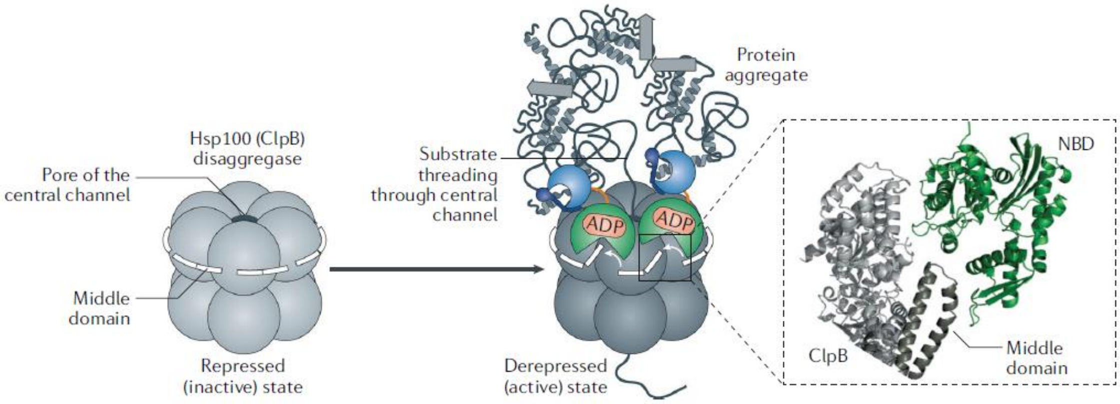


Figure I.8 A Schematic for Hsp70 and Hsp100 Association and Substrate Transfer. Hsp70 recruits Hsp100 to the aggregated substrate protein and binds to the AAA domain in the tilted conformation. Substrate is handed from Hsp70 to Hsp100 which then threads the substrate through the central pore.
Reprinted with permission from "The Hsp70 Chaperone Network" by Rina Rosenzweig, 2019, Nature Reviews Molecular Cell Biology, Copyright [2019] by Springer Nature

termini at the top of each Hsp100 hexamer. The locked conformation is associated with plugging of the central pore by the N-termini, while the titled conformation is associated with exposure of the pore upon displacement of the N-termini (92).

The conformation of the middle domain also changes when Hsp100 associates with Hsp70. Similar to ATP binding, the tilted conformation promotes Hsp70 binding while the locked conformation does not. With these regulatory mechanisms in place, Hsp100 activity is predicted to begin with Hsp70, carrying a substrate, recruiting and binding to a middle domain in the titled conformation. Through this association, the substrate is transferred from Hsp70 to Hsp100 with assistance from the Hsp100 N-termini at the top of the pore. This leads to substrate release from Hsp70 and final movement to the top of the Hsp100 pore. Once the handoff occurs, the middle domain changes to the rigid/locked conformation to promote Hsp70 disassociation. Once Hsp70 disassociation occurs, the substrate is pulled through the central pore as ATP hydrolysis causes the hydrophobic residues to change conformations and thread the substrate through the central pore (85, 93) (Figure I.8).

An efficient disassembly network is formed through the Hsp70/100 relationship. Each step is regulated to conserve energy as well as increase the overall efficiency of recognition, binding, disassembly, and refolding of the substrate protein. In *E. coli* the Hsp70/100 bi-chaperone system consists of DnaK and ClpB. The co-chaperones needed for DnaK activity include DnaJ as the J-protein and GrpE as the NEF. Together, these proteins create the KJEB disaggregase system.

The Small Heat Shock Proteins

Small heat shock proteins (sHsp) are the most ubiquitous class of molecular chaperones, appearing across all phylogeny. The number of individual sHsp present in an organism appears to correlate with cellular complexity, as most simple organisms contain one or two sHsp while plant species contain up to 50 sHsp (94, 95). However, unlike Hsp60, Hsp70, and Hsp100, the basic sHsp mechanisms of action are poorly understood due to high levels of structural and functional diversity within the family.

Like other chaperone families, the main functional goal of the sHsp is to protect the cellular proteome against stress. However unlike the Hsp60, Hsp70, and Hsp100 chaperone families, the sHsp do not require energy to function. sHsp protect against stress through tight binding of misfolded or aggregated substrate. In doing so, sHsp inhibit continued aggregation and keep any aggregates that do form in a refolding competent state (96). However, tight binding means substrate proteins are unable to refold or dissociate from the sHsp without assistance from ATP driven machines.

In prokaryotes, the sHsp and misfolded substrate complexes are dissociated by the Hsp70 and Hsp100 chaperone families. Through sHsp interaction with protein aggregates, more efficient reactivation is achieved by the Hsp70 and Hsp100 chaperones

(59). The combination of sHsp and Hsp70/Hsp100 chaperones creates a faster and more efficient stress response by the proteostasis network (63).

sHsp function drastically decreases the need for cells to degrade and resynthesize misfolded proteins, both of which are energetically expensive. Moreover, sHsp are utilized as the first responders to protein misfolding, inhibiting rapid aggregation until the Hsp70s and Hsp100s can be recruited to carry out the disaggregation processes. This response gives organisms the ability to survive changing conditions and different stressors. The sHsp thus give the ability to adapt to widely varying conditions and maintain a healthy and active proteome.

Structural Features of the sHsp

The sHsp family is structurally defined by the α-crystallin domain (ACD), with members highly variable in mass, ranging between 12-43 kDa (94). The ACD, ranging from 90-100 residues, forms a β-sandwich with each side of the sandwich containing an anti-parallel β-sheet made of three to four β-strands (97, 98) (Figure I.9). The rest of the sHsp domain profile consists of intrinsically disordered N- and C-termini that flank both sides of the ACD. The N-terminal domain/N-terminal extension (NTE) is the larger of the two termini ranging from 24 to 247 amino acids. On the other end, the C-terminal domain/C-terminal extension (CTE) is less than 20 amino acids long, with a conserved IXI/V motif, which is important for quaternary structure assembly (97) (Figure I.9).

A distinctive feature of the sHsp family is their ability to form a highly dynamic range of homo- and hetero-oligomeric structures, all built from sHsp dimers (98-102). sHsp dimers are either homo- or heterodimers depending on the sHsp family member. Some sHsp exclusively form homo- or heterodimers, while others have the ability to form both (95). To create the dimers, sHsp utilize the ACD domain in two different ways depending on the individual sHsp. The first type of dimer formation is referred to as β6 swap. This type of dimer formation is found in bacteria, archaea, fungi, and plant organisms. The β6 swap is accomplished by the β6 strand of the ACD domain in each

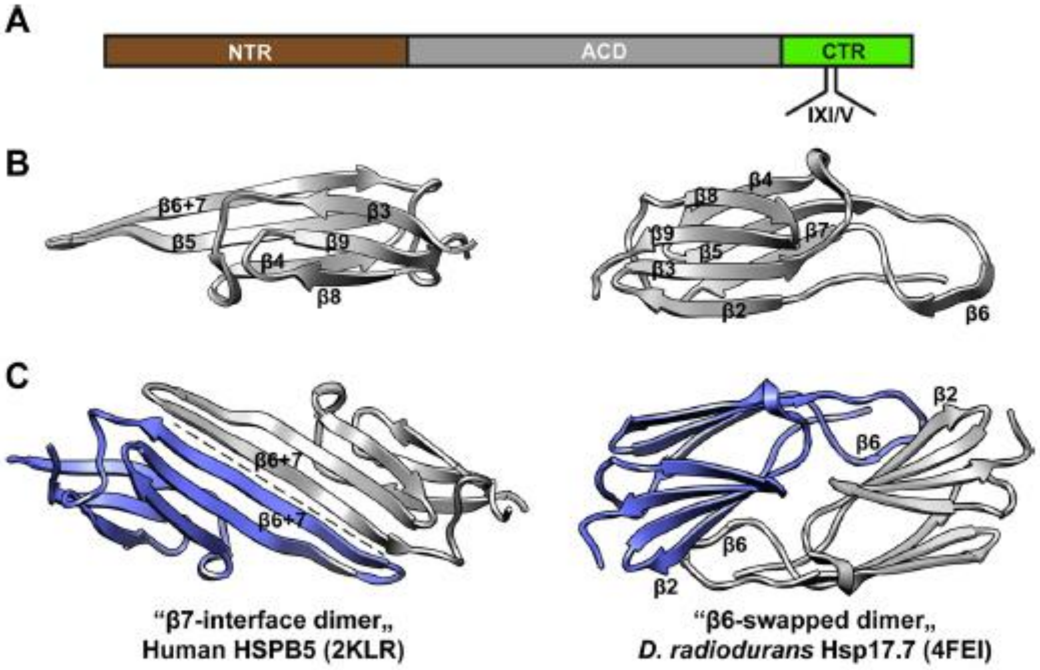


Figure I.9 sHsp Structure and Dimer Formation.
(A) Domain profile of an sHsp family member. (B) ACD of a eukaryote (left) and a bacteria (right). (C) Structures of the β7-interface ACD dimers (left) and β6-swapped ACD dimers (right).
Reprinted with permission from "Small Heat Shock Proteins: Simplicity Meets Complexity" by Martin Haslbeck, Sevil Weinkauf, and Johannes Buchner, 2018, Journal of Biological Chemistry, Copyright [2018] by American Society for Biochemistry and Molecular Biology

monomer interacting with the opposing sHsp ACD β-sandwich (Figure I.9). This interaction is possible for both homo- and heterodimeric sHsp (103, 104). The second type of dimer formation is referred to as the β7-interface and is only observed in metazoans (97). This interaction occurs between the β6 + β7-strand of each monomers ACD (Figure I.9). It is important to clarify the β6 + β7-strand in metazoans is a single β-strand, but is referred to as "β6 + β7" due to the homology with individual β6 and β7 strands of most sHsp. Therefore, each dimer with a β7-interface consists of interactions along the β6 + β7-strand of each monomer to hold the dimer together (Figure I.9).

Using the dimers as building blocks, larger homo- or hetero-oligomers form through dimer-dimer interactions. These oligomeric species are dynamic in size, with the most common oligomers ranging from 6 to 32 monomers (99), and have been shown to form hollow, sphere-like structures similar to a clatherine coat (99, 101, 105). To form the dimer-dimer interactions sHsp utilize all three domains. The most stable interaction occurs between the CTE domain and the β4 and β8 strands of the ACD domain, also referred to as the β4/ β8 groove. This groove is hydrophobic and provides an ideal environment for an interaction with the CTE IXI/V motif from another dimer. This interaction is not completely stable and can be disturbed by environmental conditions, which contribute to the dynamic nature of sHsp oligomers (102). Along with the IXI/V motif and β4/ β8 groove interaction, the sHsp oligomers are also stabilized by NTE interactions between a NTE of one dimer and the NTE or ACD of another dimer. From this ability to interact with other dimers in two different ways, the sHsp have been commonly visualized in a trimer of dimers, or six monomers, with each dimer having interactions with two other dimers. Through the same interactions, structures up to 12-32mers, or 6-16 dimers, can form. On top of this, unstable oligomeric interactions cause frequent dimer exchange within the sHsp population (95, 96, 102). These transient interactions also make all three domains readily available for substrate binding.

Classes of sHsp

Based on oligomer formation preference, three classes of sHsp have been currently defined within the sHsp family (Figure I.10). All sHsp classes are capable of interacting with misfolded or aggregated protein substrate to assist with disassembly by the disaggregase system, unless specifically addressed.

The first class of sHsp exclusively interact with themselves to form homo-oligomers. This is mostly observed at the bacterial level with examples being *S. sp.* and *D. radiodurans.* Homo-oligomers form in *S. sp.* because it expresses a single sHsp, but in *D. radiodurans* there are two sHsp which interact exclusively with themselves (95). In the same way, yeast (*S. cerevisiae*) contain two sHsp that do not interact with each other. For both yeast and *D. radiodurans,* each sHsp functions differently which suggests functional specificity drove homo-oligomer formation in these organisms.

The second class of sHsp interact exclusively with other sHsp family members to form hetero-oligomers. This class includes the sHsp from *E. coli,* IbpA and IbpB, as well as many plant sHsp. IbpA and IbpB are the only sHsp in *E. coli*, while plants contain upwards of 50 sHsp. With so many sHsp, the plant kingdom relies on internal organization within each plant speceis (95, 106). Plants organize sHsp into two classes

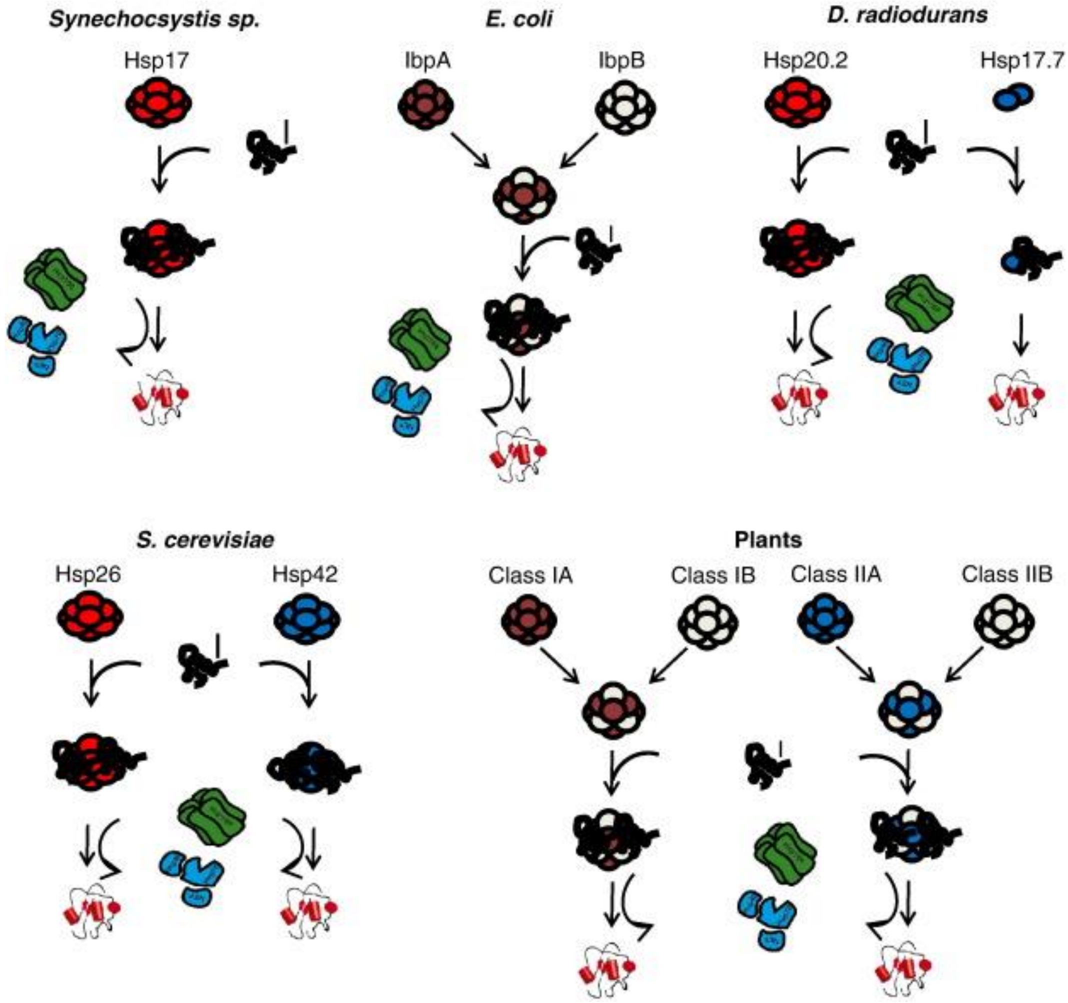


Figure I.10 Classes of sHsp are defined by oligomer formation.
Examples of sHsp that form homoligomers include Synechocsystis sp., D. radiodurans, and S. cerevisiae. Heteroligomers include E. coli and plants. Both homo-and heteroligomers interact with non-native and aggregated protein substrates to prevent continued aggregation.
Reprinted with permission from "A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis" by Martin Haslbeck, Elizabeth Vierling, 2015, Journal of Molecular Biology, Copyright [2015] by Elsevier

with each class exclusively interacting within themselves. Each class is comprised of two or more sHsp which lead to the hetero-oligomer formations. Surprisingly, both classes interact with misfolded substrate and prevent protein aggregation, but cannot interact with each other.

The final class of sHsp function as both homo- and hetero-oligomers. Currently, this class is specific to sHsp in higher eukaryotes with the main example being α-crystallins in humans (*H. sapiens)* (97). The α-crystallins are tissue specific sHsp with expression rates changing based on the tissue type (97, 107). For example, αA-crystallin and αB-crystallin are highly expressed and form hetero-oligomers in eye tissue. But, αB-crystallin is also highly expressed in cardiac tissue, while αA-crystallin is not. In this case, αB-crystallin forms homo-oligomers to carry out its function (108).

sHsp Activation

sHsp must be highly expressed and activated for defense against stress to occur. While the most common activator is elevated temperature, other stressors that activate the sHsp includes oxidative stress, protein concentration, salt concentration, and pH (109-111).

When the cell is under stress, the sHsp family expression rate increases more than any other chaperone family (112); however, the mechanisms used to control expression rates are highly variable. For example, bacteria utilize both promotion and repression for sHsp gene expression depending on the species (112). Since this dissertation is focused on the sHsp from *E. coli*, it is important to understand the gene expression and regulation mechanisms used for IbpA and IbpB (IbpAB) rather than discuss all gene regulation mechanisms.

IbpA and IbpB are located on the same operon with transcription promoted by the heat shock sigma factor, σ^{32}. Under stress, IbpAB gene induction rates increase ten-fold compared to non-stress induction rates, the highest of all *E. coli* heat shock proteins (113). Once transcribed, the RNA strand contains both IbpA and IbpB and must undergo additional processing to separate the IbpA and IbpB strand for translation. The additional processing is accomplished by RNase E which cleaves the RNA into an IbpA and IbpB strand. Further cleavage of the IbpB strand also occurs leading to a subset of nonfunctional IbpB RNA fragments (113).

Following RNA processing, further regulation of IbpAB expression occurs through the presence of an RNA thermometer within each of the IbpA and IbpB RNA strands. This RNA thermometer consists of four hairpins that inhibit ribosome binding to the Shine-Delgarno sequence when the bacteria is under ideal temperatures (~30°C). However, once temperatures rise to a more stressful level (~42°C), the hairpins relax and allow for ribosome binding (113). While the RNA thermometer is present in both the IbpA and IbpB RNA strands, current *in vivo* data suggests the thermometer is only used for IbpA regulation. For IbpB, the thermometer prevents ribosome binding *in vitro,* yet there is no known dependence *in vivo,* most likely because of the imposed regulation through IbpB RNA processing into nonfunctional fragments by RNase E (113).

After translation, the final step of IbpAB regulation, from a concentration perspective, occurs through the protein degradation pathway. Degradation of IbpA and IbpB is accomplished by the AAA+ Lon protease (114). Interestingly, the degradation rate of IbpA and IbpB are different depending on the level of IbpA and IbpB present within the cell. IbpB is degraded faster than IbpA when IbpB is alone in the cell. However, when both IbpB and IbpA are present, IbpB enhances the degradation rate of IbpA. This shows a level of interdependence for the *E. coli* sHsp that cannot be explained in the context of combatting cellular stress (114).

While regulatory mechanisms for expression are needed, the sHsp have other structural interactions that limit or induce activity. The most common structural feature related to activity is the formation and dissociation of the homo- or hetero-oligomers. Oligomers form through the interactions of all three sHsp domains with the CTE interacting with the ACD β4/ β8 groove and the NTE interacting with other NTEs or the ACD. Out of all the domains, the NTE has the highest affinity for substrate; however, substrate interaction can be limited due to the NTE involvement with oligomer formation. Therefore, most sHsp must shift from large oligomers to small oligomers in order to expose the NTE to misfolded or aggregated protein substrates (99, 102, 112). Similar to sHsp expression levels, elevated temperature is the most common environmental factor to induce sHsp to form an active smaller oligomer. Other less common stresses that induce small oligomer formation include oxidative stress, salt concentration, and pH.

While a shift in oligomer size is used by many sHsp, other activating mechanisms include conformational changes and phosphorylation. Currently, both mechanisms are only observed in eukaryotes with specific examples being Hsp27 of *S. cerevisiae*, class II sHsp of plants, and α-crystallins of *H. sapiens*. Like changes in oligomer size, conformational changes occur when the temperature is elevated. This causes the NTE to change conformational states from interacting with each other within an oligomer to being exposed to the surrounding environment (107, 115). At this time, NTE conformational changes have not been observed to cause a decrease in oligomer size. The NTE is also involved when activation occurs through phosphorylation. By phosphorylating serines in the NTE, the sHsp shift to smaller oligomers and release the NTEs into the surrounding environment. Similar to when elevated temperature causes the smaller oligomers to form, there is now a higher probability of NTE interaction with a misfolded protein substrate (115).

sHsp Substrate Recognition

Following activation, substrate recognition is required to inhibit proteins from continuous misfolding and aggregation. Similar to other chaperones, the sHsp interact with substrates through exposed hydrophobic patches or regions (102). There are no known amino acid sequences or specific hydrophobic characteristics the sHsp recognize, which supports their ambiguous nature to recognize hundreds of substrate proteins, even across species (116, 117). sHsp primarily use the NTE to bind substrate, however the ACD and CTE are able to recognize a limited number of substrates or structural features (118). This is the opposite of oligomer formation where the ACD and CTE are the major contributors while the NTE is only needed for stabilizing larger oligomer formation.

The E. coli sHsp IbpA and IbpB

Every sHsp has characteristics necessary for individual species survival. Since this work is focused on the *E. coli* sHsp IbpA and IbpB (IbpAB), it is important to describe what is known and unknown about their functional and mechanistic properties.

IbpAB were identified during heterologous protein expression in *E. coli*. During non-endogenous protein induction, *E. coli* responded by overexpressing IbpAB and were at high concentrations in inclusion bodies. This led to their naming, inclusion body protein A (IbpA) and inclusion body protein B (IbpB) (119). Eventually, they were the conserved ACD was identified in the amino acid sequence of both proteins linking them to the sHsp family.

IbpAB are able to inhibit continued aggregation and assist with protein reactivation, adhering to the main sHsp family function. Early research on IbpA and IbpB alone demonstrated both proteins are able to bind and inhibit aggregation without the other, suggesting there was no functional interdependence (120-122). However, it was eventually discovered that their ability to limit aggregation increased when both proteins were present, supporting IbpA and IbpB being synergistic (110). While limiting aggregation was functionally similar for IbpA and IbpB alone, the same could not be said about substrate reactivation by KJEB. When IbpA inhibits aggregation alone, reactivation by KJEB is not enhanced. However, when IbpA and IbpB were present together, KJEB reactivation of protein aggregates was at a higher efficiency, providing more evidence to support the interdependence of IbpAB (123, 124). Currently, these observations are supported by multiple substrate proteins including luciferase, malate dehydrogenase, and maltose binding protein (123-125).

At the same time, *in vivo* experiments explored how *E. coli* relied on IbpAB to survive. Unlike other chaperone families, it was found that IbpAB were not essential for *E. coli* survival under normal conditions. However, at elevated temperatures (50°C) and for extended periods of time (>1 hr), IbpAB were needed to completely remove protein aggregates from the cell (126). This suggested a synergistic relationship existed between IbpAB and KJEB, agreeing with the *in vitro* observations.

Once it was established that IbpAB were vital for aggregate removal during extended periods of stress, the next goal was to understand the mechanisms by which IbpAB carried out their function. Mechanistic studies first probed the activity of aggregating substrate protein when IbpAB were present. To test this, luciferase was put under aggregating conditions in the presence of IbpAB. Under these conditions, luciferase activity decreased over time showing IbpAB did not hold the substrate in a native state. Instead, luciferase was in a misfolded state that did not continuously aggregate when IbpAB were present (124). Even with this information, studies were not continued to further understand the misfolded substrate or how IbpAB were able to inhibit continued aggregation with a misfolded substrate protein.

Along with aggregate inhibition, researchers explored the mechanisms used by IbpAB to impact aggregate disassembly by KJEB. Early on, it was known IbpAB form a complex with misfolded protein substrate. Early sHsp structural data suggested a shell or coat formed around the aggregates to inhibit further growth (127, 128). By creating a coat, IbpAB would create small aggregates which would be disassembled faster and with a higher efficiency.

Following the coating model, researchers working with a homo-oligomeric sHsp proposed that the sHsp integrate within the aggregate to form a new sHsp/aggregate complex. By doing this, the sHsp would shape the aggregate structure and inhibit further aggregate growth as well as assist aggregate disassembly by making the aggregates less compact (129). However, this model lacked any data utilizing a hetero-oligomeric sHsp or, more specifically, IbpAB.

Recently, a research group proposed a model where the aggregate/IbpAB complex consists of two IbpAB populations, one population that is integrated within the aggregate complex, while the second IbpAB population surrounds the aggregate (130). To disassemble the complex, DnaK removes the outer layer of IbpAB before disaggregating the second population of substrate and IbpAB. With IbpAB both integrating and creating a coat, KJEB disassembly is faster and more efficient (130). This model is somewhat problematic as this needs a large excess of sHsp which is not physiologically relevant for any system.

Currently, all three mechanistic models are possible for how IbpAB inhibits protein aggregation to assist in enhanced disassembly by KJEB. The first model proposes IbpAB forms a coat around the protein aggregates during early aggregation, therefore inhibiting continuous aggregate growth. This model dictates that the aggregate structure is the same with and without IbpAB and the only change IbpAB impose is aggregate size. The second model proposes IbpAB integrate into the aggregate to change the aggregate interactions. This model dictates a new aggregate/IbpAB complex is created and the size of the complex does not dictate disassembly rate. Finally, the most recent data supports a combination of both models where IbpAB integrate and coat the aggregates to inhibit growth. This model dictates that IbpAB change the internal aggregate structure and form a protective coat around the complex until KJEB arrive.

A main reason all three models are possible is from the lack of IbpAB mechanistic studies utilizing techniques other than ensemble approaches and protein aggregate reactivation. This is especially true when trying to understand how IbpAB impact KJEB mediated disassembly. Many researchers assume aggregate reactivation is concurrent with aggregate disassembly. Instead, it is very plausible IbpAB assist in either aggregate disassembly or monomer reactivation. Recent data has shown sHsp can change the aggregate monomer conformation to a more native-like state (129). With this possibility, the sHsp could simply hold the monomer until it is released from the aggregate and speed up the refolding process rather than increase aggregate disassembly. While this model has not been supported by IbpAB data, it is vital to measure pure aggregate disassembly rather than protein substrate reactivation.

Furthermore, understanding how IbpAB impact structurally different protein aggregates has been lacking. While this has been studied with the homo-oligomer forming sHsp αB-crystallin, there is a need to understand if the same mechanisms are utilized by the hetero-oligomeric sHsp (118). By studying both oligomer types, important similarities or differences will be discovered for the overall sHsp family.

Problems Studying the Small Heat Shock Proteins and Aggregation

A core reason mechanistic understanding of IbpAB remains elusive is due to the inability to quantify the dynamic size range and complexity of both IbpAB oligomers and the aggregates upon which they act. To solve this problem, new technology must be developed to correctly measure and quantify these complex molecular particles. To gain a better mechanistic understanding of sHsp, we utilize a single particle fluorescence approach called burst analysis spectroscopy (BAS).

Dissertation Overview

The main research objective for this dissertation is to understand the mechanisms used by IbpAB to assist in more efficient aggregate disassembly by KJEB.

Chapter II addresses the methods used throughout the entire body of work including purification protocols, aggregation protocols, and the use of Burst Analysis Spectroscopy (BAS).

Chapter III addresses the impact of IbpAB on aggregate size, aggregate structure, and disassembly by KJEB. This work explores two different protein aggregate substrates as well as two structurally different aggregates referred to as amorphous and fibril-like.

Finally, Chapter IV addresses future directions of 1) developing new technology to study protein aggregation and protein chaperones, 2) utilizing BAS to study other sHsp mechanistic questions, and 3) understanding the interaction between IbpAB, KJEB, and protein aggregates.

CHAPTER II

METHODS*

All methods summarized in this section are used throughout the rest of the dissertation. The first section highlights protein purification, characterization, and fluorescence labeling protocols. The second section explains the protocol used to produce protein aggregates. The third section explains the protocols used for protein disaggregation. The fourth section explains protocols and implementation of ensemble fluorescence methods. The final section is an in depth explanation of Burst Analysis Spectroscopy (BAS) theory and the most recent developments.

Parts summarized with permission from “Development and Application of Multi-Color and Extended Range Burst Analysis Spectroscopy” by Daniel Shoup, Andrew Roth, Rajan Thapa, Jason Puchalla, Hays Rye, 2020, Copyright [2020], *In Prep

Protein Purifications, Characterization, and Fluorescence Labeling Protocols

Purification of WT RuBisCO and Variants

The WT, A454C, C58A/A454C, and C58A/A34C variants of *Rhodospirillum rubrum* RuBisCO were created using standard site-directed mutagenesis methods. All RuBisCO molecules were expressed in *E. coli* and purified as previously described (75, 78, 131).

In short, RuBisCO is expressed from a pTHR-based plasmid in BL21(DE3) cells at 22°C for 24-28 hrs. Cells were lysed, spun, and loaded onto an FFQ anion exchange column pre-equilibrated in FFQ buffer A (50 mM Tris, pH 7.4, 0.5 mM EDTA, 2 mM DTT). RuBisCO was eluted by a linear gradient from 7.5-25% FFQ buffer B (50 mM Tris, pH 7.4, 0.5 mM EDTA, 2 mM DTT, 2M NaCl) with fractions containing RuBisCO confirmed by SDS-PAGE. Cleanest fractions were combined and put on ice followed by a slow addition of ammonium sulfate to a final concentration of 1.2 M. The solution was filtered through a 0.45 μm syringe filter to remove precipitated protein. Following filtration, sample was loaded onto a phenyl sepharose HIC column pre-equilibrated in HIC buffer A (50 mM Tris, pH 7.4, 0.5 mM EDTA, 2 mM DTT, 1.2 M Ammonium Sulfate). RuBisCO was eluted by a linear gradient from 0-100% HIC buffer B (50 mM Tris, pH 7.4, 0.5 mM EDTA, 2 mM DTT) with fractions containing RuBisCO confirmed by SDS-PAGE. Final sample was dialyzed in storage buffer (25 mM Tris, pH 7.4, 100 mM KCl, 0.5 mM EDTA, 2 mM DTT), concentrated, supplemented with glycerol (15%

w/v), and flash frozen by liquid nitrogen. The protocol was identical for all RuBisCO variants.

Single-Site Fluorescence Labeling of RuBisCO

Labeling of RuBisCO variants were performed as previously described (75, 78, 131). In short, RuBisCO is buffer exchanged into reaction buffer (50 mM Tris, pH 7.4, 100 mM KCl, 0.5 mM EDTA, and 1mM TCEP). The sample is then placed in a glass reaction vial containing a spin vane on a stir plate. Dye was added twice to the RuBisCO sample while stirring. Each addition was at a dye to site ratio of 1.5:1 and lasted for 30 minutes. The reaction was quenched with 5 mM glutathione for 30 minutes. Once the reactions were complete, the sample was buffer exchanged and run over a PD-10 column to get rid of excess dye. The sample was then concentrated, supplemented with glycerol, and flash frozen by liquid nitrogen.

The thiol-reactive dyes used in this study were fluorescein-5-maleimide, 5-(2-acetamidoethyl) aminonaphthalene-1-sulfo-nate (IAEDANS), and Alexa647-2-maleimide. All dyes were obtained from Thermo Fisher Inc. and were prepared fresh from dry powder in anhydrous DMF immediately prior to use.

RuBisCO variants containing surface Cys residues were derivatized with reactive dyes as previously described (75, 78). In some cases, additional purification of the labeled proteins by ion exchange was conducted. Single site labeling was confirmed by denaturing ion exchange and concentration measurements via Bradford and dye molar extinction coefficient measurements.

Double-Site Fluorescence Labeling of the RuBisCO A454C Variant

Taking advantage of a significant difference in reaction rates between the 58C and 454C sites, a sample of the RuBisCO A454C variant was prepared and treated as above to completely label the 58C position. Under typical conditions, this resulted in simultaneous modification of roughly 50% of the 454C sites. The fraction of the protein labeled only at position 58 was separated from unmodified and 454-reacted molecules using high-resolution ion exchange chromatography (MonoQ, GE). Following purification, the singly labeled 58C-dye/A454C protein was then carried through the same labeling protocol to derivatize the 454C position with a fluorescent dye. Sample purity and extent and specificity of labeling were assessed as described above. The protocol was identical for the A34C RuBisCO variant.

Purification of PepQ A24C Variant

PepQ A24C was expressed and purified as previously described (82, 132).

In brief, PepQ is expressed from an inducible plasmid in *E. coli* BL21(DE3) in LB at 37 °C with the lysate loaded onto an FFQ anion exchange column (FastFlow Q, GE) at pH 7.4 and eluted with a gradient of NaCl. Fractions enriched for PepQ were identified with SDS-PAGE and pooled. The protein was then slowly stirred on ice and precipitated with gradual ammonium sulfate addition until 70% (w/v) was reached. The sample was loaded on a hydrophobic interaction column (Phenyl Sepharose FF, GE) at pH 7.4 and eluted with a decreasing ammonium sulfate gradient. Fractions enriched for PepQ were identified by SDS-PAGE and pooled. PepQ was dialyzed into storage buffer (25 mM Tris pH 7.4, 25 mM KCl, 1 µM $MnCl_2$, 2 mM DTT), concentrated, supplemented with glycerol (15–20% v/v), and snap frozen using liquid N_2.

Fluorescence Labeling of PepQ A24C Variant

A24C PepQ was labelled as previously described (82) using tetramethylrhodamine-5-iodoacetamide dihydroioidide (TMR) obtained from Invitrogen In brief, PepQ was reduced with 0.5 mM TCEP and labelled with a 10-fold excess of reactive dye, added in 1 addition for 3 h at 23°C. The reaction was quenched by adding 5 mM glutathione for 30 minutes. The sample was then run over a PD-10 gel filtration column to separate excess dye from labeled PepQ. If needed, re-purification of the labeled protein from unlabeled protein was accomplished with high-resolution ion exchange chromatography (MonoQ, GE). The extent of labelling was determined by protein quantification by the Bradford assay (Bio-Rad) and dye molar extinction coefficient measurements. (Molecular Probes).

Expression and Purification of DnaK, DnaJ, GrpE, and ClpB

DnaK was expressed from the pPROEX HTb vector in the *E. coli* BL21 cell line while DnaJ, GrpE, and ClpB were expressed from the pET 151/D-TOPO vector in *E. coli* BL21 DE3. DnaK, DnaJ, GrpE, and ClpB were all grown in baffled flasks at 37°C with DnaK, DnaJ, and GrpE grown to an O.D. of 0.6 before induction and ClpB was grown to an O.D. of 0.7 before induction. DnaK, DnaJ, GrpE, and ClpB were all induced for 3 hours at 37°C with 400 mM IPTG. DnaK, DnaJ, GrpE, and ClpB were purified identically. Cells were lysed by cell disruption and clarified by ultracentrifugation. Clarified lysates were loaded onto a Ni-NTA column (Quiagen) and washed with 15 column volumes of Ni-NTA equilibration buffer. Protein was eluted using 500 mM imidazole and dialyzed in Ni-NTA equilibration buffer with the TEV protease for 24 hours at 4°C. The 24 hour dialysis was a combination of three 8 hour exchanges in fresh Ni-NTA equilibration buffer. After cleavage, protein was again loaded onto a Ni-NTA column with the cleaved protein flowing through the column and collected. Protein was buffer exchanged into ion exchange equilibration buffer (0 M NaCl) before being loaded onto an ion exchange column (Source 30Q, GE). After a 2 column volume wash, a gradient to 500 mM NaCl was set to run over 25 column volumes. DnaK, GrpE, and ClpB were buffer exchanged into storage buffer (25 mM Tris pH: 7.4, 150 mM KCl, 0.5 mM EDTA, 2 mM DTT) and concentrated while DnaJ was exchanged into storage buffer without EDTA. Protein was supplemented with glycerol (15% v/v), snap frozen in liquid nitrogen, and stored at -80°C.

Purification of IbpA, IbpB, and Variants

IbpA and IbpB, WT and variants, were expressed and purified as previously described (110, 124, 125) with a few modifications.

In brief, IbpA or IbpB are expressed in *E. coli* BL21(DE3) with a cleavable his-tag. Cells were lysed and loaded onto a Ni-NTA Column (Qiagen) under denaturing conditions (6 M urea). Protein was eluted with 500 mM imidazole and renatured step wise in dialysis with urea concentrations of 4 M, 2 M, 1 M, followed by multiple exchanges at 0 M. The his-tag was cleaved by TEV protease at 4°C for 24 hours. Cleaved IbpA or IbpB was then denatured and loaded onto a Ni-NTA column with the cleaved IbpA or IbpB collected in the flow through. Finally, the samples were loaded onto an ion exchange column (MonoQ, GE) and eluted with a NaCl gradient. The final renaturation step occurred step wise as described above. Samples were dialyzed into storage buffer (25 mM Tris pH: 7.4, 150 mM KCl, 0.5 mM EDTA, 2 mM DTT), supplemented with glycerol, (15% v/v) and snap frozen using liquid N_2.

Fluorescence Labeling of IbpA D120C and IbpB 143C Variants

The thiol-reactive dyes used in this study were Alexa-488-C5-maleimide and Oregon Green-maleimide (ThermoFisher) with each being prepared fresh in anhydrous DMF. IbpA D120C and IbpB 143C variants were labeled following the same protocol. IbpA or IbpB were reduced with 1 mM TCEP in labeling buffer (50 mM Tris pH 7.4, 150 mM KCl, 0.5 mM EDTA) through overnight dialysis at 4°C. Labeling occurred at 23°C in a reaction vial (weaton) with 4 separate dye additions to reach an 8:1 dye to protein monomer ratio. The reaction was quenched by 5 mM glutathione for 30 minutes followed by addition of buffered urea to a final concentration of 3 M. To separate unreacted dye from protein, the samples were buffer exchanged by spin concentrator (satorius) and a PD-10 gel filtration column. Samples were then quickly diluted to 500 mM urea and put into dialysis containing 0 M urea overnight. Once samples were rid of urea, protein was concentrated using PEG 20,000. Samples were dialyzed into storage buffer (25 mM Tris pH: 7.4, 150 mM KCl, 0.5 mM EDTA, 2 mM DTT), supplemented with glycerol (15% v/v), and snap frozen with liquid N_2.

Protein Aggregation Protocols

Samples of RuBisCO and PepQ were first denatured in 8 M acid-urea buffer (25 mM glycine phosphate pH: 2.0) to a final concentration of 10 µM.

Amorphous Aggregation

RuBisCO folding buffer (50 mM Hepes pH: 7.4, 150 mM KOAc, 10 mM MgOAc, 2 mM DTT) was put on ice in a non-stick centrifuge tube with a spinvane until equilibrium was reached. The solution was placed on a stir plate at the maximum speed followed by immediate injection of denatured RuBisCO to a concentration of 200 nM monomer for 10 seconds. The solution was moved back on ice for 1 minute and 50 seconds after which 50 µL of solution was put into a pre-warmed centrifuge tube on a 23°C heating block for 5 minutes. The 50 µL solution was then diluted to 10 nM in RuBisCO folding buffer to stop aggregation.

Fibril-like Aggregation

RuBisCO folding buffer (50 mM Hepes pH:7.4, 150 mM KOAc, 10 mM MgOAc, 2 mM DTT) was warmed to 23°C on a heating block in a non-stick centrifuge tube with a spinvane until equilibrium was reached. The solution was placed on a stir plate at the maximum speed followed by immediate injection of denatured RuBisCO to a concentration of 200 nM monomer for 10 seconds. The solution was moved back onto the heating block at 23°C for 2 minutes. The solution was then diluted to 10 nM in RuBisCO folding buffer to stop aggregation.

PepQ Aggregation

PepQ folding buffer (50 mM tris pH:7.4, 50 mM KOAc, 10 mM MgOAc, 2 mM DTT) was warmed to 50°C on a heating block in a non-stick centrifuge tube with a spinvane until equilibrium was reached. The solution was placed on a stir plate at the maximum speed followed by immediate injection of denatured PepQ to a concentration of 500 nM monomer for 10 seconds. Solution was moved back onto the heating block at 50°C for 4 minutes. The solution was then diluted to 10 nM in PepQ folding buffer to stop aggregation.

Aggregation with IbpA and IbpB

Right before use IbpA and IbpB, both WT and fluorescent variants, were activated by heating together at 42°C for 10 minutes at a 1:1 ratio. Activated IbpAB were then added to each folding buffer immediately prior to the addition of the denatured RuBisCO or PepQ. Concentrations varied based on the ratio of IbpAB dimer to RuBisCO or PepQ monomer needed for the given experiment.

For addition of IbpAB later in the aggregation pathway, IbpAB were heat activated at 42°C for 10 minutes at a 1:1 ratio and then directly added to the folding buffer already containing the RuBisCO aggregates at 200 nM. This was done at 1 minute and 3 minutes after the start of the amorphous aggregation pathway and 30 seconds and 1 minute after the start of the fibril-like aggregation pathway.

Protein Disaggregation Protocol

All protein disaggregation experiments were done at a RuBisCO and PepQ aggregate monomer concentration of 10 nM in 1 mL of folding buffer with aggregates made with or without IbpAB as highlighted in the protein aggregation section. DnaK, DnaJ, GrpE, and ClpB (KJEB) were then added to the reaction at a concentration of 1 µM, 2 µM, 2 µM, and 200 nM respectively or 250 nM, 500 nM, 500 nM, and 50 nM respectively depending on the rate of disassembly needed. Following addition of KJEB, the regeneration system of creatine kinase and creatine phosphate were added followed by ATP at a concentration of 2 mM.

Ensemble Fluorescence Methods

Intermolecular FRET

Amorphous and fibril-like aggregates were made with and without IbpAB as described above with RuBisCO variants labeled with either IAEDANS or fluorescein as donor and acceptor respectively. IbpAB was added at a concentration of 1:2 and 1:5 RuBisCO monomer to IbpAB dimer, or a concentration of 400 nM and 1 µM of IbpAB, in the aggregate solution before diluting to a final concentration of 10 nM RuBisCO monomer and 20 nM or 50 nM IbpAB dimer.

Emission spectra were acquired with a PTI photon-counting spectrofluorometer equipped with a temperature-jacketed cuvette holder. For IAEDANS detection, the excitation wavelength was 336 nm with a 6 nm slit width. For each FRET experiment, a pair of emission spectra (donor only and donor-plus-acceptor) were acquired with the emission slit width also at 6 nm. The donor-side average energy transfer efficiency ($\langle E \rangle$) was then calculated by integration of the donor emission signal (at wavelengths where no acceptor emission was detectable) from donor-only (FD) and donor-plus-acceptor (FDA) experiments as $\langle E \rangle = (FD - FDA)/FD$.

Intramolecular FRET

Amorphous and fibril-like aggregates were made with and without IbpAB as described above with doubly labeled RuBisCO (58, 454) variant labeled with IAEDANS (454) and fluorescein (58) as donor and acceptor respectively. To inhibit intermolecular

FRET, doubly labeled RuBisCO was denatured with excess WT RuBisCO (75-90% total population). Spectrum were acquired in the same format and protocol as highlighted in the intermolecular FRET section.

Thioflavin T Binding

Fibril-like aggregates were made as described above with a few modifications. Denatured WT RuBisCO was injected into folding buffer containing 20 μM ThT at 23°C and let grow for 10 minutes with and without IbpAB. Sample was placed into the PTI photon-counting spectrofluorometer at 23°C to measure ThT fluorescence. Excitation was at 450 nm with emission measured at 485 nm. ThT fluorescence was normalized to amorphous aggregates under the same conditions.

Burst Analysis Spectroscopy

Introduction

Assembly and disassembly of macromolecular complexes are dynamic and heterogeneous in nature making them extremely difficult to study. As science continues to advance, new innovative techniques are needed to probe these complexes to gain a more quantitative and mechanistic understanding.

Burst analysis spectroscopy (BAS) was developed as a novel single particle fluorescence technique to study macromolecular dynamics in an *in vitro* free solution environment. BAS is distinguished by its ability to directly measure nanoparticle size population distributions as well as measure real time dynamic changes in these distributions under minimally perturbative, free solution conditions (133).

BAS Theory

BAS utilizes either advective flow or laser scanning to linearly propagate the sample through the axial detection volume of a high-sensitivity, confocal fluorescence microscope (Figure II.1). This makes the effective flow rate of the sample significantly faster than the rate of particle diffusion where the fluorescence particles, within the sample, transit the microscope detection volume perpendicular to the excitation laser. At

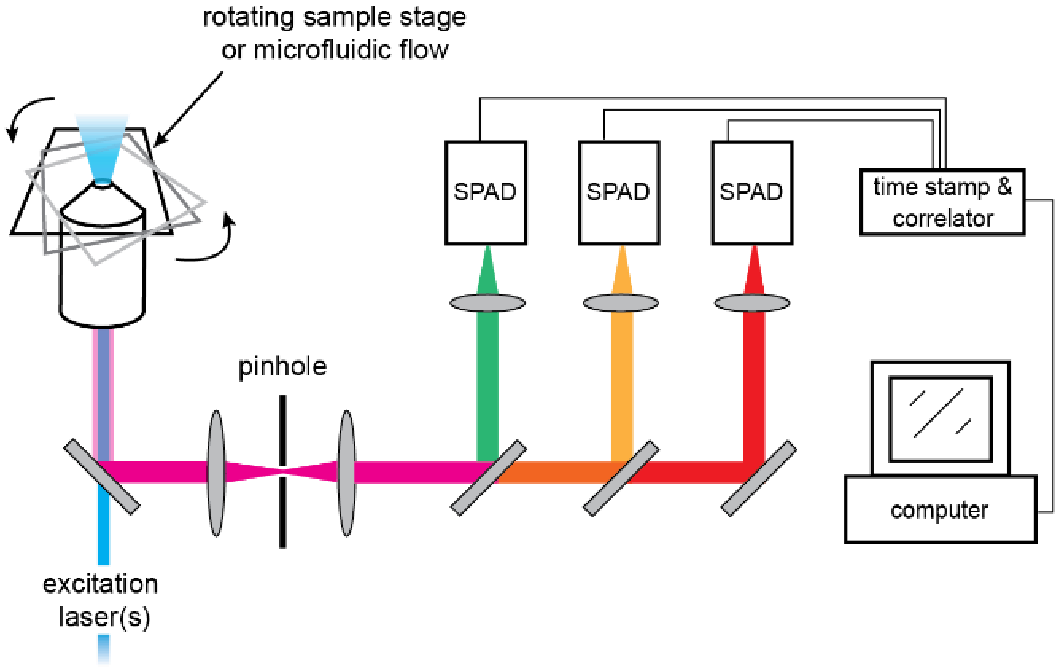


Figure II.1 Single Particle Detection Platform.
Schematic of the single particle detection platform utilized for burst analysis spectroscopy experiments. This setup uses a confocal microscope with a rotating stage with fluorescent sample excited with one or more co-aligned lasers. Fluorescent bursts are detected with single photon counting avalanche photodiodes (SPAD) and then recorded and timestamped in LabView.

low concentrations (<100pM), measured individual fluorescent bursts can be attributed to a single particle passing through the detection volume (Figure II.2A).

For a sample with a single species or particle size, the measured fluorescent bursts are a result of the particle brightness and the location where the particle crosses the detection volume (Figure II.2A, B). Together, all fluorescent bursts are combined into a logarithmic burst histogram where the bursts are binned based on their intensity (Figure II.2D). As long as the particles are equal in fluorescence labeling and quantum

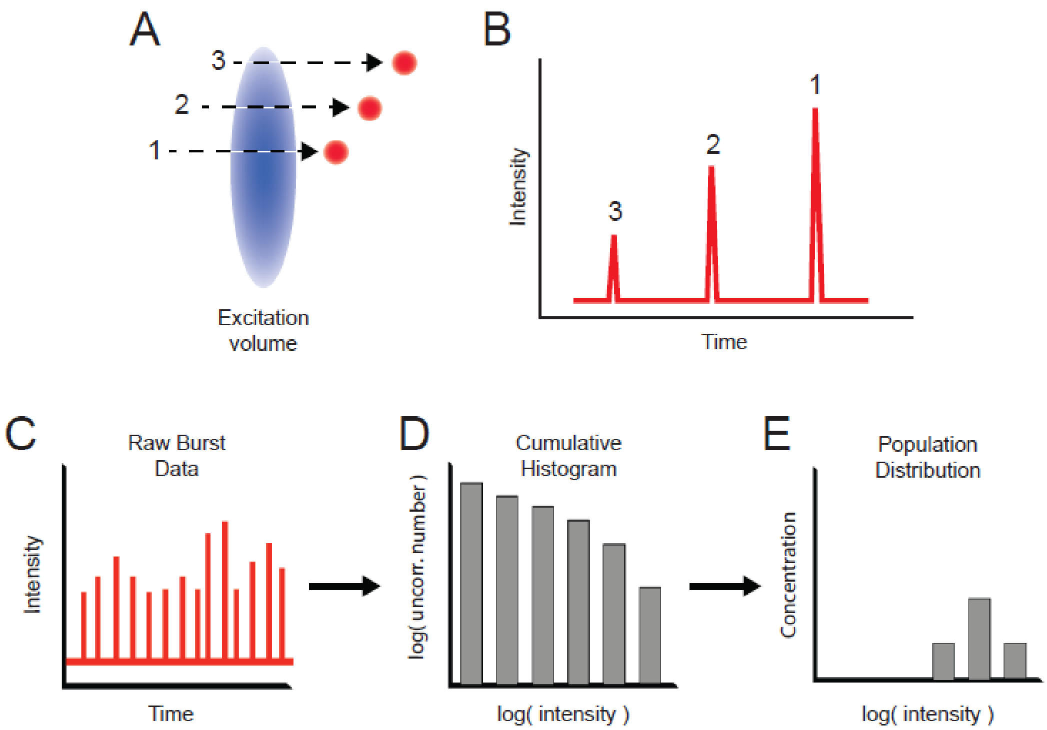


Figure II.2 Conceptual Theory of BAS.
(A) Advective flow transports a uniform population through different locations of the excitation volume resulting in different fluorescent bursts, shown in (B). (C) Example of raw fluorescent burst data. (D) Raw burst data binned into a cumulative histogram. Based on the excitation volume profile, the cumulative histogram is fit to a power law. (E) The corrected population distribution from the power law fit.

yield, the histogram for a single species can be fit to a characteristic function that is dependent to the optical setup (Figure II.2E). For a standard confocal microscope using a high-NA objective, a power law is the characteristic function that describes a cumulative histogram for a single fluorescent species.

For a heterogeneous size population, a burst histogram is still used for all the fluorescent bursts collected, but is now referred to as a cumulative burst histogram since it contains more than one sized species. To evaluate the underlying populations within the sample, BAS uses a bootstrapped reconstruction method beginning with the largest observed bursts in the sample. BAS starts with the largest bursts because it is assumed the largest intensity bins, or brightest objects, can only come from the brightest species passing near the center of the detection volume. Therefore, the number of events in the largest intensity bin gives an accurate concentration measurement for the brightest species in the sample. Using this information along with the established characteristic function of the microscope, a power law, the contribution of the brightest species to all of the lower intensity bins can be calculated. This contribution is then removed for the lower bins making any events remaining in the next brightness intensity bin the next largest species within the sample. The process of calculating the contributions of a species to the lower intensity bins followed by their removal is repeated until the corrected intensity distribution for the total population has been solved. This intensity distribution can then be converted to a particle size as long as a relationship between fluorescence intensity and particle labeling has been established.

BAS Detection Platform

BAS measurements were taken with a custom-built, multi-channel confocal microscope built on a vibrationally isolated optical table. The confocal microscope used is a Nikon Eclipse Ti-U inverted microscope base with a 60x/1.4NA CFI Plan Fluor oil immersion objective. The microscope was outfitted with a precision, 2-axis stepper motor sample stage and a custom–designed confocal optical bench with three independent detection channels. Each detection channel was configured with an optimized band-pass filter set for wavelength selection and a low-noise, single photon counting APD unit. To minimize diffusive motion during the transit of a fluorescent particle through the confocal detection volume, a sample was rotated with a tangential speed of 500 micrometers per second. Particles with a translational diffusion constant of less than 125 $\mu m^2/s$ were then dominated by advective motion through the detection volume. Fluorescence correlation spectroscopy was used to verify that this condition was met. Photon pulses were collected and time stamped with either a multichannel hardware correlator (correlater.com) or a high speed TTL counting board. Sample excitation was provided by either one or a combination of three different lasers: two diode lasers (488nm and 642nm; Omicron) and one diode-pumped solid state laser (561 nm; Lasos). The free-space beams of each laser were couple to a 3-channel fiber combiner (Gould Technologies) and the combined output was directed into the sample objective with a custom, triple-window dichroic filter (Chroma). Each laser was addressable from the integrated control and data acquisition software, custom developed using LabView (National Instruments).

Sample Preparation, Data Collection, and Data Analysis

15 µL of sample was spotted onto a BSA-blocked glass coverslip held in a custom cassette. The coverslip cassette was clamped into the high-precision, computer controlled, 2-axis translation stage connected to the customized Nikon microscope system (133, 134). Once the sample was in place, the inverted objective was moved into position on the coverslip to precisely place the detection volume into the center of the sample. Once into position, the stage holding the cassette was moved at a rate of 500 micrometers per second in a circular motion.

With the sample in place, the raw burst data was collected as a continuous data file through a custom LabView control script. These raw data files contained the amplitude and time of arrival information for bursts in both color channels, auto and cross correlations produced by the correlator and details of the experiment setup. The data files were processed via a custom Matlab analysis library. A baseline subtraction was applied to the raw burst data through a threshold filter at 5 times the data rms. Next, any burst event occurring within 1.5 milliseconds of another were removed from the dataset to avoid burst amplitude bias. Once the event amplitude and timing was determined for a single experiment, the same event analysis was performed on any experimental replicates to establish uncertainty bounds for the average experimental cumulative histogram. Subsequent analysis then proceeded as previously reported (cite).

Multi-Color BAS

A wide range of new applications and experiments are possible with the ability to utilize two or more color channels during a single BAS measurement. Through

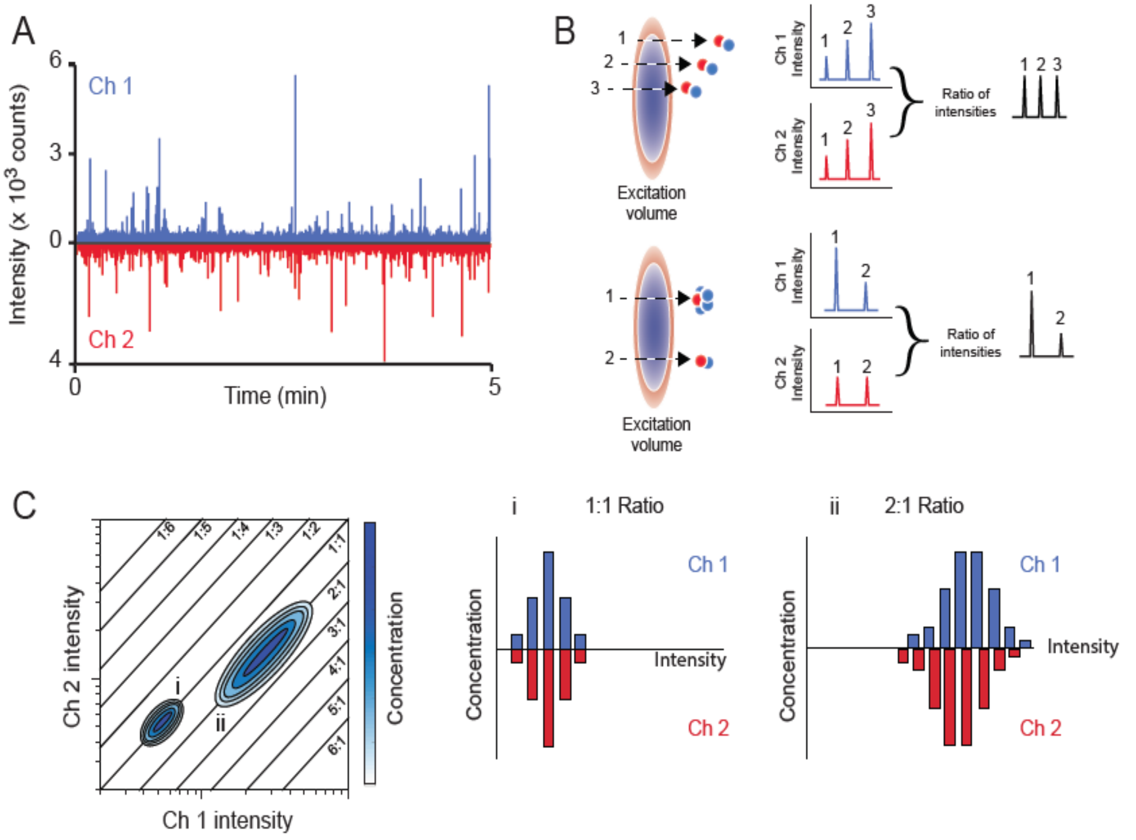


Figure II.3 Conceptual Theory of Multi-Color BAS.
(A) Raw fluorescent bursts of two fluorescently labeled objects in the same solution. (B) Schematic of interacting particles traversing the excitation volume at different locations (left). Simulated bursts based on the crossing trajectory (right) and the corresponding ratio of intensities based on the burst intensity in each wavelength channel. (C) Corrected histograms (i and ii) for a binding ratio of 1:1 (i) and 2:1 (ii). Both corrected histograms are then plotted on a heat map to represent the entire sample (right).

employing multiple channels, it is possible to simultaneously measure not only particle size, but also relative stoichiometry between multiple components within a single sample.

MC-BAS data collection employs the same measurement geometry and microscope setup as standard BAS, but utilizes two or more co-aligned excitation lasers and a matched set of detection channels. For the experiments in this dissertation, MC-BAS is limited to a two-channel design. For a MC-BAS sample, a single particle is comprised of two different macromolecules, each labeled with a different color fluorescent probe. When a doubly-labeled particle crosses the excitation volume of a BAS microscope, the fluorescence intensity from each probe is simultaneously recorded in two separate detection channels (Figure II.3A).

For data processing, events that display coincident bursts in both channels are first separated from events that show no observable coincidence. The coincident events are then sorted by the intensity ratio of the two colors within the particle (Figure II.3B). For each observed intensity ratio, a burst histogram is generated for the events in each component channel. These burst histograms are then analyzed by conventional BAS. BAS distributions for each intensity ratio are then plotted in a heat map where the diagonals specify a given intensity ratio, referred to as a MC-BAS plot (Figure II.3C). Further, in macromolecular assemblies where every component molecule is individually labeled, the emission intensity of the individual components can be used to convert the burst intensity ratios into actual molecule stoichiometries within the macromolecular assembly.

Concatenated BAS

Samples of particles that vary in brightness by more than approximately 100-fold cannot be fully examined by any single BAS measurement due to the inability to characterize the microscope optical response. Additionally, BAS requires the concentration of objects to remain below a concentration threshold (<100 pM) so that individual bursts can be reliably assigned to single particles. The concentration threshold becomes a problem when examining the disassembly of an initial tractable nanoparticle population. At the beginning of a disassembly event, the population fits within the limits, but as disassembly continues the resulting population can violate these single particle conditions.

To solve for these problems, the BAS analysis range can be expanded by combining the burst histograms from a series of diluted samples from the same experiment. To begin, an initial measurement at the highest concentration is recorded. From this sample, a dilution factor is chosen in order to extend the single particle conditions to the lower brightness, higher concentration subpopulations (Figure II.4A). For samples where the largest, brightest particles are much lower in concentration than the smallest, dimmest species, diluting the sample will effectively remove the largest particles from the diluted sample measurement. This makes the burst histogram of the diluted sample shifted toward the lower brightness populations. Importantly, the dilution sample should also yield an intensity overlap region to assist with the reconstruction of the complete dataset from the two individual parts (Figure II.4A).

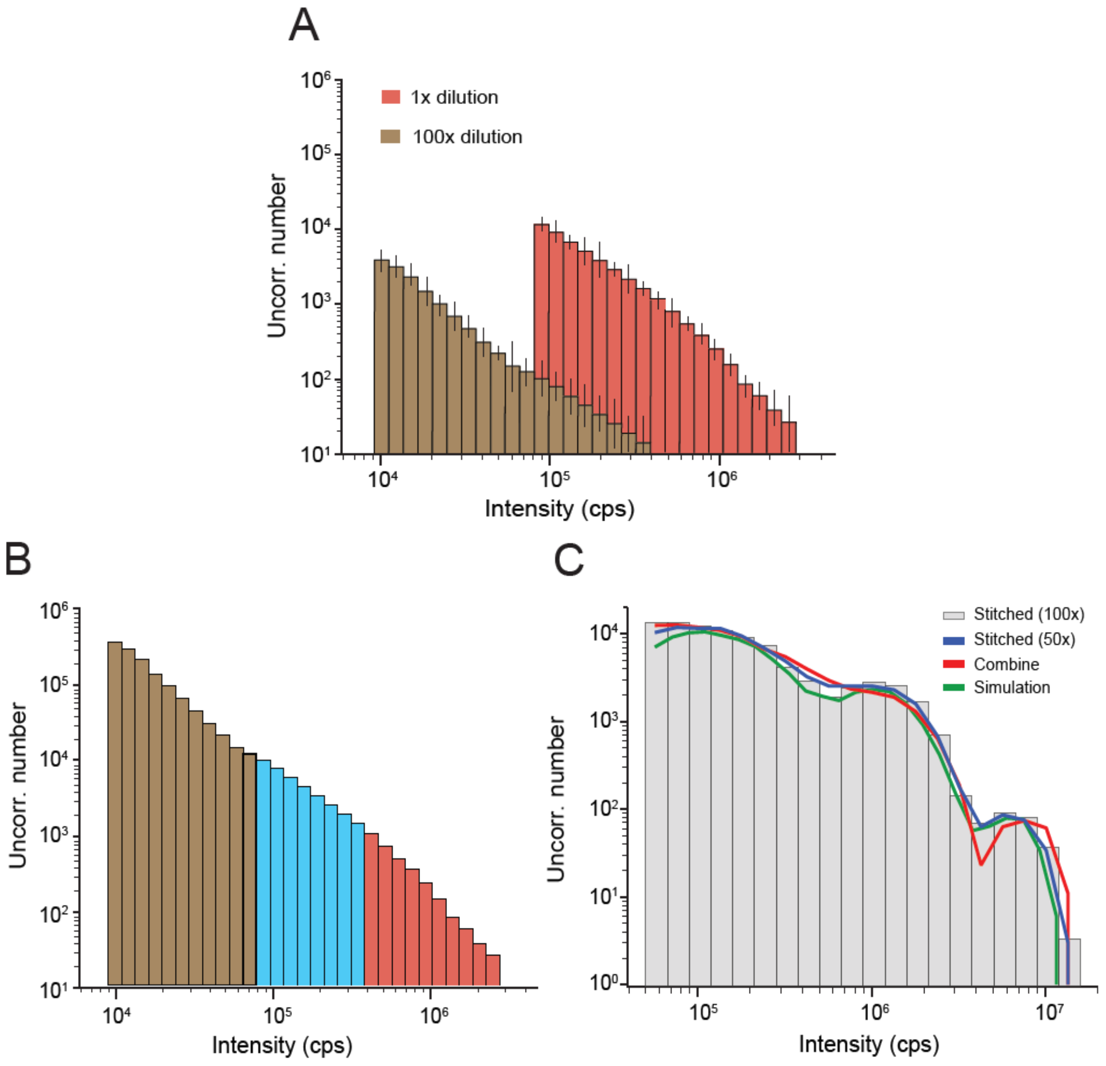


Figure II.4 Conceptual Theory of Concatenated BAS.
(A) Generated cumulative histograms of an undiluted (pink) and diluted (brown) sample. (B) Corrected cumulative histogram with the multiplied diluted region (brown), overlap region (blue), and undiluted region (pink). (C) Corrected BAS histogram comparing simulated and concatenated (stitched) data.

Once the original sample and diluted sample are measured, a best fit scale factor is determined for the overlap region. This scale factor is then applied to the diluted dataset to create a combined burst histogram were the 1) bin values to the left of the overlap region are set equal to the corresponding bin values of the diluted factor multiplied by the scale factor and 2) bin values to the right of the overlap are equal to the bin values from the undiluted sample set (Figure II.4B). The overlap section is determined by the weighted average of both the diluted and undiluted samples. This system of combining a diluted and undiluted sample results in a concatenated burst histogram which can then be analyzed using conventional BAS.

CHAPTER III

THE INFLUENCE OF IBPA AND IBPB ON PROTEIN AGGREGATION AND AGGREGATE DISASSEMBLY BY THE HSP70/100 BI-CHAPERONE SYSTEM

Introduction

To maintain cellular homeostasis, proteins must fold to their native state. However, when under extended stress the cellular environment does not provide a conducive setting for proper protein folding. These conditions result in misfolded or nonfunctional proteins and lead to protein aggregation. Once a sustained protein aggregate population forms, serious diseases arise such as cataracts, cardiomyopathy, and neurodegenerative diseases (50, 55, 135). To maintain cellular homeostasis and prevent wide spread protein misfolding, a set of highly conserved protein families evolved, called molecular chaperones, to assist in protein folding and disaggregation.

One prevalent molecular chaperone family is the small heat shock proteins (sHsp), which carry out functional roles including aggregate prevention, aggregate spatial organization, and enhanced aggregate refolding when combined with a disaggregase system. The sHsp are an ATP-independent class of chaperones containing a conserved α-crystallin domain (ACD) and flexible N- and C-terminal domains. The ACD is known to interact with itself between sHsp to from hetero- and homoeric dimers (102, 112) while the N- and C-termini interact between dimers to form large hetero- and homo-oligomeric complexes. These interactions lead to a very dynamic and complex population of sHsp oligomeric species. Due to this complexity in tertiary structure, a detailed mechanism for how sHsp carry out their basic functions has yet to be elucidated.

Here, we focus on the *E. coli* sHsp IbpA and IbpB (IbpAB) and their known function to enhance substrate protein reactivation by the *E. coli* disaggregase system made up of DnaK, DnaJ, GrpE, and ClpB (KJEB) (123-126). It was first identified *in vivo* that IbpAB could accelerate the removal of aggregates and was linked to the disaggregase system (126). Then, *in vitro* experiments utilizing common aggregation prone substrates solidified the role of IbpAB to enhanced protein reactivation by KJEB. To understand how IbpAB were able to enhance reactivation, studies found IbpAB capable of inhibiting aggregate growth as well as stabilizing protein aggregates in a refolding competent intermediate state (124).

While the functional role of IbpAB is known, several key mechanistic properties have yet to be revealed. For example, all prior IbpAB functional studies were in the context of protein substrate reactivation, making it unclear how IbpAB interact with aggregating protein. To this point, studies are unable to address if IbpAB assist with disassembly before substrate reactivation or if IbpAB only enhance substrate reactivation. Continuing, IbpAB studies using structurally different protein aggregates have been lacking, with limited understanding about how IbpAB interact with fibril-like aggregates. Changes in aggregate structure are known to have toxic traits as well as impacts on other chaperone family activity (48, 50). These known traits make it vital to study changes in aggregate structure with the sHsp family.

One reason the mechanistic properties of IbpAB remain elusive is the dynamic size range and complexity of both the IbpAB oligomers and the aggregates upon which they act. Current ensemble methods are unable to quantify these dynamic size populations or measure the specific interacting populations between IbpAB and protein aggregates. Here, we overcome these limitations by applying a single particle fluorescence-based technique called Burst Analysis Spectroscopy (133). BAS provides a powerful and highly flexible approach, permitting the real-time observation of population-resolved protein aggregation and disaggregation kinetics under non-perturbing, free solution conditions. Recently, the BAS methodologies have been expanded to increase the size detection range and simultaneously measure the size and ratio of interacting particle populations within a given particle complex (136). BAS provides a handle to measure not only IbpAB interacting populations with protein aggregates, but also the ability to measure real time disassembly of IbpAB/aggregate complexes to understand the underlying mechanisms IbpAB uses to assist in enhanced disassembly.

Three distinct mechanistic models have been proposed for sHsp inhibition and assistance in enhanced disassembly of protein aggregates. The first consists of the sHsp forming a coat around the protein aggregates. This coat will allow the formation of a smaller aggregate population and prevent the addition of other aggregates or misfolded monomers (99, 101). The second involves sHsp integrating into the aggregates to form a new sHsp/aggregate complex. In this complex, the sHsp shape the aggregate structure to inhibit further aggregate growth (129). Finally, there are recent data supporting a combination of both models where the sHsp will integrate and coat the same aggregate population to assist in aggregate inhibition and disassembly (130).

In this study, we utilize BAS, electron microscopy, and FRET to show IbpAB are capable of causing two aggregate structures to form a specific aggregate size through changing the internal aggregate conformation. By changing the internal structure, aggregate disassembly by KJEB is dramatically enhanced for two aggregate structures as well as an endogenous substrate protein. We then show KJEB release IbpAB from the IbpAB/aggregate complex in parallel with aggregate disassembly. Finally, we demonstrate IbpAB are capable of binding and inhibiting continued aggregation of an established aggregated population, but cannot fully enhance disassembly under these conditions.

Results

RuBisCO (from *R. rubrum*) has long been a model substrate for protein folding due to the reliance on GroEL/ES to properly fold (75, 78, 81, 83, 137). We have also shown RuBisCO capable of forming structurally different aggregates and the usefulness of these structures in protein aggregation and disaggregation studies with the bi-chaperone system of DnaK and ClpB. However, with RuBisCO not an endogenous *E. coli* protein, a second model substrate was needed. PepQ, an endogenous *E. coli* protein, was identified through its' dependence on GroEL/ES to properly fold as an ideal native substrate for the *E. coli* chaperone families (82, 132). Here, we show both RuBisCO and PepQ are viable substrates for IbpAB, allowing for a more robust mechanistic study of the sHsp and their interactions with protein aggregates.

Burst Analysis Spectroscopy (BAS) can measure changes in an aggregate size population with and without the small heat shock proteins IbpA and IbpB (IbpAB)

Throughout this work, protein aggregates were made utilizing denatured RuBisCO-58-Alexa-647 injected into folding buffer at 200 nM. The buffer is either at 4°C and then 23°C to form amorphous aggregates; or only at 23°C for fibril-like aggregates (Figure III.1A). To stop aggregation after the specified time, both aggregate types are diluted to 10 nM RuBisCO monomer. By labeling RuBisCO with a fluorophore, we are able to take advantage of fluorescence techniques to study the different aggregated states in the absence or presence of IbpAB.

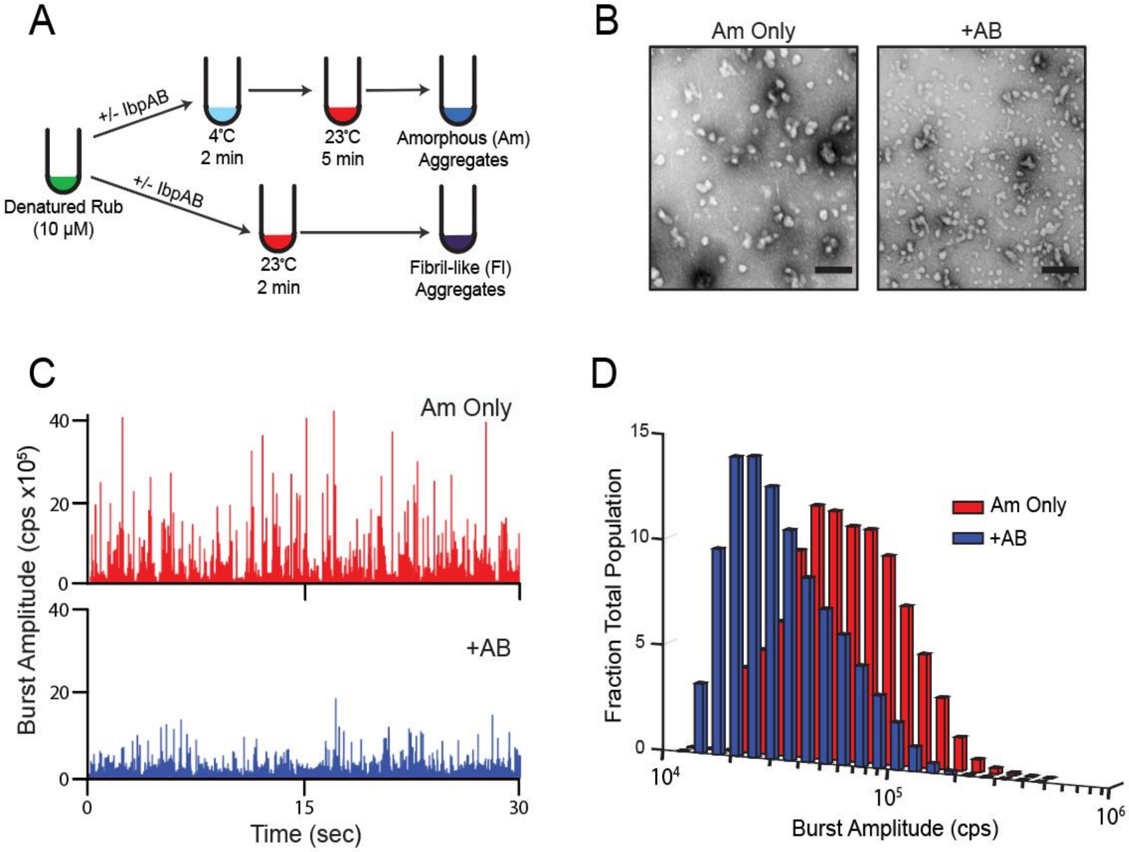


Figure III.1 The impact of IbpAB on a protein aggregate distribution can be measured with Burst Analysis Spectroscopy (BAS).

(A) Sample handling protocol used to populate either amorphous or fibril-like RuBisCO (R. rubrum) aggregates. Alexa647-labeled RuBisCO was first denatured in acid urea and then rapidly diluted 50-fold to 200 nM (final monomer concentration) in HKM buffer at either 4 °C or 23 °C, in the presence or absence of heat-activated IbpAB. Samples diluted into cold buffer were then warmed to 23 °C to populate amorphous aggregates. Direct dilution into warm (23 °C) buffer results in formation of fibril-like aggregates. Following a brief incubation period, aggregate growth was halted by dilution to a final monomer concentration of 10 nM. (B) Negative stain electron microscopy images of amorphous RuBisCO aggregate particles formed in the presence and absence of heat activated IbpAB (scale bar = 200 nm). (C) Raw photon history showing fluorescence bursts of amorphous RuBisCO-Alexa647 aggregates formed in the the presence (blue) and absence (red) of heat activated wild type IbpAB (50 nM final dimer concentration). (D) Distribution of aggregate particle sizes measured by BAS. The approximate peak of each distribution, shown as the number of RuBisCO monomers per particle, is derived from the measured effective brightness of single Alexa647-labeled RuBisCO monomers incorporated into an aggregate particle. Each BAS plot is a combination of three, independent experimental replicates.

To confirm BAS as a viable technique to measure IbpAB impact on protein aggregates, we quantitatively measured IbpAB inhibition of amorphous RuBisCO aggregates. Raw fluorescent bursts from amorphous aggregates decrease in the presence of IbpAB (Figure III.1C). From this raw data, the exact aggregate population is calculated and represented as a size distribution where smaller objects are located on the left and larger object located on the right of each BAS histogram (Figure III.1D). As expected, the aggregate population containing IbpAB is smaller than the amorphous aggregate population alone. This size distribution can be converted to monomers per aggregate based on individual fluorophore brightness in an aggregate (136). For an aggregate made of RuBisCO-Alexa 647, the brightness of an average monomer in an aggregate is 38 cps/uW/fluor. Using this information, the smallest aggregates we can measure (10^4 cps) are 5mers and the largest aggregates we can measure (10^7 cps) are 5000mers. BAS is capable of measuring an IbpAB induced change in aggregate size (Figure III.1D) and is confirmed by electron microscopy (Figure III.1B).

IbpAB inhibits aggregation of two substrates (PepQ and RuBisCO) and two structurally different aggregates (amorphous and fibril-like) to a similar size population

With BAS confirmed as a method to quantitatively measure the changes in protein aggregate size in the presence of IbpAB, the next step was to understand how IbpAB inhibits two different substrates and two different aggregate structures. To understand how IbpAB interact with another substrate, aggregation conditions were optimized for PepQ to be utilized. Both RuBisCO aggregate structures and PepQ were

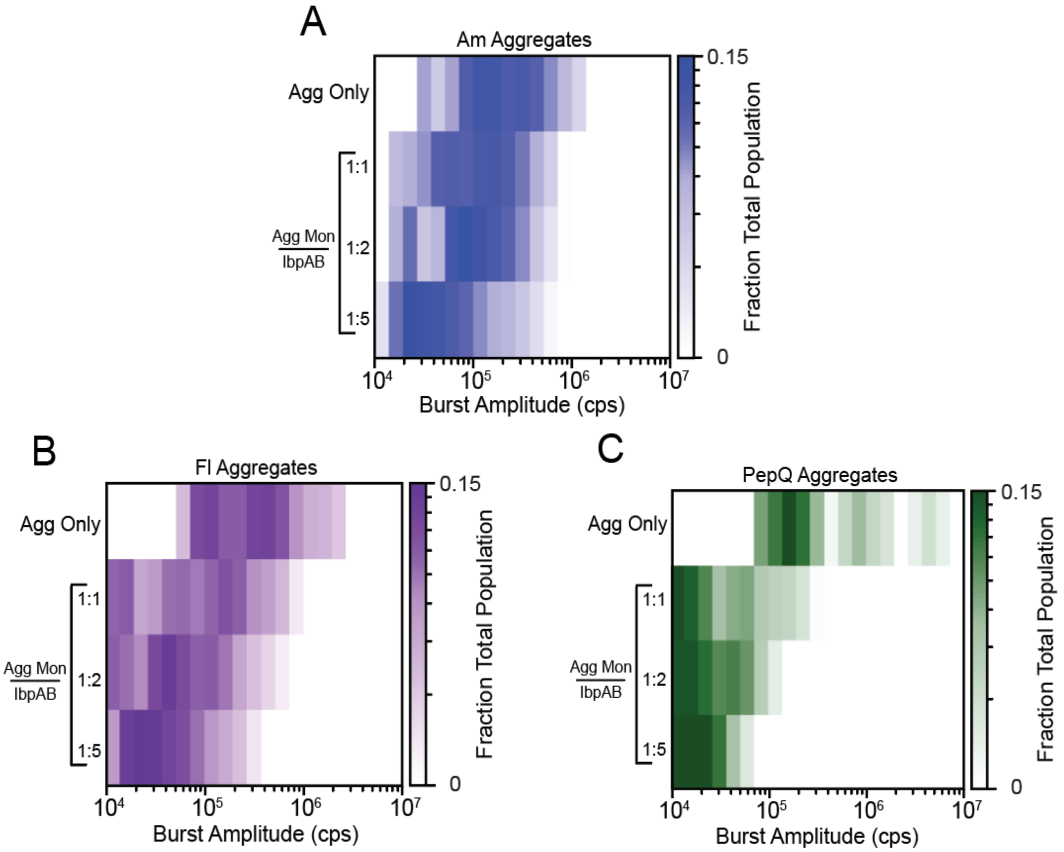


Figure III.2 IbpAB restrict the aggregation of different proteins to a similar and limited particle size range.
BAS population distributions of (A) amorphous RuBisCO, (B) fibril-like RuBisCO, and (C) PepQ aggregates formed in the presence of different amounts of heat-activated IbpAB. Aggregation of RuBisCO-Alexa647 into amorphous and fibril-like aggregates was carried out as outlined in Figure 1. TMR-labeled PepQ was denatured in acid urea and diluted 50-fold to a final monomer concentration of 500 nM in TKM buffer at 50 °C. Aggregation was halted after 4 min by rapid dilution to a final PepQ monomer concentration of 10 nM in TKM buffer at 23°C. The final mixing ratios of heat-activated IbpAB dimer to either RuBisCO or PepQ monomers (1:1, 1:2 and 1:5) are shown. Aggregate particle burst intensity is plotted on the heat map x-axis, while the color scale shows a normalized measurement of particle frequency (fraction of total events observed at each intensity value). Each BAS plot is a combination of three, independent experimental replicates.

aggregated with increasing amount of IbpAB dimer at a 1:1, 1:2, and 1:5 ratio (Figure III.2).

Without IbpAB, amorphous aggregates alone grow to a range of 20-500mers with a majority of the population between 50-200mers (Figure III.2A). As IbpAB concentration increases during amorphous aggregation, the aggregates become smaller with a majority of the population between 20-30mers and a small fraction of the population extending to 100mers. This result highlights an aggregate size dependence on IbpAB concentration.

Fibril-like aggregates grow to a range of 40-700mers when IbpAB are not present during aggregation (Figure III.2B). In the presence of increasing IbpAB concentration, fibril-like aggregates are inhibited to a majority population around 20mers with a small fraction of the population as large as 80mers. The population at 20mers is continually increased as the IbpAB concentration increases. Lower concentrations are needed of IbpAB to inhibit fibril-like aggregates to a similar population as the amorphous aggregates suggesting fibril-like are more susceptible to IbpAB action.
With PepQ, IbpAB are more efficient in preventing large aggregate formation. Without IbpAB, PepQ aggregates are highly concentrated around 60mers with a small fraction of the population extending to 2000mers (Figure III.2C). This very large population range changes drastically when IbpAB are present. With an increasing concentration of IbpAB,

the PepQ population becomes more homogeneous at 10-20mers. This result suggests IbpAB are more efficient with *in vivo* substrates.

IbpAB efficiency of inhibition is different for PepQ, amorphous, and fibril-like aggregates. However, even at the highest concentrations of IbpAB, neither RuBisCO nor PepQ become monomerized (confirmed at a 1:10 ratio of substrate to IbpAB dimer, data not shown). Instead, IbpAB creates a more homogeneous aggregate population between 20-30mers with a small fraction of the population extending to 100mers. These data show the capability of IbpAB to limit the size range of aggregates independent of aggregate structure and protein substrate.

IbpAB binds differently to amorphous versus fibril-like aggregates

With varying concentrations, IbpAB are able to inhibit both amorphous and fibril-like aggregates to the same extent. To further understand the mechanistic properties of aggregate inhibition by IbpAB, multi-color BAS (MC-BAS) was employed. MC-BAS is able to correlate two fluorescent particles bound to one another as they move through the excitation volume (Figure II.3). Then, the size of the correlated particles are analyzed and mapped out on a 2 dimensional plot where the binding ratio and sizes of each particle population can be interpreted (136). For all MC-BAS experiments RuBisCO is labeled at the 58 position with Alexa-647 while IbpA is labeled at the 20 position with Oregon Green (OG) (Figure III.3A-B). Experiments with RuBisCO and IbpB were also accomplished with Alexa-488 linked to an additional amino acid added to the C-termini (Figure III.3C-D). For each aggregation experiment a 1:1 mixing ratio of RuBisCO monomer to IbpAB dimer was utilized.

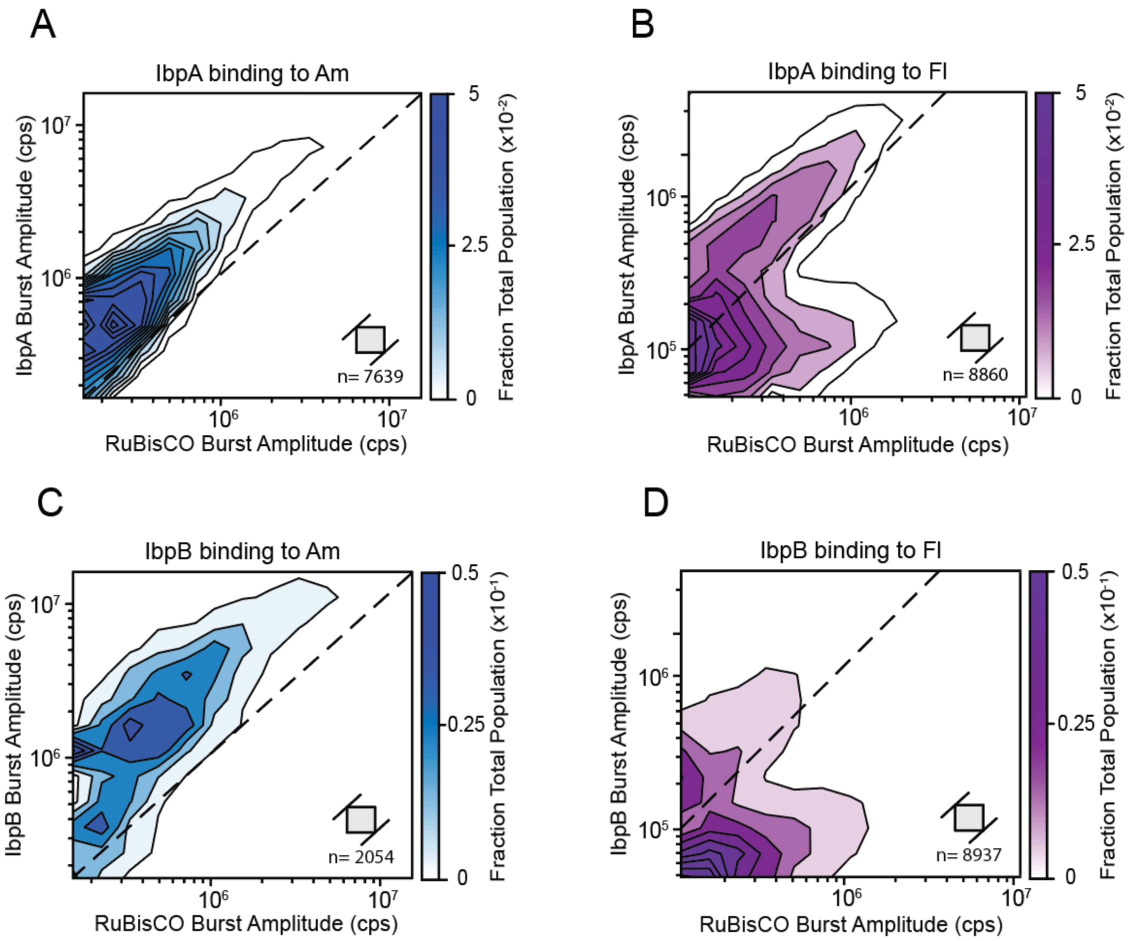


Figure III.3 The stoichiometry distributions of IbpAB bound to amorphous and fibril-like RuBisCO aggregates are distinct.

The MC-BAS distributions for (A,C) amorphous and (B,D) fibril-like RuBisCO aggregates bound to IbpAB-OG are shown. In each case, the final mixing ratio of RuBisCO monomers to IbpAB dimers was 1:1. The RuBisCO burst intensity is plotted on the x-axis and IbpAB burst intensity is plotted on the y-axis. The dashed diagonal line shows the experimentally determined 1:1 brightness equivalence for the RuBisCO- and IbpA-coupled dyes. The spread of the distributions along the positive diagonals of the plot measures the population size distribution at a given IbpAB:RuBisCO stoichiometry, while the extent of spread along the negative diagonals is proportional to the range of binding stoichiometries. Each MC-BAS plot is a combination of six, independent experimental replicates.

MC-BAS measurements show IbpAB bind differently to the different aggregate structures (Figure III.3). IbpAB appear to bind amorphous aggregates at a 1:2 ratio of aggregate monomer to IbpAB dimer across the entire aggregate size range (Figure III.3A,C), while display a bifurcated pattern when binding to fibril-like aggregates (Figure III.3B,D). In the fibril-like sample, there are two distinct IbpAB sub-populations that bind at different stoichiometries to fibril-like aggregates. In the population containing more IbpAB, IbpAB bind at a consistent ratio over the entire aggregate population. The second population contains less IbpAB, but the amount of IbpAB is consistent over the entire aggregate size range.

KJEB disassembly of both protein substrates and aggregate structures are enhanced by a concentration dependent amount of IbpAB

A common ability of the sHsp family is to cause more efficient reactivation of protein aggregates. To understand the impact IbpAB binding and aggregate inhibition had on aggregate disassembly we measured the disassembly kinetics of aggregates with and without IbpAB by BAS. To accomplish this, both RuBisCO aggregate structures and PepQ aggregates were made in the absence and presence of IbpAB at a 1:1 and 1:5 ratio of aggregate monomer to IbpAB dimer. Protein aggregate disassembly was accomplished by the *E. coli* bi-chaperone system, KJEB, at concentrations of 1 µM, 2 µM, 2 µM, and 200 nM respectively over a 30 minute time frame.

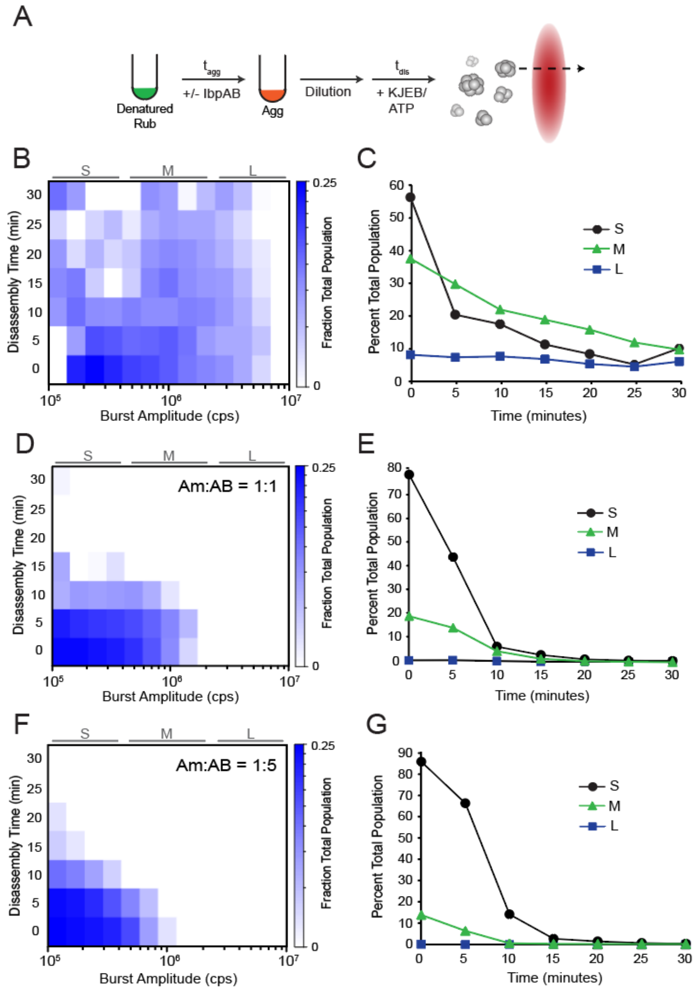


Figure III.4 IbpAB dramatically accelerates the disassembly of amorphous RuBisCO aggregates by the KJEB bi-chaperone disaggregase.

(A) Schematic of a BAS experiment designed to measure the population-resolved kinetics of protein disaggregation. Amorphous aggregates were formed in either the absence (B) or presence of IbpAB at two different RuBisCO monomer to IbpAB ratios: 1:1 (D) and 1:5 (F). Disaggregation was triggered by the addition of the KJEB bi-chaperone system (1 µM DnaK, 2 µM DnaJ, 2 µM GrpE and 200 nM ClpB), 2 mM ATP, and a creatine kinase-based ATP regeneration system. Samples were then loaded into a BAS microscope and burst data was continuously acquired for 30 min. The full experimental photon history was segmented into 5 min blocks and analysis was performed on each block. The heat maps represent a combination of three, independent experimental replicates for each aggregation condition. A zero-time measurement on each sample was collected prior to the addition of ATP. To highlight how disaggregation rates depended on aggregate size, the BAS heat maps were also coarsely binned into small (S), medium (M), and large (L) particle ranges (C, E, and G). In each case, all detected objects within a given size range were summed and plotted as a function of time following the initiation of disaggregation by KJEB.

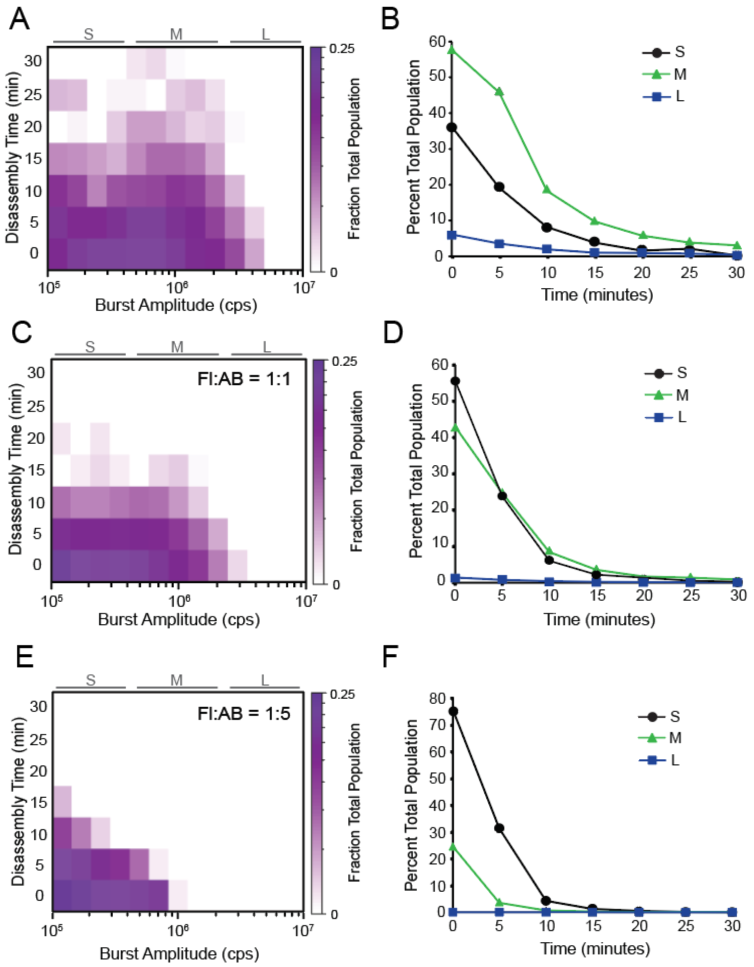


Figure III.5 IbpAB dramatically accelerates the disassembly of fibril-like RuBisCO aggregates by the KJEB bi-chaperone disaggregase.
Fibril-like aggregates were formed in either the absence (A) or presence of IbpAB at two different RuBisCO monomer to IbpAB ratios: 1:1 (C) and 1:5 (E). Disaggregation was triggered by the addition of the KJEB bi-chaperone system (1 µM DnaK, 2 µM DnaJ, 2 µM GrpE and 200 nM ClpB), 2 mM ATP, and a creatine kinase-based ATP regeneration system. Samples were then loaded onto a BAS microscope and burst data was continuously acquired for 30 min. The full experimental photon history was segmented into 5 min blocks and analysis was performed on each block. The heat maps represent a combination of three, independent experimental replicates for each aggregation condition. A zero-time measurement on each sample was collected prior to the addition of ATP. To highlight how disaggregation rates depended on aggregate size, the BAS heat maps were also coarsely binned into small (S), medium (M), and large (L) particle ranges (B, D, and F). In each case, all detected objects within a given size range were summed and plotted as a function of time following the initiation of disaggregation by KJEB.

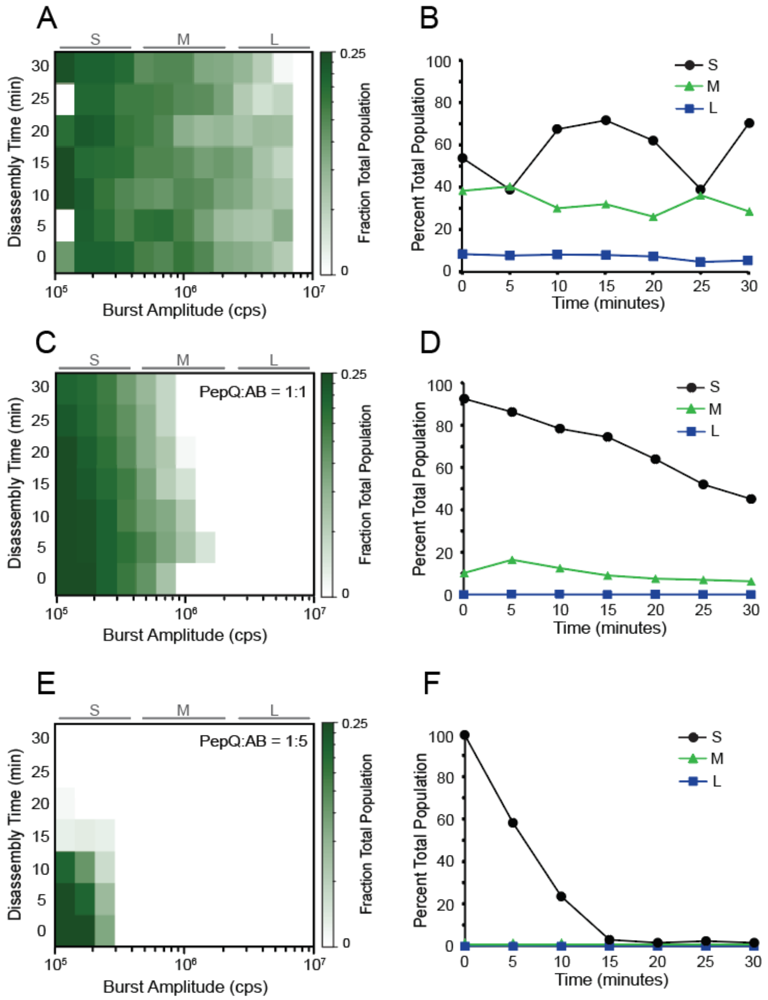


Figure III.6 IbpAB dramatically accelerates the disassembly of PepQ aggregates by the KJEB bi-chaperone disaggregase.
PepQ aggregates were formed in either the absence (A) or presence of IbpAB at two different RuBisCO monomer to IbpAB ratios: 1:1 (C) and 1:5 (E). Disaggregation was triggered by the addition of the KJEB bi-chaperone system (1 µM DnaK, 2 µM DnaJ, 2 µM GrpE and 200 nM ClpB), 2 mM ATP, and a creatine kinase-based ATP regeneration system. Samples were then loaded into a BAS microscope and burst data was continuously acquired for 30 min. The full experimental photon history was segmented into 5 min blocks and analysis was performed on each block. The heat maps represent a combination of three, independent experimental replicates for each aggregation condition. A zero-time measurement on each sample was collected prior to the addition of ATP. To highlight how disaggregation rates depended on aggregate size, the BAS heat maps were also coarsely binned into small (S), medium (M), and large (L) particle ranges (B, D, and F). In each case, all detected objects within a given size range were summed and plotted as a function of time following the initiation of disaggregation by KJEB.

As previously observed, amorphous aggregates alone are refractory to disassembly (Figure III.4A). With the addition of IbpAB at a 1:1 ratio, disassembly by KJEB increased drastically with aggregates dismantled within 15 minutes (Figure III.4C). Interestingly, when IbpAB is increased to a 1:5 ratio, rate of disassembly does not change compared to the 1:1 ratio (Figure III.4E); suggesting a limit on the ability of IbpAB to enhance disassembly.

The size of amorphous aggregates does not play a factor in disassembly rate. This is highlighted in the coarse bin plots which represent the same data in the adjacent BAS disassembly heat maps, but binned into small, medium, and large aggregates (Figure III.4B, D, F). In the coarse bin plots, a comparable concentration of medium sized aggregates are present for amorphous aggregates alone and when IbpAB are present at a 1:1 ratio. Even with a comparable concentration, the medium sized aggregates are disassembles faster when IbpAB are present compared to aggregates alone. This suggests IbpAB are utilized for more than simply limiting aggregate size.

Unlike amorphous aggregates, fibril-like aggregates without IbpAB can be disassembled in 25 minutes (Figure III.5A). Rate of disassembly is improved when IbpAB are present at a 1:1 ratio, however it is not as efficient as disassembly of amorphous aggregates containing IbpAB (Figure III.5C). This is highlighted, again, in the coarse bin plots where the kinetic rates of fibril-like aggregates alone and IbpAB at a 1:1 ratio are very similar for small and medium sized objects. However, once an excess of IbpAB are present at a 1:5 ratio, a stimulated disassembly rate for fibril-like aggregates is observed. This rate is also similar to the stimulated amorphous aggregate rate suggesting more IbpAB is needed during fibril-like aggregation to cause a comparable stimulated rate.

A similar observation is made for PepQ aggregates which also require excess IbpAB to cause a stimulated disassembly rate. IbpAB at a 1:1 ratio cannot drastically enhance PepQ aggregate disassembly (Figure III.6A, C). However, when IbpAB at a 1:5 ratio, there is a stimulated rate of disassembly similar to both amorphous aggregates with IbpAB at a 1:1 ratio and fibril-like aggregates with IbpAB at a 1:5 ratio (Figure III.6E). While fibril-like and amorphous aggregates with IbpAB are mostly medium sized during enhanced disassembly, PepQ aggregates with IbpAB must be small as well as contain excess IbpAB for enhanced disassembly to occur (Figure III.6B,D,F).

IbpAB decreases the proximity between aggregate monomers in both amorphous and fibril-like populations

After measuring IbpAB binding and the impact IbpAB has on the disassembly kinetics, we wanted to address the mechanism causing disaggregation enhancement. We started by measuring the proximity of individual aggregate monomers within each protein aggregate population with and without IbpAB. To do this we employed donor side ensemble FRET to measure the global changes in proximity between aggregate monomers in both amorphous and fibril-like aggregates. RuBisCO was labeled with the

FRET pair IEADANS and fluorescein and then mixed together at a 1:0, 1:2, and 1:5 RuBisCO monomer to IbpAB dimer ratio for both aggregate types (Figure III.7). Three FRET pairs were utilized to measure global changes between aggregate monomers.

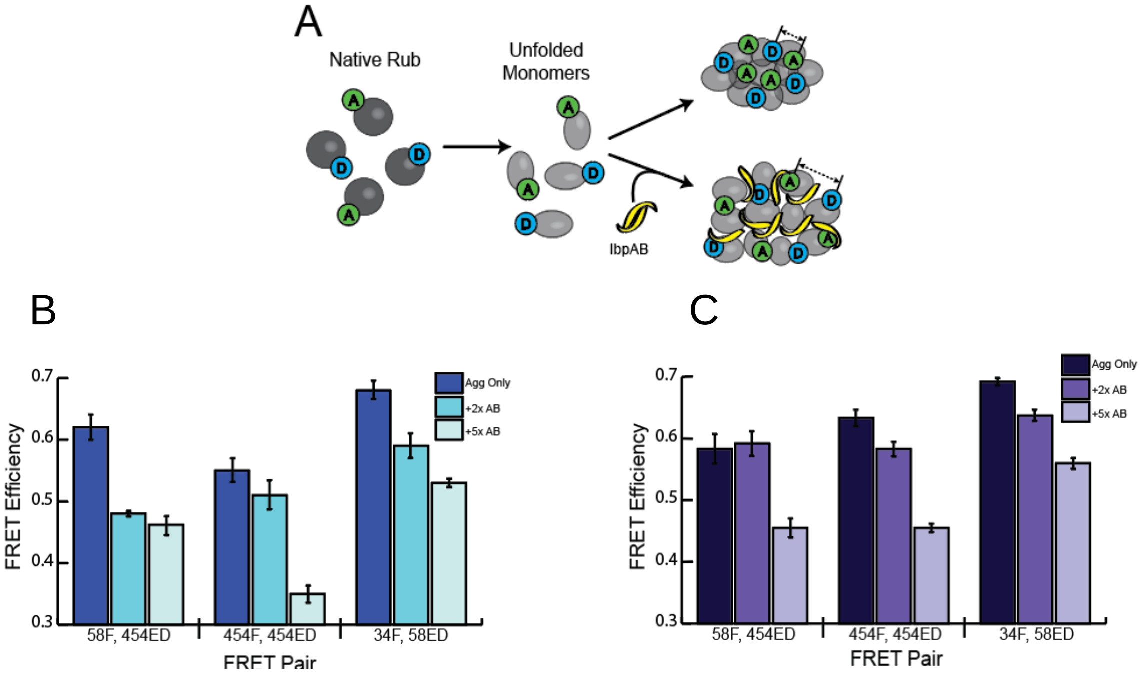


Figure III.7 IbpAB increases the proximity between aggregate monomers for both amorphous and fibril-like aggregate populations.
(A) Schematic of an experiment using ensemble inter-molecular FRET to examine the impact of IbpAB on the average relative proximity of RuBisCO monomers within aggregates.Amorphous (B) and fibril-like (C) aggregates were prepared either in the absence or presence of excess IbpAB. Samples of RuBisCO, either unlabeled, labeled with a donor fluorophore only (IAEDANS) or labeled with an acceptor fluorophore only (fluorescein) were denatured and pair-wise mixed at 1:1 ratios prior to initiation of aggregation in the presence or absence of either a 2-fold or 5-fold excess of IbpAB. Three concentration-matched samples were thus created for each aggregate type, both with and without IbpAB: donor-only, acceptor-only or donor-acceptor. The same protocol was replicated using three different paired labeling sites on the RuBisCO monomer (58F + 454ED; 454F + 454ED; 34F + 58ED; refs). FRET efficiencies are calculated from the magnitude of the corrected donor-side quenching, while enhanced acceptor fluorescence (not shown) was used to confirm Förster coupling.

At a 1:2 ratio, IbpAB causes a global decrease in FRET efficiency for both amorphous and fibril-like aggregates. The change is more significant for amorphous aggregates, but once IbpAB are increased to 1:5, the drop in FRET efficiency is very

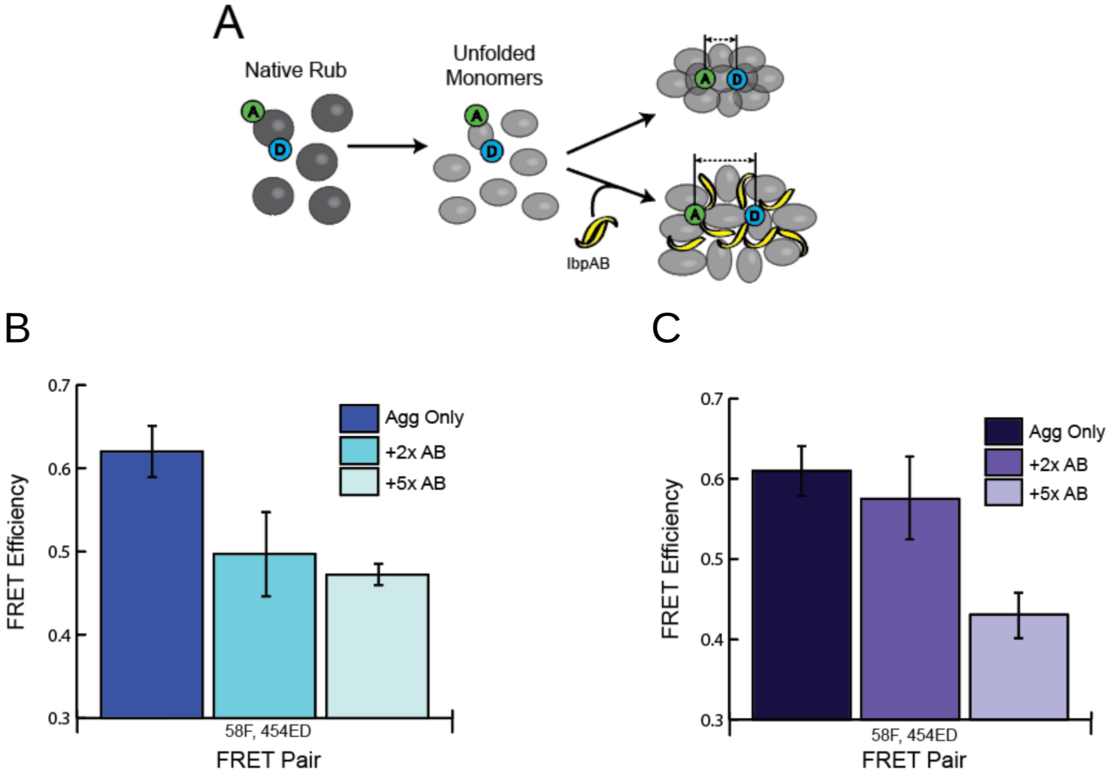


Figure III.8 IbpAB changes the internal structure of a misfolded monomer within both amorphous and fibril-like aggregate populations.
(A) Schematic of an experiment using ensemble intra-molecular FRET to examine the impact of IbpAB on the average conformation of aggregate-incorporated RuBisCO monomers. Amorphous (B) and fibril-like (C) RuBisCO aggregates were created by mixing a denatured, doubly labeled RuBisCO monomer (58F/454ED; ref) into a large excess of denatured, unlabeled RuBisCO (1:9), prior to the initiation of aggregation, in the presence or absence of a 2-fold or 5-fold excess of IbpAB. Under these conditions, Förster coupling between different labeled monomers within the same aggregate particle is minimal, so that the observed FRET signal is dominated by coupling between the probes attached to the same monomer. Concentration-matched reference samples using donor-only (454ED) and acceptor-only (58F) RuBisCO were also prepared and used to calculate the donor-side FRET efficiency, as well as confirm the present of Förster coupling. In all cases, error bars show the s.d. of three independent experimental replicates.

similar for both aggregate types (Figure III.7). This means IbpAB are capable of decreasing proximity between individual monomers within the overall aggregate population while also inhibiting aggregate growth.

IbpAB changes the conformation of a misfolded monomer within both amorphous and fibril-like populations

In addition to changing the proximity between aggregate monomers, IbpAB could also change the internal aggregate monomer conformation to assist with enhanced disassembly. To test this, donor side ensemble intramolecular FRET was employed to measure the ability of IbpAB to change individual aggregate monomer conformation. RuBisCO was doubly labeled with IAEDANS at the 454 position and fluorescein at the 58 position. Labeled RuBisCO was mixed with WT RuBisCO at a 1:9 ratio so the final aggregate population contained 10% labeled protein to limit intermolecular FRET (Figure III.8). Both amorphous and fibril-like aggregates were made at a 1:0, 1:2, and 1:5 RuBisCO monomer to IbpAB dimer ratio.

At a 1:2 ratio, a decrease in intramolecular FRET is measured for amorphous aggregates, while no significant change for fibril-like aggregates occurs. Then, when IbpAB are increased to 1:5, a similar decrease in intramolecular FRET is measured for fibril-like aggregates and amorphous aggregates (Figure III.8). Interestingly, the amorphous aggregate signal does not change between a 1:2 and 1:5 ratio of IbpAB dimer. This result differentiates intermolecular FRET from intramolecular FRET with intermolecular FRET signals decreasing in a similar manner for both aggregate types, while intramolecular FRET signal has a difference between the two aggregate types at a 1:2 ratio.

Both intermolecular FRET and intramolecular FRET suggest IbpAB are capable of changing the proximity between aggregate monomers and the internal conformation for both amorphous and fibril-like aggregates. However, a higher amount of IbpAB are needed for fibril-like aggregates to match the internal change of amorphous aggregates.

IbpAB removal from the aggregate/IbpAB complex coincides with disassembly

To directly measure if IbpAB are coating or integrating into the protein aggregates, we utilized MC-BAS to measure the coincidence of IbpAB and RuBisCO aggregates during KJEB disassembly. If IbpAB are integrated into the protein aggregate, then loss of coincident signal between IbpAB and aggregate molecules will coincide

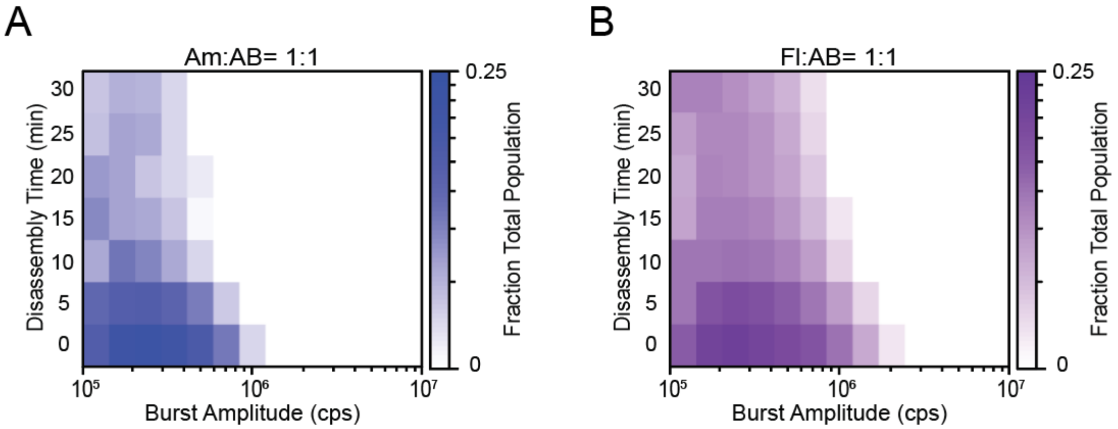


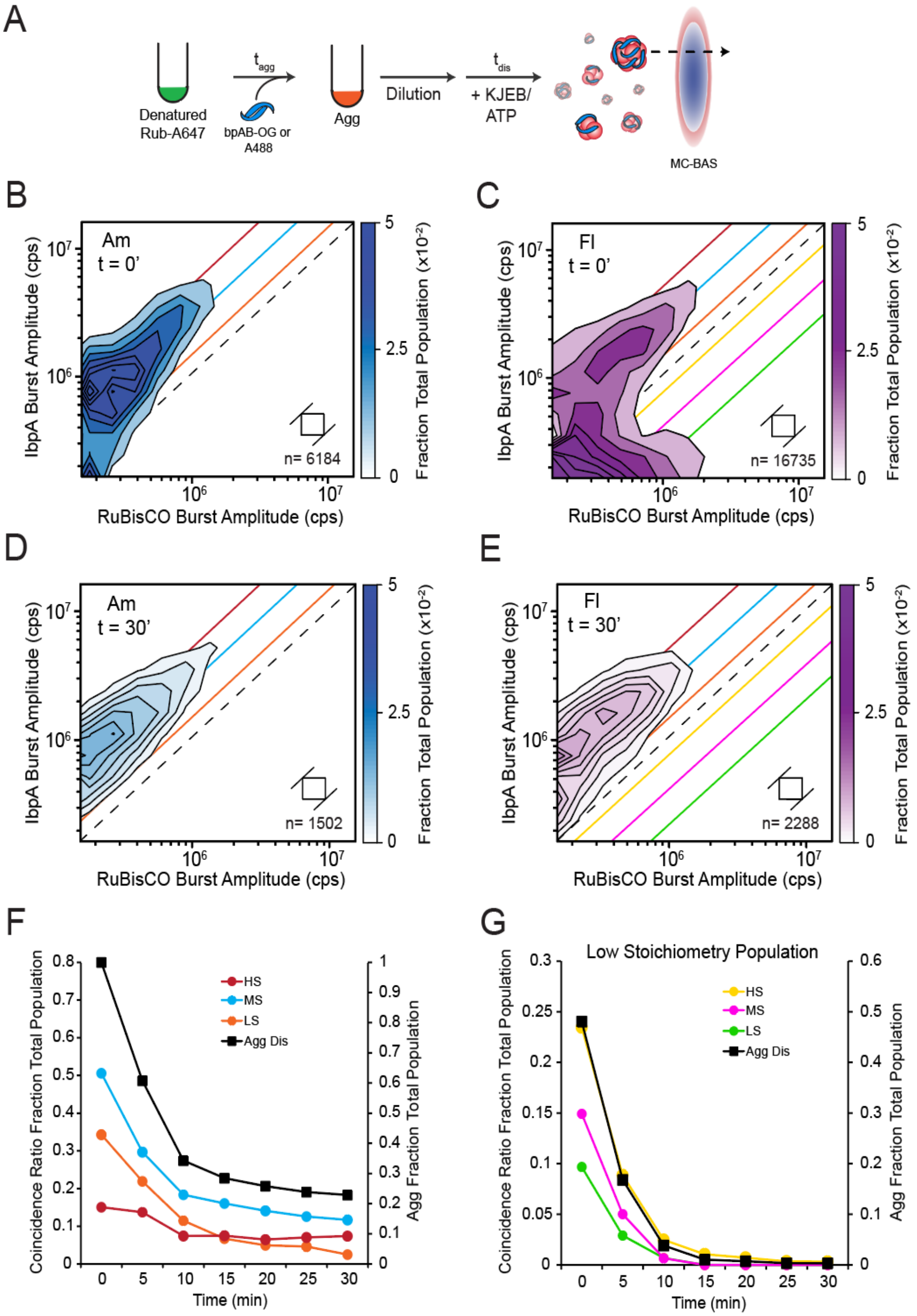
Figure III.9 Slowed Aggregate Disassembly.
Amorphous (A) and fibril-like (B) aggregation in complex with IbpAB disassembled by KJEB at a concentration of 250 nM, 500 nM, 500 nM, and 50 nM, respectively. Each plot represents three experimental replicates.

with disassembly. However, if IbpAB are forming a coat around the aggregates, then loss of coincident signal will precede aggregate disassembly.

MC-BAS experiments were carried out using RuBisCO labeled with Alexa-647, IbpA labeled with Oregon Green, and WT IbpB (Figure III.9). Aggregates were made with IbpAB at a 1:1 ratio followed by the addition of KJEB. Due to IbpAB enhanced disassembly rate, a lower KJEB concentration was used to slow aggregation and improve MC-BAS analysis (Figure III.9). For robust MC-BAS analysis, KJEB concentrations were 250 nM, 500 nM, 500 nM, and 50 nM, respectively.

Before ATP initiated KJEB activity, IbpAB binding to amorphous aggregates displayed the same 1:2 linear relationship over the entire aggregate population (Figure III.3A and III.10A). Once disassembly is initiated, a decrease in coincidence occurs over 30 minutes and corresponds with aggregate disassembly over the same time period. On a more detailed level, a simultaneous loss of coincidence occurs across the entire aggregate population with no single aggregate size being removed faster or slower than the other aggregate sizes. Continuing, IbpAB and aggregate particles lost coincidence at the same rate showing KJEB do not prioritize or prefer IbpAB removal or aggregate disassembly.

Figure III.10 IbpAB removal from RuBisCO aggregates closely tracks overall particle disassembly.
(A) Schematic of RuBisCO aggregates with IbpA-OG and IbpB-WT dimer at a 1:1 ratio. Disaggregation was triggered by the addition of the KJEB bi-chaperone system (250 nM DnaK, 500 nM DnaJ, 500 nM GrpE and 50 nM ClpB), 2 mM ATP, and a creatine kinase-based ATP regeneration system. Samples were then loaded onto a BAS microscope and burst data was continuously acquired for 30 min. Each MC-BAS plot is a combination of six, independent experimental replicates. (B) A zero-time MC-BAS measurement was collected for amorphous aggregates prior to the addition of ATP. Colored diagonals represent coarse binning of low binding stoichiometry (orange), moderate binding stoichiometry (blue), and high binding stoichiometry (red). Each bin was then utilized in (F) to represent loss of coincident signal in relation to aggregate disassembly. (C) A zero-time MC-BAS measurement was collected for fibril-like aggregates prior to the addition of ATP. The colored diagonals represent binding stoichiometry however, with the existence of two different binding populations, three additional colored diagonals were utilized to represent the lower binding stoichiometry population. (D,E) End point MC-BAS plots of both RuBisCO aggregate types show coincidence is not completely lost over the 30 minute observation window. (F) Total amorphous aggregate population versus loss of coincidence over 30 minutes of disassembly. High stoichiometry (HS), moderate stoichiometry (MS), and low stoichiometry (LS) correspond to the colored coarse bins in (B) and (D). (G) Low stoichiometry aggregate population versus loss of coincidence over 30 minutes of disassembly. High stoichiometry (HS), moderate stoichiometry (MS), and low stoichiometry (LS) correspond to the colored coarse bins in (C) and (E).



Before ATP initiated KJEB activity, fibril-like aggregates also displayed the same IbpAB binding pattern with two distinct IbpAB binding populations (Figure III.3B and III.10E). Once disassembly is initiated, the aggregate population containing the lower concentration of IbpAB are the first to be disassembled and are gone within the first 15 minutes. The population containing more IbpAB are disassembled slower and in a similar pattern as the amorphous aggregate/IbpAB complex. Also, like the amorphous aggregates, the disassembly rate of fibril-like aggregates are comparable to the loss of coincidence observed between IbpAB and the aggregates. Again, there is no evidence for KJEB removing an IbpAB coat before fibril-like disassembly or KJEB prioritizing IbpAB removal over aggregate disassembly.

IbpAB binds and stops aggregation when added later in the aggregation pathway, but does not reflect stimulated disassembly comparable to being present at aggregation initiation

To understand if IbpAB induced conformational change of aggregates is required for stimulated disassembly, IbpAB were added later in the aggregation pathway to force IbpAB to bind to a small aggregate population rather than a monomeric population. This allowed IbpAB to still limit aggregate size, but not to change any monomeric conformational state or small aggregate interactions. MC-BAS was employed to measure IbpAB binding when added later in the aggregation pathway. IbpAB were added at a 1:1 ratio to amorphous aggregates after 1 minute of aggregation and fibril-like aggregates after 30 seconds of aggregation.

When IbpAB are added 1 minute later in the amorphous aggregation pathway, IbpAB binding to the aggregates are no longer a 1:2 ratio like when IbpAB are added at

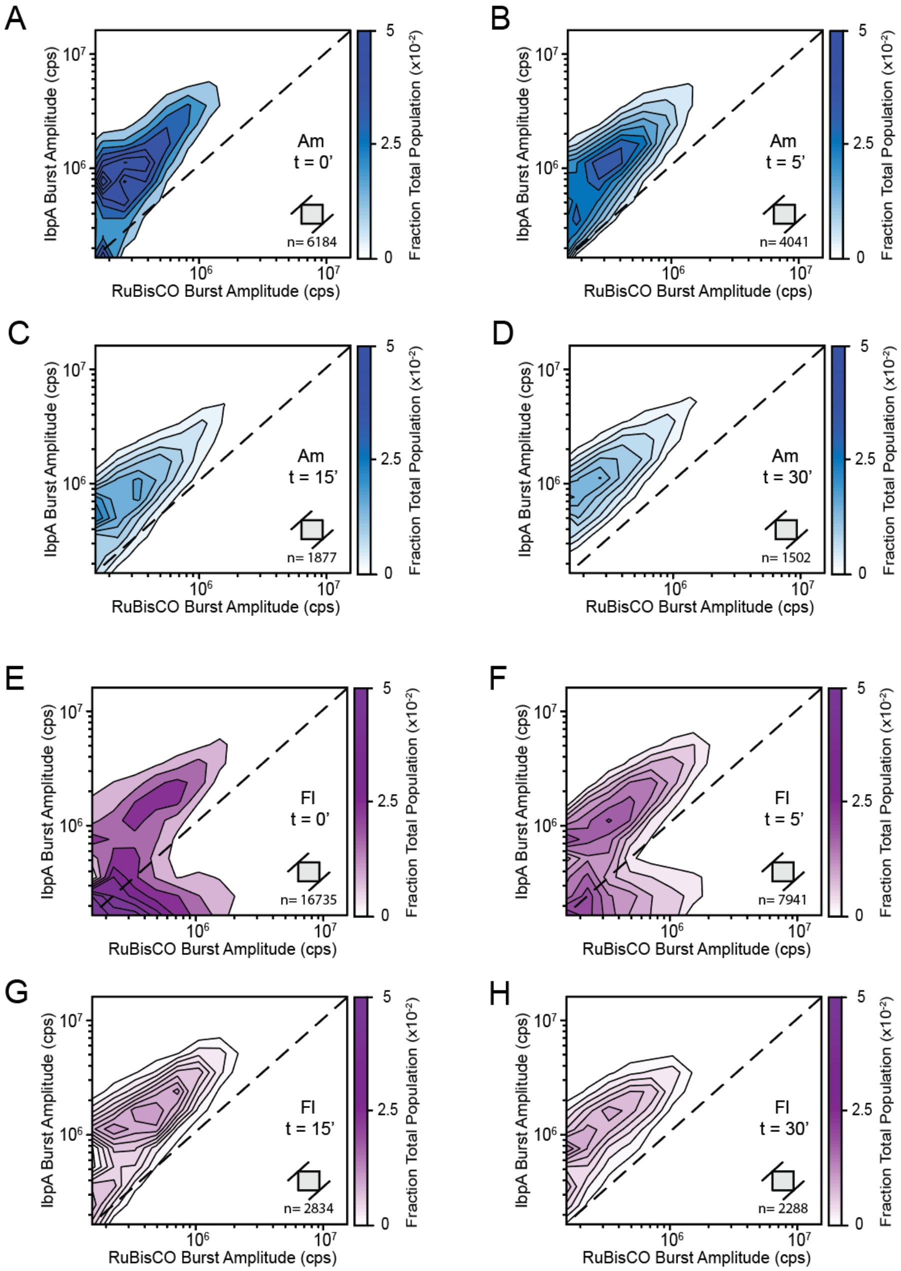
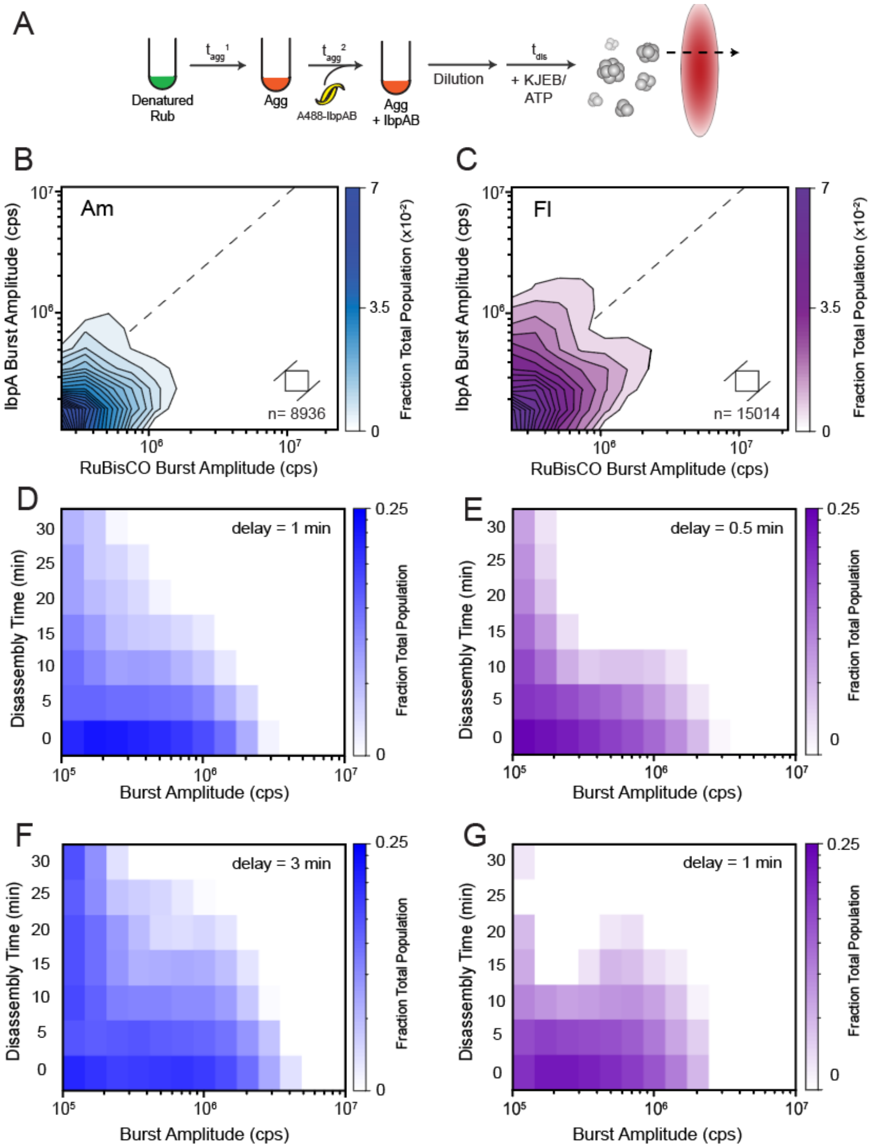


Figure III.11 IbpAB removal from RuBisCO aggregates closely tracks overall particle disassembly.

Purposefully slowed disassembly of (A-D) amorphous (blue) and (E-H) fibril-like (purple) aggregates in a complex with labeled IbpAB was measured by MC-BAS over 30 minutes. Both RuBisCO aggregate types were made at a 1:1 ratio with IbpA-OG and IbpB-WT dimer and exposed to KJEB at concentrations of 250 nM, 500 nM, 500 nM, and 50 nM respectively. MC-BAS analysis is shown at (A,E) 0, (B,F) 5, (C,G) 15, and (D,H) 30 minutes after KJEB addition. Each MC-BAS plot represents six experimental replicates.

the beginning of aggregation (Figure III.11A). Instead, the majority of the aggregate population contains less than a 1:1 ratio of IbpAB bound with no correlation between aggregate size and the amount of bound IbpAB.

Figure III.12 Delayed addition of IbpAB to RuBisCO aggregates prevents further particle growth but results in loss of stimulated disassembly.
(A) Sample handling protocol for experiments where the addition of IbpAB to an aggregating sample of RuBisCO is delayed relative to the initiation of aggregation. MC-BAS was used to examine the binding distribution of IbpAB-OG on amorphous (B) and fibril-like (C) RuBisCO-Alexa647 aggregates upon delayed IbpAB addition (tpre) in the absence of KJEB-mediated disassembly. For amorphous aggregates, tpre = 1 min and for fibril-like aggregates, tpre = 30 sec. In both cases, the final RuBisCO monomer:IbpAB mixing ratio was 1:1 and no KJEB disaggregase was added. Each MC-BAS plot is a combination of six, independent experimental replicates. The population-resolved kinetics of disaggregation by KJEB were examined with BAS following delayed addition of wild type IbpAB to amorphous (D and F) and fibril-like (E and G) RuBisCO aggregates. For amorphous aggregates, tpre = 1 min (D) or 3 min (F) at a final RuBisCO monomer to IbpAB mixing ratio of 1:1, while for the fibril-like aggregates a final mixing ratio of 1:5 and tpre = of 30 sec (E) and 1 min (G) were used. Disassembly conditions and KJEB concentrations were identical to those used in Figure 4. Each heat map is a combination of three, independent experimental replicates.



When IbpAB are added 30 seconds later in the fibril-like aggregation pathway, the bi-furcated IbpAB binding pattern is absent (Figure III.11B). Instead, the appearance of a single, broad binding population distribution appeared. This pattern is similar to IbpAB binding to amorphous aggregates when added at a later time point (Figure III.11B). The majority of the aggregate population contains less than a 1:1 ratio of IbpAB to fibril-like monomer, but more striking is the loss of the fibril-like aggregate population containing a higher concentration of IbpAB. Overall when IbpAB are added later in the aggregation pathway the binding pattern for both fibril-like and amorphous aggregates are similar, a sharp contrast to when IbpAB are present at the beginning of aggregation.

After measuring the changes in IbpAB binding to both aggregate structures, the next goal was to understand how late IbpAB addition changed the disassembly kinetics. For these experiments, IbpAB concentrations were selected based on the previous data when IbpAB were added at the beginning of aggregation (Figure III.4). This allowed for comparison between IbpAB added at the beginning of and later into aggregation. To fit this criteria, a 1:1 ratio of amorphous monomer to IbpAB dimer and a 1:5 ratio of fibril-like monomer to IbpAB dimer were selected. Furthermore, to test disassembly of different aggregate sizes, IbpAB addition occurred at an early and late time point in the aggregation pathway (Figure III.13). The early time points were 1 minute into the amorphous aggregation pathway and 30 seconds into the fibril-like aggregation pathway. The late time points were 3 minutes into the amorphous pathway and 1 minute into the fibril-like pathway. Disassembly was accomplished by KJEB at concentrations of 1 µM, 2 µM, 2 µM, and 200 nM respectively.

When added later in the aggregation pathway IbpAB are still able to inhibit continued amorphous and fibril-like aggregate growth. However, they cannot reverse aggregation that has already occurred (Figure III.13). If IbpAB are present at a 1:1 ratio at initiation, amorphous aggregates are inhibited to a majority population between 20-

200mers. When IbpAB are added after 1 minute of amorphous aggregation, the aggregate range broadens to 50-1000mers (Figure III.12C). It is even more significant when IbpAB addition is delayed by 3 minutes, with the aggregate range increasing to 50-2000mers (Figure III.12E). Fibril-like aggregates follow a similar trend. When IbpAB are added at a 1:5 ratio at the beginning of fibril-like aggregation, a majority aggregate population is between 10-50mers. When IbpAB are added after 30 seconds of aggregation, a broad population forms between 50-1000mers (Figure III.12D). When IbpAB are added after 1 minute of aggregation, the same population is measured between 50-1000mers (Figure III.12F).

When IbpAB are added 1 minute after amorphous aggregation, disassembly is stimulated when comparing to the disassembly of amorphous aggregates alone. However, the disassembly rate when IbpAB are added after 1 minute does not match the disassembly efficiency when IbpAB are added at the beginning of aggregation. Instead, a small aggregate population (around a 50mer) are resistant to KJEB and do not disassemble during the 30 minute observation window. Furthermore, once IbpAB addition is delayed to 3 minutes, the disassembly rate is more impaired coupled with an increase in the resistant aggregate population. This suggests amorphous aggregates are capable of forming a core that are resistant to KJEB disassembly unless IbpAB are present to change the internal structure.

When IbpAB are added after 30 seconds of fibril-like aggregation, disassembly is completed at a similar rate when IbpAB are present at initiation of aggregation. The exception is a disassembly resistant population similar to the amorphous aggregates

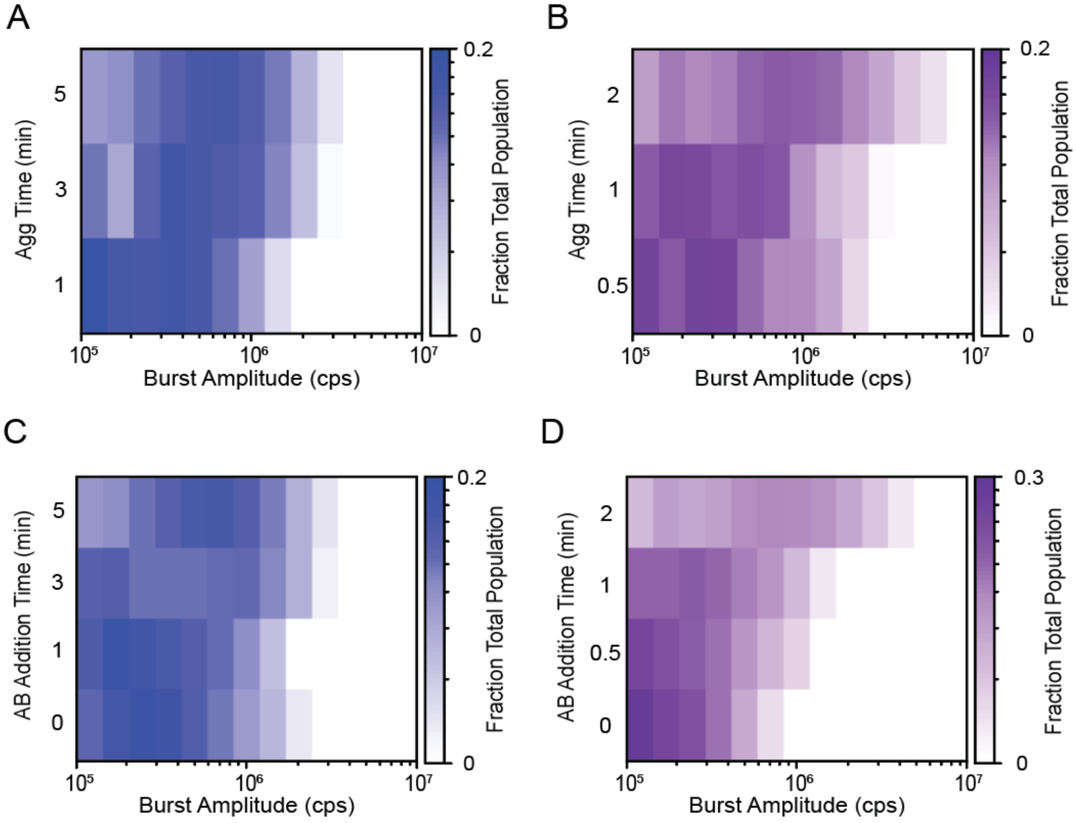


Figure III.13 Aggregate and IbpAB Addition Time Course.
Amorphous (A) and fibril-like (B) aggregate population over time measured by BAS. Each aggregate population was grown to the time on the y-axis for (A) and (B) then diluted to stop aggregation. IbpAB were added at different time points during amorphous (C) and fibril-like (D) aggregation. Once IbpAB were added, the aggregation pathway was taken to 5 minutes for amorphous aggregates and 2 minutes for fibril-like aggregates.

when IbpAB are added late. The trend does not hold when IbpAB are added 1 minute after aggregation. Disassembly under these conditions appears similar to disassembly when IbpAB are present at a 1:1 ratio, not a 1:5 ratio.

Discussion

The sHsp are identified through their association with misfolded and aggregated proteins when cells are under stress. This association inhibits aggregate size while also assisting in accelerated and more efficient substrate reactivation by the bi-chaperone disaggregase system. A key question is whether sHsp achieve this function by coating the aggregate to exclusively limit aggregate size, integrating into the aggregate to change the aggregate structure, or a combination where two populations are used to integrate and coat an aggregate population. We examined this issue with the *E. coli* sHsp IbpA and IbpB with two substrate proteins RuBisCO and PepQ. First, we found IbpAB limits aggregate size of both aggregate substrates and two aggregate structures, consistent with other observations. Furthermore, our data supports and expands the integration model where IbpAB are capable and required to change the aggregate structure to enhance aggregate disassembly.

Early images of sHsp oligomers and experimental results made a strong argument for sHsp utilizing a coat to inhibit aggregate size. The coating model continued to be supported when evidence suggested disaggregase chaperones had faster disaggregation rates for smaller aggregates compared to larger aggregates. By extension, this proposed sHsp were needed to exclusively inhibit aggregate growth to assist the disaggregase chaperones.

More in depth studies followed with a focus on the homogeneous sHsp family members. These studies suggested sHsp were able to integrate within the aggregate and hold the substrate in a conformation similar to the native state, leading to the sHsp integration mechanism of action.

The work laid out here with IbpAB and RuBisCO are consistent with the studies on the homogeneous sHsp. We measured the capabilities of IbpAB to decrease the proximity of the fluorescent probes between individual monomers as well as within the monomer itself for both aggregate types. This demonstrates a key similarity between homogenous sHsp and heterogeneous sHsp, as well as supports the theory that sHsp can change an aggregate conformation rather than exclusively inhibit aggregate size.
We also observed IbpAB inhibiting aggregates to a more homogenous and specific size, independent of aggregate structure or protein substrate. The majority of these aggregates are 15-20mers with the aggregate population distribution heavily dependent on IbpAB concentration. Importantly, even at a high concentration, IbpAB are unable to drive an aggregate population to monomers, but instead increase the homogeneity of the aggregate population.

To make this single aggregate population, IbpAB bind at a ratio of 1:2 RuBisCO monomer to IbpAB dimer over the entire amorphous population. This binding pattern is possible through two mechanisms. The first mechanism is IbpAB having a consistent binding affinity to aggregates over the entire size range, which means IbpAB have the same affinity for aggregated substrate regardless of aggregate size. The second mechanism is IbpAB preferably binds monomers or small aggregates which come together to create the entire aggregate population. To test the possible models, we added IbpAB later in the aggregation pathway. If IbpAB preferably bind small amorphous aggregates, then adding IbpAB later in the aggregation pathway will cause a shift in the binding ratio of amorphous aggregate to IbpAB dimer. If IbpAB have a consistent binding affinity to any aggregate size, then we will not see a change in binding ratio when IbpAB are added later in the aggregation pathway. Indeed we measure a decrease in the binding ratio as highlighted in figure 8. This supports IbpAB preferentially binding monomeric or small aggregates which then come together to form the entire aggregate population.

For fibril-like aggregates, IbpAB display two binding populations where one population consists of a 1:2 ratio of fibril-like monomer to IbpAB dimer and the other population contains the same amount of IbpAB dimer across the entire aggregate size population. These two binding populations suggest there are either two distinct aggregate populations form during fibril-like aggregation or this is an intermediate state where IbpAB are actively converting one population into the other population. In both cases the higher binding population resembles IbpAB binding to amorphous aggregates while the lower binding population contains the same amount of IbpAB binding sites across the entire population and is consistent with an end capping distribution. To test these models, IbpAB was added later in the fibril-like aggregation pathway. If there are two distinct aggregate populations within the fibril-like aggregates, then we will measure the same bi-furcation when IbpAB are added later in the pathway since IbpAB are not present to force a change in aggregate structure. If this is an intermediate state, then we will observe a loss of one of the populations since IbpAB are not present to force an aggregate structural change to create the second aggregate population. In our experiments we measure a loss of the second population which suggests fibril-like aggregates are a single population and are forced into a new aggregate structure when IbpAB are present. This, in turn, would make an aggregate population similar to the amorphous aggregates which are more disassembly competent.

Further evidence supporting the integration model is observed during disassembly. We utilized MC-BAS to directly measure the coincidence of aggregates and IbpAB during disassembly. If IbpAB coat the aggregates, then we would measure a loss of coincidence between IbpAB and aggregates before disassembly occurs, as there would be a loss of the IbpAB coat before KJEB could disassemble the aggregates. If IbpAB integrated into the aggregate, then we would measure a loss of coincidence between IbpAB and aggregates simultaneously with disassembly. Indeed, this is what we observed where disassembly of both amorphous and fibril-like aggregates coincided with IbpAB release from the IbpAB/aggregate complex. If IbpAB were not integrated into the aggregate, there would be a shift towards the x-axis on the MC-BAS plot signaling a loss of IbpAB while the aggregates stay the same size. This would be followed by disassembly of the protein aggregates. We did not measure a release of an IbpAB coat followed by disassembly of amorphous or fibril-like aggregates. Instead, aggregate disassembly with IbpAB is a cooperative process without a preference given to a specific aggregate size. Together, these data argue IbpAB are integrated into the aggregates to form a mixed complex where a new aggregate structure is created to stimulate disassembly.

While these observations agree with the integration model, it was unclear if stimulated disassembly was caused by IbpAB changing the aggregate conformation or from IbpAB binding to the aggregate without changing the conformation. To address this, IbpAB were introduced into the aggregation pathway after the formation of misfolded monomer intermediates and small aggregates. This enabled IbpAB to bind and limit aggregation, but not change any monomer folded state or small aggregate interactions. For both amorphous and fibril-like aggregates, adding IbpAB late does enhance disassembly rates when compared to the rates without any IbpAB present. However, the disassembly rates when IbpAB are added late do not match the disassembly rates when IbpAB are present at the start of aggregation. Taking into account the disassembly and MC-BAS data, IbpAB are capable of binding small aggregate populations and inhibiting them from growing or at the least, slowing down continued aggregation. However, IbpAB are more efficient when they are present at the start of aggregation and can change both the monomeric conformational state and the interactions between monomers in order to fully enhance disassembly.

These findings are consistent with a heteromeric sHsp population that integrates with a protein aggregate population to change the monomeric interactions and conformations, resulting in stimulated aggregate disassembly by KJEB. An important next step is to understand how sHsp utilize different domains or interfaces during aggregate inhibition. It is also important to understand how IbpAB interact with KJEB during the disassembly process.

CHAPTER IV

CONCLUSIONS AND DISCUSSION

Conclusions

In the past, members of the sHsp chaperone family were identified to inhibit protein aggregation and increase protein reactivation by the disaggregase system (123-127). However, the mechanisms used to carry out these functions have not been elucidated. This is a result of the sHsp forming a wide range of oligomer sizes and interactions. To understand the sHsp mechanisms and role of oligomeric complexes we selected the *E. coli* sHsp IbpA and IbpB as a model.

Three prevalent models describe how the sHsp create smaller aggregates and enhance disassembly: (1) creation of a coat around the aggregates (99, 101), (2) integration within the aggregates, changing the internal aggregate structure (129), and (3) a combination of both models where two distinct sHsp populations are simultaneously coating the aggregate and integrating within the aggregate (130). In models (1) and (3), the Hsp70 chaperone must release the sHsp coat from the aggregate before disassembly can occur.

Extensions of Burst Analysis Spectroscopy

To test these proposed models, our laboratory developed a single particle fluorescence method capable of measuring dynamic heterogeneous nanoparticle populations in free solution, referred to as burst analysis spectroscopy (BAS). This initial platform was ideal for single heterogeneous populations (133), but additional development was needed to quantify two interacting nanoparticle populations within the solution as well as expand the quantifiable size range. Multicolor-BAS, or MC-BAS, was developed to measure up to three interacting heterogeneous populations and quantify the binding ratio of the interacting nanoparticles. The second BAS expansion was developed to measure and quantify an additional order of magnitude in population size. This development is called concatenated-BAS, or C-BAS. C-BAS allows aggregate populations to be quantified for four orders of magnitude, as opposed to the two orders of magnitude using conventional BAS. These two developments made BAS an ideal technique to distinguish between the three sHsp mechanistic models.

IbpAB restrict aggregate growth of different aggregate types to a common particle size.

Using BAS, we measured IbpAB inhibiting aggregates to a more homogenous and specific size, independent of aggregate structure or protein substrate. As IbpAB concentration was increased, the majority of each aggregate population were driven to a size range between 20-30mers. For RuBisCO, the fibril-like aggregates were more susceptible to IbpAB inhibition compared to amorphous aggregates. For substrates, PepQ was more susceptible to IbpAB inhibition than RuBisCO aggregates. A majority of the PepQ aggregate population reached a range of 15-20mers at a 1:1 ratio of PepQ monomer to IbpAB dimer while RuBisCO aggregates did not reach this size range until a 1:5 ratio of RuBisCO monomer to IbpAB dimer was utilized. This is the first study showing the capabilities of IbpAB with a native substrate and the first indication of how effective IbpAB are *in vivo*.

Even with IbpAB capable of limiting protein aggregation, they were not able to drive an aggregate population to a monomeric population. Instead, an increase in IbpAB concentration creates a more homogeneous aggregate population. This suggests aggregate disassembly relies on structural or size homogeneity rather than creating a small aggregate population. While these observations validate BAS as a useful tool, the results do not distinguish a difference between the three mechanistic models.

IbpAB display a distinct binding behavior toward amorphous and fibril-like aggregates.

With IbpAB capable of inhibiting both amorphous and fibril-like aggregates, we wanted to test how IbpAB were binding to the aggregates. To do this, we utilized MC-BAS to measure the binding pattern of IbpAB on both amorphous and fibril-like aggregates.

For amorphous aggregates, IbpAB have a consistent 1:2 binding ratio of RuBisCO monomer to IbpAB dimer over the entire aggregate size population. There are two possibilities for how IbpAB bind to amorphous aggregates to create the observed linear binding pattern. The first possibility is IbpAB bind or capture non-native monomers or small amorphous aggregates to create an IbpAB/aggregate complex across the entire size population. This aligns with the integration model where IbpAB changes the aggregate structure across the entire size population. The second possibility is IbpAB bind to the surface of the amorphous aggregates in an ordered manner. This possibility is consistent with a coat being created around the aggregate and requires the binding ratio to be dependent on the available surface binding sites. This availability would scale with the aggregate size therefore, causing the linear binding population.

To test these possibilities and subsequently aligned models, we added IbpAB later in the aggregation pathway. If IbpAB are integrated into the aggregates, then later addition in the aggregation pathway will cause the binding ratio to change. This is predicted because IbpAB are not able to bind to the non-native monomers or very small aggregates, but instead IbpAB would bind to a medium sized aggregate population. This would increase the RuBisCO amount in each IbpAB/aggregate complex and move the ratio away from 1:2. If IbpAB are coating the aggregate, then later IbpAB addition would result in a similar binding pattern to when IbpAB are added at the beginning of aggregation. This is because the aggregate surface area dictates the IbpAB binding ratio rather than requiring IbpAB to be integrated into the aggregate. The surface area would stay consistent in both experiments, therefore causing the IbpAB binding ratio will be consistent.

IbpAB added later in the amorphous aggregation pathway resulted in the binding ratio changing. This result supports IbpAB integrating into the amorphous aggregates through binding non-native monomers or very small aggregates.

For fibril-like aggregates, IbpAB displays a different binding pattern compared to the binding pattern with amorphous aggregates. Instead of a single binding ratio over the entire aggregate size range, IbpAB have two distinct binding populations for the same sized aggregates. One of these populations consists of a 1:2 ratio of fibril-like monomer to IbpAB dimer, while the other population has a consistent amount of IbpAB dimer across the entire aggregate size range. The 1:2 ratio population resembles the amorphous population; while the lower IbpAB binding population resembles fibril aggregate end-capping with the only available binding sites on the end of the fibril-like aggregates. Therefore, as the aggregates grow, the only possible way to stop continued growth is for IbpAB to bind to the exposed ends.

There are two possibilities that would lead to simultaneous formation of two binding populations. The first possibility is IbpAB are able to force the fibril-like aggregates into a new aggregate structure and we are measuring a transition point where two structural populations exist. This possibility aligns with the integration model because IbpAB cause a change in aggregate structure instead of simply inhibiting aggregate size. The second possibility is the fibril-like aggregates made in this study contain two structurally different aggregates, with each population differing in the amount of IbpAB binding sites. This possibility aligns with the coating model because the aggregate structure dictates the amount of binding rather than IbpAB inducing changes in the aggregate structure.

To test both possibilities, we added IbpAB later in the aggregation pathway, similar to the amorphous aggregates. If there are two distinct aggregate populations within the fibril-like aggregate sample, then we should measure the same bi-furcation pattern. If IbpAB are forcing a change in structure, then we should lose one of the binding populations since IbpAB would not be present to force the formation of a new aggregate structure.

IbpAB added later in the fibril-like aggregation pathway resulted in losing the 1:2 binding ratio population. This result supports IbpAB changing the fibril-like aggregate structure to one similar to the IbpAB/amorphous aggregate complex. This result also supports the integration model where IbpAB are capable of changing the aggregate structure to inhibit further aggregation.

IbpAB binding impacts the internal structure of both amorphous and fibril-like aggregates

To directly measure if IbpAB are capable of changing aggregate structure, we utilized established FRET pairs on the RuBisCO molecule to measure the average change in proximity between monomers in an aggregate population. If IbpAB were capable of integrating within the aggregate, then a change in the average FRET signal would be measured. On the other hand, if IbpAB coated the outside of the aggregate, then the average FRET signal would not change. For both amorphous and fibril-like aggregates, we measured a decrease in FRET efficiency when IbpAB were present. Importantly, this decrease was measured for all FRET pairs during both amorphous and fibril-like aggregation. This suggests IbpAB are capable of integrating within the aggregate and changing the intermolecular space between individual monomers.

Since IbpAB are able to change the proximity between aggregate monomers, we also tested if IbpAB induce a change in the aggregate monomer conformation. Similar to the intermolecular FRET results, the intramolecular FRET signal decreased for both aggregate types in the presence of IbpAB. While we only tested a single FRET pair, these data suggest that IbpAB cause the individual monomers to be in a more extended conformation compared to when IbpAB are no present. The more extended conformation would assist in faster aggregate disassembly by the KJEB system because less energy would be required to completely unfold the aggregated substrate.

Overall, these data supports IbpAB causing a global change in aggregate structure. We speculate that IbpAB inhibit aggregate growth by changing the non-native monomer conformations and interactions to form a newly structured aggregated/IbpAB complex.

Amorphous and fibril-like aggregates disassemble at different rates

The ability of KJE and KJEB to disassemble two structurally different RuBisCO aggregates has not been explored. In collaboration with a previous graduate student, vital questions were answered about the impact RuBisCO aggregate structure and time of aggregation has on KJE and KJEB disassembly.

Amorphous aggregates grown for more than two minutes require the full KJEB system to be completely disassembled. However, after 5 minutes of aggregation, KJEB were not able to completely disassemble the aggregate population. This solidifies an amorphous aggregate growth timeframe in which KJEB disassembly can be studied.

For fibril-like aggregates, KJE were able to drastically reduce the amount of aggregates grown for 30 seconds while KJEB were sufficient for complete disassembly. However, when aggregates were grown for 2 minutes, KJE were highly ineffective with KJEB being more effective but unable to completely disassemble the aggregates.

Through this work, parameters were established for a relevant RuBisCO aggregate size range, aggregation timeframe, and amount of disaggregation machinery required for RuBisCO aggregate disassembly. Using these parameters, the ability of IbpAB to assist in aggregate disassembly was explored.

IbpAB stimulates the disassembly rate and efficiency of both amorphous and fibril-like aggregates as well as the endogenous substrate, PepQ

After establishing the disassembly parameters of amorphous and fibril-like RuBisCO aggregates, the next step was to understand if IbpAB enhanced the efficiency and rate of KJEB disassembly for both aggregate structures and PepQ aggregates. Using BAS, we exhibited IbpAB enhancement of the efficiency and rate of disassembly for PepQ, amorphous, and fibril-like aggregates. The disassembly kinetics for each reaction were heavily dependent on IbpAB concentration and did not correlate with how efficient IbpAB inhibited each aggregate type. The non-correlation was highlighted by IbpAB interaction with amorphous aggregates. IbpAB did not drastically inhibit amorphous aggregate size when present at a 1:1 ratio; however for disassembly, a 1:1 ratio drastically stimulated and enhanced amorphous disassembly compared to amorphous aggregates alone. The opposite was measured when IbpAB were present during PepQ aggregation. At a 1:1 ratio, IbpAB were able to inhibit PepQ aggregate size more than both RuBisCO aggregate structures. However for PepQ disassembly, enhanced disassembly did not occur until IbpAB were present at a 1:5 ratio.

These observations support aggregate structure, or the ability for IbpAB to change the aggregate structure, dictating disassembly rates and efficiency, not the size of the aggregate. Combining the aggregate disassembly and FRET data, amorphous aggregates seem to require a, relatively, lower concentration of IbpAB to change monomer-monomer interactions and monomeric structure. These structural changes then lead to an increased disassembly rate and higher efficiency. Fibril-like aggregates require a higher IbpAB concentration to equal the structural changes and disassembly rates of amorphous aggregates. Once the FRET efficiency is similar to the amorphous aggregate FRET efficiency, the disassembly rate and efficiency are also similar.

IbpAB integrates into the aggregates to stimulate disassembly

To directly measure a difference between IbpAB coating the aggregates and IbpAB integrating into the aggregates, we utilized MC-BAS to measure the coincidence of protein aggregates and IbpAB during real time disassembly. If IbpAB coat the aggregates, then we would measure a loss of coincidence between IbpAB and aggregate particles before disassembly. This would occur because of the mandatory removal of the IbpAB coat by KJEB before aggregate disassembly. If IbpAB are integrated into the aggregate, then we would measure a loss of coincidence between IbpAB and aggregates simultaneous to KJEB aggregate disassembly. For both amorphous and fibril-like aggregates we measured simultaneous disassembly and loss of IbpAB and aggregate coincidence. These measurements support the FRET and single color disassembly results where IbpAB are integrating into the aggregate to assist with enhanced disassembly.

IbpAB must change both the monomeric and aggregate structure to stimulate disassembly

While, these data agree with IbpAB integrating into the aggregates, it was unclear if IbpAB were required to change the internal aggregate structure or if they could simply bind to the aggregate to assist in aggregate disassembly. To probe this, IbpAB were introduced later in the aggregation pathway after small aggregates and misfolded monomer intermediates had formed. This enabled IbpAB to bind to the aggregates and limit aggregation, but not force any changes in the monomer folded state or small aggregate interactions.

For both amorphous and fibril-like aggregates, late IbpAB addition enhanced disassembly when compared to disassembly without IbpAB, but could not match the disassembly rate when IbpAB were present at the start of aggregation. These data support a model where IbpAB need to change both the monomeric folded state and the interactions between monomers to assist in fully enhanced disassembly. These data do not support IbpAB forming a coat around an aggregate to enhance disassembly. This is because during late IbpAB addition, aggregate size was still inhibited yet KJEB disassembly rate was not fully enhanced. If IbpAB were needed to simply form a coat around the aggregates and inhibit aggregate size, then the same enhanced disassembly rate would be measured.

Final Model

This body of work supports a model where IbpAB (or sHsp) first bind non-native substrate in a different fold/conformation which inhibits continued protein aggregation. The IbpAB binding leads to the formation of a homogenous aggregate/IbpAB complex with the aggregate monomers separated from each other. This change in substrate conformation and separation from other monomers leads to enhanced disassembly by the bi-chaperone system. Finally, during enhanced disassembly IbpAB and aggregate monomers are removed in unison from the complex.

Future Directions

Further BAS expansions

Burst Analysis Spectroscopy is capable of additional particle population analysis through the incorporation of single molecule FRET (smFRET). By integrating smFRET, this would allow for distance/proximity measurements to be made within a nanoparticle population. An example to understand the added value smFRET would provide are the ensemble FRET experiments used to understand the impact IbpAB have on aggregate structure. In the ensemble FRET measurements, we assumed the changes in FRET efficiency induced by IbpAB could be applied across the entire aggregate population. This is because we cannot measure FRET efficiencies for individual aggregate sizes. By integrating smFRET into the BAS setup, it would allow for simultaneous FRET and BAS measurements to help specify if IbpAB caused a single aggregate size to undergo a major structural change, or if the entire particle population is undergoing the same structural change. By integrating smFRET, it would be possible to get a more detailed picture of the aggregate population under different conditions.

A more advanced BAS develop would be the incorporation of a microfluidics system for enhanced sample mixing and data collection. Currently, samples are prepared and placed on a coverslip before data collection. During this process, a new coverslip is needed for each experimental replicate along with resetting the microscope objective and scanning volume. Because of this, there is a limit on early experimental time point measurements during population resolved kinetic experiments. On top of this, the current sampling method (a rotating stage) requires the acquisition timeframe to be over three minutes to collect the required number of events for robust statistical analysis. Because of the required three minutes of acquisition time, we are unable to measure time points under three minutes and becomes a problem when the kinetic rates are faster than the minimum acquisition timeframe.

Using microfluidics, it is possible to overcome these barriers by using automated sample mixing and sample flow in a microfluidic chamber. Automated mixing will replace sample and microscope preparation which allows for collecting early kinetic time points. Sample flow through a microfluidic chamber removes the limit on acquisition time since we could position the scanning volume at any time point along the kinetic pathway while the sample moves through the chamber.

To summarize, microfluidics would allow for unlimited acquisition time at any experimental time point for any BAS technique. A simple aggregate disassembly assay could be measured within seconds of starting the disassembly process compared to the five minute time point to which we are currently limited. Also, with no limit on the acquisition time, more time points throughout the disassembly process would be accessible and lead to more robust disassembly kinetic information. Not to mention, this could occur while the experimenter is off at lunch or completing other tasks, a substantial increase in productivity.

IbpAB induced DnaK binding to aggregates

The mechanisms used by KJEB to enhance disassembly of aggregate/IbpAB complexes are not well known. While this work contains compelling evidence that IbpAB change the aggregate structure, it is unknown how the structural change makes KJEB more efficient. The two models proposed to make KJEB more efficient are: 1) IbpAB induced structural change increases DnaJ binding sites or 2) IbpAB induced structural change creates a more disassembly susceptible aggregate state that requires less energy to disassemble.

To test these models, MC-BAS can be used to measure DnaK binding to an aggregate with or without IbpAB. If IbpAB increase the amount of DnaJ binding sites, then we will measure an increase in DnaK binding to the aggregate/IbpAB complex compared to aggregates alone. If IbpAB create a more disassembly susceptable aggregate, then DnaK binding will not increase in the presence of an aggregate/IbpAB complex compared to aggregates alone.

Mapping the locations of IbpAB and KJEB on protein aggregates

After measuring the difference in IbpAB binding to amorphous versus fibril-like aggregates, a next step is to understand spatial location of IbpAB on the aggregates. While this data might not be useful for amorphous aggregates, the spatial location of IbpAB on fibril-like aggregates is needed. A technique that can be used to understand spatial locations of interacting particles is Tip-Enhanced Raman Spectroscopy (TERS) and would be accomplished through a collaboration with Dr. Dmitry Kurouski. TERS is capable of providing nanometer level spatial resolution of particle complexes (138) through differentiation of particle specific chemical signatures. Following our protocols, the RuBisCO and PepQ aggregates are large enough to fit the resolution requirements, however protein aggregates are notorious for being heterogeneous and do not contain predominant chemical signatures. Not only is this true for aggregates, but also for the sHsp, which are structurally dominated by intrinsically disordered domains.

To solve this problem, we plan on utilizing the same fluorescent tags attached to the substrate and IbpAB molecules during BAS experiments to obtain the needed specific chemical signatures. As long as the fluorescence profiles do not interfere with the raman profiles, utilizing the fluorescent tags would allow us to differentiate between protein aggregates, sHsp, and other chaperone families.

The ability to visualize spatial locations of chaperones on protein aggregates would allow us to differentiate between the sHsp coating model and the sHsp integration model. For the sHsp integration model, the TERS results would show a mixture of protein aggregates and IbpAB throughout the entire particle population. While the sHsp coating model would result in IbpAB surrounding the aggregate population and give a strong IbpAB signal. However, we would not be able to differentiate the coating model from the model with two IbpAB populations.

Finally, by utilizing TERS we could explore the spatial locations of KJEB during aggregate disassembly with or without IbpAB. KJEB disassembly of a protein aggregate could occur through two mechanisms. In the first mechanism, KJEB would cover the entire aggregate though utilization of multiple disassembly sites or multiple substrate recognition sites. In the second mechanism, KJEB would locate to specific disassembly loci where all the disassembly machinery congregate, creating a very high local concentration of chaperones. Both mechanisms would be distinguishable utilizing TERS and would help to understand the mechanisms used by KJEB for aggregate disassembly as well as the role IbpAB plays in this process.

Mutation or deletion induced changes in IbpAB substrate binding

Functional and structural diversity among the sHsp family members has let to different domains being required to inhibit or interact with specific protein substrates and aggregate structures depending on the individual sHsp (97, 118). However, little progress has been made to understand the mechanisms behind sHsp domain dependence for certain functions. This is mostly due to the utilization of ensemble techniques like aggregate turbidity and protein reactivation, which are unable to measure sHsp binding and aggregate disassembly. Using BAS, we can uncover mechanisms used by certain sHsp domains to inhibit aggregation or bind to specific aggregate structures. This would provide insight into whether certain mutations or entire domains might utilize a different binding pattern to inhibit aggregates compared to WT sHsp. These answers would be highly valuable in the sHsp community as we start to address sHsp hyperactivity and explore sHsp engineering for possible aggregate prevention *in vivo*.

Using BAS with human sHsp

This work established BAS capable of measuring IbpAB hetero-oligomer populations alone and in a complex with protein substrates. With IbpAB as a model, unknown mechanism of additional sHsp family members are now within reach. A current project is understanding the mechanisms utilized by the human sHsp αB-crystallin (αB) during assembly of the intermediate filament protein, desmin (139-141). It is known αB binds and stabilizes many oligomeric proteins that form filaments, tubules, and fibrils, however the mechanisms used have not been highly studied. Desmin requires αB stabilization during assembly into intermediate filaments. Once fully formed, desmin stabilizes and organizes sarcomeres within muscle tissue. In the absence of αB, or when certain αB mutations are present, desmin filaments are destabilized and result in irreversible desmin aggregation. This aggregation leads to skeletal and cardiac myopathies, referred to as desmin related myopathy (DRM) (141).

Currently, researchers do not have the capabilities to study real time desmin filament assembly or αB interaction during the assembly process. Using BAS, we want to understand αB and desmin interaction during filament formation and why αB is needed during the process. This analysis could serve as a platform to understand why desmin filament formation goes awry when certain αB mutations are present.

REFERENCES

1. Fruton JS. Early theories of protein structure. Annals of the New York Academy of Sciences. 1979;325:xiv, 1-18. Epub 1979/05/31. doi: 10.1111/j.1749-6632.1979.tb14125.x. PubMed PMID: 378063.

2. Sanger F. The free amino groups of insulin. The Biochemical journal. 1945;39(5):507-15. Epub 1945/01/01. doi: 10.1042/bj0390507. PubMed PMID: 16747948; PMCID: PMC1258275.

3. Stretton AO. The first sequence. Fred Sanger and insulin. Genetics. 2002;162(2):527-32. Epub 2002/10/26. PubMed PMID: 12399368; PMCID: PMC1462286.

4. Mirsky AE, Pauling L. On the Structure of Native, Denatured, and Coagulated Proteins. Proceedings of the National Academy of Sciences of the United States of America. 1936;22(7):439-47. Epub 1936/07/01. doi: 10.1073/pnas.22.7.439. PubMed PMID: 16577722; PMCID: PMC1076802.

5. Molecular Weight Analysis In Centrifugal Fields. Science. 1934;79(2050):327-32. Epub 1934/04/13. doi: 10.1126/science.79.2050.327. PubMed PMID: 17780055.

6. Pauling L, Corey RB. The Configuration of Polypeptide Chains in Proteins. In: Zechmeister L, editor. Fortschritte der Chemie Organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products / Progrés dans la Chimie des Substances Organiques Naturelles. Vienna: Springer Vienna; 1954. p. 180-239.

7. Pauling L, Corey RB, Branson HR. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proceedings of the National Academy of Sciences of the United States of America. 1951;37(4):205-11. Epub 1951/04/01. doi: 10.1073/pnas.37.4.205. PubMed PMID: 14816373; PMCID: PMC1063337.

8. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958;181(4610):662-6. Epub 1958/03/08. doi: 10.1038/181662a0. PubMed PMID: 13517261.

9. Dobb MG, Fraser RD, Macrae TP. The fine structure of silk fibroin. The Journal of cell biology. 1967;32(2):289-95. Epub 1967/02/01. doi: 10.1083/jcb.32.2.289. PubMed PMID: 10976222; PMCID: PMC2107248.

10. SCHERAGA HA. Protein Structure and Function, From a Colloidal to a Molecular View. Carlsberg Res Commun. 1983;49:1-55.

11. Scheraga HA, Nemethy G, Steinberg IZ. The contribution of hydrophobic bonds to the thermal stability of protein conformations. The Journal of biological chemistry. 1962;237:2506-8. Epub 1962/08/01. PubMed PMID: 14498257.

12. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973;181(4096):223-30. Epub 1973/07/20. doi: 10.1126/science.181.4096.223. PubMed PMID: 4124164.

13. Cheung MS, Garcia AE, Onuchic JN. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(2):685-90. Epub 2002/01/24. doi: 10.1073/pnas.022387699. PubMed PMID: 11805324; PMCID: PMC117366.

14. Anfinsen CB, Haber E. Studies on the reduction and re-formation of protein disulfide bonds. The Journal of biological chemistry. 1961;236:1361-3. Epub 1961/05/01. PubMed PMID: 13683523.

15. Anfinsen CB, Haber E, Sela M, White FH, Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proceedings of the National Academy of Sciences of the United States of America. 1961;47:1309-14. Epub 1961/09/15. doi: 10.1073/pnas.47.9.1309. PubMed PMID: 13683522; PMCID: PMC223141.

16. Haber E, Anfinsen CB. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. The Journal of biological chemistry. 1961;236:422-4. Epub 1961/02/01. PubMed PMID: 13710229.

17. Haber E, Anfinsen CB. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. The Journal of biological chemistry. 1962;237:1839-44. Epub 1962/06/01. PubMed PMID: 13903380.

18. Goldberger RF, Epstein CJ, Anfinsen CB. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. The Journal of biological chemistry. 1963;238:628-35. Epub 1963/02/01. PubMed PMID: 13948694.

19. Potts JT, Jr., Young DM, Anfinsen CB. Reconstitution of fully active RNase S by carboxypeptidase-degraded RNase S-peptide. The Journal of biological chemistry. 1963;238:2593-4. Epub 1963/07/01. PubMed PMID: 13985757.

20. Sela M, Anfinsen CB, Harrington WF. The correlation of ribonuclease activity with specific aspects of tertiary structure. Biochimica et biophysica acta. 1957;26(3):502-12. Epub 1957/12/01. doi: 10.1016/0006-3002(57)90096-3. PubMed PMID: 13499407.

21. Sali A, Shakhnovich E, Karplus M. How does a protein fold? Nature. 1994;369(6477):248-51. Epub 1994/05/19. doi: 10.1038/369248a0. PubMed PMID: 7710478.

22. Levinthal C. Are There Pathways for Protein Folding? Extrait du Journal de Chimie Physique. 1968;65(1):44.

23. Kuwajima K. A folding model of alpha-lactalbumin deduced from the three-state denaturation mechanism. Journal of molecular biology. 1977;114(2):241-58. Epub 1977/08/05. doi: 10.1016/0022-2836(77)90208-x. PubMed PMID: 909087.

24. Zetina CR, Goldberg ME. Kinetics of renaturation and self-assembly of intermediates on the pathway of folding of the beta 2-subunit of Escherichia coli tryptophan-synthetase. Journal of molecular biology. 1982;157(1):133-48. Epub 1982/05/05. doi: 10.1016/0022-2836(82)90516-2. PubMed PMID: 7050392.

25. Kim PS, Baldwin RL. Intermediates in the folding reactions of small proteins. Annu Rev Biochem. 1990;59:631-60. Epub 1990/01/01. doi: 10.1146/annurev.bi.59.070190.003215. PubMed PMID: 2197986.

26. Karplus M, Weaver DL. Protein-folding dynamics. Nature. 1976;260(5550):404-6. Epub 1976/04/01. doi: 10.1038/260404a0. PubMed PMID: 1256583.

27. Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem. 1982;51:459-89. Epub 1982/01/01. doi: 10.1146/annurev.bi.51.070182.002331. PubMed PMID: 6287919.

28. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science. 1993;262(5135):892-6. Epub 1993/11/05. doi: 10.1126/science.8235610. PubMed PMID: 8235610.

29. Matthews CR. Pathways of protein folding. Annu Rev Biochem. 1993;62:653-83. Epub 1993/01/01. doi: 10.1146/annurev.bi.62.070193.003253. PubMed PMID: 8352599.

30. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annual review of physical chemistry. 1997;48:545-600. Epub 1997/01/01. doi: 10.1146/annurev.physchem.48.1.545. PubMed PMID: 9348663.

31. Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. Protein science : a publication of the Protein Society. 2000;9(3):452-65. Epub 2000/04/07. doi: 10.1110/ps.9.3.452. PubMed PMID: 10752607; PMCID: PMC2144570.

32. Brockwell DJ, Radford SE. Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol. 2007;17(1):30-7. Epub 2007/01/24. doi: 10.1016/j.sbi.2007.01.003. PubMed PMID: 17239580; PMCID: PMC2706323.

33. Thomas PJ, Qu BH, Pedersen PL. Defective protein folding as a basis of human disease. Trends Biochem Sci. 1995;20(11):456-9. Epub 1995/11/01. doi: 10.1016/s0968-0004(00)89100-8. PubMed PMID: 8578588.

34. Finkelstein AV, Badretdin AJ, Galzitskaya OV, Ivankov DN, Bogatyreva NS, Garbuzynskiy SO. There and back again: Two views on the protein folding puzzle. Physics of life reviews. 2017;21:56-71. Epub 2017/02/14. doi: 10.1016/j.plrev.2017.01.025. PubMed PMID: 28190683.

35. Diamond DL, Randall LL. Kinetic partitioning. Poising SecB to favor association with a rapidly folding ligand. The Journal of biological chemistry. 1997;272(46):28994-8. Epub 1997/11/20. doi: 10.1074/jbc.272.46.28994. PubMed PMID: 9360972.

36. Dobson CM. Protein folding and misfolding. Nature. 2003;426(6968):884-90. Epub 2003/12/20. doi: 10.1038/nature02261. PubMed PMID: 14685248.

37. Englander SW, Mayne L, Krishna MM. Protein folding and misfolding: mechanism and principles. Q Rev Biophys. 2007;40(4):287-326. Epub 2008/04/15. doi: 10.1017/s0033583508004654. PubMed PMID: 18405419; PMCID: PMC3433742.

38. Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol. 2007;17(1):21-9. Epub 2007/01/26. doi: 10.1016/j.sbi.2007.01.008. PubMed PMID: 17251003.

39. Lukacs GL, Verkman AS. CFTR: folding, misfolding and correcting the DeltaF508 conformational defect. Trends in molecular medicine. 2012;18(2):81-91. Epub 2011/12/06. doi: 10.1016/j.molmed.2011.10.003. PubMed PMID: 22138491; PMCID: PMC3643519.

40. Fraser-Pitt D, O'Neil D. Cystic fibrosis - a multiorgan protein misfolding disease. Future science OA. 2015;1(2):Fso57. Epub 2015/09/01. doi: 10.4155/fso.15.57. PubMed PMID: 28031875; PMCID: PMC5137970.

41. Farinha CM, Canato S. From the endoplasmic reticulum to the plasma membrane: mechanisms of CFTR folding and trafficking. Cellular and molecular life

sciences : CMLS. 2017;74(1):39-55. Epub 2016/10/05. doi: 10.1007/s00018-016-2387-7. PubMed PMID: 27699454.

42. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. Nature structural & molecular biology. 2009;16(6):574-81. Epub 2009/06/06. doi: 10.1038/nsmb.1591. PubMed PMID: 19491934.

43. Frieden C. Protein aggregation processes: In search of the mechanism. Protein science : a publication of the Protein Society. 2007;16(11):2334-44. Epub 2007/10/27. doi: 10.1110/ps.073164107. PubMed PMID: 17962399; PMCID: PMC2211696.

44. Chi EY, Krishnan S, Randolph TW, Carpenter JF. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharmaceutical research. 2003;20(9):1325-36. Epub 2003/10/22. doi: 10.1023/a:1025771421906. PubMed PMID: 14567625.

45. Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Folding & design. 1998;3(1):R9-23. Epub 1998/03/21. doi: 10.1016/s1359-0278(98)00002-9. PubMed PMID: 9502314.

46. Zhuravlev PI, Reddy G, Straub JE, Thirumalai D. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. Journal of molecular biology. 2014;426(14):2653-66. Epub 2014/05/23. doi: 10.1016/j.jmb.2014.05.007. PubMed PMID: 24846645; PMCID: PMC4100209.

47. Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. Nature reviews Neuroscience. 2003;4(1):49-60. Epub 2003/01/04. doi: 10.1038/nrn1007. PubMed PMID: 12511861.

48. Stefani M, Dobson CM. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. Journal of molecular medicine (Berlin, Germany). 2003;81(11):678-99. Epub 2003/08/28. doi: 10.1007/s00109-003-0464-5. PubMed PMID: 12942175.

49. Speed MA, Wang DI, King J. Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. Nature biotechnology. 1996;14(10):1283-7. Epub 1996/10/01. doi: 10.1038/nbt1096-1283. PubMed PMID: 9631094.

50. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem. 2006;75:333-66. Epub 2006/06/08. doi: 10.1146/annurev.biochem.75.101304.123901. PubMed PMID: 16756495.

51. Wulf MA, Senatore A, Aguzzi A. The biological function of the cellular prion protein: an update. BMC biology. 2017;15(1):34. Epub 2017/05/04. doi: 10.1186/s12915-017-0375-5. PubMed PMID: 28464931; PMCID: PMC5412054.

52. Silveira JR, Raymond GJ, Hughson AG, Race RE, Sim VL, Hayes SF, Caughey B. The most infectious prion protein particles. Nature. 2005;437(7056):257-61. Epub 2005/09/09. doi: 10.1038/nature03989. PubMed PMID: 16148934; PMCID: PMC1513539.

53. Kupfer L, Hinrichs W, Groschup MH. Prion protein misfolding. Current molecular medicine. 2009;9(7):826-35. Epub 2009/10/29. doi: 10.2174/156652409789105543. PubMed PMID: 19860662; PMCID: PMC3330701.

54. Kuffer A, Lakkaraju AK, Mogha A, Petersen SC, Airich K, Doucerain C, Marpakwar R, Bakirci P, Senatore A, Monnard A, Schiavi C, Nuvolone M, Grosshans B, Hornemann S, Bassilana F, Monk KR, Aguzzi A. The prion protein is an agonistic ligand of the G protein-coupled receptor Adgrg6. Nature. 2016;536(7617):464-8. Epub 2016/08/09. doi: 10.1038/nature19312. PubMed PMID: 27501152; PMCID: PMC5499706.

55. Mattson MP. Pathways towards and away from Alzheimer's disease. Nature. 2004;430(7000):631-9. Epub 2004/08/06. doi: 10.1038/nature02621. PubMed PMID: 15295589; PMCID: PMC3091392.

56. Lesne S, Koh MT, Kotilinek L, Kayed R, Glabe CG, Yang A, Gallagher M, Ashe KH. A specific amyloid-beta protein assembly in the brain impairs memory. Nature. 2006;440(7082):352-7. Epub 2006/03/17. doi: 10.1038/nature04533. PubMed PMID: 16541076.

57. Jucker M, Walker LC. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature. 2013;501(7465):45-51. Epub 2013/09/06. doi: 10.1038/nature12481. PubMed PMID: 24005412; PMCID: PMC3963807.

58. Takalo M, Salminen A, Soininen H, Hiltunen M, Haapasalo A. Protein aggregation and degradation mechanisms in neurodegenerative diseases. American journal of neurodegenerative disease. 2013;2(1):1-14. Epub 2013/03/22. PubMed PMID: 23516262; PMCID: PMC3601466.

59. Walter S, Buchner J. Molecular chaperones--cellular machines for protein folding. Angewandte Chemie (International ed in English). 2002;41(7):1098-113. Epub 2002/12/20. doi: 10.1002/1521-3773(20020402)41:7<1098::aid-anie1098>3.0.co;2-9. PubMed PMID: 12491239.

60. Zietkiewicz S, Krzewska J, Liberek K. Successive and synergistic action of the Hsp70 and Hsp100 chaperones in protein disaggregation. The Journal of biological chemistry. 2004;279(43):44376-83. Epub 2004/08/11. doi: 10.1074/jbc.M402405200. PubMed PMID: 15302880.

61. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem. 2013;82:323-55. Epub 2013/06/12. doi: 10.1146/annurev-biochem-060208-092442. PubMed PMID: 23746257.

62. Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. Nature reviews Molecular cell biology. 2013;14(10):630-42. Epub 2013/09/13. doi: 10.1038/nrm3658. PubMed PMID: 24026055; PMCID: PMC4340576.

63. Mogk A, Bukau B, Kampinga HH. Cellular Handling of Protein Aggregates by Disaggregation Machines. Molecular cell. 2018;69(2):214-26. Epub 2018/01/21. doi: 10.1016/j.molcel.2018.01.004. PubMed PMID: 29351843.

64. Ellis J. Proteins as molecular chaperones. Nature. 1987;328(6129):378-9. Epub 1987/07/05. doi: 10.1038/328378a0. PubMed PMID: 3112578.

65. Rothman JE. Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. Cell. 1989;59(4):591-601. Epub 1989/11/17. doi: 10.1016/0092-8674(89)90005-6. PubMed PMID: 2573430.

66. Hendrick JP, Hartl FU. Molecular chaperone functions of heat-shock proteins. Annu Rev Biochem. 1993;62:349-84. Epub 1993/01/01. doi: 10.1146/annurev.bi.62.070193.002025. PubMed PMID: 8102520.

67. Kim S, Schilke B, Craig EA, Horwich AL. Folding in vivo of a newly translated yeast cytosolic enzyme is mediated by the SSA class of cytosolic yeast Hsp70 proteins. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(22):12860-5. PubMed PMID: 9789005; PMCID: PMC23633.

68. 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(6745):693-6. Epub 1999/08/24. doi: 10.1038/23301. PubMed PMID: 10458167.

69. Goloubinoff P, Christeller JT, Gatenby AA, Lorimer GH. Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP. Nature. 1989;342(6252):884-9. Epub 1989/12/21. doi: 10.1038/342884a0. PubMed PMID: 10532860.

70. Goloubinoff P, Gatenby AA, Lorimer GH. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. Nature. 1989;337(6202):44-7. doi: 10.1038/337044a0. PubMed PMID: 2562907.

71. Horwich AL, Low KB, Fenton WA, Hirshfield IN, Furtak K. Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. Cell. 1993;74(5):909-17. PubMed PMID: 8104102.

72. Weissman JS, Kashi Y, Fenton WA, Horwich AL. GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. Cell. 1994;78(4):693-702. PubMed PMID: 7915201.

73. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell. 1996;84(3):481-90. PubMed PMID: 8608602.

74. Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL. Structure and function in GroEL-mediated protein folding. Annu Rev Biochem. 1998;67:581-608. doi: 10.1146/annurev.biochem.67.1.581. PubMed PMID: 9759498.

75. Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, Horwich AL. GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. Cell. 1999;97(3):325-38. Epub 1999/05/13. doi: 10.1016/s0092-8674(00)80742-4. PubMed PMID: 10319813.

76. Chaudhuri TK, Farr GW, Fenton WA, Rospert S, Horwich AL. GroEL/GroES-mediated folding of a protein too large to be encapsulated. Cell. 2001;107(2):235-46. PubMed PMID: 11672530.

77. Hayer-Hartl M, Bracher A, Hartl FU. The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. Trends Biochem Sci. 2016;41(1):62-76. doi: 10.1016/j.tibs.2015.07.009. PubMed PMID: 26422689.

78. Lin Z, Rye HS. Expansion and compression of a protein folding intermediate by GroEL. Molecular cell. 2004;16(1):23-34. doi: 10.1016/j.molcel.2004.09.003. PubMed PMID: 15469819; PMCID: PMC3759401.

79. Saibil HR, Fenton WA, Clare DK, Horwich AL. Structure and allostery of the chaperonin GroEL. Journal of molecular biology. 2013;425(9):1476-87. doi: 10.1016/j.jmb.2012.11.028. PubMed PMID: 23183375.

80. Apetri AC, Horwich AL. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proceedings of the National Academy

of Sciences of the United States of America. 2008;105(45):17351-5. Epub 2008/11/07. doi: 10.1073/pnas.0809794105. PubMed PMID: 18987317; PMCID: PMC2579888.

81. Weaver J, Rye HS. The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. The Journal of biological chemistry. 2014;289(33):23219-32. doi: 10.1074/jbc.M114.577205. PubMed PMID: 24970895; PMCID: PMC4132819.

82. Weaver J, Jiang M, Roth A, Puchalla J, Zhang J, Rye HS. GroEL actively stimulates folding of the endogenous substrate protein PepQ. Nat Commun. 2017;8:15934. doi: 10.1038/ncomms15934. PubMed PMID: 28665408; PMCID: PMC5497066.

83. Lin Z, Madan D, Rye HS. GroEL stimulates protein folding through forced unfolding. Nature structural & molecular biology. 2008;15(3):303-11. Epub 2008/03/04. doi: 10.1038/nsmb.1394. PubMed PMID: 18311152; PMCID: PMC3744391.

84. Horwich AL, Fenton WA. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q Rev Biophys. 2009;42(2):83-116. doi: 10.1017/S0033583509004764. PubMed PMID: 19638247.

85. Rosenzweig R, Nillegoda NB, Mayer MP, Bukau B. The Hsp70 chaperone network. Nature Reviews Molecular Cell Biology. 2019. doi: 10.1038/s41580-019-0133-3.

86. Mayer MP, Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. Cellular and molecular life sciences : CMLS. 2005;62(6):670-84. Epub 2005/03/17. doi: 10.1007/s00018-004-4464-6. PubMed PMID: 15770419; PMCID: PMC2773841.

87. Goloubinoff P, De Los Rios P. The mechanism of Hsp70 chaperones: (entropic) pulling the models together. Trends Biochem Sci. 2007;32(8):372-80. Epub 2007/07/17. doi: 10.1016/j.tibs.2007.06.008. PubMed PMID: 17629485.

88. Mayer MP. Hsp70 chaperone dynamics and molecular mechanism. Trends Biochem Sci. 2013;38(10):507-14. Epub 2013/09/10. doi: 10.1016/j.tibs.2013.08.001. PubMed PMID: 24012426.

89. Slepenkov SV, Witt SN. The unfolding story of the Escherichia coli Hsp70 DnaK: is DnaK a holdase or an unfoldase? Molecular microbiology. 2002;45(5):1197-206. Epub 2002/09/05. doi: 10.1046/j.1365-2958.2002.03093.x. PubMed PMID: 12207689.

90. Kampinga HH, Craig EA. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nature reviews Molecular cell biology. 2010;11(8):579-92. doi: 10.1038/nrm2941. PubMed PMID: 20651708; PMCID: PMC3003299.

91. Alderson TR, Kim JH, Markley JL. Dynamical Structures of Hsp70 and Hsp70-Hsp40 Complexes. Structure. 2016;24(7):1014-30. Epub 2016/06/28. doi: 10.1016/j.str.2016.05.011. PubMed PMID: 27345933; PMCID: PMC4938735.

92. Deville C, Carroni M, Franke KB, Topf M, Bukau B, Mogk A, Saibil HR. Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase. Science advances. 2017;3(8):e1701726. Epub 2017/08/12. doi: 10.1126/sciadv.1701726. PubMed PMID: 28798962; PMCID: PMC5544394.

93. Mogk A, Kummer E, Bukau B. Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. Frontiers in molecular biosciences. 2015;2:22. Epub 2015/06/05. doi: 10.3389/fmolb.2015.00022. PubMed PMID: 26042222; PMCID: PMC4436881.

94. Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. Some like it hot: the structure and function of small heat-shock proteins. Nature structural & molecular biology. 2005;12(10):842-6. doi: 10.1038/nsmb993. PubMed PMID: 16205709.

95. Haslbeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. Journal of molecular biology. 2015;427(7):1537-48. doi: 10.1016/j.jmb.2015.02.002. PubMed PMID: 25681016; PMCID: 4360138.

96. Strauch A, Haslbeck M. The function of small heat-shock proteins and their implication in proteostasis. Essays in biochemistry. 2016;60(2):163-72. Epub 2016/10/17. doi: 10.1042/ebc20160010. PubMed PMID: 27744332.

97. Haslbeck M, Peschek J, Buchner J, Weinkauf S. Structure and function of alpha-crystallins: Traversing from in vitro to in vivo. Biochimica et biophysica acta. 2016;1860(1 Pt B):149-66. doi: 10.1016/j.bbagen.2015.06.008. PubMed PMID: 26116912.

98. McHaourab HS, Lin YL, Spiller BW. Crystal structure of an activated variant of small heat shock protein Hsp16.5. Biochemistry. 2012;51(25):5105-12. Epub 2012/06/08. doi: 10.1021/bi300525x. PubMed PMID: 22670769; PMCID: PMC3384710.

99. White HE, Orlova EV, Chen S, Wang L, Ignatiou A, Gowen B, Stromer T, Franzmann TM, Haslbeck M, Buchner J, Saibil HR. Multiple distinct assemblies reveal conformational flexibility in the small heat shock protein Hsp26. Structure.

2006;14(7):1197-204. Epub 2006/07/18. doi: 10.1016/j.str.2006.05.021. PubMed PMID: 16843901.

100. Hanazono Y, Takeda K, Oka T, Abe T, Tomonari T, Akiyama N, Aikawa Y, Yohda M, Miki K. Nonequivalence observed for the 16-meric structure of a small heat shock protein, SpHsp16.0, from Schizosaccharomyces pombe. Structure. 2013;21(2):220-8. Epub 2013/01/01. doi: 10.1016/j.str.2012.11.015. PubMed PMID: 23273429.

101. Mani N, Bhandari S, Moreno R, Hu L, Prasad BVV, Suguna K. Multiple oligomeric structures of a bacterial small heat shock protein. Scientific reports. 2016;6:24019. Epub 2016/04/08. doi: 10.1038/srep24019. PubMed PMID: 27053150; PMCID: PMC4823740.

102. Haslbeck M, Weinkauf S, Buchner J. Small heat shock proteins: Simplicity meets complexity. The Journal of biological chemistry. 2019;294(6):2121-32. doi: 10.1074/jbc.REV118.002809. PubMed PMID: 30385502; PMCID: PMC6369295.

103. Basha E, Jones C, Blackwell AE, Cheng G, Waters ER, Samsel KA, Siddique M, Pett V, Wysocki V, Vierling E. An unusual dimeric small heat shock protein provides insight into the mechanism of this class of chaperones. Journal of molecular biology. 2013;425(10):1683-96. Epub 2013/02/19. doi: 10.1016/j.jmb.2013.02.011. PubMed PMID: 23416558; PMCID: PMC3646915.

104. Liu L, Chen J, Yang B, Wang Y. Crystal structure and function of an unusual dimeric Hsp20.1 provide insight into the thermal protection mechanism of small heat shock proteins. Biochemical and biophysical research communications. 2015;458(2):429-34. Epub 2015/02/11. doi: 10.1016/j.bbrc.2015.01.134. PubMed PMID: 25660449.

105. Carra S, Alberti S, Benesch JLP, Boelens W, Buchner J, Carver JA, Cecconi C, Ecroyd H, Gusev N, Hightower LE, Klevit RE, Lee HO, Liberek K, Lockwood B, Poletti A, Timmerman V, Toth ME, Vierling E, Wu T, Tanguay RM. Small heat shock proteins: multifaceted proteins with important implications for life. Cell Stress Chaperones. 2019;24(2):295-308. Epub 2019/02/14. doi: 10.1007/s12192-019-00979-z. PubMed PMID: 30758704; PMCID: PMC6439001.

106. Kuczynska-Wisnik D, Kedzierska S, Matuszewska E, Lund P, Taylor A, Lipinska B, Laskowska E. The Escherichia coli small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. Microbiology. 2002;148(Pt 6):1757-65. doi: 10.1099/00221287-148-6-1757. PubMed PMID: 12055295.

107. Peschek J, Braun N, Franzmann TM, Georgalis Y, Haslbeck M, Weinkauf S, Buchner J. The eye lens chaperone alpha-crystallin forms defined globular assemblies. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(32):13272-7. doi: 10.1073/pnas.0902651106. PubMed PMID: 19651604; PMCID: PMC2726422.

108. Federico Bennardini AW, Michele Chiesi. aB crystallin in Cardiac Tissues assocation with actin and desmin filaments. Circulation Research. 1992;71(2):288-94.

109. Fleckenstein T, Kastenmuller A, Stein ML, Peters C, Daake M, Krause M, Weinfurtner D, Haslbeck M, Weinkauf S, Groll M, Buchner J. The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH-Dependent Conformational Changes. Molecular cell. 2015;58(6):1067-78. doi: 10.1016/j.molcel.2015.04.019. PubMed PMID: 26009280.

110. Kitagawa M, Miyakawa M, Matsumura Y, Tsuchido T. Escherichia colismall heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. European Journal of Biochemistry. 2002;269(12):2907-17. doi: 10.1046/j.1432-1033.2002.02958.x.

111. Teresa M. Treweek SM, Heath Ecroyd, John A. Carver. Small heat-shock proteins-important players in regulating cellular proteostasis Cellular and Molecular Life Sciences. 2015(72):429-51.

112. Mogk A, Ruger-Herreros C, Bukau B. Cellular Functions and Mechanisms of Action of Small Heat Shock Proteins. Annual review of microbiology. 2019;73:89-110. Epub 2019/05/16. doi: 10.1146/annurev-micro-020518-115515. PubMed PMID: 31091419.

113. Gaubig LC, Waldminghaus T, Narberhaus F. Multiple layers of control govern expression of the Escherichia coli ibpAB heat-shock operon. Microbiology. 2011;157(Pt 1):66-76. Epub 2010/09/25. doi: 10.1099/mic.0.043802-0. PubMed PMID: 20864473.

114. Shi X, Yan L, Zhang H, Sun K, Chang Z, Fu X. Differential degradation for small heat shock proteins IbpA and IbpB is synchronized in Escherichia coli: implications for their functional cooperation in substrate refolding. Biochemical and biophysical research communications. 2014;452(3):402-7. doi: 10.1016/j.bbrc.2014.08.084. PubMed PMID: 25173932.

115. Ito H, Kamei K, Iwamoto I, Inaguma Y, Nohara D, Kato K. Phosphorylation-induced change of the oligomerization state of alpha B-crystallin. The Journal of biological chemistry. 2001;276(7):5346-52. doi: 10.1074/jbc.M009004200. PubMed PMID: 11096101.

116. Jaya N, Garcia V, Vierling E. Substrate binding site flexibility of the small heat shock protein molecular chaperones. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(37):15604-9. doi: 10.1073/pnas.0902177106. PubMed PMID: 19717454; PMCID: PMC2773522.

117. Fu X, Shi X, Yin L, Liu J, Joo K, Lee J, Chang Z. Small heat shock protein IbpB acts as a robust chaperone in living cells by hierarchically activating its multi-type substrate-binding residues. The Journal of biological chemistry. 2013;288(17):11897-906. Epub 2013/03/15. doi: 10.1074/jbc.M113.450437. PubMed PMID: 23486475; PMCID: PMC3636877.

118. Mainz A, Peschek J, Stavropoulou M, Back KC, Bardiaux B, Asami S, Prade E, Peters C, Weinkauf S, Buchner J, Reif B. The chaperone alphaB-crystallin uses different interfaces to capture an amorphous and an amyloid client. Nature structural & molecular biology. 2015;22(11):898-905. doi: 10.1038/nsmb.3108. PubMed PMID: 26458046.

119. Steven P. Allen JOP, James K. Gierse, Alan M. Easton. Two Novel Heat Shock Genes Encoding Proteins Produced in Response to Heterologous Protein Expression in Escherichia coli. Journal of bacteriology. 1992;174(21):6938-47.

120. Veinger L, Diamant, S., Buchner, J., Goloubinoff, P. sHsp IbpB from e coli stabilizes stress denatured proteins for subsequent refolding by a multichaperone network. Journal of Biological Chemistry. 1998;273(18):11032-7.

121. Veinger L, Diamant S, Buchner J, Goloubinoff P. The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. The Journal of biological chemistry. 1998;273(18):11032-7. PubMed PMID: 9556585.

122. Shearstone JR, Baneyx F. Biochemical characterization of the small heat shock protein IbpB from Escherichia coli. The Journal of biological chemistry. 1999;274(15):9937-45. PubMed PMID: 10187768.

123. Mogk A, Schlieker C, Friedrich KL, Schonfeld HJ, Vierling E, Bukau B. Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. The Journal of biological chemistry. 2003;278(33):31033-42. doi: 10.1074/jbc.M303587200. PubMed PMID: 12788951.

124. Matuszewska M, Kuczynska-Wisnik D, Laskowska E, Liberek K. The small heat shock protein IbpA of Escherichia coli cooperates with IbpB in stabilization of thermally aggregated proteins in a disaggregation competent state. The Journal of biological chemistry. 2005;280(13):12292-8. Epub 2005/01/25. doi: 10.1074/jbc.M412706200. PubMed PMID: 15665332.

125. Ratajczak E, Zietkiewicz S, Liberek K. Distinct activities of Escherichia coli small heat shock proteins IbpA and IbpB promote efficient protein disaggregation. Journal of molecular biology. 2009;386(1):178-89. doi: 10.1016/j.jmb.2008.12.009. PubMed PMID: 19101567.

126. Mogk A, Deuerling E, Vorderwulbecke S, Vierling E, Bukau B. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. Molecular microbiology. 2003;50(2):585-95. PubMed PMID: 14617181.

127. Lee GJ. A small heat shock protein stably binds heat denatured model substrates and can maintain a substrate in a folding competent state. The EMBO journal. 1997;16(3):657-71.

128. Augusteyn RC. alpha crystallin polymers and polymerization- the view from down under. International Journal of Biological Macromolecules. 1998(22):253-62.

129. Ungelenk S, Moayed F, Ho CT, Grousl T, Scharf A, Mashaghi A, Tans S, Mayer MP, Mogk A, Bukau B. Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. Nat Commun. 2016;7:13673. doi: 10.1038/ncomms13673. PubMed PMID: 27901028; PMCID: PMC5141385.

130. Zwirowski S, Klosowska A, Obuchowski I, Nillegoda NB, Pirog A, Zietkiewicz S, Bukau B, Mogk A, Liberek K. Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. The EMBO journal. 2017;36(6):783-96. doi: 10.15252/embj.201593378. PubMed PMID: 28219929; PMCID: PMC5350560.

131. Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, Sigler PB, Horwich AL. Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. Nature. 1997;388(6644):792-8. doi: 10.1038/42047. PubMed PMID: 9285593.

132. Weaver J, Watts T, Li P, Rye HS. Structural basis of substrate selectivity of E. coli prolidase. PloS one. 2014;9(10):e111531. Epub 2014/10/30. doi: 10.1371/journal.pone.0111531. PubMed PMID: 25354344; PMCID: PMC4213023.

133. Puchalla J, Krantz K, Austin R, Rye H. Burst analysis spectroscopy: a versatile single-particle approach for studying distributions of protein aggregates and fluorescent assemblies. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(38):14400-5. doi: 10.1073/pnas.0805969105. PubMed PMID: 18780782; PMCID: 2567176.

134. Brooks A, Shoup D, Kustigian L, Puchalla J, Carr CM, Rye HS. Single particle fluorescence burst analysis of epsin induced membrane fission. PloS one.

2015;10(3):e0119563. doi: 10.1371/journal.pone.0119563. PubMed PMID: 25799353; PMCID: 4370887.

135. Kumar V, Sami N, Kashav T, Islam A, Ahmad F, Hassan MI. Protein aggregation and neurodegenerative diseases: From theory to therapy. European journal of medicinal chemistry. 2016;124:1105-20. Epub 2016/10/22. doi: 10.1016/j.ejmech.2016.07.054. PubMed PMID: 27486076.

136. Daniel Shoup AR, Rajan Thapa, Jason Puchalla, Hays S. Rye Development and applications of extended range and multi-dimensional burst analysis spectroscopy (in preparation). 2017.

137. Madan D, Lin Z, Rye HS. Triggering protein folding within the GroEL-GroES complex. The Journal of biological chemistry. 2008;283(46):32003-13. Epub 2008/09/11. doi: 10.1074/jbc.M802898200. PubMed PMID: 18782766; PMCID: PMC2581556.

138. Bonhommeau S, Lecomte S. Tip-Enhanced Raman Spectroscopy: A Tool for Nanoscale Chemical and Structural Characterization of Biomolecules. Chemphyschem : a European journal of chemical physics and physical chemistry. 2018;19(1):8-18. Epub 2017/11/07. doi: 10.1002/cphc.201701067. PubMed PMID: 29106771.

139. Perng MD, Wen SF, van den IP, Prescott AR, Quinlan RA. Desmin aggregate formation by R120G alphaB-crystallin is caused by altered filament interactions and is dependent upon network status in cells. Molecular biology of the cell. 2004;15(5):2335-46. doi: 10.1091/mbc.E03-12-0893. PubMed PMID: 15004226; PMCID: PMC404027.

140. Houck SA, Landsbury A, Clark JI, Quinlan RA. Multiple sites in alphaB-crystallin modulate its interactions with desmin filaments assembled in vitro. PloS one. 2011;6(11):e25859. doi: 10.1371/journal.pone.0025859. PubMed PMID: 22096479; PMCID: PMC3212511.

141. Subhalakshmi Ganguly AMaSS. Role of a-crystallin B in regulation of stress induced cardiomyocyte apoptosis. Cardiovascular & Hematological Agents in Medicinal Chemistry. 2014.

142. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB. The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature. 1994;371(6498):578-86. Epub 1994/10/13. doi: 10.1038/371578a0. PubMed PMID: 7935790.

143. Chen L, Sigler PB. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. Cell. 1999;99(7):757-68. Epub 2000/01/05. doi: 10.1016/s0092-8674(00)81673-6. PubMed PMID: 10619429.

144. Fenton WA, Kashi Y, Furtak K, Horwich AL. Residues in chaperonin GroEL required for polypeptide binding and release. Nature. 1994;371(6498):614-9. Epub 1994/10/13. doi: 10.1038/371614a0. PubMed PMID: 7935796.

145. Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, Beckmann R, Hayer-Hartl M. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. Nature. 2010;463(7278):197-202. Epub 2010/01/16. doi: 10.1038/nature08651. PubMed PMID: 20075914.

146. Mayhew M, da Silva AC, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU. Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature. 1996;379(6564):420-6. Epub 1996/02/01. doi: 10.1038/379420a0. PubMed PMID: 8559246.

147. Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, Saibil HR, Fenton WA, Horwich AL. Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. Cell. 1995;83(4):577-87. Epub 1995/11/17. doi: 10.1016/0092-8674(95)90098-5. PubMed PMID: 7585961.

148. Todd MJ, Viitanen PV, Lorimer GH. Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. Science. 1994;265(5172):659-66. Epub 1994/07/29. doi: 10.1126/science.7913555. PubMed PMID: 7913555.

149. Jewett AI, Shea JE. Reconciling theories of chaperonin accelerated folding with experimental evidence. Cellular and molecular life sciences : CMLS. 2010;67(2):255-76. Epub 2009/10/24. doi: 10.1007/s00018-009-0164-6. PubMed PMID: 19851829.

150. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324-32. Epub 2011/07/22. doi: 10.1038/nature10317. PubMed PMID: 21776078.

151. Betancourt MR, Thirumalai D. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. Journal of molecular biology. 1999;287(3):627-44. Epub 1999/03/27. doi: 10.1006/jmbi.1999.2591. PubMed PMID: 10092464.

152. Todd MJ, Lorimer GH, Thirumalai D. Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(9):4030-5. Epub 1996/04/30. doi: 10.1073/pnas.93.9.4030. PubMed PMID: 8633011; PMCID: PMC39481.

153. Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, Jiang G, Lamb DC, Hartl FU, Hayer-Hartl M. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell. 2010;142(1):112-22. Epub 2010/07/07. doi: 10.1016/j.cell.2010.05.027. PubMed PMID: 20603018.

154. Lin Z, Puchalla J, Shoup D, Rye HS. Repetitive protein unfolding by the trans ring of the GroEL-GroES chaperonin complex stimulates folding. The Journal of biological chemistry. 2013;288(43):30944-55. Epub 2013/09/12. doi: 10.1074/jbc.M113.480178. PubMed PMID: 24022487; PMCID: PMC3829408.

155. Sharma S, Chakraborty K, Muller BK, Astola N, Tang YC, Lamb DC, Hayer-Hartl M, Hartl FU. Monitoring protein conformation along the pathway of chaperonin-assisted folding. Cell. 2008;133(1):142-53. Epub 2008/04/09. doi: 10.1016/j.cell.2008.01.048. PubMed PMID: 18394994.

156. Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, Hayer-Hartl M, Hartl FU. GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. Cell. 2014;157(4):922-34. Epub 2014/05/13. doi: 10.1016/j.cell.2014.03.038. PubMed PMID: 24813614; PMCID: PMC4071350.

157. Ambrose A, Fenton W, Mason DJ, Chapman E, Horwich AL. Unfolded DapA forms aggregates when diluted into free solution, confounding comparison with folding by the GroEL/GroES chaperonin system. FEBS letters. 2015;589(4):497-9. Epub 2015/01/21. doi: 10.1016/j.febslet.2015.01.008. PubMed PMID: 25601566; PMCID: PMC4410871.

158. Park MS, Hill CM, Li Y, Hardy RK, Khanna H, Khang YH, Raushel FM. Catalytic properties of the PepQ prolidase from Escherichia coli. Archives of biochemistry and biophysics. 2004;429(2):224-30. Epub 2004/08/18. doi: 10.1016/j.abb.2004.06.022. PubMed PMID: 15313226.

159. Bazan JF, Weaver LH, Roderick SL, Huber R, Matthews BW. Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(7):2473-7. Epub 1994/03/29. doi: 10.1073/pnas.91.7.2473. PubMed PMID: 8146141; PMCID: PMC43391.

160. Lowther WT, Matthews BW. Structure and function of the methionine aminopeptidases. Biochimica et biophysica acta. 2000;1477(1-2):157-67. Epub 2000/03/10. doi: 10.1016/s0167-4838(99)00271-x. PubMed PMID: 10708856.

161. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. A systematic survey of in vivo obligate chaperonin-dependent substrates. The EMBO journal.

2010;29(9):1552-64. Epub 2010/04/03. doi: 10.1038/emboj.2010.52. PubMed PMID: 20360681; PMCID: PMC3212837.

162. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell. 2005;122(2):209-20. Epub 2005/07/30. doi: 10.1016/j.cell.2005.05.028. PubMed PMID: 16051146.

163. Braman J, Papworth C, Greener A. Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods in molecular biology (Clifton, NJ). 1996;57:31-44. Epub 1996/01/01. doi: 10.1385/0-89603-332-5:31. PubMed PMID: 8849992.

164. Rye HS. Application of fluorescence resonance energy transfer to the GroEL-GroES chaperonin reaction. Methods (San Diego, Calif). 2001;24(3):278-88. Epub 2001/06/14. doi: 10.1006/meth.2001.1188. PubMed PMID: 11403576; PMCID: PMC3744193.

165. Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, Kerner MJ, Hartl FU, Hayer-Hartl M. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell. 2006;125(5):903-14. Epub 2006/06/06. doi: 10.1016/j.cell.2006.04.027. PubMed PMID: 16751100.

166. Pal N, Dev Verma S, Singh MK, Sen S. Fluorescence correlation spectroscopy: an efficient tool for measuring size, size-distribution and polydispersity of microemulsion droplets in solution. Analytical chemistry. 2011;83(20):7736-44. Epub 2011/09/09. doi: 10.1021/ac2012637. PubMed PMID: 21899251.

167. Sengupta P, Garai K, Balaji J, Periasamy N, Maiti S. Measuring size distribution in highly heterogeneous systems with fluorescence correlation spectroscopy. Biophysical journal. 2003;84(3):1977-84. Epub 2003/03/01. doi: 10.1016/s0006-3495(03)75006-1. PubMed PMID: 12609900; PMCID: PMC1302767.

168. Widengren J, Rigler R, Mets U. Triplet-state monitoring by fluorescence correlation spectroscopy. Journal of fluorescence. 1994;4(3):255-8. Epub 1994/09/01. doi: 10.1007/bf01878460. PubMed PMID: 24233457.

169. Kreuzer KN, Jongeneel CV. Escherichia coli phage T4 topoisomerase. Methods Enzymol. 1983;100:144-60. Epub 1983/01/01. doi: 10.1016/0076-6879(83)00051-8. PubMed PMID: 6312256.

170. Poso D, Clarke AR, Burston SG. A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. Journal of molecular biology.

2004;338(5):969-77. Epub 2004/04/28. doi: 10.1016/j.jmb.2004.03.010. PubMed PMID: 15111060.

171. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature methods. 2013;10(6):584-90. Epub 2013/05/07. doi: 10.1038/nmeth.2472. PubMed PMID: 23644547; PMCID: PMC3684049.

172. Grant T, Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. eLife. 2015;4:e06980. Epub 2015/05/30. doi: 10.7554/eLife.06980. PubMed PMID: 26023829; PMCID: PMC4471936.

173. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. EMAN2: an extensible image processing suite for electron microscopy. Journal of structural biology. 2007;157(1):38-46. Epub 2006/07/25. doi: 10.1016/j.jsb.2006.05.009. PubMed PMID: 16859925.

174. Scheres SH, Chen S. Prevention of overfitting in cryo-EM structure determination. Nature methods. 2012;9(9):853-4. Epub 2012/07/31. doi: 10.1038/nmeth.2115. PubMed PMID: 22842542; PMCID: PMC4912033.

175. Heymann JB, Cardone G, Winkler DC, Steven AC. Computational resources for cryo-electron tomography in Bsoft. Journal of structural biology. 2008;161(3):232-42. Epub 2007/09/18. doi: 10.1016/j.jsb.2007.08.002. PubMed PMID: 17869539; PMCID: PMC2409064.

176. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004;25(13):1605-12. Epub 2004/07/21. doi: 10.1002/jcc.20084. PubMed PMID: 15264254.

177. Grallert H, Rutkat K, Buchner J. Limits of protein folding inside GroE complexes. The Journal of biological chemistry. 2000;275(27):20424-30. Epub 2000/04/26. doi: 10.1074/jbc.M002243200. PubMed PMID: 10779510.

178. van der Vies SM, Viitanen PV, Gatenby AA, Lorimer GH, Jaenicke R. Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. Biochemistry. 1992;31(14):3635-44. Epub 1992/04/14. doi: 10.1021/bi00129a012. PubMed PMID: 1348956.

179. Burston SG, Ranson NA, Clarke AR. The origins and consequences of asymmetry in the chaperonin reaction cycle. Journal of molecular biology.

1995;249(1):138-52. Epub 1995/05/26. doi: 10.1006/jmbi.1995.0285. PubMed PMID: 7776368.

180. Chen DH, Madan D, Weaver J, Lin Z, Schroder GF, Chiu W, Rye HS. Visualizing GroEL/ES in the act of encapsulating a folding protein. Cell. 2013;153(6):1354-65. Epub 2013/06/12. doi: 10.1016/j.cell.2013.04.052. PubMed PMID: 23746846; PMCID: PMC3695626.

181. Tang YC, Chang HC, Chakraborty K, Hartl FU, Hayer-Hartl M. Essential role of the chaperonin folding compartment in vivo. The EMBO journal. 2008;27(10):1458-68. Epub 2008/04/18. doi: 10.1038/emboj.2008.77. PubMed PMID: 18418386; PMCID: PMC2396394.

182. McLennan NF, Girshovich AS, Lissin NM, Charters Y, Masters M. The strongly conserved carboxyl-terminus glycine-methionine motif of the Escherichia coli GroEL chaperonin is dispensable. Molecular microbiology. 1993;7(1):49-58. Epub 1993/01/01. doi: 10.1111/j.1365-2958.1993.tb01096.x. PubMed PMID: 8094879.

183. Burnett BP, Horwich AL, Low KB. A carboxy-terminal deletion impairs the assembly of GroEL and confers a pleiotropic phenotype in Escherichia coli K-12. Journal of bacteriology. 1994;176(22):6980-5. Epub 1994/11/01. doi: 10.1128/jb.176.22.6980-6985.1994. PubMed PMID: 7961461; PMCID: PMC197070.

184. Farr GW, Fenton WA, Horwich AL. Perturbed ATPase activity and not "close confinement" of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(13):5342-7. Epub 2007/03/21. doi: 10.1073/pnas.0700820104. PubMed PMID: 17372195; PMCID: PMC1828711.

185. Brocchieri L, Karlin S. Conservation among HSP60 sequences in relation to structure, function, and evolution. Protein science : a publication of the Protein Society. 2000;9(3):476-86. Epub 2000/04/07. doi: 10.1110/ps.9.3.476. PubMed PMID: 10752609; PMCID: PMC2144576.

186. Lund PA. Multiple chaperonins in bacteria--why so many? FEMS microbiology reviews. 2009;33(4):785-800. Epub 2009/05/07. doi: 10.1111/j.1574-6976.2009.00178.x. PubMed PMID: 19416363.

187. Gruber R, Horovitz A. Allosteric Mechanisms in Chaperonin Machines. Chem Rev. 2016;116(11):6588-606. Epub 2016/01/05. doi: 10.1021/acs.chemrev.5b00556. PubMed PMID: 26726755.

188. Grason JP, Gresham JS, Lorimer GH. Setting the chaperonin timer: a two-stroke, two-speed, protein machine. Proceedings of the National Academy of Sciences of the

United States of America. 2008;105(45):17339-44. Epub 2008/11/08. doi: 10.1073/pnas.0807418105. PubMed PMID: 18988739; PMCID: PMC2580751.

189. Falke S, Tama F, Brooks CL, 3rd, Gogol EP, Fisher MT. The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. Journal of molecular biology. 2005;348(1):219-30. Epub 2005/04/06. doi: 10.1016/j.jmb.2005.02.027. PubMed PMID: 15808865.

190. Horwich AL, Apetri AC, Fenton WA. The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS letters. 2009;583(16):2654-62. Epub 2009/07/07. doi: 10.1016/j.febslet.2009.06.049. PubMed PMID: 19577567; PMCID: PMC2759771.

191. Libich DS, Tugarinov V, Clore GM. Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(29):8817-23. Epub 2015/07/01. doi: 10.1073/pnas.1510083112. PubMed PMID: 26124125; PMCID: PMC4517251.

192. Voziyan PA, Fisher MT. Chaperonin-assisted folding of glutamine synthetase under nonpermissive conditions: off-pathway aggregation propensity does not determine the co-chaperonin requirement. Protein science : a publication of the Protein Society. 2000;9(12):2405-12. Epub 2001/02/24. doi: 10.1110/ps.9.12.2405. PubMed PMID: 11206062; PMCID: PMC2144532.

193. Chen S, Roseman AM, Hunter AS, Wood SP, Burston SG, Ranson NA, Clarke AR, Saibil HR. Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. Nature. 1994;371(6494):261-4. Epub 1994/09/15. doi: 10.1038/371261a0. PubMed PMID: 7915827.

194. Roseman AM, Chen S, White H, Braig K, Saibil HR. The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. Cell. 1996;87(2):241-51. Epub 1996/10/18. doi: 10.1016/s0092-8674(00)81342-2. PubMed PMID: 8861908.

195. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature. 1997;388(6644):741-50. Epub 1997/08/21. doi: 10.1038/41944. PubMed PMID: 9285585.

196. Motojima F, Chaudhry C, Fenton WA, Farr GW, Horwich AL. Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(42):15005-12. Epub 2004/10/14. doi: 10.1073/pnas.0406132101. PubMed PMID: 15479763; PMCID: PMC523455.

197. Dalton KM, Frydman J, Pande VS. The dynamic conformational cycle of the group I chaperonin C-termini revealed via molecular dynamics simulation. PloS one. 2015;10(3):e0117724. Epub 2015/03/31. doi: 10.1371/journal.pone.0117724. PubMed PMID: 25822285; PMCID: PMC4379175.

198. Avellaneda MJ, Koers EJ, Naqvi MM, Tans SJ. The chaperone toolbox at the single-molecule level: From clamping to confining. Protein science : a publication of the Protein Society. 2017;26(7):1291-302. Epub 2017/03/28. doi: 10.1002/pro.3161. PubMed PMID: 28342267; PMCID: PMC5477537.

199. Lin Z, Rye HS. GroEL-mediated protein folding: making the impossible, possible. Critical reviews in biochemistry and molecular biology. 2006;41(4):211-39. Epub 2006/07/20. doi: 10.1080/10409230600760382. PubMed PMID: 16849107; PMCID: PMC3783267.

200. Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. Cell. 2001;107(2):223-33. Epub 2001/10/24. doi: 10.1016/s0092-8674(01)00517-7. PubMed PMID: 11672529.

201. Gupta AJ, Haldar S, Milicic G, Hartl FU, Hayer-Hartl M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. Journal of molecular biology. 2014;426(15):2739-54. Epub 2014/05/13. doi: 10.1016/j.jmb.2014.04.018. PubMed PMID: 24816391.

202. Chan HS, Dill KA. A simple model of chaperonin-mediated protein folding. Proteins. 1996;24(3):345-51. Epub 1996/03/01. doi: 10.1002/(sici)1097-0134(199603)24:3<345::aid-prot7>3.0.co;2-f. PubMed PMID: 8778781.

203. Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. Annual review of biophysics and biomolecular structure. 2001;30:245-69. Epub 2001/05/08. doi: 10.1146/annurev.biophys.30.1.245. PubMed PMID: 11340060.

204. Ellis RJ. Protein folding: importance of the Anfinsen cage. Current biology : CB. 2003;13(22):R881-3. Epub 2003/11/15. doi: 10.1016/j.cub.2003.10.051. PubMed PMID: 14614845.

205. Motojima F, Motojima-Miyazaki Y, Yoshida M. Revisiting the contribution of negative charges on the chaperonin cage wall to the acceleration of protein folding. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(39):15740-5. Epub 2012/09/11. doi: 10.1073/pnas.1204547109. PubMed PMID: 22961256; PMCID: PMC3465421.

206. Tyagi NK, Fenton WA, Deniz AA, Horwich AL. Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. FEBS letters. 2011;585(12):1969-72. Epub 2011/05/26. doi: 10.1016/j.febslet.2011.05.031. PubMed PMID: 21609718; PMCID: PMC3144026.

207. Bechtluft P, van Leeuwen RG, Tyreman M, Tomkiewicz D, Nouwen N, Tepper HL, Driessen AJ, Tans SJ. Direct observation of chaperone-induced changes in a protein folding pathway. Science. 2007;318(5855):1458-61. Epub 2007/12/01. doi: 10.1126/science.1144972. PubMed PMID: 18048690.

208. Mashaghi A, Bezrukavnikov S, Minde DP, Wentink AS, Kityk R, Zachmann-Brand B, Mayer MP, Kramer G, Bukau B, Tans SJ. Alternative modes of client binding enable functional plasticity of Hsp70. Nature. 2016;539(7629):448-51. Epub 2016/11/04. doi: 10.1038/nature20137. PubMed PMID: 27783598.

209. Mashaghi A, Kramer G, Bechtluft P, Zachmann-Brand B, Driessen AJ, Bukau B, Tans SJ. Reshaping of the conformational search of a protein by the chaperone trigger factor. Nature. 2013;500(7460):98-101. Epub 2013/07/09. doi: 10.1038/nature12293. PubMed PMID: 23831649.

210. Ueno T, Taguchi H, Tadakuma H, Yoshida M, Funatsu T. GroEL mediates protein folding with a two successive timer mechanism. Molecular cell. 2004;14(4):423-34. Epub 2004/05/20. doi: 10.1016/s1097-2765(04)00261-8. PubMed PMID: 15149592.

211. Cecconi C, Shank EA, Bustamante C, Marqusee S. Direct observation of the three-state folding of a single protein molecule. Science. 2005;309(5743):2057-60. Epub 2005/09/24. doi: 10.1126/science.1116702. PubMed PMID: 16179479.

212. Naqvi MM, Heidarsson PO, Otazo MR, Mossa A, Kragelund BB, Cecconi C. Single-molecule folding mechanisms of the apo- and Mg(2+)-bound states of human neuronal calcium sensor-1. Biophysical journal. 2015;109(1):113-23. Epub 2015/07/15. doi: 10.1016/j.bpj.2015.05.028. PubMed PMID: 26153708; PMCID: PMC4572569.

213. Stigler J, Ziegler F, Gieseke A, Gebhardt JC, Rief M. The complex folding network of single calmodulin molecules. Science. 2011;334(6055):512-6. Epub 2011/10/29. doi: 10.1126/science.1207598. PubMed PMID: 22034433.

214. Dobson CM, Sali A, Karplus M. Protein Folding: A Perspective from Theory and Experiment. Angewandte Chemie (International ed in English). 1998;37(7):868-93. Epub 1998/04/20. doi: 10.1002/(sici)1521-3773(19980420)37:7<868::aid-anie868>3.0.co;2-h. PubMed PMID: 29711488.

215. Walther KA, Grater F, Dougan L, Badilla CL, Berne BJ, Fernandez JM. Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(19):7916-21. Epub 2007/05/02. doi: 10.1073/pnas.0702179104. PubMed PMID: 17470816; PMCID: PMC1876547.

216. Priya S, Sharma SK, Sood V, Mattoo RU, Finka A, Azem A, De Los Rios P, Goloubinoff P. GroEL and CCT are catalytic unfoldases mediating out-of-cage polypeptide refolding without ATP. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(18):7199-204. Epub 2013/04/16. doi: 10.1073/pnas.1219867110. PubMed PMID: 23584019; PMCID: PMC3645539.

217. Motojima F, Yoshida M. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. The EMBO journal. 2010;29(23):4008-19. Epub 2010/10/21. doi: 10.1038/emboj.2010.262. PubMed PMID: 20959808; PMCID: PMC3020636.

APPENDIX A

GROEL ACTIVELY STIMULATES FOLDING OF THE ENDOGENOUS SUBSTRATE PROTEIN PEPQ*

Summary

Many essential proteins cannot fold without help from chaperonins, like the GroELS system of *Escherichia coli*. How chaperonins accelerate protein folding remains controversial. Here we test key predictions of both passive and active models of GroELS-stimulated folding, using the endogenous *E. coli* metalloprotease PepQ. While GroELS increases the folding rate of PepQ by over 15-fold, we demonstrate that slow spontaneous folding of PepQ is not caused by aggregation. Fluorescence measurements suggest that, when folding inside the GroEL-GroES cavity, PepQ populates conformations not observed during spontaneous folding in free solution. Using cryo-electron microscopy, we show that the GroEL C-termini make physical contact with the PepQ folding intermediate and help retain it deep within the GroEL cavity, resulting in reduced compactness of the PepQ monomer. Our findings strongly support an active model of chaperonin-mediated protein folding, where partial unfolding of misfolded intermediates plays a key role.

***Reprinted with permission from “GroEL actively stimulates folding of the endogenous substrate protein PepQ” by Jeremy Weaver, Mengqiu Jiang, Andrew Roth, Jason Puchalla, Junjie Zhang & Hays S. Rye, 2017, Nature Communications, 8, Copyright [2017] by Springer Nature**

Introduction

Folding is a highly error prone process for many large and essential cellular proteins. Misfolding and aggregation often overwhelm the delicate thermodynamic balance that drives a protein toward its native state. Throughout evolutionary history, living systems have solved this problem with specialized, ATP-powered machines known as molecular chaperones (61). The Hsp60s or chaperonins are a central and essential family of the molecular chaperones, and the GroELS chaperonin system of *Escherichia coli* is one of the best studied examples (77, 84). GroEL is a homo-oligomer of 14, 57 kDa subunits that is arranged in two, seven membered rings stacked back-to-back. Each ring contains a large, open, solvent-filled cavity (142). The inner cavity surface of the uppermost domain (the apical domain) is lined with hydrophobic amino acids that capture non-native substrate proteins (143, 144). Substrate proteins that strictly depend upon GroEL for folding (so-called stringent substrate proteins) must be briefly enclosed within a complex formed by a GroEL ring and the smaller, ring-shaped co-chaperonin GroEL (73, 145-147). Formation of the GroEL- GroES complex first requires that a GroEL ring bind ATP, which triggers a series of conformational rearrangements of the GroEL ring, permitting GroES to bind and resulting in the encapsulation of the substrate protein. Enclosure of the substrate protein beneath GroES results in ejection and confinement of the protein inside the enlarged GroEL-GroES chamber (a *cis* complex) and initiation of protein folding. Folding continues within the isolated GroEL-GroES cavity for a brief period, until the complex is disassembled and the substrate protein, folded or not, is released back into free solution (72, 73, 146-

148).

Despite this detailed structural and functional knowledge, current models of GroEL-assisted folding remain divided into two general types based upon whether GroEL is presumed to act passively or actively (77, 84, 149). Passive models, like the Anfinsen cage or infinite dilution model, postulate that protein folding is only enhanced by GroELS because folding intermediates are prevented from aggregating by isolating them within the protective environment of the GroELS chamber (84, 149). Purely passive models implicitly assume that the folding of GroEL- dependent proteins are constrained only by the aggregation propensity of on-pathway folding intermediates. Active GroEL folding models, by contrast, assume that stringent GroEL- substrate proteins can and do populate off-pathway, kinetically trapped states. In this view, GroELS stimulates protein folding because these kinetically trapped intermediates benefit not only from protection against aggregation but also from additional, and essential, corrective actions provided by the chaperonin (150). The mechanism of this corrective action remains controversial, but has been suggested to come from either (1) repetitive unfolding and iterative annealing (151, 152) or (2) smoothing of a substrate protein's free energy landscape as a result of confinement inside the GroEL-GroES cavity, where either steric constraints and/or interactions within the chamber prevent unproductive folding pathways in favour of productive ones (84, 149).

Several stringent substrate proteins have been shown to display folding behaviour that is consistent with one or more predictions of active GroEL folding models

(78, 153, 154). Some of the most detailed analysis to date has been conducted with ribulose-1, 5-bisphosphate carboxylase oxygenase (RuBisCO) from *R. rubrum* and a double mutant of *E. coli* maltose binding protein (MBP) (78, 83, 153-155). While highly suggestive, these studies nonetheless leave the importance of active folding unclear. General conclusions about the impact of active folding cannot be robustly drawn from such a small number of examples. In addition, in the case of RuBisCO, the mismatch between the biological source of the substrate protein (*Rhodospirillum rubrum*) and the chaperonin (*E. coli*), leaves the biological consequences of these findings open to interpretation. Similarly, in the case MBP, it was necessary to employ an engineered double-mutant of this protein in order to study GroEL-stimulated folding, because wild-type MBP neither interacts with, nor needs the chaperonin for folding in its natural biological context. Thus, a convincing demonstration of active folding assistance by GroEL of a stringent, endogenous *E. coli* substrate protein has remained elusive. A recent study on the assisted folding of the *E. coli* HTP synthase/lyase DapA sought to address this problem (156). The results of this work suggested that DapA requires an active GroEL folding mechanism. However, a more recent study of DapA folding called key elements of this work into question (157).

In order to test the central predictions of passive and active models of chaperonin-mediated folding, we have re-examined the mechanism of GroELS-assisted protein folding using the biologically relevant, endogenous *E. coli* prolidase enzyme, PepQ. PepQ catalyses the hydrolysis of dipeptides that contain C-terminal proline residues (132, 158). It forms a homodimer, with each monomer (B50 kDa) built from two domains: a small, mixed a/b N-terminal domain and a pita-bread fold (159, 160) C-terminal domain that contains the active site (Figure A.1a; (132)). Two independent proteomics studies predicted that PepQ requires the assistance of GroEL- GroES for folding *in vivo* (161, 162). In addition, PepQ is a member of a protein structural family that is not represented among the well- characterized GroEL-substrate proteins. Here, using a combination of enzymatic assays, single-molecule fluorescence techniques, and cryo-electron microscopy (EM), we demonstrate that GroEL actively alters the folding of PepQ. Initial capture of a kinetically trapped PepQ monomer by a GroEL ring results in substantial unfolding, a process that relies in part on a direct, physical interaction between the PepQ folding intermediate and the unstructured GroEL C-terminal tails. Subsequent encapsulation of the partially unfolded folding PepQ monomer within the GroEL/ES chamber fundamentally alters the folding trajectory of the protein, resulting in a faster and more efficient search for the native state.

Methods

Bacterial strains

All bacterial strains used in this work were originally obtained from the laboratory of Dr. Arthur Horwich at Yale University Medical School.

Bacterial strains employed:

BL21—*E. coli* B dcm ompT hsdS(rB-mB-) gal BL21DE3—*E. coli* B dcm ompT hsdS(rB-mB-) gal [**1**DE3]

DH5a—*E. coli* fhuA2 lac(del)U169 phoA glnV44 F80’ lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17.

Proteins

Wild type and variants of GroEL (SR1 and C-terminal truncation mutants), GroES and wild type *E. coli* PepQ were all expressed and purified as described previously (78, 81, 83, 137, 154). The cysteine mutant of PepQ, A24C, was generated via standard site-directed mutagenesis (163) and the sequence was verified by DNA sequencing. This mutant was expressed and purified following the protocol for wild type PepQ.

In brief, GroEL was expressed from an inducible plasmid in *E. coli* BL21 in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation (142,000 *g*), followed by anion exchange chromatography (FastFlow Q, GE) at pH 7.4. GroEL fractions were concentrated by 70% (w/v) ammonium sulfate precipitation. This precipitate was solubilized and dialyzed against buffer at pH 6.8 containing 25% (wild-type GroEL) or 12.5% (all GroEL mutants) methanol. A second round of strong anion exchange (FastFlow Q, GE), run in the same methanol-containing buffer at pH 6.8, was used to strip co-purifying small proteins and peptides from the GroEL oligomers. To further remove contaminating proteins and peptides that remain tightly associated through prior stages of purification, GroEL fractions were gently agitated in the same methanol-containing buffer and Affi Blue Gel resin overnight at 4 °C under an argon atmosphere. The final sample was dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated, and snap frozen using liquid N_2.

GroES was expressed from an inducible plasmid in *E. coli* BL21(DE3) in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation (142,000 *g*), followed by acidification with sodium acetate, and anion exchange chromatography at pH 4.6 (FastFlow Q, GE). The sample was dialysed against buffer at pH 7.4 and applied to a strong anion exchange column (Source Q, GE). GroES was eluted with NaCl and enriched fractions were pooled. The sample was dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated and snap frozen using liquid N_2.

PepQ and PepQ mutants were expressed from an inducible plasmid in *E. coli* BL21(DE3) in LB at 37 °C. After cell disruption, the crude lysate was clarified by ultracentrifugation. The supernatant was applied to a strong anion exchange column (FastFlow Q, GE) at pH 7.4 and eluted with a gradient of NaCl. Fractions enriched for PepQ were pooled, and the protein was precipitated with 70% (w/v) ammonium sulfate. The sample was loaded on a hydrophobic interaction column (Phenyl Sepharose FF, GE) at pH 7.4 and eluted with a decreasing ammonium sulfate gradient. Fractions enriched for PepQ were pooled, dialysed into storage buffer (pH 7.4), supplemented with glycerol (15–20% v/v), concentrated and snap frozen using liquid N_2.

Labelling of PepQ

A24C PepQ was labelled using either 5-iodoacetamido- fluorescein (fluorescein, F), 5-(2-acetamidoethyl) aminonaphthalene 1-sulfonate (EDANS, ED), tetramethylrhodamine-5-iodoacetamide dihydroiodide (tetramethylrhoadmine, TMR), or Oregon Green 488 iodoacetamide (Oregon Green, OG). All dyes were obtained from Invitrogen (Molecular Probes). PepQ (B10 mg ml^{-1} in 50 mM Tris, pH 8, 100 mM KCl, 1 mM $MgCl_2$) was reduced with 0.5 mM tris(2- carboxyethyl)phosphine (TCEP) TCEP and labelled with a 10-fold excess of reactive dye, added in 1 addition for 3 h at 23°C. The reaction was quenched by adding glutathione (5 mM), and the labelled PepQ was first separated from unbound dye by gel filtration (PD-10 column, Pharmacia), followed by re-purification of the labelled protein with high-resolution ion exchange chromatography (MonoQ, GE). The extent of labelling was determined by protein quantification by the Bradford assay (Bio-Rad) and dye quantification under denaturing conditions using known dye molar extinction coefficients (37, 63). Unique labelling of a single cysteine was verified by both denaturing anion exchange chromatography (MonoQ, GE) in 8 M urea buffer and by detection of a single major and fluorescent tryptic peptide upon separation by C8 reverse-phase chromatography (164).

Folding assays

PepQ refolding assays were conducted in TKM buffer (50 mM Tris-HCl, pH 7.4, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 2 mM DTT) using a protocol similar to that employed previously for RuBisCO (75, 78, 83, 137, 164), with differences in the folding buffer composition, duration of post-reaction incubation, and the detailed assay method (132). All folding assays were conducted using PepQ that was diluted at least 40-fold into 8 M urea, 25 mM glycine phosphate, pH 2, and incubated at room temperature for at least 20 min before further use. CD spectra show a complete loss of secondary structure under these conditions. Spontaneous refolding of PepQ was initiated by a 50-fold dilution from denaturant into TKM buffer (50 mM Tris-HCl, pH 7.4, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 2 mM DTT) and quenched through the addition of excess GroEL. Chaperonin-mediated folding reactions using either wild-type or mutant tetradecameric GroEL began with a 50-fold dilution of denatured PepQ into TKM buffer containing a particular GroEL variant. GroES and ATP were added to initiate the reaction cycle and the reaction was quenched with hexokinase and glucose (81, 83, 154). Folding reactions in single-ring mutants of GroEL were done similarly, except quenching was accomplished by the simultaneous addition of EDTA and incubation of the sample at 0 °C (165). After quenching, all samples were incubated for 60 min at room temperature to allow for dimerization. The enzyme activity of all samples was measured through an NAD-coupled reaction using alanine dehydrogenase from *B. subtilis* (132).

Measuring PepQ persistence in solution

Fluorescein-labelled PepQ (24F) was allowed to refold spontaneously or in the presence of the chaperone system (as in Fig. 1b, see Folding assays in Methods section). Samples were taken after 60 min and run on 10% SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Gels were imaged with a Typhoon Trio (GE Healthcare) and quantified with ImageJ.

Fluorescence and light scattering

Light scattering and fluorescence measurements were conducted with a T-format fluorometer (PTI), equipped with a jacketed cuvette holder for temperature control with a high-precision circulating water bath (Neslab). For both types of experiments, the assays were initiated by diluting acid-urea denatured PepQ at least 50-fold into temperature-equilibrated TKM buffer (23 °C) in the presence or absence of GroEL. Tryptophan fluorescence was monitored with excitation at 295±4 nm and emission at 340±4 nm. The excitation and emission wavelengths were both 340±1 nm for light scattering experiments.

Stopped-flow fluorescence

Stopped-flow experiments were conducted using an SFM-400 rapid mixing unit (BioLogic) equipped with a custom-designed two-channel fluorescence detection system (78, 83, 132, 164). Mixing was done using equal volume injections from two syringes, one containing GroEL-PepQ complexes and one containing GroES and ATP. Each solution was in TKM buffer. Measurements were taken every 150 ms.

Steady-state FRET

Steady-state fluorescence measurements were conducted with a T-format fluorometer (PTI), equipped with a jacketed cuvette holder for temperature control with a high-precision circulating water bath (Neslab). FRET efficiencies were calculated from the changes in donor-side fluorescence of matched donor only and donor plus acceptor labelled molecules (75, 164).

Single-molecule fluorescence microscope

Built on a research quality, vibrationally isolated optical table, the system is constructed around a Nikon Eclipse Ti-U inverted microscope base using an x 60/1.4NA CFI Plan Fluor oil immersion objective. The microscope base is outfitted with a precision, 2-axis stepper motor sample stage (Optiscan II; Prior) and a custom-designed confocal optical bench with three, independent detection channels. Each detection channel is configured with an optimized band-pass filter set for wavelength selection and a low-noise, single photon counting APD unit (SPCM-AQRH-15; Excelitas). Photon pulses are collected and time stamped with either a multichannel hardware correlator (correlator.com) or high speed TTL counting board (NI9402; National Instruments). Sample excitation is provided by either one or a combination of three lasers: two diode lasers (488 and 642 nm; Omicron) and one diode-pumped solid state laser (561 nm; Lasos). The free-space beams of each laser are each coupled to a three-channel fibre combiner (PSK-000843; Gould Technologies) and the combined output is directed into the sample objective with a custom, triple-window dichroic filter (Chroma). Each laser is addressable from the integrated control and data acquisition software, custom developed using LabView (National Instruments).

PepQ refolding by fluorescence correlation spectroscopy

PepQ-24TMR was diluted greater than 40-fold (to 5 mM) into 8 M urea, 25 mM glycine phosphate, pH 2 and incubated for 20 min at room temperature. For spontaneous folding reactions, this PepQ-TMR was then diluted to 100 nM in the same solution. The folding reaction was initiated by dilution of PepQ to 2 nM in TKM buffer. Folding was quenched by the addition of 50 ml of the refolding reaction to 50 ml of 1 mM GroEL in TKM buffer. For GroEL-mediated folding, 5 mM denatured PepQ-24TMR was diluted to 100 nM in TKM buffer containing GroEL (200 nM final tetradecamer concentration). After a 10 min incubation at room temperature, this solution was diluted into TKM buffer containing GroEL, GroES and an ATP-regeneration system (154). Folding was then initiated by the addition of ATP. The final concentration of ATP was 2 mM, GroEL was 1 mM, and GroES was 2 mM. Folding was quenched by the addition of 20 ml of the reaction mixture with an equal volume of hexokinase and glucose. Dimerization was not observable in refolding assays conducted at 1–2 nM PepQ, based on a reproducible lack of detectable enzymatic activity, even with up to eight hours of incubation at 23°C. PepQ enzymatic activity is, however, detectable when the native dimer is diluted to 1–2 nM. Fluorescence correlation spectroscopy (FCS) data were collected by placing 10 ml of the quenched reaction mixtures onto BSA-blocked coverslips mounted on a custom-built, inverted confocal microscope and covered with a humidified chamber to prevent evaporation.

Autocorrelation curves were collected for 2 min using 50 μW continuous input power from a 561 nm diode-pumped solid state laser. Autocorrelation curves were normalized in mean amplitude between 10^{-6} and 10^{-5} s for display purposes. As standards, the autocorrelation curves of PepQ fully bound to GroEL (obtained by not adding ATP to a folding reaction), as a native dimer (obtained by diluting native PepQ-24TMR in buffer), and as a native monomer (obtained by allowing a GroEL-mediated folding reaction with 1 nM PepQ to continue for an hour) were also determined. Each autocorrelation curve was fit using a multi-component model (166, 167) to account for populations of freely diffusing and GroEL-bound PepQ. Curve fitting was conducted using two different approaches. First, the diffusion coefficient of each population was fixed and the fractional population was allowed to vary. Second, the average diffusion coefficient of the entire population was determined. The resulting refolding curves obtained from these two methods were statistically equivalent.

Two-colour single-molecule co-incidence detection

The 24OG and 24TMR PepQ variants were each diluted to 5 mM in acid-urea and allowed to unfold at room temperature. For spontaneous folding, the two solutions were diluted together into acid-urea to a concentration of 50 nM each. Folding was initiated by a 50-fold dilution into TKM buffer containing 0.1 mg ml $^{-1}$ BSA. BSA has no effect on the folding rate of PepQ (Figure A.10b), but was necessary to prevent loss of protein at very-low concentrations to liquid handling equipment. After 10 min at 23 °C, samples were diluted 20-fold into the same buffer and 10 ml samples were immediately placed on a BSA-blocked, optical glass coverslip mounted in a custom holder, fitted on the microscope stage. Samples were covered with a humidifier cap to prevent evaporation. For chaperone-mediated folding, the two 5 mM solutions of denatured, labelled PepQ were diluted together to 50 nM each into TKM containing GroEL and GroES, and folding was initiated by the addition of ATP (2 mM) to a solution containing: 50 nM PepQ-24OG, 50 nM PepQ-24TMR, 200 nM GroEL, 400 nM GroES). After incubated at 23 °C for at least one hour, the reaction was halted by a 50-fold dilution into TKM. This sample was then diluted a further 20-fold into TKM containing 0.1 mg ml $^{-1}$ BSA and immediately assayed. Fluorescence burst data were collected for each sample over a 1 min window using 100 ms sampling bins. Simultaneous excitation was provided from two co-aligned lasers (488 and 561 nm), each providing 200 mW of power at the sample.

To quantify the formation of native PepQ dimers resulting from productive GroEL folding, as well as the formation of PepQ aggregates during spontaneous folding, we developed a cross-correlation statistic (plotted in Figure A.3) that evaluates the percent photon arrival time overlap between two time streams. To begin, each time stream was normalized so the maximum spike intensity amplitude had a value of one. A threshold filter was then applied (5*r.m.s.) to both color channels to isolate spike activity and remove low-level detector noise. The filtered time streams were used to create a binary mask of spike events. On the basis of a particle transit time through the excitation volume of about 1 ms, both binary time streams were re-binned in 1 ms time bins. The cross-correlation versus time lag between two time streams T_1 and T_2, each with a total of N_1 and N_2 non-zero time bins, was then generated for the þ EL/ES/ATP and spontaneous activity data.

With this normalization, the autocorrelation of any time stream had a value of 1, while the minimum cross-correlation value was bounded at zero. Due to the non- zero probability of photons randomly arriving at two detectors at the same time, the minimal cross-correlation value was not zero.

To assess our cross-correlation measure we used the photon arrival data from either EL/ES/ATP or spontaneous PepQ activity to generate an expected baseline activity (that is, zero significant co-incidence). Each time stream was compared to a 5 s cyclically shifted version of itself to examine the correlation between two nominally uncorrelated time streams of identical photon rate and noise (denoted as No Overlap, Figure A.3c,d). In both of these baseline cases, there is approximately a 1% cross-correlation independent of lag time or detector channel (red line). We then calibrated the cross-correlation statistic for several data streams of a known and fixed amount of similarity. To do this, we replaced a segment of a time stream T1 with an equal length segment of a time stream T2 at a random location. The ratio of the segment length to total length then corresponded to the percent overlap. The original data stream T2 and the altered data stream T1 then represented two data streams of known overlap and whose cross-correlation could be used for comparison. This process was repeated 20 times for each percent overlap. The cross-correlation results were averaged and the uncertainty in the mean for each lag was monitored. The resulting family of cross-correlation curves (Figure A.3a,b) was then used to assess the level of overlap between two-colour channels that had not been shifted in time.

We then tested the null hypothesis that the pairwise differences between the cross-correlations values of the spontaneous data and various possible overlap simulations had a mean equal to zero. The resulting P values from this family of t-tests indicated the most likely zero null hypothesis occurred for an overlap of 0.75% (Figure A.3b, inset).

Triplet state conversion of fluorescent PepQ variants

Native, TMR- or Oregon Green-labelled PepQ dimer (24TMR and 24OG, respectively) was diluted to 1 nM (dimer) in TKM buffer with supplemental BSA (0.1 mg ml $^{-1}$). An amount of 10 ml of sample was pipetted onto a BSA-blocked coverslip, mounted as described for Fig. 2, using a 561 nm laser for 24TMR and a 488 nm laser for 24OG. For each sample, data were collected for 2 min, with 500 ms collection bins, at each laser power level. The entire, non-normalized data set for each dye was fit globally using IgorPro (Wavemetrics) to an autocorrelation function that included a correction factor for the effect of a triplet state population (168).

Protease protection

The protease sensitivity of non-native PepQ bound to a GroEL ring was conducted as described previously for the substrate protein RuBisCO (83, 137). Briefly, denatured PepQ (100 nM) labelled at position 24 with fluorescein (PepQ-24F) was bound to asymmetric GroEL-GroES-ADP bullets (120 nM) (83). Chymotrypsin (0.3 mg ml^{-1}) was added, and time points were taken, with the reaction stopped by addition of phenylmethylsulfonyl fluoride (PMSF) (1 mM). Samples were run on 10% SDS–PAGE and imaged using a Typhoon Trio (GE Healthcare).

GroEL ATPase activity

The ATPase activity of GroEL was assayed using an NADH-coupled reaction. In brief, the GroEL (200 nM) ATPase cycle was monitored in the presence of ATP (2 mM) and GroES (400 nM) using 10 U ml^{-1} pyruvate kinase, 10 U ml^{-1} lactate dehydrogenase, 1 mM phosphoenolpyruvate and 0.2 mM NADH. This system regenerates ATP, maintaining it at a constant concentration, and produces a decrease in the absorbance at 340 nm as NADH is consumed (81, 137, 154, 169, 170). The rate of spontaneous ATP hydrolysis under each condition without GroEL was also determined to control for effects on the coupling system.

Cryo-electron microscopy

PepQ was denatured in 8 M urea, 25 mM glycine phosphate, at pH 2 and incubated for 30 min at room temperature. Denatured PepQ (50 mM), in droplets of 4.6

ml (2.3 mM per addition) was titrated into solutions of either GroEL or D526 (8 mM tetradecamers, 100 ml) in TKM buffer, followed by rapid, repeated mixing and then incubated at room temperature for 5 min. The final concentration of PepQ was 7 mM. 3 ml of this PepQ/GroEL mixture was applied to a C-Flat 1.2/1.3 400 mesh holey carbon grid at 20 °C with 100% relative humidity and vitrified using a Vitrobot (Mark III, FEI company, Netherlands). The thin-ice areas that showed clear and mono-dispersed particles were imaged under an FEI Tecnai F20 electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 KV. Data were collected on a Gatan K2 Summit direct detection camera (Gatan, Pleasanton CA) in electron counting mode (171) at a nominal magnification of x 19,000, yielding a pixel size of 1.85 Å. The dose rate was 10 e^{-} pixel^{-2} s^{-1} at the camera. A 33-frame movie stack was recorded for each micrograph, for a total exposure time of 6.6 s. The total dose onto the specimen was 19 e^{-} Å^{-2}.

Image processing and map visualization

For the wild-type GroEL-PepQ complex, 1,450 micrographs were collected and aligned iteratively and filtered based on electron dose using Unblur (172). Sum images (1,109) showing strong power spectra were selected for particle picking in EMAN2 (173), yielding 217,317 particles with a box size of 160 160 pixels (2). These particles were processed in Relion (174) with a downscaling factor of 2, respectively. Two runs of reference-free 2D classification were performed to produce a 'cleaner' data set containing 170,639 particles, which were separated into four classes in 3D classification with no symmetry applied. Particles belonging to apo GroEL and PepQ-bound GroEL were used for final asymmetric 3D refinement, respectively. The processing procedure used for the D526-PepQ complex was the same as used for the wild type GroEL complex. Briefly, 847 movie stacks were collected and 703 aligned images were picked, yielding 224,696 particles. 117,040 clean particles were screened after two runs of reference-free 2D classification. 3D classification and asymmetric refinement was performed similar as wild-type GroEL-PepQ complex (Supplementary Fig. 6). The final resolutions of the 3D density maps for both wild-type GroEL and D526 were 7.9 Å for the apo states and 8.3 Å for the PepQ-bound states, assessed with the gold-standard criterion at 0.143 Fourier shell correlation (174). Local resolutions were estimated using Blocres (175). The unwrapping of the maps was done with ‘e2unwrap3d.py’ in EMAN2. Visualization and fitting of atomic models into the cryo-EM density maps, were done in UCSF Chimera (176).

Circular dichroism spectroscopy

PepQ was diluted $>$100-fold into 25 mM sodium phosphate, pH 7.2, 100 mM $MnCl_2$ (native) or 25 mM sodium phosphate, pH 2.1, 8 M Urea (denatured) to a concentration of 0.1 mg ml $^{-1}$. Following an equilibration at 23 °C for 15 min, samples were loaded into a 1 cm path length cuvette and the circular dichroism (CD) of the sample measured in the far ultra-violet region using an Aviv 202:CD spectrometer (Aviv Biomedical). The sample temperature was equilibrated in the spectrometer to 25 °C before the initiation of measurements and was maintained at this temperature throughout. The CD signal at each wavelength was averaged for 30 s, using 1 nm wavelength steps. Scans of each sample buffer were used as blanks.

Encapsulation of PepQ by GroEL

PepQ-24F (100 nM) was denatured in 8 M urea, 25 mM glycine phosphate, pH 2 and bound to wild-type or D526GroEL-GroES-ADP bullets (200 nM) or full-length or D526 single-ring GroEL (300 nM) supplemented with GroES (600 nM). A single turnover was initiated by the addition of ATP (2 mM) followed by quenching with hexokinase and glucose after 10 s. Unencapsulated PepQ was digested with Proteinase K (0.5 mg ml^{-1}) for 10 min before quenching with PMSF (1 mM). Samples were run on SDS–PAGE and scanned for fluorescein fluorescence using a Typhoon Trio (GE Healthcare). Band intensity was measured with ImageJ.

Results

Slow spontaneous PepQ folding is not caused by aggregation

Upon dilution from chemical denaturant, PepQ folds spontaneously at room temperature (23 °C) to a final yield of 50–60% with an observed half-time of B20 min (Figure IV.1b). However, in the presence of the cycling GroEL-GroES system, PepQ folds with an observed half-time of B1 min to a final yield of 80–90% (Figure A.1b). Encapsulation of PepQ within a non-cycling chaperonin complex, composed of the GroEL single-ring mutant SR1 and GroES, also results in accelerated refolding, consistent with previous observations from other GroEL-substrate proteins (Figure

A.1b;) (73, 156, 165, 177) Inside the static SR1-ES cavity, PepQ folds at a rate similar to that observed with cycling wild type GroEL at 23 °C, although it displays a consistently lower yield. Thus, while PepQ does not require GroEL to fold, the chaperonin accelerates the folding rate of the enzyme by 15–20-fold, while increasing the native state yield by 40%.

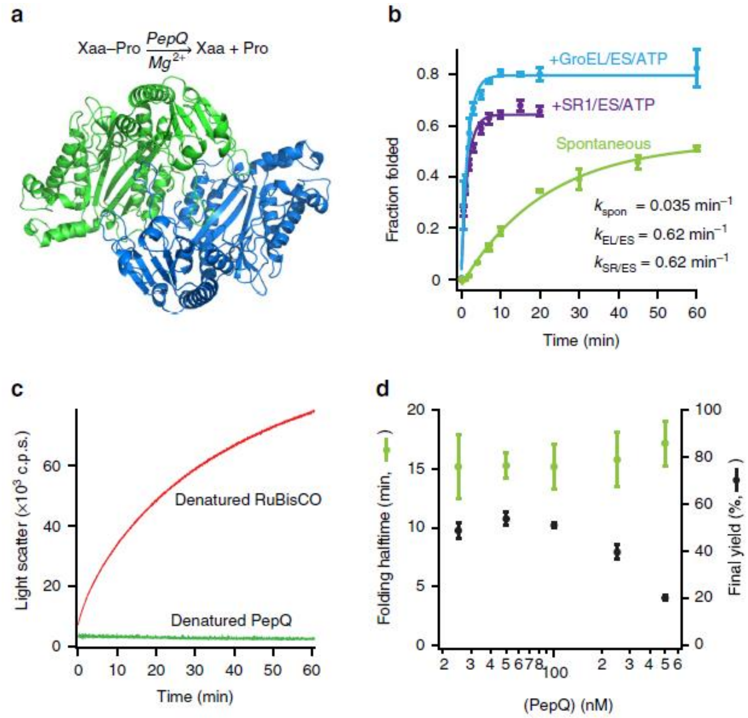


Figure A.1 Stimulated folding of PepQ by GroEL does not depend on large-scale suppression of aggregation.

(a) The E. coli metalloprotease PepQ catalyses the hydrolysis of dipeptides containing C-terminal proline residues. The structure of the native PepQ homodimer (PDB ID: 4QR8) is shown, illustrating the pita-bread fold common to this enzyme family. (b) Refolding of PepQ was monitored by the recovery of enzymatic activity. PepQ was denatured in acid-urea and then diluted into either buffer alone (100 nM; spontaneous, green) or buffer containing GroEL (200 nM). The GroEL-PepQ binary complex was then supplemented with GroES (400 nM) and ATP (2 mM) to initiate folding (þ GroEL/ES/ATP, blue). In a parallel experiment, denatured PepQ was bound to the single-ring mutant of GroEL, SR1 (300 nM), and refolded in the presence of GroES (600 nM) and ATP (2 mM; þ SR1/ES/ATP, purple). Data were fit to a single-exponential rate law (solid lines), resulting in observed folding rate constants of 0.62±0.05 min^{-1} for GroEL, 0.62±0.09 min^{-1-} for SR1 and 0.035±0.005 min^{-1} for the spontaneous reaction. Error bars show the standard deviation of three independent experiments. (c) Large-scale aggregation of PepQ and RuBisCO was examined by static light scattering at 340 nm. PepQ (green) and RuBisCO (red) were each denatured in acid-urea and then separately diluted into buffer at 23°C (100 nM final monomer). Each trace is the average of three separate experiments. (d) The rate and yield of spontaneous PepQ folding as a function of enzyme concentration is shown. Chemically denatured, wild type PepQ was diluted 50-fold into buffer to yield spontaneous folding reactions at the indicated final monomer concentration. The folding rate at each protein concentration (green) and native state yield (black) are shown. Error bars show the s.d. of three independent folding experiments.

The observation of slow spontaneous folding, in combination with a decreased native state yield, suggests that PepQ folding is inhibited by non-productive side reactions like misfolding or aggregation. We therefore examined the fate of the PepQ that

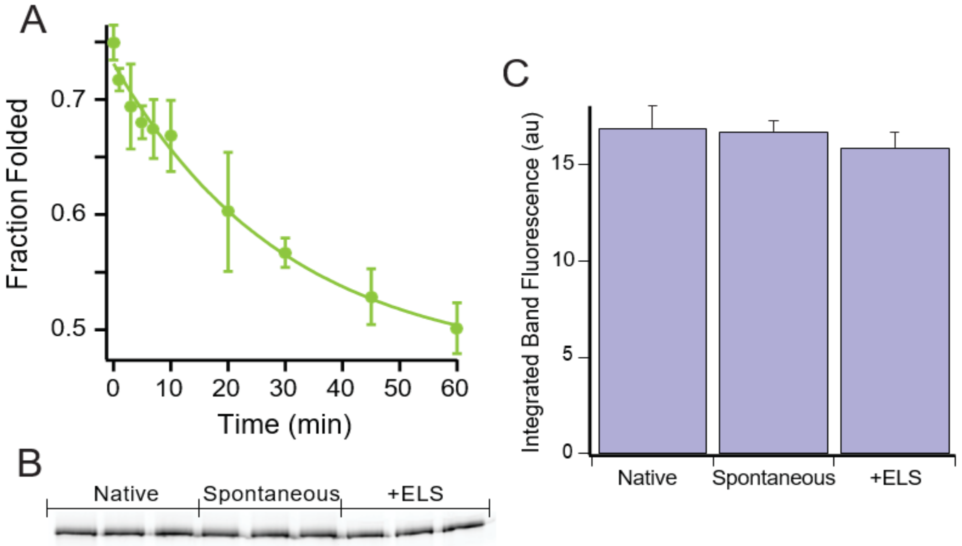
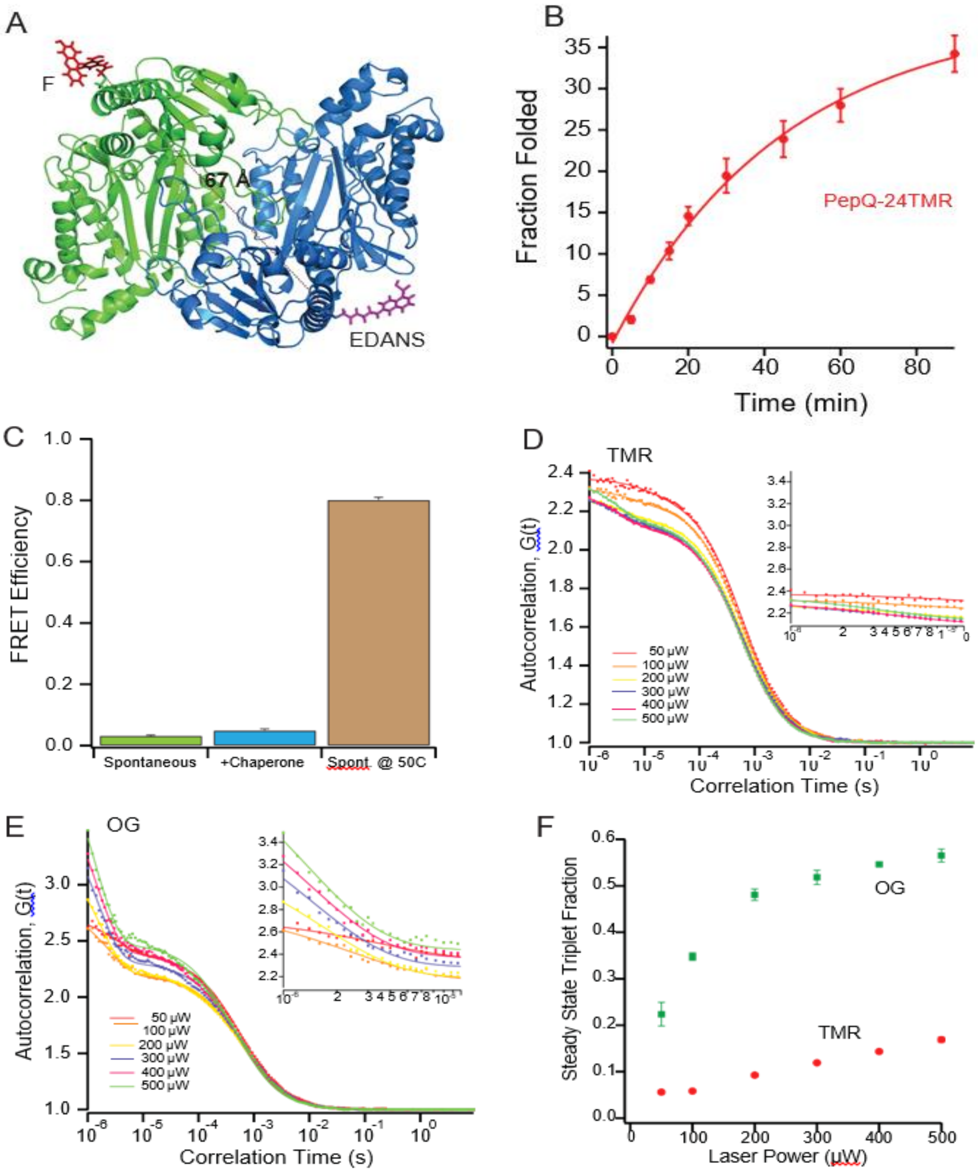


Figure A.2 PepQ slowly and spontaneously populates a misfolded state that cannot be rescued by GroEL, but which remains completely soluble.
(A) The rate at which PepQ becomes refractory to refolding by GroEL was measured by examining the loss of recoverable enzymatic activity over time. PepQ was denatured in acid-urea, diluted directly into buffer (100 nM) and then incubated at 23 °C. At the indicated times, samples were removed and mixed with an equal volume of a solution containing GroEL (400 nM) and GroES (800 nM), followed by addition of ATP (2mM). Each GroEL-supplemented sample was incubated for 60 min to permit assisted folding to reach completion; the total amount of native PepQ was determined by enzymatic assay. The observed final PepQ yields were fit to a single-exponential rate law (solid line), resulting in an observed rate of 0.033 ± 0.006 min-1. Error bars show the standard deviation of three independent experiments. (B) The amount of PepQ in solution during the folding reaction was measured to determine if the decreased yield in spontaneous folding results from loss to the microcentrifuge tube surface. PepQ-24F was denatured in acid-urea and then diluted into either buffer alone (100 nM; *Spontaneous*) or buffer containing GroEL (200 nM). For the GroEL sample, GroES (400 nM) and ATP (2 mM) was added to initiate folding (+*ELS*). Native PepQ-24F was also diluted into buffer (100 nM; *Native*). Samples were taken after 60 minutes and analyzed by SDS-PAGE with fluorescent gel scanning. Each lane shows an independent folding experiment. (C) Gel bands from (B) were quantified with ImageJ. Error bars show the standard deviation of three independent experiments.

fails to reach the native state. While this material remains fully in solution, over time it loses the ability to fold productively, even with assistance from GroEL (Figure A.2). Approximately half of the PepQ population becomes refractory to GroEL-mediated folding with a time constant that is similar to that observed for productive spontaneous folding. Because the non-native states of many chaperonin-dependent proteins are highly prone to aggregation, we sought to determine whether inefficient PepQ folding was due to aggregation. We first examined the static light scattering of a spontaneous folding reaction in which PepQ was rapidly diluted from denaturant into refolding buffer. Surprisingly, PepQ displayed no significant increase in light scattering, even after 1 hr of incubation at 23 ° (Figure A.1c). By contrast, *R. rubrum* RuBisCO, a stringent GroEL substrate protein well known to aggregate at 23 °C (69, 78, 178) showed a rapid and substantial increase in light scattering under the same conditions (Figure A.1c). These observations indicate that denatured PepQ does not form high concentrations of large aggregates, at least under the conditions of the spontaneous folding assay. However, PepQ could form inhibitory aggregates that are too small or rare to be well detected by light scattering. If true, the observed rate of spontaneous PepQ folding should be a sensitive function of the total protein concentration. Strikingly, over a concentration range from 25 to 500 nM, the observed half-time of PepQ folding remained unchanged, although we did observe a decrease in the native state yield as the protein concentration was increased above 250 nM (Figure A.1d). The concentration independence of the PepQ folding rate suggested that the slowness of spontaneous folding is not caused by inhibitory aggregation. To further test this conclusion, we examined PepQ folding at low protein concentrations using a set of fluorescence-based assays. We first introduced a surface-exposed Cys residue into the first helix of the PepQ N-terminal domain (A24C), which permitted unique attachment of exogenous fluorescent probes (Figure A.3A). Importantly, PepQ labelled at position 24 with small dyes like IAEDANS (PepQ-24ED), fluorescein (PepQ-24F), Oregon Green (PepQ-24OG) or tetramethyl rhodamine (PepQ-24TMR) displayed no apparent alteration in enzymatic activity or stability. Spontaneous folding of the PepQ-24ED variant displayed no significant difference compared to wild type PepQ, and the PepQ-24F, PepQ-24OG, PepQ-24TMR variants folded only slightly more slowly (Figure A.3B). We used these labelled PepQ variants in an intermolecular Förster Resonance Energy Transfer (FRET) assay designed to examine aggregate formation during spontaneous PepQ folding (132). In this assay, two differently labelled PepQ monomers were employed: PepQ-24ED as the donor and PepQ-24F as the acceptor. In the native PepQ dimer, these sites are positioned too far apart for Förster coupling (Figure A.3A), so that any observed FRET signal should report primarily on aggregate formation. When the two PepQ samples were mixed, denatured and diluted together into refolding buffer at 50 °C, formation of PepQ aggregates was readily observed as a robust FRET signal (80% FRET efficiency; Figure A.3C). Surprisingly, when the same experiment was conducted under spontaneous folding conditions at 23 °C, the observed FRET efficiency was less than 4%, suggesting a lack of significant aggregation (Figure A.3C).

Figure A.3 Characterization of PepQ variants labeled with fluorescent dyes. (A) Structure of the PepQ dimer (PDB ID: 4QR8) showing the position of an engineered Cys residue at position 24 that was used for the attachment of the exogenous fluorescent dyes like EDANS and fluorescein (*F*). For the purpose of illustration, dye structures were modeled into the PepQ structure using PyMol. While the dye orientations shown should only be considered approximate, the distance between the Cys sulfur atoms across the native dimer interface should be ~ 67 Å. (B) Spontaneous folding of PepQ-24TMR folding was monitored by the recovery of enzymatic activity. PepQ was denatured in acid-urea and then diluted into buffer (100 nM). Data were fit to a single-exponential rate law (solid lines), resulting in observed folding rates of 0.022 ± 0.004 min-1 for the spontaneous reaction. Error bars show the standard deviation of three independent experiments. Folding of PepQ-24C labeled with the other exogenous fluorescent probes used in this study were all very similar. (C) PepQ aggregation was monitored with intermolecular FRET using two different, denatured populations of labeled PepQ. For the donor sample, PepQ was labeled at position 24 with IAEDANS (PepQ-24ED). The acceptor sample was created by labeling position 24 with fluorescein (PepQ-24F). Acid-urea denatured, donor-labeled PepQ (50 nM) and either denatured, unlabeled PepQ (50 nM) or denatured, acceptor-labeled PepQ (50 nM) were mixed together and diluted into buffer alone at 23 °C (*green*), 55 °C (*brown*), or into buffer containing GroEL (200 nM) at 23 °C, followed by the addition of GroES (400 nM) and ATP (2 mM; *blue*). The observed donor-side, steady state FRET efficiency was calculated from the donor fluorescence emission spectra of matched donor-only and donor-acceptor samples. Error bars show the standard deviation of three experimental replicates. In the absence of aggregation, efficient folding should result in no detectable FRET from a mixed PepQ sample, assuming a folding efficiency of 80-90% (Figure 1B), and noting that the distance between labeled positions is at least 20 Å greater than the Förster distance typically observed with this pair of probes 2 (Figure S2A). The FRET efficiency observed following productive folding with GroEL is 4.9 ± 0.5%, while for spontaneous folding the observed FRET efficiency is 3.2 ± 0.5%. PepQ can be forced to form aggregates that are easily detectable by FRET (80 ± 0.8% efficiency) by raising the sample temperature to 50 °C during the initial dilution of PepQ from denaturant. (D-F) The propensity of the fluorescent probes, TMR and OG, to convert to the triple state was measured through changes in the FCS curve of PepQ labeled with each probe as a function of excitation power. The autocorrelation of native PepQ-24TMR (D) or PepQ-25OG (E) (1 nM dimer) was measured at different input laser powers (50-500 uW; 488 nm for OG and 561 nm for TMR). In each case, the FCS curve was acquired three times at each power setting; one representative trace for each power setting is shown for each dye. Data were fit using an autocorrelation function incorporating a term for triplet state formation by the dye 3,4. (F) The steady state triplet fraction of each dye from the fits in (D) and (E) was plotted against the corresponding laser power. Error bars represent the average of three traces.



We next examined PepQ folding and aggregation at extremely low protein concentrations using single-molecule detection techniques. First, samples of PepQ-24TMR were denatured in acid-urea and spontaneous folding was initiated by rapid dilution (50-fold) into refolding buffer at 23 °C, yielding a final monomer concentration of 2 nM. This sample was allowed to fold spontaneously at 23 °C and samples were removed and mixed with a large excess of GroEL at different time points. Excess GroEL was added to both quench the folding reaction and increase the effective diffusion time of uncommitted PepQ monomers, which were bound by the much larger GroEL tetradecamer (154, 156). Fluorescence correlation spectra (FCS) were then acquired for each time point and the fraction of folded versus non-native PepQ was extracted from each autocorrelation curve by comparison with two reference states: non-native PepQ-24TMR bound to GroEL and native PepQ-24TMR. The normalized autocorrelation

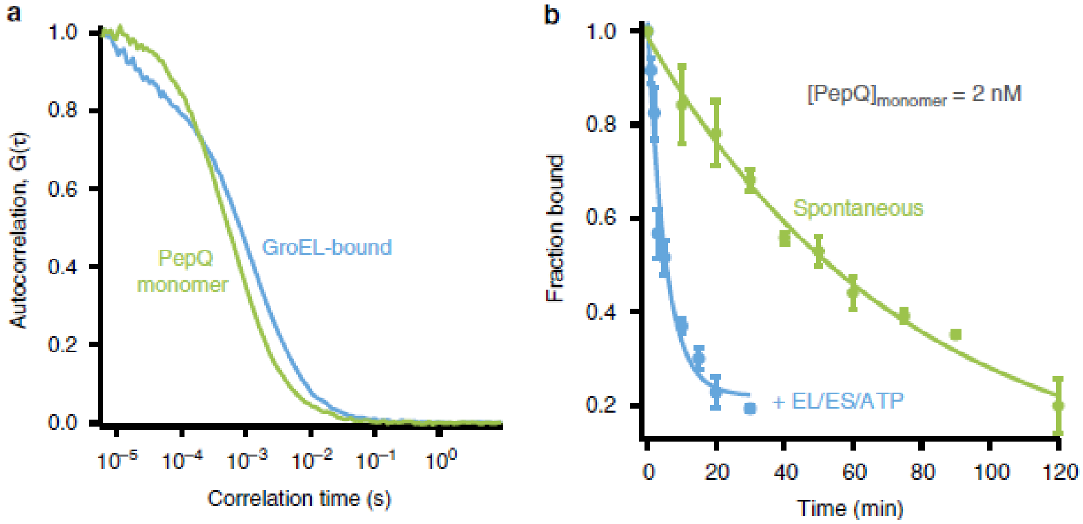


Figure A.4 GroEL accelerates folding of PepQ at very low protein concentrations.

a) The difference in diffusion time of the PepQ monomer in free solution, versus bound to a GroEL tetradecamer, can be detected by FCS. The observed FCS curves of the PepQ-24TMR monomer (2 nM), either alone (green) or bound to GroEL (1 mM; blue) are shown. (b) Refolding of PepQ-24TMR was monitored by FCS, using the observed shift in diffusion time shown in a. PepQ-24TMR was denatured in acid-urea and diluted either directly into buffer (2 nM; spontaneous, green) or into buffer containing wild type GroEL (1 mM). Refolding with GroEL was initiated by addition of GroES (2 mM) and ATP (2 mM; þ EL/ES/ATP, blue). At the indicated times, GroEL-mediated folding was quenched by depletion of ATP before FCS measurement, while samples of the spontaneous reaction were mixed with GroEL alone (1 mM) before FCS measurement in order to quench folding and shift the diffusion time of any uncommitted PepQ monomer. The observed fractional change in diffusion time was fit to a single-exponential rate law (solid lines), resulting in rate constants of 0.19±0.04 min^{-1} for GroEL-mediated folding and 0.013±0.002 min^{-1} for spontaneous folding. Error bars show the s.d. of three experimental replicates.

curves of these two reference states are shown in Fig. 2a. The rate of spontaneous PepQ-24TMR folding, measured at 2 nM by FCS, closely recapitulates the rate of folding of the protein observed at 100 nM (Figure A.4b). More importantly, when the same experiment was conducted with fully cycling GroEL-GroES, folding of PepQ-24TMR was stimulated by the same 15–20-fold observed at higher concentrations (Figure IV.4b). Using single-molecule, two-color co-incidence detection we next probed the assembly status of PepQ during spontaneous folding at 2 nM. As a control, we first examined formation of PepQ-24TMR was denatured and refolded at a total PepQ concentration of 100 nM in the presence of the active GroEL- GroES system, in order to permit formation of PepQ dimers carrying both probes. This sample was then diluted to 100 pM PepQ and fluorescence bursts were collected using using a two-channel, confocal-type single-molecule microscope (Figure A.5a, inset). The native PepQ dimer was readily detectable as a robust fraction of coincident events (Figure A.5a). Notably, the observed coincident fraction (B10%) was lower than the theoretically expected value of B50% for a 1:1 mixture of PepQ-24OG and PepQ-24TMR. This difference is most likely due to the much greater tendency of OG to convert to a long-lived dark (triplet) state, relative to TMR (Figure A.3D–F), which results in a substantial decrease in observed co-incidence.

To examine PepQ monomer assembly during spontaneous folding, samples of PepQ-24OG and PepQ-24TMR were mixed at 1:1, denatured in acid-urea and then rapidly diluted (50-fold) into refolding buffer at 23 °C to initiate spontaneous folding at a final protein concentration of 2 nM monomer. This sample was incubated at 23 °C for 10 min, then diluted another 20-fold to a final PepQ monomer concentration of 100 pM.

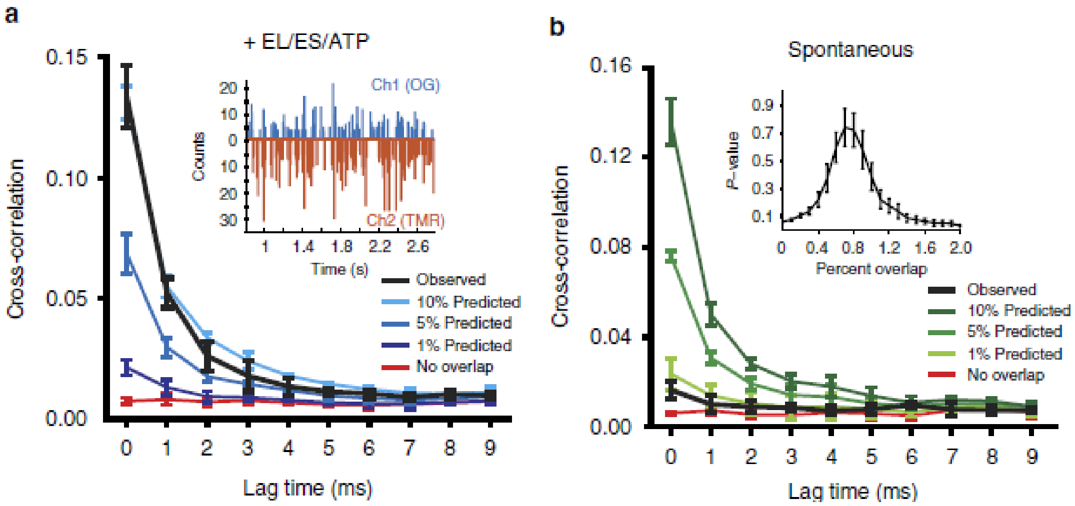


Figure A.5 Non-native PepQ does not aggregate at very low concentrations. (a) The formation of the native PepQ dimer was examined with single-molecule, two-colour co-incidence detection. Samples of PepQ-24TMR and PepQ-24OG were denatured, mixed at a stoichiometry of 1:1 (100 nM total PepQ monomer), and then refolded with GroEL, GroES and ATP for 20 min. The sample was diluted 1,000-fold and examined for fluorescence bursts. Examples of the photon history from each detection channel are shown in the inset. Fluorescence burst co-incidence was examined by cross-correlation analysis of the experimental burst data (black). The cross-correlation of numerically generated burst data with known levels of co-incidence (10%, 5%, 1% and no overlap) are also shown. (b) Two-colour co-incidence analysis of spontaneous PepQ folding. Samples of PepQ-24TMR and PepQ-24OG were denatured, mixed at a stoichiometry of 1:1 and directly diluted in buffer to a final monomer concentration of 2 nM. The protein was allowed to fold spontaneously at 23°C for 10 min. The sample was then diluted 20-fold and examined for fluorescence bursts. Cross-correlation analysis of the experimental data set (black), in comparison with numerically generated burst data at known co-incidence levels, are shown. The inset illustrates the P value distribution for fitting of the experimental data to numerically generated data sets of known co-incidence, yielding a maximum co-incidence likelihood of <1%.

The native PepQ dimer does not readily form at a monomer concentration of 2 nM. However, it is possible that low-order, non-native aggregates stabilized by much larger contact surfaces might still form (69). We therefore anticipated that any co-incidence observed between the two labelled PepQ monomers would have to result from such low-order aggregates. Importantly, the observed co-incidence was less than 1% (Figure A.5b). Even taking into account the reduced sensitivity caused by the differences in triplet state conversion of the OG and TMR dyes, these measurements indicate that, at most, 4–5% of the PepQ monomers could be found in an assembled state of any kind, including the smallest possible aggregates (non-native dimers), during spontaneous

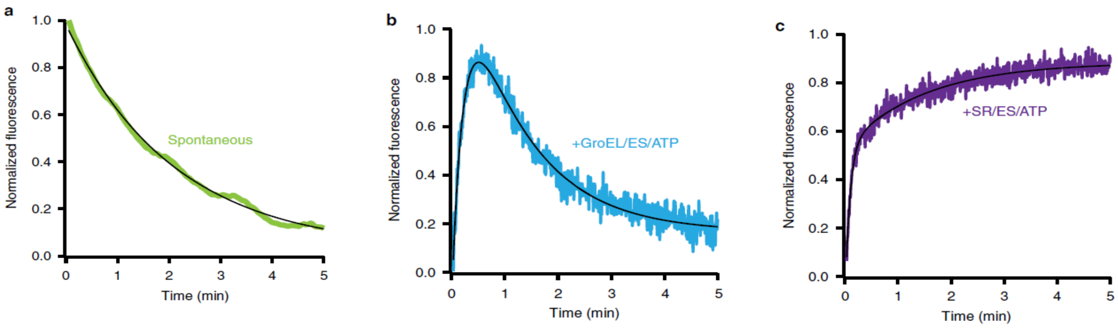


Figure A.6 GroEL alters the folding trajectory of the PepQ monomer. Folding of PepQ was monitored by changes in intrinsic tryptophan fluorescence during (a) spontaneous folding, (b) folding by the fully cycling GroEL-GroES system (c) folding after a single round of encapsulation within the SR1-GroES complex. For spontaneous folding, wild-type PepQ was first denatured in acid-urea then diluted directly into buffer (100 nM). For GroEL-GroES folding, acid-urea denatured PepQ (100 nM) was bound to wild-type GroEL (200 nM) and refolded in the presence of GroES (400 nM) and ATP (2 mM). For SR1-GroES folding, acid-urea denatured PepQ (100 nM) was bound to SR1 (300 nM) and refolded in the presence of GroES (600 nM) and ATP (2 mM). In all cases, the traces shown represent the average of 10 independent experimental replicates. All traces were fit (solid lines) to either a single-exponential rate law (spontaneous) or a sum of exponentials (GroEL-GroES and SR1-GroES). The observed rate constants were ~ 0.477±0.003 min^{-1} for spontaneous folding, 4.63±0.05 min^{-1} and ~ 0.826±0.007 min^{-1} for GroEL-GroES folding and 11.5±0.2 min^{-1} and 0.669±0.010 min^{-1} for SR1-GroES folding.

folding at 2 nM. In total, these observations demonstrate that slow spontaneous folding of PepQ cannot be due to inhibitory aggregation, but instead must result from the inherently inefficient conformational search of the PepQ monomer. Our data suggest that the PepQ monomers that do not reach the native state during spontaneous folding at low protein concentrations likely persist as kinetically trapped monomers.

GroEL alters the folding trajectory of the PepQ monomer

To achieve the large folding stimulation observed with PepQ, in the absence of aggregation, GroEL must actively alter how the protein folds. To investigate the nature of this alteration, we exploited the intrinsic tryptophan fluorescence of PepQ. Importantly, PepQ has multiple tryptophan residues, while GroEL and GroES are devoid of this amino acid. During spontaneous folding, the tryptophan fluorescence of PepQ displays a single, downward transition with a time constant of ~125 sec (Figure A.6a). The rate of this fluorescence decrease is substantially faster than the limiting rate at which PepQ spontaneously commits to the native state (Figure A.1b). This suggest that, at least for spontaneous folding, the observed shifts in tryptophan fluorescence report on transitions that precede the committed step of PepQ folding. By contrast, assisted folding of PepQ with the cycling GroEL-GroES system results in a rapid, early increase in tryptophan fluorescence (t= ~13 sec), which is followed by a subsequent decrease in fluorescence with a time constant of ~73 sec (Figure IV.6b). The large increase in fluorescence observed with GroEL most likely reports on an early folding transition that occurs after the PepQ folding intermediate has been released into the GroEL-GroES cavity. It is unlikely that the early fluorescence rise is due to either GroES binding and encapsulation alone, or to simple release of the PepQ monomer into the cavity, as these events occur much faster than the observed rate of the PepQ fluorescence change (73, 75,

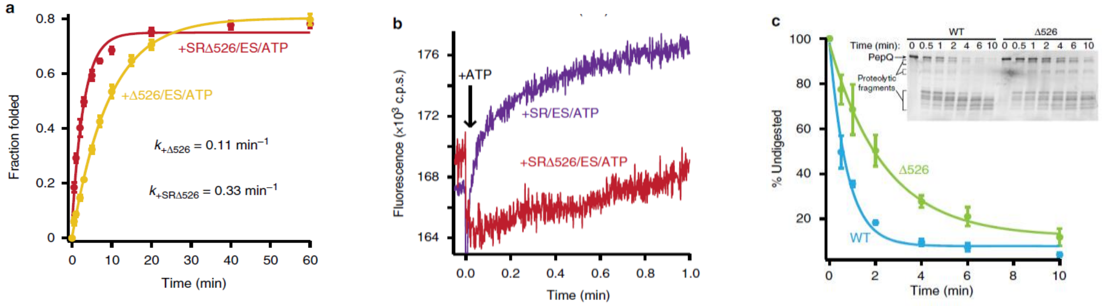


Figure A.7 The GroEL C-termini alter the conformation and folding of the PepQ monomer.

(a) Acid-urea denatured PepQ was bound to a C-terminal truncation mutant of tetradecamer GroEL, D526 (200 nM, red) or the single-ring truncation mutant, SRD526 (300 nM, orange) and refolded in the presence of GroES (400 and 600 nM, respectively) and ATP (2 mM). In each case, the observed regain in enzymatic activity was fit to a single-exponential rate law (solid lines), resulting in observed rate constants of 0.106±0.003 min^{-1} for D526-mediated folding and 0.332±0.038 min^{-1} for SRD526-mediated folding. (b) Intra-cavity folding of PepQ at early times was monitored by changes in tryptophan fluorescence following addition of GroES and ATP to complexes of non-native PepQ bound to SR1 (blue) or SRD526 (green). Acid-urea denatured PepQ (100 nM) was first bound to either SR1 or SRD526 (300 nM in both cases), and then rapidly mixed with an equal volume of GroES (600 nM) and ATP (2 mM) in a stopped-flow apparatus. The traces shown represent the average of 20 experimental replicates. (c) Residual structure in a GroEL-bound PepQ folding intermediate was examined by protease susceptibility. PepQ- 24F (100 nM) was denatured in acid-urea and bound to the trans ring of either wild-type GroEL-GroES or D526-GroES ADP complexes (ref. 23; 120 nM) and then treated with chymotrypsin for the indicated times before quenching with phenylmethylsulfonyl fluoride (PMSF) (1 mM). Samples were analysed by SDS–PAGE and laser-excited fluorescence gel scanning (inset). The migration position of full-length PepQ, as well as the position of three dominant groups of proteolytic fragments, are indicated. The amount of full-length PepQ was quantified by densitometry. The data were fit to a single-exponential rate law, with a half-time for the digestion of PepQ bound to the open ring of a wild type GroEL-GroES-ADP complex of 0.53±0.06 min (EL, blue) and 1.66±0.17 min for the D526-GroES ADP complex (D526, green). In all cases, error bars show the s.d. of three experimental replicates.

78). Although PepQ folding with the cycling GroEL-GroES system rapidly becomes asynchronous, the transition between the increasing and decreasing fluorescence phases occurs after roughly one cavity lifetime at 23 °C (75, 154, 179). This observation supports the idea that the increase in fluorescence occurs inside the GroEL-ES cavity. To directly test this conclusion, we employed SR1 to examine a single round of PepQ encapsulation and folding inside the GroEL-GroES cavity. Notably, PepQ confined within the SR1-GroES cavity also displays a rapid increase in fluorescence, but no subsequent decrease (Figure A.64c), confirming the conclusion that the early increase in PepQ fluorescence occurs within the GroEL-GroES cavity.

To further define how the folding behavior of PepQ is altered by GroEL, we examined the impact of the GroEL C-termini on PepQ folding. We previously showed that a tailless GroEL variant (D526 GroEL) has a significantly reduced ability to assist the folding of the classically stringent GroEL-substrate protein, RuBisCO from *R. rubrum* (81). Interestingly, removal of the C-termini has an even more pronounced negative impact on PepQ folding. Deletion of the C-termini from a cycling GroEL tetradecamer (D526) causes a nearly sixfold reduction in the observed PepQ folding rate (Fig. 5a), versus an approximate twofold reduction with RuBisCO (81). By contrast, removal of the C-termini from the non-cycling SR1 GroEL variant (SRD526) results in a more modest twofold decrease in the PepQ folding rate (Figure A.7a). Because deletion of the C-termini can, in some cases, result in premature substrate protein release before GroES binding and encapsulation (180), we considered whether the observed drop in

PepQ folding rate with the D526 variants is simply due to a trivial decrease in encapsulation efficiency. However, D526 displays no substantial premature release of PepQ relative to full-length GroEL and the early escape of PepQ from SRD526, compared to SR1, is no greater than 10% (Figure A.8). While consistent with previous observations with RuBisCO (180), this minor drop in encapsulation efficiency is too small to explain the reduction in observed folding rate.

The stimulation of PepQ folding by the GroEL C-termini could, in principle, result from: (1) enhanced unfolding of PepQ by the C-termini (78, 81, 154, 155); (2) stimulation of productive folding transitions, or blockage of inhibitory ones, by the tails

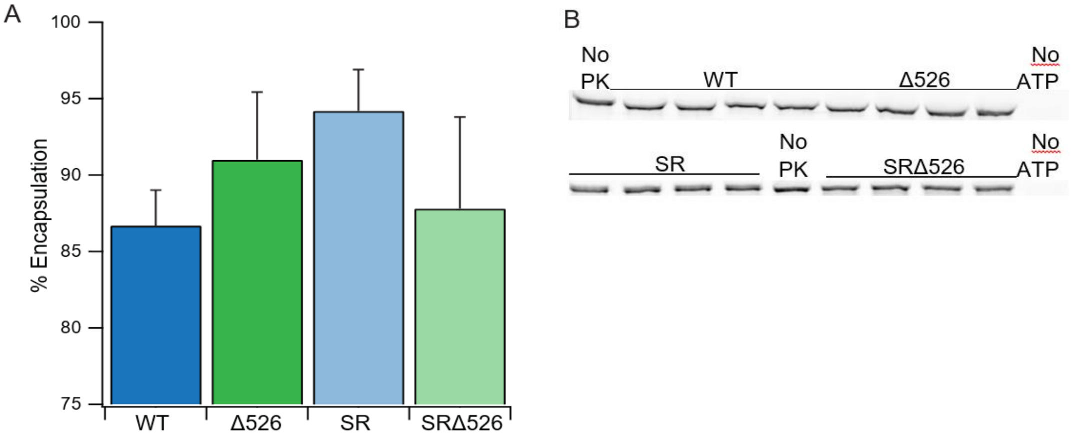


Figure A.8 Reduced folding rate in GroEL truncation mutants is not caused by a severe encapsulation deficiency.
(A) PepQ-24F (100nM) was bound to wild-type or Δ526 GroEL-ES-ADP bullets (200nM) or full-length single ring GroEL (SR) or Δ526 single ring GroEL (SRΔ526, 300nM) and supplemented with GroES (600nM). A single turnover was initiated by the addition of ATP (2mM) followed by quenching with hexokinase and glucose after 10 seconds. Un-encapsulated PepQ was digested with Proteinase K (0.5 g/mL). Samples were run on SDS-PAGE and analyzed by fluorescent gel scanning, seen in (B). Samples were normalized to undigested controls. Error bars show the deviation of n = 4 independent samples.

during intra-cavity folding (165, 181); or (3) a combination of both unfolding and confinement effects. Importantly, these models all predict that the folding trajectory of a PepQ monomer inside the GroEL-GroES cavity should change upon C-terminal tail removal. To test this prediction, we exploited the tryptophan fluorescence properties of PepQ to examine a single round of encapsulation inside both the full-length SR1-GroES cavity and the truncated SRD526-GroES cavity. Strikingly, the rapid and early rise in tryptophan fluorescence that is observed when PepQ folds inside the SR1-GroES cavity, completely disappears when PepQ is encapsulated inside a SRD526-GroES cavity Figure A.7b). These observations are highly consistent with the idea that GroEL

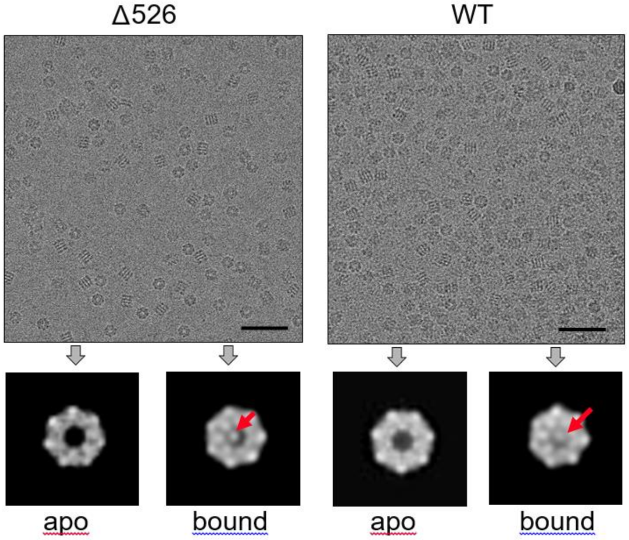


Figure A.9 Reference-free 2D class-averages of Δ526 GroEL and wild type GroEL tetradecamers incubated with nonnative PepQ.
Panel A and B show representative, raw data micrographs for Δ526 GroEL (*Δ526*) and wild type GroEL (*WT*) incubated with non-native PepQ. Panels C, D, E, F show the reference-free 2D classifications of tetradecamers that either did (*bound*), or did not (*apo*), bind PepQ. The black scale bars denote 100nm. The apo states of both wild type and Δ526 tetradecamers display perfect seven fold symmetry, with an empty central cavity. By contrast, the PepQ-bound states of both tetradecamers show a very clear extra density inside the central cavity, highlighted by the red arrows.

promotes conformations of the PepQ monomer that are not, or at least not well, populated during spontaneous folding in free solution and that the C-terminal tails are at least partially involved in this process.

Cryo-EM observation of PepQ unfolding by the GroEL C-termini

We previously demonstrated that GroEL helps stimulate productive folding of a kinetically trapped RuBisCO monomer through partial unfolding (78, 81, 83, 154). In addition, we showed that maximal RuBisCO unfolding requires the GroEL C-terminal tails (180). If structural disruption of the misfolded substrate proteins is a general feature of GroEL-stimulated folding, then GroEL could also be expected to unfold the kinetically trapped PepQ monomer. To test this proposition, we first examined the protease susceptibility of a PepQ folding intermediate bound to both wild-type GroEL and D526. Chemically denatured PepQ was first bound to the open, *trans* ring of an

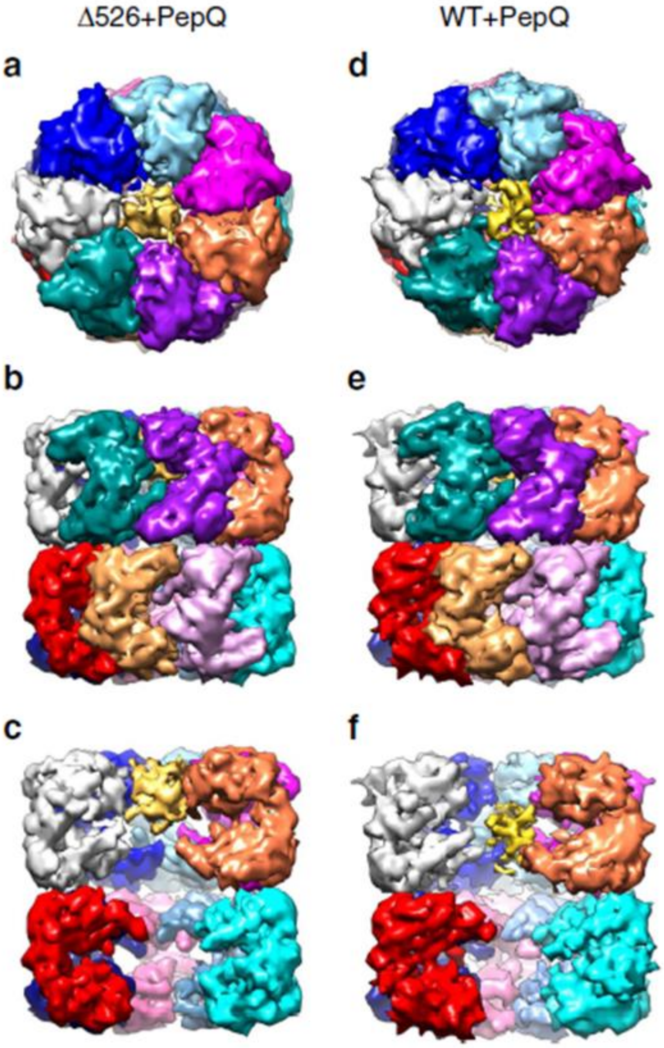


Figure A.10 The impact of the GroEL C-termini on a bound PepQ monomer. The conformation of a non-native PepQ monomer bound to either a wild-type GroEL tetradecamer (WT) or to the C-terminal truncation variant (D526) was examined by cryo-EM. The map of the D526-PepQ complex is shown from the top (a), side (b) and cutaway (c) views. The map of the wild-type GroEL-PepQ complex is also shown from the top (d), side (e) and cutaway (f) views. Each subunit of GroEL is coloured differently while the PepQ monomer is coloured yellow. In the D526 GroEL, the density attributed to the PepQ monomer is observed near the top of the GroEL cavity in one ring (c). However, in the wild-type GroEL complex, the PepQ density is observed near the centre of the cavity (f) in one ring.

asymmetric GroEL-GroES complex created with either wild-type GroEL or D526, then treated with limiting amounts of chymotrypsin (83). Consistent with our previous

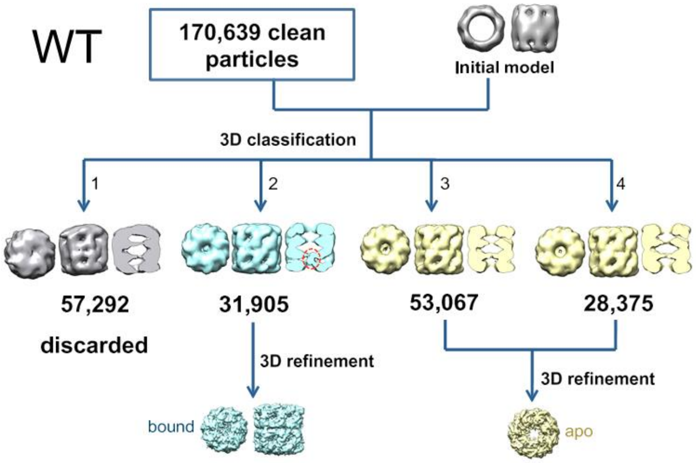


Figure A.11 Image processing for the wild type GroEL complexes.
Two runs of reference free 2D classification resulted in 170,639 clean particles. The apo state of *E.coli* GroEL was used as an initial model to generate four 3D classes with C1 symmetry. The number of particles in each 3D class is shown. In Class 2, a thick extra density was readily visible in the central cavity of a well-shaped tetradecamer, which is highlighted by the dashed red circle. Particles of this class were used to generate the refined density map of the PepQ-bound wild type GroEL complex at 8.3 Å resolution. Class 3 and 4 possess particle shapes consistent with an empty GroEL tetradecamer and were used to generate the refined density map of the apo Δ526 GroEL at 7.9 Å resolution. Because of its anomalous particle shape, Class 1 was discarded. The initial refinement model employed was derived from the crystal structure of *E. coli* GroEL (PDB ID: 4HEL), blurred to a resolution of 40 Å.

RuBisCO observations, PepQ bound to a full-length GroEL ring was degraded ~2.5-fold faster than PepQ bound to the D526 ring (Figure A.7c).

To develop a more detailed picture of the interaction between PepQ and GroEL, we employed cryo-EM to examine the structures of both wild type GroEL and D526 tetradecamers bound to non-native PepQ. Chemically denatured PepQ was first mixed with unliganded (apo) GroEL or D526 tetradecamers, then vitrified in thin ice and imaged with single-particle cryo-EM. Reference-free two-dimensional (2D) image

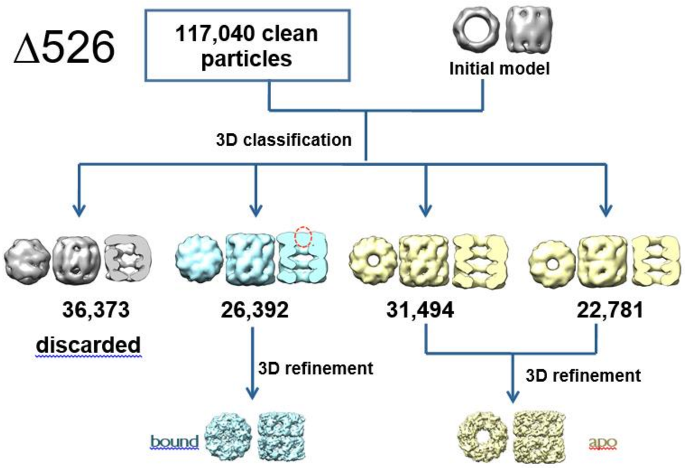


Figure A.12 Image processing for the Δ526 complexes.
Two runs of reference free 2D classification resulted in 117,040 clean particles. The apo state of *E.coli* GroEL was used as an initial model to generate four 3D classes with C1 symmetry. The number of particles in each 3D class is shown. In Class 2, a thick extra density was readily visible in the central cavity of a well-shaped tetradecamer, which is highlighted by the dashed red circle. Particles of this class were used to generate the refined density map of the PepQ-bound Δ526 GroEL complex at 8.3 Å resolution. Class 3 and 4 possess particle shapes consistent with an empty GroEL tetradecamer and were used to generate the refined density map of the apo Δ526 GroEL at 7.9 Å resolution. Because of its anomalous particle shape, Class 1 was discarded. The initial refinement model employed was derived from the crystal structure of *E. coli* GroEL (PDB ID: 4HEL), blurred to a resolution of 40 Å.

classification revealed a robust population of GroEL tetradecamer complexes with substantial density visible in the central cavity of a major 2D class-average for both wild-type GroEL and D526 (Figure A.9). The observed central density is highly consistent with the expected binding position of the non-native PepQ monomer. Further, three-dimensional (3D) classification and map refinement, without any applied symmetry, revealed both apo and PepQ-bound states of the tetradecamer for both wild-type GroEL and D526. (Figure A.10 and Figure A.11 and A.12). On the basis of the gold-standard Fourier shell correlation, the overall resolution for the apo states of both wild-type

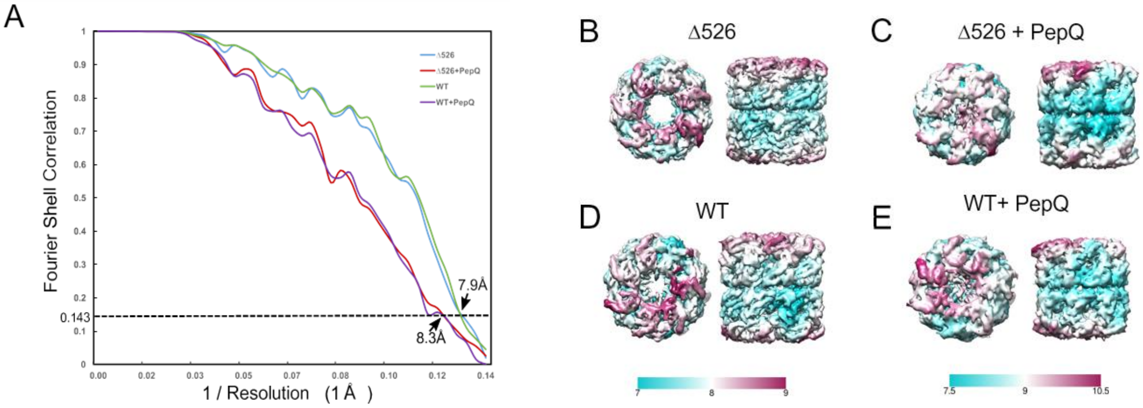


Figure A.13 Overall and local resolutions of GroEL complexes.
Panel A shows the final resolution of the refined maps for all complexes determined by the gold- standard criteria and measured at 0.143 FSC. The curves for both Δ526 GroEL apo (*blue*) and PepQ-bound (*red*), as well as the wild type GroEL apo (*green*) and PepQ- bound (*purple*), are shown. Both apo state structures have a resolution of 7.9 Å resolution, while both PepQ-bound complexes have a resolution of 8.3 Å. Panels B and D show the local resolution of the Δ526 GroEL and wild type GroEL apo structures. Panels C and E show the local resolution for the PepQ-bond Δ526 GroEL and wild type GroEL structures. In each case, the colors shown indicate the local resolution in each structure, with blue being higher resolution and pink being lower. The color map corresponding to the resolution range for each structure is shown at the bottom (in Å). In all cases, the GroEL apical domains display a lower estimated resolution (~ 9 Å), compared to the equatorial domains (~ 7 Å). The PepQ density inside chamber displayed the lowest local resolution in both maps, most likely due to the conformational heterogeneity of the non-native PepQ monomer.

GroEL and D526 was 7.9 Å, while the overall resolution of the pepQ-bound states for both wild-type GroEL and D526 was 8.3 Å (Figure A.13A). Importantly, the resolution obtained was not uniformly distributed over the entire structure of either complex, but was significantly higher in the equatorial and intermediate domains, and lower in the apical domains (Figure A.13B–E). The lowest local resolution was observed for density associated with the PepQ monomer in the central cavity, as expected for a non-native protein folding intermediate, which likely populates a mixture of conformations.

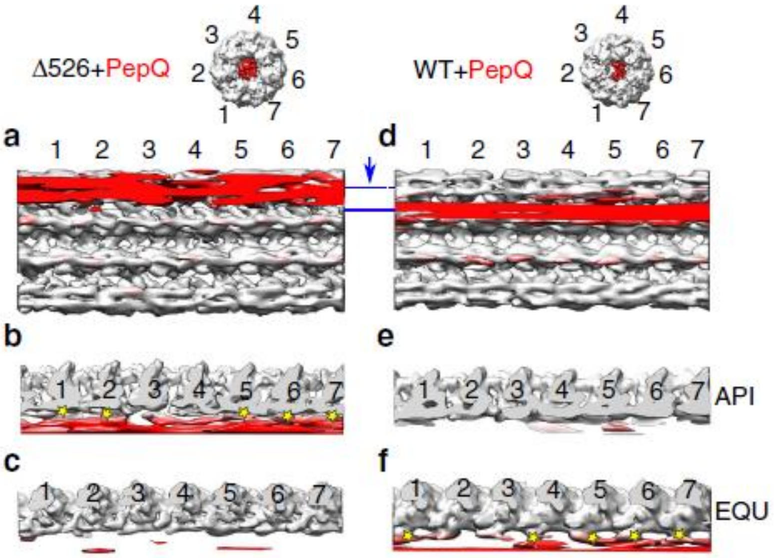


Figure A.14 The PepQ monomer interacts with multiple GroEL subunits. GroEL tetradecamers with bound PepQ are shown in unwrapped, planar displays as viewed from inside the ring for the D526 (a) and wild-type (d) GroEL complexes. A top view slice of the planar map, through the apical (API) domains (dashed blue line), is shown in b for D526 and in e for wild-type GroEL. Interactions between the non-native PepQ monomer and the D526 GroEL apical domains of subunits 1, 2, 5, 6 and 7 are highlighted (b; yellow stars). A top view slice of the planar map, through the equatorial (EQU) domains (solid blue line), is shown in c for D526 and in f for wild-type GroEL. Contacts between the non-native PepQ and the C-termini of subunits 1, 4, 5, 6 and 7 of the wild-type GroEL are highlighted (f; yellow stars). The isosurface threshold for b and c is 1.74s and is 1.65s for e and f.

The cryo-EM structures reveal the striking impact of the GroEL C-termini on both the conformation and the binding position of the PepQ folding intermediate. In the absence of the C-terminal tails, the PepQ monomer appears as a strong extra density associated with the upper, inner surface of the apical domains of the GroEL cavity (Figure A.10c and 15d). By contrast, on a wild-type GroEL ring, the PepQ folding intermediate shifts to a much lower average position in the cavity, moving towards the base of the cavity and in the direction of the C-termini. At the same time, the density of the PepQ intermediate decreases significantly (Figure A.10f and 15c), which indicates a

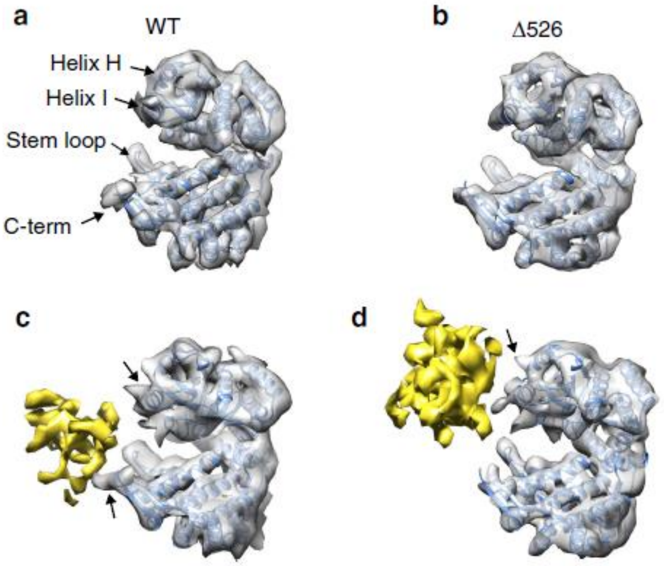


Figure A.15 The GroEL C-termini helps retain and unfold the PepQ monomer. (a,b) A single subunit of the apo GroEL atomic model (PDB ID: 4HEL) fit into the cryo-EM densities of unoccupied wild-type GroEL and D526 GroEL tetradecamers. The positions of the H and I helices of the GroEL apical domain, as well as the equatorial stem loop (D41-P47) and the C-terminus are labelled on the wild type GroEL structure. (c,d) Single subunit of the apo GroEL atomic model fit into the cryo-EM densities of PepQ-bound wild-type GroEL and D526 GroEL tetradecamers. The density from the non-native PepQ monomer is coloured yellow. Black arrows (c,d) indicate the interactions between a GroEL subunit and the PepQ monomer. When the density volumes of the D526 and wild-type GroEL tetradecamers are matched (B61,000 Å3 for a single GroEL subunit in both cases), the observed density volumes for the PepQ monomer are 8,564 Å3 in the D526 and 2,696 Å3 in the wild type GroEL complex, consistent with the PepQ monomer being more unfolded when bound to the wild type GroEL ring.

more unfolded and heterogenous conformational ensemble of the PepQ, leading to its weaker density in the cryo-EM map. The non-native PepQ monomer can also be seen to make contact with multiple GroEL subunits on both a wild-type GroEL and D526 ring (Figure A.14). However, the location of the contacts between the PepQ folding intermediate and the GroEL subunits changes dramatically when the C-terminal tails are removed. In the tailless D526 ring, the PepQ monomer appears to make exclusive contact with the central face of the apical domains, in the region of helices H and I (Figure A.14a–c and 15d). By contrast, on a wild-type GroEL ring, the PepQ monomer shifts to a significantly lower position at the base of the apical domain, and is accompanied by a set of new, strong contacts that localize in the region of the GroEL C-terminal tails (Figure A.14d–f and 15c). To confirm that the density observed at the base

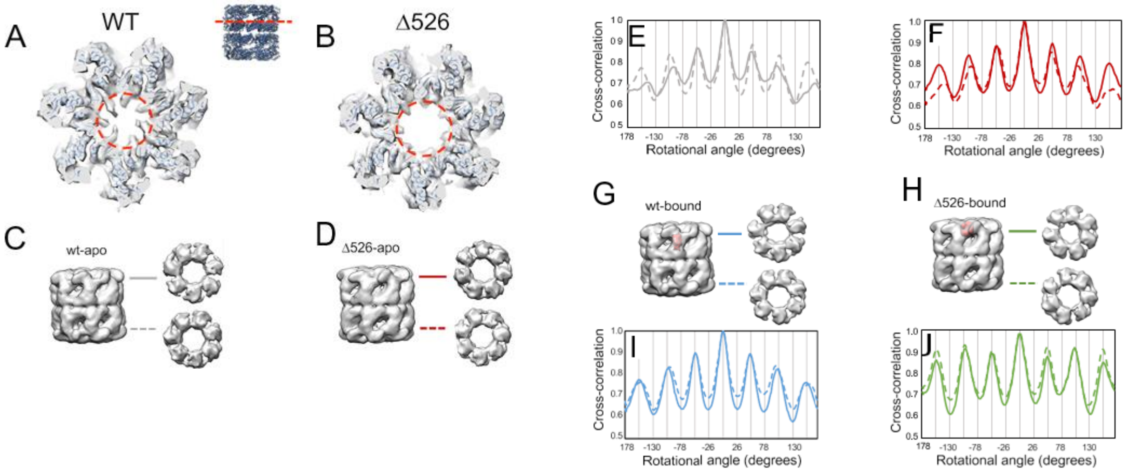


Figure A.16 Difference in C-termini densities between wild type and Δ526 GroEL, and the change of apical domain symmetry upon PepQ binding. The GroEL C-termini are partially resolved in the empty ring of the wild type GroEL complex,but are missing in Δ526 rings. The atomic coordinates of GroEL (PDB ID: 4HEL) were fit into the density maps of wild type GroEL and Δ526. An end-on, top view of an empty wild type GroEL ring (A) and Δ526 (B) ring are shown, at the slice position indicated by the dashed red line in the inset side view. Notably, the GroEL C-termini cannot be seen in GroEL crystal structures, most likely due to their high flexibility and conformational heterogeneity. However, a substantial fraction of the C-termini are visible in the wild type GroEL structure presented here, indicated by the significant extra density visible within the dashed red circle in panel A. The observation of the C-termini in this structure is likely due to the use of C1 symmetry during model refinement. Importantly, and as expected, the same density is missing in the Δ526 ring. Rotational cross-correlation of the apical domains from both rings of wild type apo GroEL (C and E), apo Δ526 (D and F), PepQ-bound wild type GroEL (G and I) and PepQ-bound Δ526 (H and J), respectively. Solid and dashed lines indicate the two different rings of the same tetradecamer complex. The apical domains within a single ring do not show perfect 7- fold symmetry, as indicated by the differences in the height of the peaks in the rotational cross-correlation curves (E,F,I and J). However, PepQ binding induces a much larger.

of the wild-type GroEL cavity does, in fact, originate from the C-termini, we examined this region in the empty wild-type GroEL and D526 tetradecamers. As expected, the

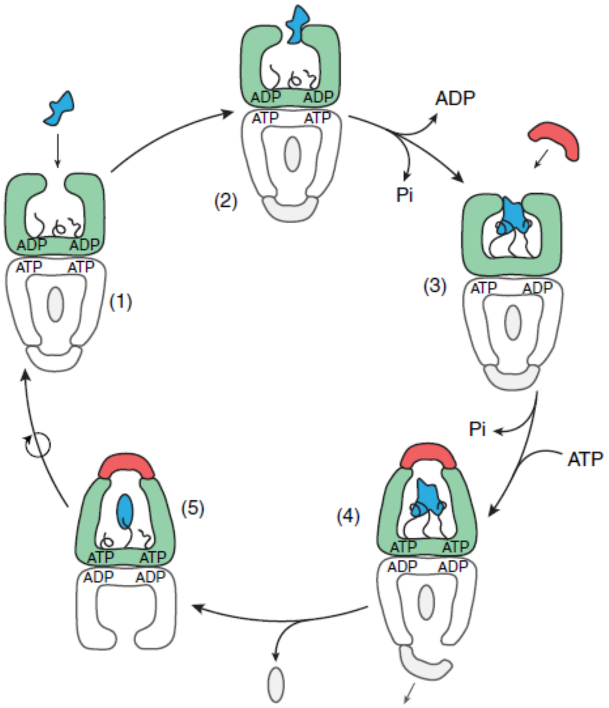


Figure A.17 Schematic of the GroEL-GroES reaction cycle.
(1) A non-native substrate protein (irregular blue shape) enters the GroEL reaction cycle on the open trans ring (green) of the ATP bullet complex22. (2) Substrate protein binding accelerates both the release of ADP from the trans ring and ATP hydrolysis in the opposite, cis ring (grey; refs 50,51). (3) Binding of the non-native substrate protein by the C-terminal tails (black), helps retain the substrate protein deep within the GroEL cavity and, in combination with additional binding by multiple apical domains, results in substrate protein unfolding (results here and refs 19,22,24,39,40,55). (4) Assembly of the new folding cavity on the trans ring causes both forced unfolding and compaction of the substrate protein, and is directly coupled to the disassembly of the folding cavity on the opposite ring, potentially through a transient, symmetric intermediate (22,23,37,77–79). (5) A subsequent allosteric shift of the GroEL-GroES complex results in full ejection of the substrate protein into the enclosed GroEL-GroES cavity and the initiation of folding before ATP hydrolysis50. Because ATP hydrolysis is the rate limiting step of the reaction cycle, increased binding of substrate proteins to the open trans ring (2) results in more rapid cycling of the GroEL-GroES system and a shorter folding cavity lifetime (22,50,51,80).

density observed at the base of the wild-type GroEL ring, projecting from the precise position expected for the C-termini, is missing in the tailless D526 ring (Figure A.16A,B). In total, these result strongly support the idea that the non-native PepQ monomer is significantly more unfolded when bound to a GroEL ring with intact C-terminal tails and this unfolding has a direct functional impact on the efficiency of productive folding.

Discussion

Fundamentally, chaperonins like GroELS function as kinetic editors of protein folding reactions, altering how folding intermediates partition between available conformational states. A key question, however, is whether chaperonins achieve this editing action by actively altering the conformational space available to their substrate proteins, or by exclusively working as passive aggregation inhibitors. We examined this issue from a new angle by characterizing the folding of the *E. coli* metalloprotease PepQ, a stringent, *in vivo* GroEL-substrate protein. We found that slow spontaneous folding of PepQ is not caused by inhibitory aggregation. The capture of this kinetically trapped PepQ folding intermediate by a GroEL ring results in conformational perturbations that are consistent with unfolding. In addition, the intrinsically unstructured C-terminal tails of the GroEL subunits play a central role in this process (Figure A.17).

Determining the function of the flexible C-terminal tails in chaperonin-assisted protein folding has been challenging. Early studies showed that the tails play no role in tetradecamer assembly or stability (182). At the same time, removal of the C-termini was found to have negligible impacts on *E. coli* growth under standard laboratory conditions, leading to the suggestion that the tails do not play any important role in assisted folding (182, 183). Other studies, however, demonstrated that removal of, or large alterations to, the C-termini can have serious negative consequences in *in vitro* protein folding assays (81, 165, 181, 182, 184). In addition, *E. coli* strains possessing C-terminally truncated GroEL genes display substantially compromised fitness in competition with wild-type strains (182). These observations, in combination with the extensive, although not quite universal, conservation of the chaperonin C-terminal tails over much of phylogeny (185, 186) suggest that the C-termini do play an important role in assisted protein folding. Our prior work with RuBisCO supported this conclusion, implicating the C-termini in substrate protein capture, retention and unfolding during GroES binding (81, 180). The observations we present here with PepQ strengthen and extend these conclusions, showing that the unstructured C-termini make physical contact with a non-native substrate protein before ATP or GroES binding. In addition, we have visualized the consequences of this interaction, demonstrating simultaneous engagement of a folding intermediate by both the inner apical face and the unstructured tails of multiple GroEL subunits. This multi-level binding mode both retains the folding intermediate deeper inside the GroEL cavity and assists in partial unfolding of the misfolded PepQ monomer.

Interestingly, we observe a single, well populated class of the PepQ folding intermediate bound to a GroEL ring, both in the presence and absence of the GroEL C-termini. This contrasts with a previous cryo-EM study conducted with the smaller substrate protein malate dehydrogenase (MDH), in which asymmetric model refinement suggested multiple potential binding modes of the MDH folding intermediate. While the C-termini were not resolved in this prior study, and the resolution of these MDH structures is several angstroms lower than the PepQ structures we report here, two sub-populations of the MDH folding intermediate appear to be bound in a deep internal position within the GroEL cavity, consistent with the binding position we observe with PepQ. A third sub-population of the bound MDH monomer appeared to be bound in a more elevated position near the upper, exterior surface of the GroEL apical domains. At the same time, the MDH folding intermediate displayed a substantially smaller contact surface with the GroEL ring in comparison to what we observe with PepQ (Figure A.14) This observation is consistent with the difference in relative mass of MDH (33 kDa) compared to PepQ (52 kDa) and suggests that the smaller MDH protein could be bound more weakly or sample a larger potential range of bound states.

In principle, the cryo-EM structure of the PepQ-GroEL complex could also reveal conformational changes of the GroEL tetradecamer that are coupled to substrate protein capture. Overall, the conformations of the GroEL tetradecamer in the presence and absence of the PepQ folding intermediate are similar. Notable breaks in the rotational symmetry of the PepQ-occupied GroEL ring, both in the presence and absence of the C-termini, are apparent (Figure A.14). However, significant deviations from ideal rotational symmetry are also observed in the unoccupied apo GroEL and D526 rings (Figure A.16C–J). Strikingly, PepQ binding induces a dramatic increase in the rotational symmetry of the D526 apical domains, in both the bound and second, unoccupied rings (Figure A.16H,J). The structural changes seen in the second, unoccupied ring are most likely a consequence of allosteric coupling between the GroEL rings. The coordinated binding and release of nucleotides, GroES and substrate proteins are well established and essential features of the functional GroEL reaction cycle (187). Many of the structural details of this allosteric coupling remain poorly understood, however. In particular, it remains unclear how substrate protein binding forces ADP out of one GroEL ring while simultaneously accelerating the release of GroES from the opposite ring(75, 137, 188). Previous work suggested that this allosteric response may involve counter-clockwise movements of the GroEL apical domains, in both the substrate occupied and second, empty ring (189). Our structural observations with PepQ suggest that a shift in the rotational symmetry of the GroEL apical domains likely also plays a role. In addition, the C-termini appear to be intimately involved in modulating this structural shift. This observation is consistent with our prior observations that removal of the C-termini attenuates negative cooperativity in ATP binding between the two GroEL rings (81).

Overall, our observations with PepQ are not consistent with an exclusively passive, aggregation-blocking role for roEL in stimulated protein folding (80, 190). These observations are, however, fully consistent with our previous demonstration that GroEL plays an active role in the assisted folding of *R. rubrum* RuBisCO (78, 81, 83, 154). They are also consistent with observations from other groups on other substrate proteins, including another endogenous *E. coli* enzyme DapA (153, 155, 156, 191, 192). Our observation with PepQ also suggest that active participation by GroELS in stimulated protein folding is likely to be a general mechanistic feature of these chaperonin machines. DapA, like RuBisCO, is a member of the TIM-barrel family of proteins, a canonical a/b-fold that is highly represented in the subset of *E. coli* proteins that depend on GroEL for folding (161, 162). By contrast, PepQ is a member the so-called pita-bread proteins (158-160), a protein fold that is fundamentally distinct from the TIM-barrel fold (132). To date, no pita-bread fold has been examined in detail as a GroEL-substrate protein. The addition of PepQ to the list of *E. coli* proteins that derive a large, active folding enhancement from GroEL strengthens the argument that similar mechanisms are likely to stimulate the folding of many stringent substrate proteins.

Interestingly, PepQ appears to have no ready access to fast and productive folding pathways in free solution. At the same time, persistent misfolding produces PepQ monomers that, although they do not aggregate, cannot reach the native state even with assistance from GroEL. This suggests that the conformational search of the non-native PepQ monomer, at least in free solution, is dominated by deep and inhibitory kinetic wells that GroEL helps the protein to avoid. Whether the iterative annealing or confinement-based models most accurately describe this active folding mechanism of GroEL remains controversial (77). Importantly, these mechanisms make distinct predictions about what should happen to PepQ folding when the GroEL cycling rate is altered. If unfolding of kinetically trapped intermediates is important for stimulated folding of PepQ, it should be possible for the cycling GroEL-GroES system to achieve a stimulated folding rate that exceeds the limiting, intra-cavity folding rate observed with SR1-GroES. By contrast, if confinement is most important for PepQ folding, then the non-cycling SR1-GroES system should display the maximum possible enhanced folding rate, a rate that the cycling system could approach but never exceed (154).

To test these predictions, we examined the folding rate of PepQ under conditions where the GroEL-GroES cycling rate was systematically increased. Modulation of the GroEL ATPase rate was accomplished by addition of bovine serum albumin (BSA), which interacts only weakly with GroEL. Because progression of the GroEL ATPase cycle is linked to ADP release, which is in turn coupled to binding of proteins to the open, post-hydrolysis *trans* GroEL ring (Figure A.17) and (75, 137, 188), BSA can be used to accelerate the GroEL-GroES ATPase cycle (Figure A.18a). However, because the interaction between BSA and GroEL is weak, BSA only poorly competes with PepQ for binding to GroEL. At concentrations up to 0.1 mg ml^{-1}, BSA has a small, negative impact on the observed PepQ folding rate observed with cycling GroEL-GroES (Figure A.18b). Strikingly, however, while the addition of BSA has no impact on either

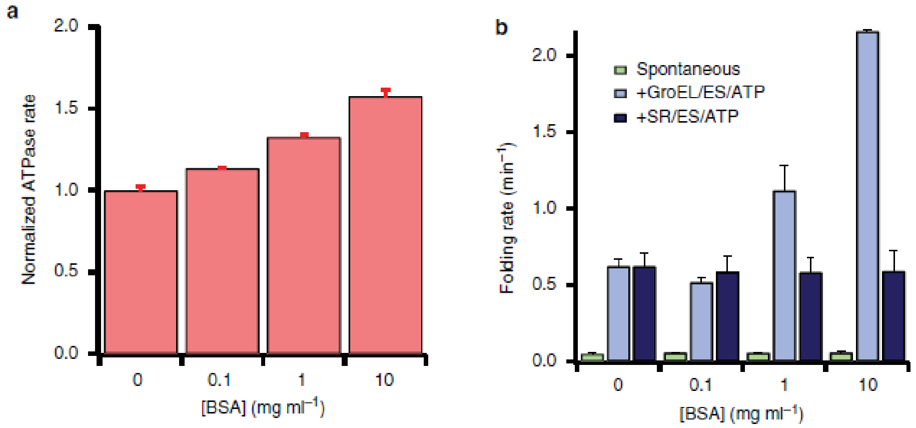


Figure A.18 Cycling GroEL-GroES can fold PepQ faster than confinement alone.
(a) The rate of ATP hydrolysis by GroEL-GroES is stimulated in the presence of BSA. The steady-state rate of ATP hydrolysis by GroEL (200 nM) in the presence of GroES (400 nM) and ATP (2 mM) was measured with varying concentrations of [BSA]. Error bars show the standard deviation of three independent experiments. (b) Addition of BSA to a cycling GroEL-GroES system substantially accelerates the rate of assisted PepQ folding. The rate of spontaneous PepQ folding (green), intra-cavity folding with SR1-GroES (dark blue), and folding with fully the cycling wild type GroEL-GroES system (light blue) was examined in the presence of different concentrations of BSA. Experimental conditions were the same as Fig. 1, with the exception that native BSA was present in the buffer. Error bars show the s.d. of three experimental replicates.

spontaneous PepQ folding or SR1-GroES mediated folding, higher BSA concentrations substantially enhance the PepQ folding rate achieved with cycling GroEL-GroES (Figure A.18b). Importantly, the magnitude of this effect increases as the concentration of BSA increases, mirroring the impact of BSA on the steady-state rate of ATP turnover by GroEL-GroES (Figure A.18b). This response is very similar to our prior observations with RuBisCO (154), where an ~40% increase in the steady-state GroEL-GroES ATPase rate yielded a 2.5–3-fold enhancement of the observed RuBisCO folding rate. These observations suggest that repetitive unfolding by GroEL is required to achieve maximally stimulated folding of many stringent substrate proteins.

When considering the stimulatory impact of partial unfolding, it is important to note that GroEL unfolds substrate protein in two distinct phases. The first is associated with the capture of a folding intermediates by the GroEL ring, where a binding- driven expansion of the substrate protein can, in some cases, result is substantial conformational disruption (this study and (78, 81, 83, 154, 156, 191).This unfolding event occurs both during and immediately after capture of a folding intermediate, but before ATP binding. Most likely, binding-associated unfolding is similar to surface-catalysed denaturation, where the substrate protein becomes splayed across the multiple interaction surfaces of the apical domains as well as, we suggest, the C-termini (78, 81, 83, 154, 191).

GroEL also imposes a second, directed unfolding process that is impelled by ATP (78, 81, 83, 154, 155). When a GroEL ring binds ATP, the apical domains are driven through a large-scale, rigid body rearrangement that both rotates and elevates them (193-195). While these shifts are necessary for GroES binding and substrate encapsulation (72, 145-148), previous observations have also demonstrated that (1) substrate proteins remain associated with the apical domain as they initiate their movement, imposing a substantial load on their motion (196) and (2) apical domain movement can simultaneously impart a rapid, forced unfolding event on the substrate protein (81, 83, 154, 155). While our studies with PepQ were not designed to detect forced unfolding, it is striking that both binding-driven unfolding of PepQ and RuBisCO (this study (81)) and forced unfolding of RuBisCO (81) are attenuated when the C-termini are removed. These observations suggest that the C-termini represent a secondary binding platform at the base of the GroEL cavity that is important both for the initial capture and unfolding of the substrate protein, as well as retention of the folding intermediate within the GroEL cavity during the process of apical domain movement and GroES binding (180). It remains an open question how the C-termini are induced to release the substrate protein upon the initiation of folding. However, both experimental (81) and computational studies (197) indicate that the C-termini are coupled to the GroELS allosteric cycle, suggesting that modulation of the interaction between the C-termini and a folding intermediate might be controlled by the GroEL ATPase cycle in a manner that parallels the behavior of the apical domains.

Fundamentally, the iterative annealing and confinement mechanisms are not mutually exclusive. A combined mechanism, in which kinetically trapped folding intermediates are first partially unfolded, then briefly confined within the privileged environment of the GroEL-GroES cavity where re-population of misfolded conformations is discouraged, might well yield a maximally efficient strategy for accelerating the folding of especially recalcitrant proteins. Several of our observations with PepQ are consistent with such a mechanism. In the presence of either single-ring or double-ring GroEL variants, PepQ displays a sizable fluorescent burst phase that is completely absent during spontaneous folding (Figure IV.6). This observation suggests that the PepQ monomer, while confined within the GroEL-GroES cavity, populates at least one conformational state (or ensemble of states) with ready access to the native state. During spontaneous folding; however, this state is either very rarely populated, or not populated at all. At the same time, removal of the GroEL C-terminal tails slows overall PepQ folding and completely eliminates the fluorescence burst phase (Figure IV.7). This behavior is strikingly similar to the impact of C-terminal tail removal on RuBisCO folding, where the formation of a rapidly folding, burst phase intermediate depends upon both partial unfolding and encapsulation within the GroEL-GroES cavity (81, 83, 180). As with RuBisCO, C-terminal tail removal also has a more profound impact on PepQ folding with the cycling, tailless D526 tetradecamer than it does on the tailless single-ring SRD526 (Figure IV.1 and 7). For both RuBisCO and PepQ, however, long-term confinement within the chaperonin cavity, even when partial unfolding is reduced through C-terminal tail removal (for example, SR1-GroES versus SRD526-GroES) results in substantially enhanced folding in comparison to the free solution folding of both proteins (Figure IV.1 and 7 and (81, 83). In total, these observations are consistent with an active chaperonin mechanism in which partial unfolding and confinement lead to optimal stimulation of folding for highly dependent substrate proteins. It is worth noting that in a living *E. coli* cell, additional chaperone systems (for example, the Hsp70s and Hsp100s) can engage a folding intermediate before its processing by GroELS. Learning how these additional chaperone systems impact the folding of GroELS substrates will be an important next step towards understanding the mechanism of chaperone and chaperonin-mediated folding pathways.

APPENDIX B

GROEL MEDIATED ACCELERATION OF PROTEIN FOLDING BY ENHANCED COLLAPSE*

Summary

The discovery of chaperonins introduced the very idea that one protein may assist another to fold. But how their stereotypical internal chamber could help folding has remained an enigma, owing in part to the challenges of detecting forces exerted on polypeptide substrates. Here, using optical tweezers combined with fluorescence imaging, we show that the chaperonin GroEL induces attractive forces that drive the compaction and folding of polypeptide chains. This enhanced chain collapse does not strictly require, but is aided by GroES binding, and is stimulated by nucleotide binding and the GroEL C-terminal tails. Our findings directly demonstrate GroEL-mediated folding acceleration and provide an underlying principle distinct from previous models. We conjecture that limited protein chain collapse is a general folding impediment that the GroEL-ES system helps resolve.

***"GroEL-mediated acceleration of protein folding by enhanced collapse" Naqvi, M., Avellandeda, M.J., Roth, A., Koers, E., Sunderlikova, V., Kramer, G., Rye, H. S., and Tans, S. J. (2020) *Submitted*.**

Introduction

Amino acid chains display notoriously complex conformational search processes, which can produce highly intricate protein structures and functions, but also engender perilous delays that cause toxic aggregation(36). A longstanding question is whether such conformational searches are accelerated in the cellular context, and if so, which physicochemical principle could enable it. The archetypal chaperone GroEL-ES is generally believed most capable of achieving this effect, even as proposed mechanisms have remained controversial (77, 198, 199). GroEL-ES is speculated to accelerate folding by sterically confining protein chains within its closed chamber and thus lowering their entropy (79, 200, 201), or by partially unfolding kinetically trapped conformations (82, 83, 202, 203). However, GroEL-ES may alternatively act only as a passive Anfinsen cage that limits aggregation interactions between proteins (80, 204). Testing these models is challenging, as the distinguishing driving forces and conformational effects exerted by GroEL-ES are difficult to measure. Moreover, accelerated folding of individual proteins can be confounded by suppressed aggregation and heterogeneity in the population (205, 206). Establishing the principles of folding pathway acceleration is important for understanding protein folding and malfunction, and the function of chaperonins and chaperones in general.

Methods

Expression and purification of MBP, dmMBP and rhodanese

MBP and dmMBP were overexpressed in T7 competent cells (NEB laboratories) in LB medium supplemented with 0.2% glucose and 50 µg/ml kanamycin at 30°C until OD_{600}~0.6, induced with 0.4 µM IPTG (Sigma) and incubated at 18°C overnight. The culture was harvested by centrifugation at 5000 g for 20 minutes at 4 °C. All following steps were carried out at 4 °C. The pellet was resuspended in ice-cold buffer A (50 mM phosphate buffer pH 7.5, 200 mM NaCl, 10 mMEDTA, 50 mM Glutamic Acid–Arginine (Sigma) and 3 mM ß-mercaptoethanol (Sigma)) and lysed using an Emulsiflex homogenizer. The lysate was cleared from cell debris by centrifugation at 50000 g for 1 hour followed by incubation with Amylose resin (NEB) for 1 hour. After extensive washing with buffer A, the proteins were eluted using buffer A supplemented with 20 mM maltose. For rhodanese, the pellet was resuspended in buffer B (100 mM Tris-HCl pH 7.0, 5 mMEDTA, 20 mM $Na_2S_2O_3$, 2 mM ß-mercaptoethanol) and lysed as described above. Thelysate was mixed with Protino™ Ni-NTA Agarose (Macherey-Nagel) and incubated for 1hour. After washing, the protein was eluted with buffer B supplemented with 250 mMimidazole.

Purification of GroEL, GroES and their variants

GroEL was expressed from an inducible plasmid in *E. coli* BL21 in LB at 37 °C (36). After cell disruption, the crude lysate was clarified by ultracentrifugation (142000 rcf), followed by anion exchange chromatography (FastFlowQ, GE) equilibrated in buffer C (50 mM Tris pH 7.4, 0.5 mM EDTA, 2 mM DTT) and eluted by linear gradient from 7.5% to 35% with buffer D (50 mM Tris pH 7.4, 0.5 mM EDTA, 2 MNaCl, 2 mM DTT). GroEL fractions were concentrated by 70% (w/v) ammonium sulfateprecipitation. This precipitate was solubilized and dialyzed against 50 mM Bis-Tris pH 6.0,50 mM KCl, 0.5 mM EDTA, 2 mM DTT containing 25% (wild-type GroEL) or 12.5% (all GroEL mutants) methanol. A second round of strong anion exchange (FastFlow Q, GE), run in the same methanol-containing buffer at pH 6.0, was used to strip co-purifying small proteins and peptides from the GroEL oligomers. To further remove contaminating proteins and peptides that remain tightly associated through prior stages of purification, GroEL fractions were gently agitated in the same methanol-containing buffer and Affi Blue Gel (BioRad) resin overnight at 4°C under an argon atmosphere. The final sample was dialyzed into storage buffer (25 mM Tris pH 7.4, 100 mM KCl, 0.5 mM EDTA, 2 mM DTT),supplemented with glycerol (15–20% v/v), concentrated, and snap frozen using liquid nitrogen.

GroES was expressed from an inducible plasmid in *E. coli* BL21 (DE3) in LB at 37°C. After cell disruption, the crude lysate was clarified by ultracentrifugation (142,000 rcf), followed by acidification with sodium acetate, and cation exchange chromatography (FastFlow S, GE) equilibrated in buffer E (50 mM NaOAc pH 4.6, 0.5 mM EDTA, 2 mM DTT) and eluted by linear gradient from 0% to 25% buffer F (50 mM NaOAc pH 4.6, 0.5 mM EDTA, 2 M NaCl, 2mM DTT). The sample was dialyzed against 25 mM Tris pH 7.4, 0.5 mM EDTA, 50 mM KCl, 2 mM DTT and applied to a strong anion exchange column (Source Q, GE). GroES was eluted with NaCl and enriched fractions were pooled. The sample was dialyzed into storage buffer supplemented with glycerol (15 – 20% v/v), concentrated, and snap frozen using liquid nitrogen.

For the expression of Single Ring GroEL (SR1), *E. coli* BL21 DE3 transformed with pSR1 was grown in LB-Ampicillin (100 µg/ml) at 30°C to an OD_{600}=0.5. Overexpression was induced by adding 1 mM IPTG and growth was continued for 3 hours. Cells were harvested by centrifugation and stored at -70°C after flash freezing in liquid nitrogen. Frozen cells were resuspended in 20 mM Tris-HCl pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, lysed using a French Press and cell debris were removed by centrifugation. SR1 was enriched by fractionated $(NH_4)_2SO_4$ precipitation between 35% and 45% saturation. Following dialysis in 50 mM Tris-HCl pH 8, 1 mM EDTA at 4 °C, the protein solution was fractionated using a DEAE Sepharose Fast Flow anion exchange chromatography resin (GEhealthcare) eluting with a gradient from 0 to 1 M NaCl and further fractionated by size-exclusion chromatography using a HiPrep™ 26/60 Sephacryl® S-500 HR column. SR1containing fractions were pooled, concentrated using Amicon Ultra centrifugal filters (Merck), frozen in liquid nitrogen and stored at -70 °C.

GroEL labeling

The GroEL variant (EL315C) was labeled with Atto-532 maleimide (Sigma). Reactive dyes were prepared fresh from dry powderin anhydrous dimethylformamide (DMF) immediately prior to use. All proteins were first buffer exchanged 300-400x the original volume by a Vivaspin Turbo 15 (Satorious) into 50 mM Tris buffer pH 7.4, 100 mM KCl, 0.5 mM EDTA, and 1 mM TCEP. The proteins were then run over gel filtration (PD-10 column; Pharmacia) equilibrated in reaction buffer (50 mM Tris pH 7.4, 100mM KCl, 0.5 mM EDTA, 0.5 mM TCEP). EL315C was concentrated to a final concentration of 70 µM (monomer) in a volume of 5 mL. Protein samples were added to individual 5 mL conical Weaton reaction vials, followed by two sequential reactive dye additions. Freshly prepared Atto-532 maleimide in DMF was added at a molar ratio of 1:6.5 to EL315C monomer. Following each addition, the sample was incubated for 45 minutes in the dark at 23°C. Following the full 1.5 hours reaction time, the sample was quenched by addition of 5 mM glutathione. The labeled EL315-Atto532 were separated from unreacted dye by four rounds of dilution and concentration in a Vivaspin Turbo 15 (Sartorious), followed by gel filtration (PD-10 column; Pharmacia). The labeled proteins were then supplemented with glycerol (15-20%) and snap frozen using liquid nitrogen. Protein concentration was determined using a calibrated Bradford assay, in which the protein standard was from a sample of wild-type GroEL whose concentration had been previously established. Conjugated dye concentrations are determined by absorption spectroscopy of the denatured proteins (in 6 M Gdm buffer) using the following corrected extinction coefficient: Atto-532, 11,5000 $M^{-1}cm^{-1}$. GroEL-Atto 532 activity was confirmed by MESG ATPase activity assay (EnzChek, Molecular Probes) and native gel filtration (Superdex 200, GE).

GroES mobile loops and GroEL C-tails

The GroES mobile loops (198) ETKSAGGIVLTGS and GroEL C-tails (GGM)4M were ordered from Genscript. GroES mobile loops and C-tails were dissolved in MQ water and snap frozen using liquid nitrogen. Prior to measurements the samples were dissolved in HMK buffer (50 mM HEPES, pH 7.5, 5 mM MgCl2, 100 mMKCl). The GroES mobile loops were added in fivefold molar excess to GroEL during optical tweezers experiments (Figure B.3 and Figure B.8).

Protein-DNA constructs

The cysteines at the N and C termini of proteins were coupled with 20 bp maleimide ssDNA oligos at 37°C for one hour. 2.5 kbp and 1.3 kbp DNA tethers were generated by PCR from pUC19 plasmid (NEB) with a double digoxigenin- or biotin-labeled primer on one side and a phosphoprimer on the other side. Purification was done with the QIAquick PCR purification kit (Qiagen). The phosphorylated strand was digested by Lambda exonuclease (NEB) for 2 hours at 37°C and purified using an Amicon 30 kDa MWCOfilter (Merck). Deep Vent exo- DNA polymerase (NEB) and a 20 nt more upstream primer than the phosphor primer from the PCR was used for the fill up of the second DNA strand creating a 20 nt overhang. This overhang is complementary to the 20 nt oligonucleotide sequence coupled to the termini of proteins. The overhang DNA was added to the protein-oligo chimera together with T4 ligase (NEB) and incubated for 30 min at 16°C followed by 30min on ice. The resulting protein-DNA hybrid was flash frozen and stored at -80°C until measurement.

Optical tweezers experiments

Neutravidin coated beads (2.1 µm) were purchased from Spherotech and stored at 4°C until use. Anti-digoxigenin beads were prepared by coating carboxylated polystyrene beads (2.1 µm, Spherotec) with anti-digoxigenin antibodies from Sigma-Aldrich using a carbodiimide reaction (Poly-Link Protein Coupling Kit, PolysciencesInc.) The protein coated beads were prepared by mixing 50 ng of MBP, dmMBP orrhodanese constructs with anti-digoxigenin beads in 10 µl HMK buffer. The mixture was then incubated at 4°C for 30 min on a rotary mixer. Next, the beads were dissolved in 400 µl HMK buffer for optical tweezers experiments. Optical tweezers measurements were done in HMK buffer. ADP and ATP solutions were prepared by dissolving ADP and ATP sodium salt from Sigma Aldrich in HMK buffer. Experiments in GroEL ADP conditions were verified using ultra-pure ADP (99.9%, Gentaur).

Stretch-relax experiments were performed on two optical tweezers setups. The first was a custom-built single trap instrument. A substrate-coated anti-digoxigenin bead was held in theoptical trap and a NeutrAvidin bead was placed on the end of a micropipette tip. The two beads were brought in close contact, allowing a tether between the beads to form. Proteins were stretched and relaxed by moving the flow-cell and micropipette with a nanopositioning piezo stage at 50 nm/s speed which corresponds to a pulling rate of 5 pN/s. The deflection of the bead in the trap was measured using quadrant photodiode at 50 Hz. The data were filtered with a 5th order Butterworth filter at 20 Hz. The optical traps were calibrated by recording the power spectrum of the Brownian motion of the beads yielding stiffnesses ranging from 120 – 170 pN/μm.

The second setup was a dual trap optical tweezers instrument (C-trap from Lumicks). As described above, tethers were formed by bringing similarly prepared construct-coated and NeutrAvidin beads in close proximity. The protein was stretched and relaxed at a constant velocity of 50 nm/s, by moving one of the traps. The data was acquired at 500 kHz and averaged to 500 Hz. For constant force measurements, tension was held at 2 pN on average for 30 s using a proportional–integral–derivative (PID) feedback loop, before pulling again at constant velocity (Figure B.1E). In figure B.3C, the distance between the traps is constant, while the extension of the protein is monitored as it changes conformation. Note that the beads can change position within the traps. For fluorescence measurements in combination with stretch-relax experiments (Figure B.4A-D), Atto-532 labelled GroEL proteins were visualized using a green excitation laser (532 nm), with 2 mM Trolox and 4 mM ß-mercaptoethanol in the buffer. The excitation beam was used to scan along the tethered construct at 10 Hz during the force-spectroscopy measurements, generating fluorescence kymographs that were aligned to the force signal using ImageJ.

Data Analysis

Several checks were performed to confirm that the data corresponded to a proper single tether, which include comparing the total measured unfolded length to the expected length, consistency with the WLC model (at higher forces), overstretching at 67pN, and final tether breakage in one clean step. The unfolding forces (FU), contour lengths (LC), refolding forces (F*) and compaction energies (EC) were quantified from force extension data, using an open source MATLAB code (5) after modifications. FU was determined from stretching traces as the force required to fully unfold a protein (Figure B.1C left, E, Figure B.2C to E, Figure B.4, and Figure B.8). For stretching traces in which the protein did fully unfold below the maximum force that could be applied (67 pN, corresponding to the DNA overstretching plateau), FU was determined as 67 pN, the maximally sustained force (Figure B.11, Figure B.2C). The contour lengths (Lc) of refolded states were determined from the force-extension data of the stretching curve before the first unfolding transition, using the WLC model (Figure B.2A, B). The persistence lengths of the DNA (45 nm) and protein (1.5 nm), and the stretch modulus of DNA (1200 pN) were fixed parameters in the WLC model. In Fig 4A and 12C, the instantaneous protein contour length was calculated using the same WLC model. Compaction energy (EC) was calculated by quantifying the area under the relaxation curve and then subtracting the area under the WLC curve for fully unfolded protein

(Figure B.6B, C, Figure B.8, and Figure B.3C). P* was determined as the fraction of relaxation traces that show (one or more) steps in Lc of more than 15 nm (Figure B.2C and Figure B.8). F* was quantified as the measured force just before such a step in Lc (during relaxation, Figure B.4C, Figure B.8, and Figure B.10). The folding probability (Pc) was quantified as the fraction of relax-stretch cycles showing refolding to the core MBP state (Figure B.4C, Figure B.10, Figure B.2, and Figure B.8).

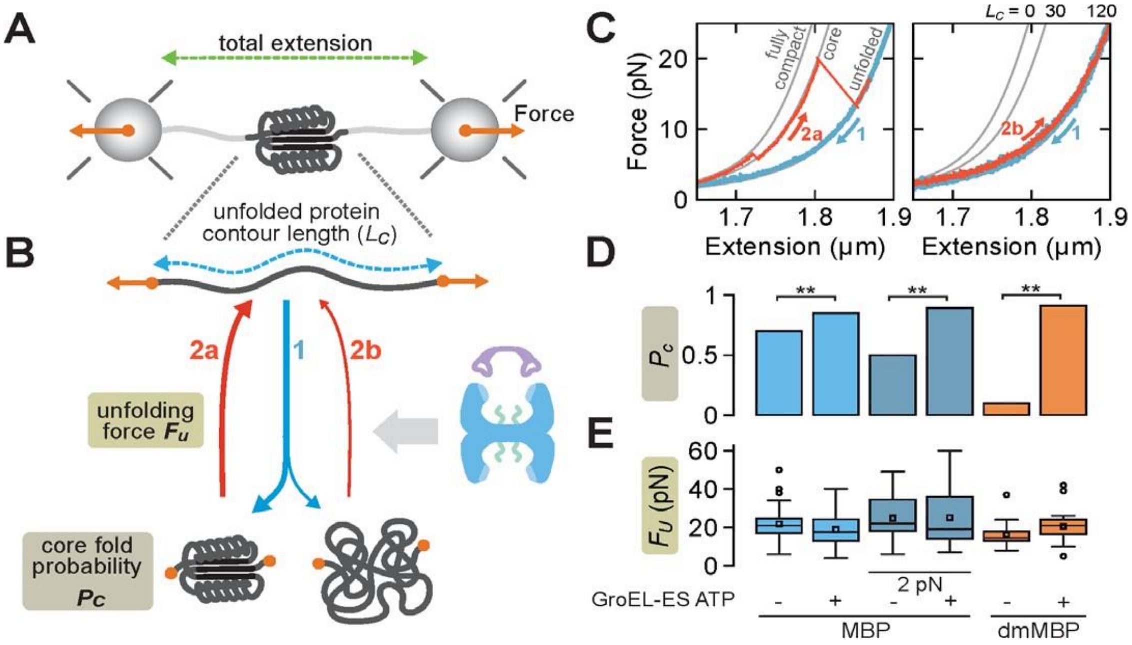


Figure B.1 GroEL-ES accelerates folding.
(A) Optical tweezers experiments. (B) Relax-wait-stretch cycles to quantify MBP core refolding. (C) Resulting example force-extension data without GroEL-ES, showing refolding (right) or not (left). Gray: predicted worm-like-chain (WLC) curves. (D) Fraction of cycles showing core refolding (P_c), with or without 200 nM GroEL, 500 nM GroES, 1 mM ATP. Two stars: significant difference (p<0.05). (E) Unfolding force F_U during stretching. For dmMBP alone, $F_U$of the first stretching curve is displayed because of the low refolding rate.

Statistical Analysis

The statistical significance of differences in folding probability (P_c) and refolding at force probability ($P*$) between experimental conditions was calculated using one tailed two proportion z-test. The statistical significance of differences in compaction energy (E_c) and maximally sustained forces (F_U) between experimental conditions was calculated using two sample assuming unequal variance t-Test. Test results are mentioned as p values in the main text. In box charts, whiskers indicate 90% and 10% extreme values, the inner line represents the median, the length of the box indicate interquartile range and the inner small square the mean of the population.

Results

We first aimed to test whether GroEL-ES mediates folding acceleration by following individual proteins in time (Figure B.1A-C). We tethered Maltose Binding Proteins (MBP), exposed them to relax-wait-stretch cycles using optical tweezers, and quantified the fraction of cycles P_c showing core refolding (207-209). When GroEL, GroES and ATP were flown in, P_c increased modestly from 0.7 to 0.85 (Figure B.1D, 20, $p<0.05$), while the unfolding force F_u remained similar (Figure B.1E). A reported slow-folding mutant dmMBP (206) displayed a steeperP_c increase from 0.06 to 0.9 in these experiments (Figure B.1D and Figure B.2). To study a different folding limitation, we maintained a low force on the MBP chains during the waiting period. The addition of GroEL-ES and ATP again increased P_c (from 0.5 to 0.9, Figure B.1D, $p<0.05$).We

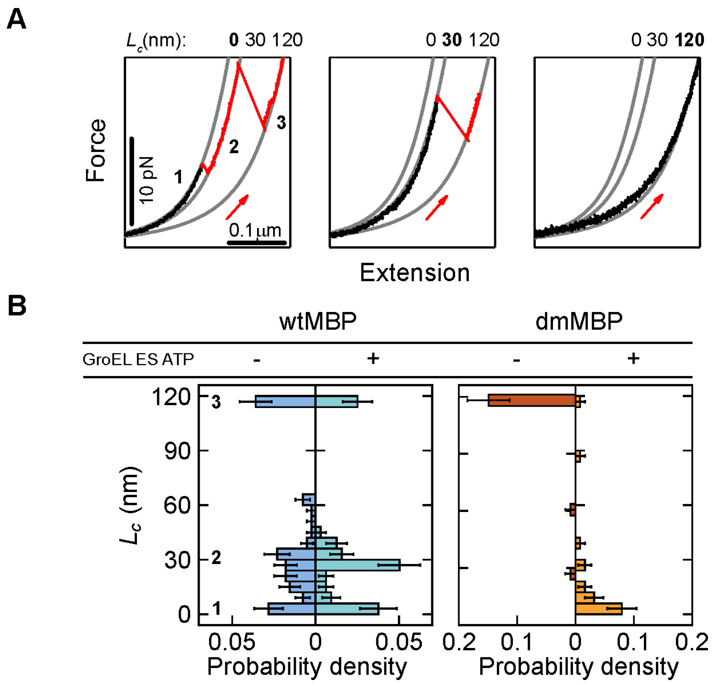


Figure B.2 Contour lengths of refolded states.

(A) Determination of contour lengths (L_c) from force-extension data taken during stretching. Displayed stretching curves are taken after relaxation of unfolded chains and waiting for 5 s at 0 pN to allow refolding. The protein states after this refolding window are characterized by their contour lengths (L_c), which represents the length of the non-compacted part of the protein chain. This L_c of refolded states is determined based on the first part of the stretching data, where no (detectable) unfolding has yet occurred (in black). This data is described by a single mean L_c value that is determined using the worm-like chain (WLC) model (gray curves). Gray curves: force-extension behavior of DNA tethers attached to an unfolded protein chain of three different contour lengths L_c. Indicated are WLC curves for fully compacted (L_c=0 nm), MBP core state(L_c = 30 nm), and fully unfolded state (L_c = 120 nm). Panels indicate different example stretching curves that are observed for MBP with 200 nM GroEL, 500 nM GroES, and 1 mM ATP. Left panel: refolded state L_c = 0 nm (1), followed by unfolding to core state (2), and to fully unfolded state (3). Middle panel: refolded state L_c = 30 nm (core state), which then unfolds to fully unfolded state. Right panel: refolded state L_c = 120 nm (fully unfolded state). For traces that show no unfolding transitions, like the latter example, L_c is measured at 10 pN. (B) Probability density (P.D.) of contour lengths (L_c) of refolded states. Determination of L_c as described in panel A, for MBP (blue), and dmMBP (orange) in the presence and absence of GroEL-ES and ATP. This analysis is also performed to quantify the frequencies (P_c) of refolding to the core MBP structure. Also note that refolded states larger than the core structure always displays unfolding to (via) the core state, which involves detachment of MBP c-terminal helices from the core structure. Error bars are s.d.

thought that the DNA tethers attached to the proteins could, at least in principle, have prevented folding-stimulation by hindering closure of the GroEL-ES cavity. Yet, the findings indicated accelerated formation of tertiary structure nonetheless. The mechanism of acceleration remained unresolved, however to reveal possible underlying induced forces and conformational changes, we focused on GroEL alone in its different nucleotide-bound states, without GroES. We now observed sudden switches to unfolded states that persisted over multiple cycles until the tether broke. These switches occurred frequently for the APO state, and less so for the ATP (mimicked using the slowly hydrolyzing GroEL398A (75) and ADP states (50, 30, and 20% of tethers, respectively, Figure B.3). These findings fully agree with the known stable binding of

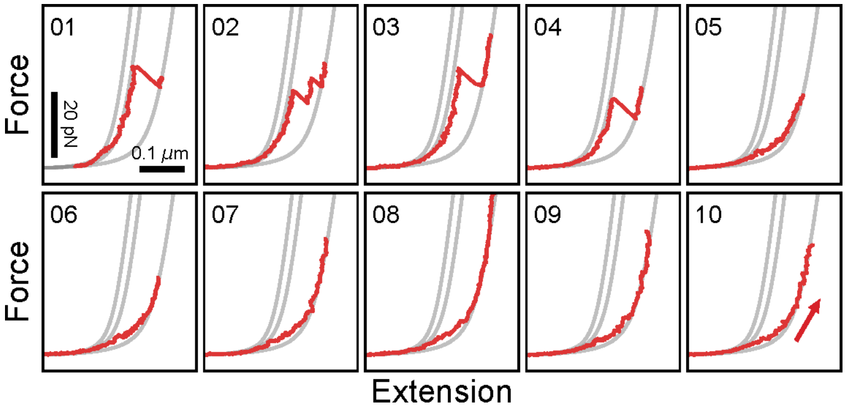


Figure B.3 Irreversible switching to unfolded states in presence of GroEL. Force extension traces of MBP showing a sudden switch to a stable unfolded state. Successive stretching traces from relax-stretch cycles for MBP in the presence of 200 nM GroEL and 1 mM ADP, which initially (cycles 1 to 4) show the data following the worm-like chain (WLC) curve of the MBP core state (middle gray line) and thus indicating core refolding, followed by progressive unfolding to the fully unfolded state (right gray line). However, the data follows the WLC curve of the unfolded state and shows no unfolding transitions in the subsequent cycles (5 to 10), indicating the chain remained stabilized in the unfolded state. The data was close to the WLC curve corresponding to the unfolded state, though the deviation at lower forces suggested that a compacted yet non-folded state continued to be formed and disrupted.

unfolded substrates to the apical domains at the rim of the GroEL cavity (210). In addition, however, polypeptide chains now displayed gradual decreases in contour length L_c during relaxation, down to dimensions of folded states, for all three nucleotide states (Figure B.4A). The chain was thus gradually compacted, as L_c measures the extended (non-compact) part of the chain (Figure B.4B). In principle, such a compaction could be the result of one chain binding many GroEL apical domain sites. A number of features indicated a different scenario however.

First, the gradual compaction was accompanied by sudden step-wise compaction events (Figure B.4A, stars). These steps were suggestive of folding

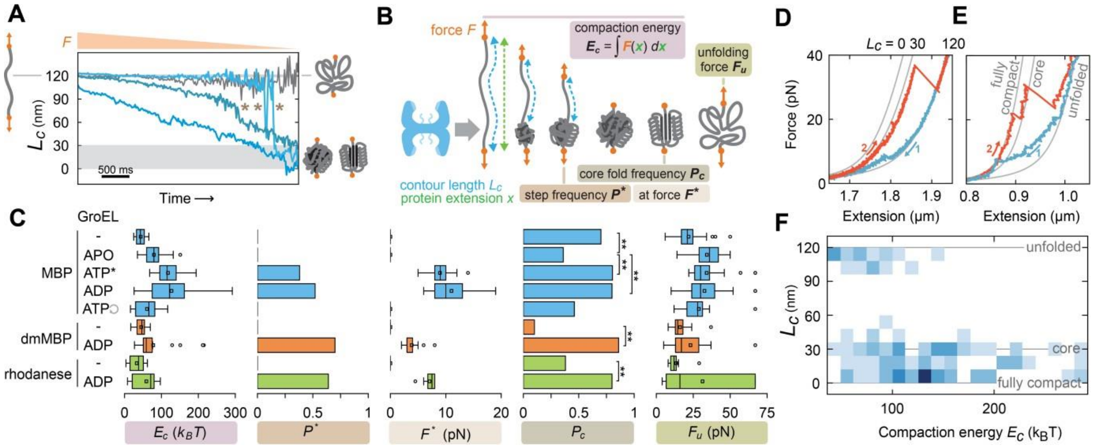


Figure B.4 An open GroEL cavity can enhance protein chain collapse and folding.
(A) Unfolded chain contour length (Lc) during relaxation, from force-extension data (above 2 pN only because of noise). Blue: dmMBP with GroEL-ADP. Stars: Lc steps. Gray: example to illustrate no detectable compaction above 2 pN (dmMBP alone). (B) Relax-wait-stretch cycle. (C) Quantification of relax-wait-stretch cycles (see panel B). ATP*: GroEL398-ATP. ATP-cycle: GroEL-ATP. Two stars: $p<0.05$. (Dand E) Example force-extension data of relax-wait-stretch cycle, MBP with GroEL-ADP. (F) Lc of refolded states detected during stretching (fig. S1), against Ec of previous relaxation, for GroEL-ADP and GroEL398-ATP.

transitions rather than stable binding: they were large in size (up to nearly the total chain length, Figure B.5B), occurred at high forces (up to 19 pN, Figure B.4A-C), and exhibited reversible 'hopping' (211-213) (Figure B.4A and Figure B.5C). Second, compaction was strongest for the ATP and ADP states, when the irreversible stabilization of unfolded states was weakest. The former was quantified by high fraction $P^*$of traces showing steps, and a high total compaction energy E_c for the full relaxation process (Figure B.4B-C and Figure B.6). Third, after relaxation and waiting,

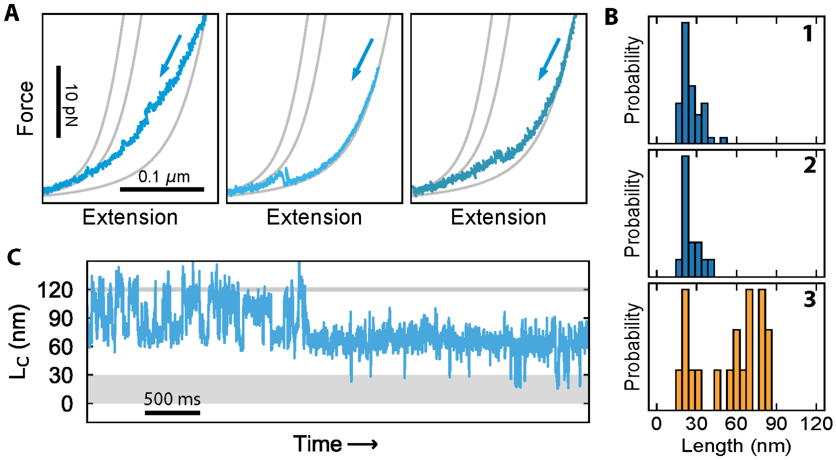
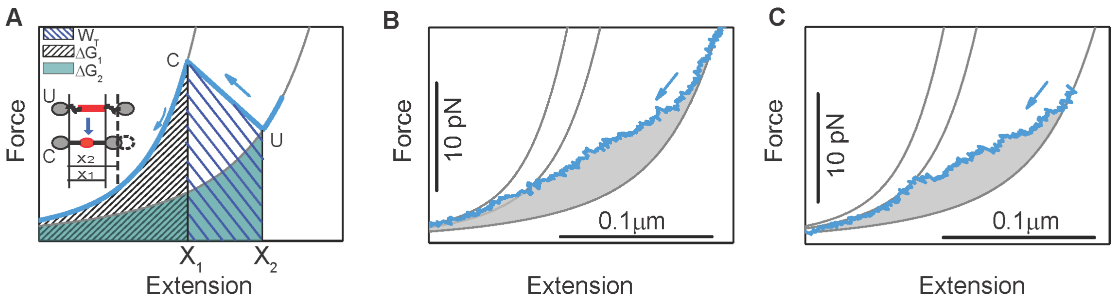


Figure B.5 GroEL-mediated chain compaction and hopping transitions.
(A) Three example relaxation curves, displayed in force-distance graph, showing gradual and step-wise compaction, ford mMBP in the presence of 200 nM GroEL and 1 mM ADP condition. (B) Histograms of contour length changes of step-wise compactions during relaxation (see panel a) for MBP, 200 nM GroEL and 1 mMADP (1, N = 52), MBP, 200 nM GroEL398 and 1mM ATP (2, N = 14) and dmMBP, 200 nM GroEL and1 mM ADP (3, N = 19) conditions (C) Contour length vs time trace, showing repeated step-wise transitions (hopping) between states for dmMBP, mediated by GroEL. Data is taken with both traps at a constant position, in the presence of 200 nM GroEL and 1mM ADP. We stress that the details of these data, such as the folding step-sizes, are specific to this experiment and not generally observed. The latter may be expected. For hopping transitions of isolated proteins without chaperones, the energy landscape is defined only by the tethered protein that is in principle the same for different experiments. In contrast, here the energy landscape is also defined by GroEL, and how and where it is interacting with the substrate, which has a random aspect, and hence will produce differences between experiments.

chains were often found in the core state, with stretching data following the core WLC curve (Figure B.4D-F), as was seen before for autonomous and GroEL-ES-assisted

Figure B.6 Determination of compaction energy.
A) Schematic diagram showing how the compaction energy (Ec) is determined from force extension cycles. We consider an unfolded protein chain during relaxation that suddenly compacts fully, but the derivation can also be used for gradual compaction composed of multiple smaller compaction events. At this single sudden event (U->C), the measured distance between beads (Extension, along the x-axis) suddenly decreases from x2 to x1, and the measured force F(x) that acts on these beads and throughout the DNA-protein-DNA tether (Force, along the Y-axis) increases from F(x2) to F(x1), because the tether is now effectively shorter and hence its tension higher (see inset in panel a). In the case of slow relaxation where the system is in equilibrium and there is no heat dissipation, energy is conserved, and hence the corresponding increase in potential energy of the beads equals to the work done (WT) by DNA-protein-DNA tether. WT is then estimated as the area under the force-extension curve from x2 to x1 that quantifies the increase in bead potential energy. Note that displacing an object over distance dx against a force costs an amount of energy F·dx. Thus, WT is quantified by $\int x1\ (x)$ (wide-hash region). Which can be decomposed into two contributions: the work done to compact the protein (Ec) and to increase the tension in the DNA-protein-DNA tether (Wext), hence: WT = Wext + Ec. Wext can be calculated using the WLC model. To use this, we write Wext = ΔG2 – ΔG1, where ΔG2 is the work done in extending the unfolded protein chain and DNA linkers from extension 0 to x2 (green region), which is calculated with the WLC model. ΔG1 is the work done in extending only DNA from 0 to x1 (narrow-hash region), and is also calculated with the WLC model. ΔG1 has no contribution from the protein chain because it is fully compacted in this state. The work done in compacting the protein (Ec) can then be calculated as WT + ΔG1 – ΔG2. Thus, in graphical terms, Ec equals the size of the wide-hash (WT) plus narrow-hash (ΔG1) regions minus the size of the green region (ΔG2). In more simple terms, this is thus the size of the area under the measured curve F(x) minus the size of the area under the WLC curve for the unfolded protein (right gray curve), as illustrated for measured data in panels B and C. Note that in the latter one can integrate from x = 0 to any x > x2, as beyond x2 the chain is fully unfolded and hence there are no further area contributions. Perhaps counterintuitively, Ec is thus determined not only by F(x) for x in between x1 and x2, but also by F(x) for x in between 0 and x1. Note that while the compacted chain (C) may in principle be deformed for x < x1, the length changes as well as energies are negligible, owing to its high stiffness compared to the DNA, while the force is identical throughout the chain. Finally, we note that this estimate of Ec is a lower-bound, given that the system is not fully in equilibrium. (B to C) For gradual collapse and collapse with refolding jumps, the compaction energy Ec is thus determined by the size of the indicated gray area.



folding. Notably, the fraction of cycles showing refolded cores was particularly high for the ATP and ADP states (P_c=0.8 for both, Figure B.4C).Moreover, P_c even increased beyond 0.95 when the previous compaction was strong (E_c>100 k_BT, Figure B.4F and Figure B.7). These findings showed that compaction was distinct from the stabilization of unfolded states, and played a role in stimulating folding instead.

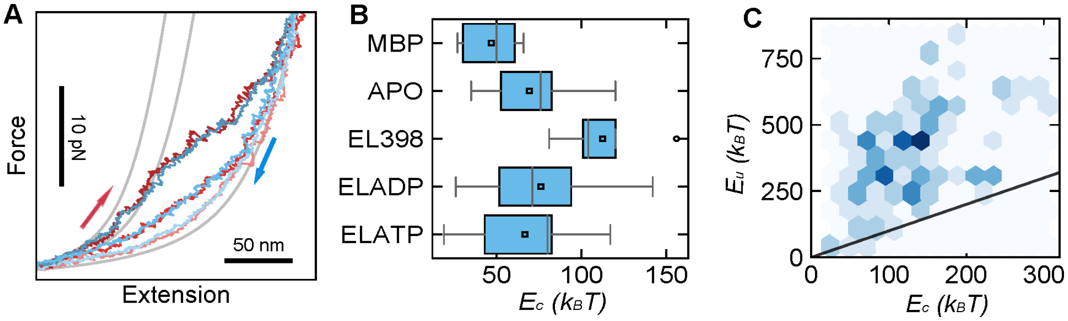


Figure B.7 Collapse without folding, and stretching energy.
(A) Force-Extension traces that show compaction during relaxation (blue), and after waiting at 0 pN for 5 s, are followed by stretching (red) that do not show unfolding features. These cycles indicate that significant compaction can occur without the formation of detectable folded states. Note that most cycles in these conditions rather show refolding instead (See Fig. 2 main text). The displayed data is for MBP in presence of 200 nM GroEL and 1 mM ADP. (B) Corresponding compaction energy (Ec) of relaxation traces that do not produce detectable refolding (for examples see panel a) in different conditions. Ec is highest for EL398ATP and ELADP (as in Fig. 2C). (C) Area under the stretching curve (stretching energy Eu) against the compaction energy (Ec) from the prior relaxation trace for MBP with 200 nM GroEL and 1mM ADP condition. The data here includes both cycles that do show refolding and cycles that do not. Black line indicates Eu = Ec. Relax-stretch cycles producing no folding (panels A and B) are close to this line.

GroEL thus displayed two interaction modes. In the first, unfolded substrates were bound, immobilized, and stabilized. In the second, they were compacted by attractive forces while preserving the necessary mobility to form tertiary structure. These compaction forces, which oppose and equal the applied forces (Figure B.4B, orange arrows), reached well above 10 pN (Figure B.4C, E). Such a collapse process, in which chains compact and may form some secondary and tertiary structure, is considered key in autonomous folding, and occurs gradually for isolated chains under decreasing tension (207, 214, 215). Here, we found that E_c,P*, and P_c for MBP increased when interacting with GroEL (Figure B.4C). Such an enhancement of the chain collapse was also observed for dmMBP and rhodanese (Figure B.4C and Figure B.8).

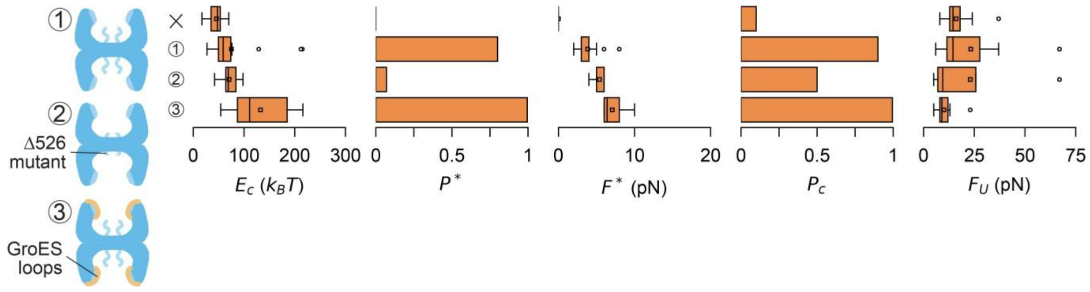


Figure B.8 Roles of GroEL apical domains and cavity, dmMBP data. From dmMBP relax-stretch cycles, we quantified: total compaction energy during relaxation (E_c), probability (P*) at force (F*) of steps during relaxation, core refold probability after 5 s. at 0 pN (P_c), unfolding force (F_U). Conditions are, from top to bottom: No Chaperone (x, *N*=19), 200nM GroEL and 1mM ADP (1, *N*=21), 200nMELΔ526 and 1mMADP (2, *N*=11) and 200nM GroEL, 1mM ADP and 1μM loops (3, *N*=16). Ford mMBP without GroEL (x), F_U of first pulls is displayed because of the low refolding rate.

To disentangle GroEL apical domain contributions to both modes, we mixed polypeptides representing the unstructured GroES loops, which compete strongly for the apical domain substrate binding sites (216), in the highest P^* condition (GroEL-ADP). MBP and dmMBP were no longer irreversibly stabilized in unfolded states, consistent with reduced apical domain binding. Interestingly, the refolded fractions P_c and P^* were now higher than for GroEL-ADP only (Figure B.10 and Figure B.9). These increases were not caused by the GroES-loops, which alone did not yield increases as expected (Figure B.10). Thus, the apical domains appeared to antagonize the second mode while promoting the first. The data also suggested another effect. The unfolding force F_u of MBP, had increased with GroEL-ADP and decreased back to MBP-only levels in the presence of GroES loops (22, 33, 23 pN respectively, Figure

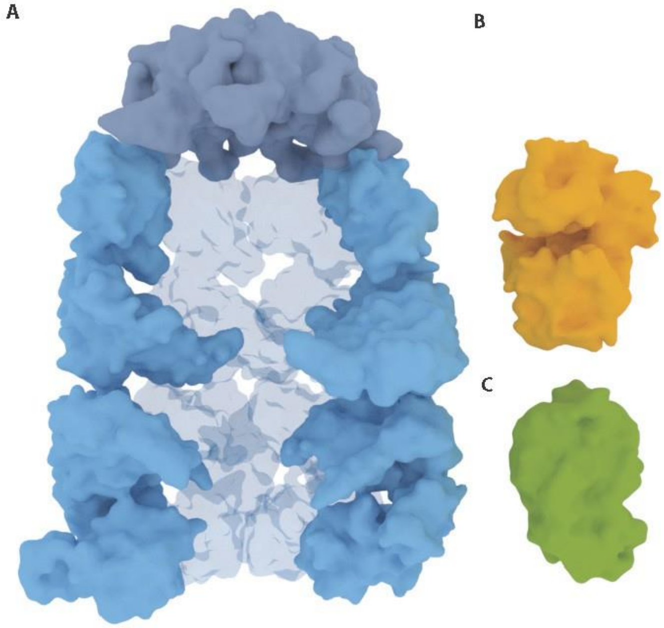


Figure B.9 Protein structures.
(A) A medial slice of GroEL-GroES ADP bullet, side view (PDB: 1PF9).(B) MBP (PDB ID: 2MV0) in orange. (C) Rhodanese (PDB ID: 1rhs) in green. Proteins are displayed in the same scale to compare their relative sizes.

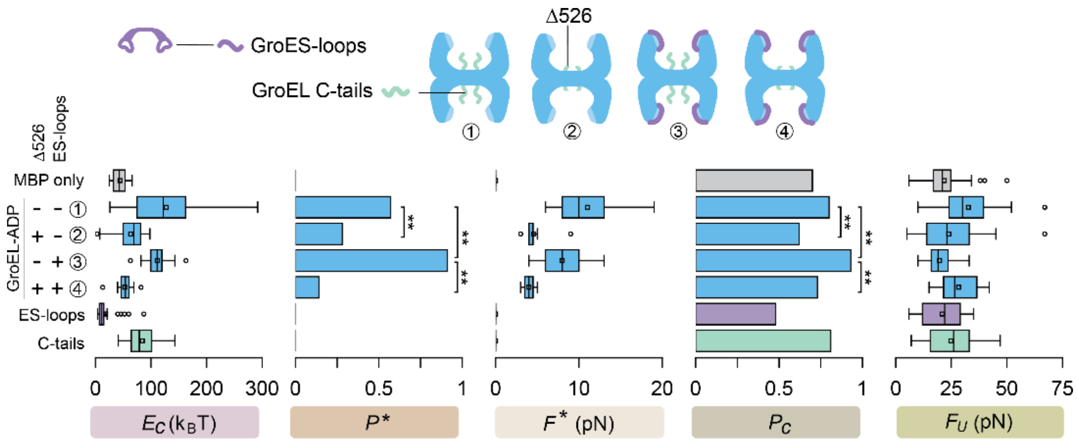


Figure B.10 Roles of GroEL apical domains and cavity.
Quantification of relax-wait-stretch cycles (Fig.2B), for MBP and two GroEL modifications: With polypeptides representing GroES-loops that bind theGroEL apical domains, and by truncating the C-terminal tails at the GroEL cavity bottom (GroELΔ526). Two stars: $p<0.05$

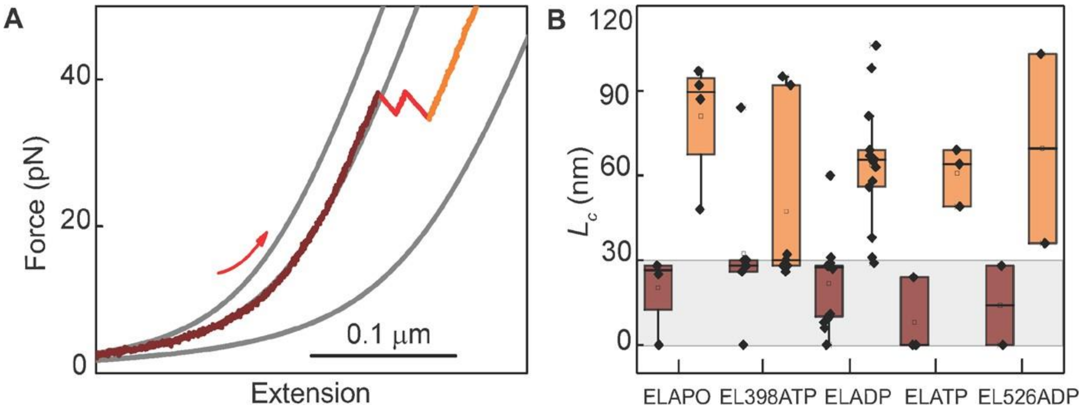


Figure B.11 Lengths of stabilized partially folded structures.
(A) Stretching curve showing stabilization of partially folded MBP states against forced unfolding, in the presence of GroEL-ADP.Displayed data initially follows worm-like chain (WLC) curve of MBP core state (dark red), then unfolds partially in two steps (red), to a partially folded state that is stable against high applied forces (orange). (B) Distributions of contour lengths L_c. Dark orange: L_c of initial MBP stretching data (panela, dark red). Light orange: L_c of observed MBP structures that are stable against forces over 40 pN (panel A, orange). 40 pN is the maximum force in the absence of chaperonin; see Fig. 2 C main text. The data indicates that the GroEL-stabilized MBP structures are typically smaller than the core state, in different nucleotide conditions.

B.10 and Figure B.11). dmMBP showed a similar trend (Figure B.8). GroEL thus can stabilize (partially) refolded states against forced unfolding, in addition to stabilizing unfolded states. Binding of the GroES-loops suppressed both modes. Yet, compaction and folding remained stimulated (Figure B.10).

If not the apical domains, could the GroEL cavity then play a role in folding stimulation, even without closure by GroES? To test this, we truncated the unstructured C-terminal tails at the cavity bottom (GroELΔ526) (81). P_c and P^* were indeed lower for GroELΔ526-ADP than for GroEL-ADP, by more than two-fold, both with and without GroES-loops present (Figure B.10 and Figure B.9). We found that even alone,

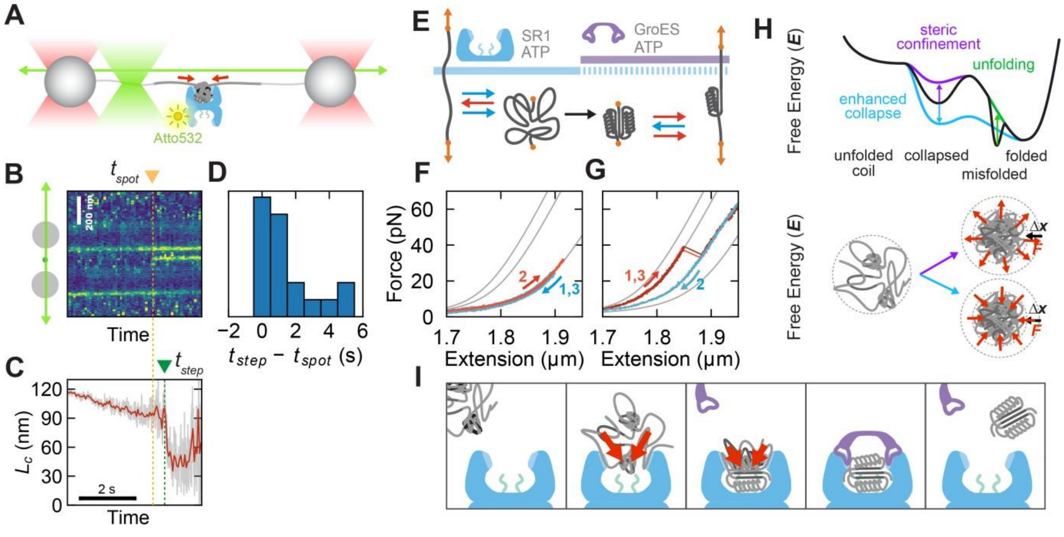


Figure B.12 Compaction and folding in a single GroEL tetradecamer.
(A-C) Optical tweezers and single-molecule fluorescence. (B) Fluorescence kymograph shows single GroEL binding during relaxation. (C) Simultaneous optical tweezers data shows folding step. (D) Time between GroEL binding and folding step (*N* = 19). (E-G) Ternary complex (SR1-GroES-MBP) test. (F) Relax-wait-stretch cycles with SR1 and ATP, indicating SR1 binding and unfolded-state stabilization. (G) Flow-in of GroES and ATP yields core refolding. (H) Energy landscape cartoons for different acceleration models. (I) Event sequence suggested from data.

the C-tails could promote some compaction (Figure B.10). Overall, the data thus showed that GroEL-mediated collapse and folding depended on the C-tails in the GroEL cavity. The capacity to enhance chain collapse while maintaining folding dynamics is notable, and suggests a balance between different cavity properties (81). Finally, we aimed to verify two key interactions in the substrate-chaperone complex, which required different approaches. To directly visualize GroEL binding, we scanned a fluorescence excitation beam along MBP during relax-wait-stretch cycles (Figure IV.30A). ADP and Atto532-labeled GroEL were present, but at a lower concentration to limit background fluorescence. The appearance of a fluorescent spot between the beads indicated binding of a single GroEL tetradecamer (Figure B.12B). Consistently, during relaxation, such GroEL binding events always occurred first, and folding steps afterwards (Figure B.12C, D), thus confirming stimulated folding transitions in substrates complexed with GroEL.

Second, we used a buffer-exchange protocol to verify ternary complex (GroEL-GroES-MBP) formation in the optical tweezers assay. Unfolded MBP was first complexed to a single-ring GroEL variant (SR1) with ATP (*29*). SR1 binding was evidenced by unfolded-state stabilization of MBP (Figure B.12E to F). GroES is known to bind SR1 very tightly, thus trapping it in the ADP-bound state, and displacing substrates from the apical domains into the SR1-GroES cavity (77, 199). The relaxed MBP-SR1 complex was then exposed to GroES and ATP, which thus washed away unbound SR1. This exposure triggered refolding to the core state that did not unfold fully below 67 pN (Figure B.12G). These data strongly suggested a SR1-ES-MBP complex: if not bound by SR1, MBP core states unfold below 40 pN (Figure B.3C), and if only SR1 and not GroES is bound, MBP remains unfolded (Figure B.12F). Because the MBP has DNA linkers attached, these data suggested the SR1-GroES cavity is not hermetically sealed (217). It is thus possible that GroES does not require intimate contact with all sevenSR1 monomers to form a stable complex capable of initiating folding.

Discussion

The enhanced-collapse model of folding acceleration that our data suggests is distinct from current models (77, 199) (Figure B.12H, I). In confinement models, steric repulsion forces exerted by a closed cavity decrease the chain entropy and thus increase its free energy, which lowers the folding barrier. In unfolding models, pulling forces on misfolds also increase the chain free energy and allow escape to proper folds. Here, we measured attractive forces that compact protein chains, mediated by open GroEL (not closed by GroES), which rather suggested a decreased chain free energy, while the observed higher folding probabilities indicated a lowered folding barrier. Consistently, collapse is thought to lower barriers in autonomous folding (214), as chain segments are overall closer together. There are many similar aspects to previous work as well. Though the compaction mechanism and resulting compacted states are different, chain segments are also brought together in confinement models. Consequently, collapse enhancement is consistent with diverse experimental findings (199, 201). As seen before

(199), chain segments were bound by GroEL apical domains in a nucleotide-dependent manner. Segments that were not bound, or released by GroES, were free to collapse and fold partially, mediated by the attractive forces (Figure B.12I). Enhanced collapse canal so act in conjunction with other mechanisms. Notably, the attractive forces may help to unfold misfolded chains that are apically bound, while the steric constraints of open and closed cavities can help define compacted states.

The capability to manipulate protein chain collapse may be vital for cells to support a broad range of protein sequences with different collapse tendencies, and thus expand the space of foldable proteins. Indeed, GroEL-ES assists highly diverse proteins, which raises the question of whether limited chain collapse is a general folding obstacle. Collapse enhancement may also provide a driving force for substrate transfer from Hsp70 to GroEL, limit the lifetime of aggregation-prone collapsed states, enable control over conformational transitions in intrinsically disordered proteins, and could be exploited by type II chaperonins such as TRiC/CCT, as well as by other chaperones.