ENHANCEMENT OF CHAPERONE-MEDIATED PROTEIN FOLDING THROUGH SUBSTRATE PROTEIN INTERACTIONS WITH THE GROEL C-TERMINI

A Dissertation

by

JEREMY SCOTT WEAVER

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 S. Rye
Committee Members,	Tatyana Igumenova
	Steve Lockless
	Ry Young
Head of Department,	Gregory D. Reinhart

May 2015

Major Subject: Biochemistry

ABSTRACT

Many essential proteins require the assistance of molecular chaperones to achieve and maintain their native, folded conformations. The *E. coli* GroEL-ES chaperonin system is capable of aiding the folding of the cellular proteome through several distinct mechanisms, including the blocking of intermolecular aggregation, the confinement of substrate proteins inside the GroEL-ES cavity, and the forced unfolding of substrate proteins to reinitiate the folding process. This study describes the role of the C-terminal residues of the GroEL protein enhancing of the folding of substrate proteins.

The 23 C-terminal residues of the GroEL monomer partially consist of four tandem repeats of a Gly-Gly-Met motif, leading to an intrinsically disordered conformation. Visualized using cryo-electron microscopy, these residues extend from the bottom of a GroEL cavity and interact with substrate proteins both before and during the folding process; removal of these residues leads to deficiencies in substrate protein encapsulation and folding. Interactions between the C-terminal tails of GroEL and the substrate protein lead to changes in both the conformation of the substrate protein and the binding position on the GroEL cavity surface before folding begins. These changes result in the substrate protein adopting a more unfolded state and migrating deeper into the cavity.

These effects were observed with two distinct substrate proteins, the carbon-fixing enzyme RuBisCO and the metalloprotease PepQ, which have unrelated, dissimilar structural folds. Changes in the subsequent intra-cavity folding relative to folding in the

absence of the GroEL C-termini were also observed for both proteins, indicating altered utilization of folding pathways or intermediates. Significantly, these alterations of substrate protein conformation and folding, which lead to the enhancement of folding rate, are independent of any passive mechanism of preventing aggregation, as PepQ has no propensity for aggregation under the employed conditions.

The results of this study show that GroEL can actively enhance the folding of proteins by altering the conformation of the substrate protein, thus changing the folding pathway and directing it towards the native-state.

DEDICATION

This dissertation is dedicated to every person that enters a graduate program and does not finish. I hope that their new paths allow them to fulfill their dreams and aspirations.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Hays Rye, for his expertise, guidance and support, and my committee members for their involvement in overseeing my progress throughout my graduate career. I would like to thank my collaborators for supporting our endeavors with their biophysical techniques. I would like to thank the random panel members at the NIH study group who recommended funding Hays' grant and the NIH for following that recommendation. I would like to thank the many friends I have made in graduate school, in particular Ben Kaster and Tim Hou, with whom I have made many lasting memories. Finally, I would like to thank the giant family that is the Department of Biochemistry and Biophysics for its support and assistance over the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

A History of Protein Folding

Proteins are synthesized by the ribosome as unstructured chains of amino acids that often need to adopt stable, three-dimensional structures. These three-dimensional structures are required for the activity of many proteins, as the structure of each protein creates specific binding sites for ligands or positions amino acid residues correctly for the catalysis of chemical reactions. For viability, cells require thousands of different processes to occur; therefore, the proteins that perform these processes have evolved many different structures. From barrels to propeller-like structures, each protein starts from an unfolded form. The process of a protein transitioning from an unstructured to a structured conformation is commonly referred to as protein folding.

'Protein' is a common term in current times, both inside and outside of the scientific community. This term originated in the early 1800s in a letter sent from the Swedish chemist Jöns Jacob Berzelius to the Dutch chemist Gerardus Johannes Mulder [1]. In his examination of proteins, Mulder determined that these molecules were significantly larger than the molecules that chemists at the time generally examined, with proteins having an average molecular weight of ~8.8 kDa [2,3]. Mulder believed that all proteins were nearly identical in chemical composition [2,3], and the determination that several proteins had similar compositions suggested that all proteins were related macromolecules [4]. Although it took nearly 100 years, this

concept was partially confirmed when all of the common amino acids were finally discovered [5,6]. Eventually, the peptide theory was introduced by both Fischer and Hofmeister [7-9]; this theory stated that proteins are comprised of chains of amino acids. Although proteins were known to be comprised of one or more peptides, the assembly of these peptides into proteins remained a point of contention [9]. Methodologies were developed for determining the amino acid composition of proteins, as well as the sequence of those amino acids in the polypeptide chain [10-19]. These advances led to the initial characterization of many proteins [20-22], thus demonstrating the validity of the peptide theory and revealing the polymeric nature of proteins.

Proteins were known to adopt stable, compact structures [23,24], but the forces that held proteins together were not known. Understanding the chemical properties of the amino acids allowed for predictions to be made concerning the interactions that stabilize protein structure. Although hydrogen bonds were presented as potentially ubiquitous, stabilizing interactions inside of proteins [25], there was little direct evidence supporting this concept. However, having been observed in many proteins and with a known importance in insulin [14,26,27], disulfide bonds were considered to be an important, stabilizing, structural element in proteins; many early theories were explained or rebutted only by the observed effects on sulfhydryl groups [28]. The mechanisms leading to the altered accessibility of these groups upon the exposure of a protein to various chemicals (e.g. urea, gaunidinium, and detergents) were unknown, but it was known that the structure of the protein changed into a more open state under these conditions [29]. The inactivation of protein activity, structure, and other observable

properties with exposure to changed temperature, pH or the presence of the aforementioned chemicals was found to be reversible for many proteins [28,30-34]. This demonstrated that proteins could adopt multiple conformations, and that proteins were designed to return to the functional state. It was postulated that this could serve as a control mechanism for regulating protein function *in vivo* [28]. These observations involving the renaturation of proteins after exposure to chemical or thermal denaturation are considered the first developments in the field of protein folding.

The advancement of protein crystallography revolutionized many aspects of biology and biochemistry. Although the first crystal structures of whole proteins were not solved until the late 1950s [35-39], earlier structures existed of amino acids, small peptides, and protein fragments [40-43]. The observed bond lengths and angles in these structures allowed scientists to begin formulating hypotheses about protein structure with greater detail and accuracy. Linus Pauling cemented his previous predictions on a role for hydrogen bonding in the accurate description of alpha-helices and beta-sheets [44,45]. The importance of hydrophobic interactions in the core of proteins was also gaining prominence [46-48]. With a greater understanding of the non-covalent forces that stabilize protein structure, protein chemists theorized that the unfolding and refolding of proteins caused by altered temperature or the presence of chemical denaturants was dominated by the loss or gain of protein tertiary structure [49]; the extent of secondary structure loss in protein denaturation was realized fully with the advancement of spectrophotometric methods [50-54]. This greater knowledge of protein

structure and the observed reversibility of protein denaturation are the cornerstones of the modern theory of protein folding.

The process of protein folding was deduced thus: hydrogen bonding drives the formation of secondary structures, which then fold upon themselves and are stabilized by disulfide bonds and other interactions between physically adjacent side chains [49]. Because proteins could undergo this process spontaneously and repeatedly, it was postulated that the primary structure of the protein encoded the instructions for the folding of the protein [49]. Although the reversibility of denaturation had been shown in many proteins, the strongest support for this hypothesis was found in the detailed studies on the refolding of ribonucleases led by Christian Anfinsen [54-61]. Anfinsen found that bovine pancreatic ribonuclease could efficiently and independently adopt its native conformation after being completely denatured, as long as incorrect disulfide bond formation could be reversed. This demonstrated that other proteins or macromolecules were not required for the correct folding of ribonuclease. Anfinsen expanded the theory of protein folding with his "thermodynamic hypothesis," stating,

> "the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment" [62].

This hypothesis eventually became known as Anfinsen's Dogma. Despite the ability of ribonuclease to fold independently, the refolding process occurred on the timescale of hours [56,57]. Although this result seemed too slow for Anfinsen [59], Cyrus Levinthal noted that if protein folding occurred through a random sampling of conformations, the time required for a small protein to sample every possible conformation would be longer than the lifetime of the universe [63]. "Levinthal's Paradox" suggests that proteins must fold in a piecemeal and sequential manner [64-66], where small domains or individual secondary structural elements of the protein fold independently and then assemble together to form the final, native conformation.

One prediction of a stepwise folding process is the formation of intermediate conformations. Earlier denaturation studies, however, suggested that proteins fold and unfold in a two-state process: in the presence of a denaturant, a protein is either folded or unfolded, and it does not adopt intermediately-folded states [28,62,67]. Although this two-state assumption applied to many small proteins, biophysical studies of many proteins revealed intermediate conformations and multi-step folding pathways [65,67-74]. For the folding of larger proteins, the molten globule theory was suggested, where in the earliest stages of folding, the protein collapses into a loosely-associated globule-like conformation [75-77]. This collapse is partially compelled by the entropic drive of hydrophobic residues away from water and the enthalpically favorable van der Waals interactions that are made among those residues [75,78]. These energetically favorable changes offset the reduction of conformational flexibility (and therefore, entropy) in the protein upon early secondary structure formation. This compact intermediate has fewer

degrees of freedom in sampling the interactions between residues, which reduces the time required to sample and adopt the native structure [79]. With models suggesting globules and pathways in the folding of proteins, it became clear that most large proteins do not always fold through a defined set of intermediates.

Despite the consideration of this globule conformation as the first intermediate in a folding pathway, the "molten" aspect of this intermediate led to the drastic expansion of pathway models. The molten globule itself was not considered as one static structure, but rather as a collection of similar conformations [78]. Likewise, every intermediate in the folding pathway could represent a number of similar conformations. Due to the differences in these conformations, the available interactions that could be sampled in the progression towards other intermediate conformations could lead to a branching of the folding pathway [78,80,81]. The theory of protein folding, which was originally thought to occur through two-state transitions, had progressed to folding funnels and landscapes, where a myriad of conformations exist between the unfolded and native states [81-85]. While multiple pathways down these funnels can yield the native conformation of the protein, other intermediates can proceed into non-productive conformations. These non-native proteins not only lack the function of the native protein but can also have other effects on the cell.

Troubled Folding and Protein Folding Diseases

While thermodynamic forces drive protein folding, many challenges can occur that lead a protein to a non-native structure. Despite the prediction of Anfinsen's dogma

that the primary structure of a protein is sufficient for determining the structure of the native conformation, Anfinsen's hypothesis failed to account for the kinetic variables of protein folding. In the course of his studies, Anfinsen observed both incorrect folding (also referred to as misfolding) and slow folding of RNase A [55,57,58], complications he attributed to incorrect disulfide bond formation [62]. These misfolded proteins have adopted a non-native conformation in a low-energy state that is kinetically trapped, preventing the return to a previous conformation through the nearly insurmountable energy requirements of covalent bond rearrangement [58,59,61]. Although he was able to observe the fast, productive folding of proteins without disulfide bonds [68,86], the observation of non-native states with RNase A was an early demonstration that the lowest-energy state may not always be the fastest-forming conformation.

Unproductive folding and the formation of non-native products are not issues limited solely to RNase A or disulfide-containing proteins. Many proteins from many organisms have been shown to be unable, in whole or in part, to obtain their native conformations independently [30,87-90]. Included in this group are globally essential proteins, such as the carbon-fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [87,91], and proteins implicated in human disease, such as cystic fibrosis transmembrane conductance regulator (CFTR) [92-94]. Although only 20-30% of wild-type (unaltered) CFTR is correctly folded and trafficked inside the cell (with the remaining 70-80% being degraded) [92-94], the fraction of CFTR that is functional is sufficient for viability. Variations in the amino acids sequence (commonly referred to as mutations) in CFTR lead to a nearly complete failure of this protein to fold

and be further processed by the cell [94-96]. For example, in the most common disease-associated mutant of CFTR (ΔF508), the protein is trapped in an early folding intermediate that is subsequently degraded [94]. The loss of the CFTR protein leads to an imbalance in salt and water in tissues, and in the lungs, the result is a build-up of mucus that promotes bacterial and viral infections [97]. Such infections are commonly the cause of death for those bearing mutations in CFTR [97,98].

The inability to fold productively is at the core of not only cystic fibrosis but also many other diseases such as short-chain acyl-CoA dehydrogenase deficiency and phenylketonuria [99-101]. In many cases, however, non-native proteins interact with each other, forming non-native oligomers that are commonly referred to as aggregates. Aggregation is driven by compatible binding regions on the surfaces of non-native proteins; these surfaces could be comprised of entire folded domains or of exposed hydrophobic residues that are normally buried within the protein structure [102-107]. Aggregation can occur from any state in the folding process, including the unfolded state, intermediate conformations or even after a protein has acquired the native fold [105,108-110]. Not only does aggregation lead to a loss of the protein, but these aggregates can also interfere with other cellular functions, such as membrane stability [111]. Both the direct and indirect effects of aggregation can cause disease states [112-116]. Protein aggregation is the basis of many neurodegenerative diseases, including Alzheimer's, Parkinson's, and Creutzfeldt–Jakob disease [115,117-119].

Although Alzheimer's and cystic fibrosis are very different pathologically, both of these diseases, as well as many others, stem from a failure of proteins to reach and maintain a native state conformation.

GroEL: the Essential Folding Machine

Although protein folding diseases predominantly affect human tissues and organs, these complications originate from protein misfolding and aggregation inside individual cells. At the cellular level, preventing and repairing these two problems does not serve to prevent disease, but rather to recover proteins with the goal of restoring the proteins to their intended function or reducing them to individual amino acids to be used elsewhere. The process of preventing and repairing protein misfolding and aggregation is tasked to a set of proteins in the cell known as molecular chaperones [120-122]. These proteins have different structures, specificities, and capabilities, but all work together to maintain cellular proteostasis [120,123-127]. A central hub in the network of molecular chaperones is the Hsp60, a protein responsible for folding unfolded or refolding misfolded substrate proteins.

The first Hsp60 to be discovered and identified was the protein GroEL from *Escherichia coli*. In the pursuit of identifying important host interactions, phage biologists isolated *E. coli* colonies that were non-permissive for bacteriophage lambda due to mutations in genes other than *lamB*, which encodes the lambda receptor [128]. Identification of the *E. coli* gene responsible for this phenotype relied on a complementation experiment done by phage-mediated, specialized transduction. In the

case of the Hsp60, plaque-forming isolates contained the segment of the *E. coli* genome containing the *groE* gene, which was named for the requirement of this gene for lambda growth and for its role in the proper assembly of the major capsid protein (encoded in the phage gene *E*) [128-130]. Restriction mapping and further complementation experiments led to the isolation of *groE*. Through other phage-insensitive *E. coli* mutants and additional complementation experiments, the *groE* locus was found to contain two genes designated as *groEL* and *groES* (for large and small respectively) [130-132]. These proteins were subsequently identified to be essential for not only the propagation of other bacteriophage, including T4 and P22, but for the host *E. coli* as well [133-140].

Identified as host genes required for phage development, *groEL* and *groES* remained without a characterized function. Electron microscopy revealed that in bacteria carrying mutations in *groEL* and *groES*, misshapen phage particles were produced [128,132], suggesting a role in capsid assembly. However, the defect in phage morphogenesis was not identical across the different phage types, as some phages had blocked head production while others had defects in tail assembly [128,136-139]. Despite considerable research connecting the *groE* genes with bacteriophage, the role of GroEL and GroES was determined in other biological systems. Several important breakthroughs were made in examining the influence of GroEL on β-lactamase, a periplasmic protein that is used by bacteria to inactivate penicillin-like antibiotics. Research revealed that GroEL interacts with precursor β-lactamase (pre-Bla), an unfolded polypeptide in the cytoplasm that must be exported before maturation [141].

This was the first report of GroEL interacting with unfolded proteins. Furthermore, the interaction between the pre-Bla and GroEL was stable for extended periods of time, maintaining pre-Bla in a state competent for later export and maturation [141-143]. Addition of adenosine triphosphate (ATP) abolished the interaction between the two proteins, liberating pre-Bla for transport across the cytoplasmic membrane [144]. Beyond establishing a role for GroEL in binding unfolded proteins, the initial study with pre-Bla also advanced the characterization of GroEL. GroEL was shown by electron microscopy to form tetradecamers that adopt a ring-like structure [141]. This study with pre-Bla represented a major step forward in understanding both the structure and function of GroEL.

The observed interaction between pre-Bla and GroEL lead to further studies exploring the interaction between GroEL and other exported proteins. GroEL was found to interact with proOmpA and prePhoE, both of which are precursor forms of outer membrane proteins of *E. coli* [145]. Because GroEL could maintain exported proteins in a soluble form before being trafficked to the membrane, it was named a chaperone protein. While GroEL was found to bind various unfolded, non-native proteins, it was found to be incapable of binding soluble, folded proteins [145]. The gene product of the second component of *groE*, GroES, was found to form heptamers with no affinity for either folded or unfolded proteins [131,145,146]. The notion that GroEL and GroES work in concert to support the folding of proteins was first hypothesized after a study of plant RuBisCO assembly, because in chloroplasts, RuBisCO requires a GroEL homolog known as the RuBisCO-binding protein [91]. Further research demonstrated that GroEL

or its homologs can enhance the folding of a number of proteins, including bacterial RuBisCO, chicken DHFR, and bovine liver rhodanese, providing additional evidence for a general role for GroEL in folding proteins [87,147-151]. The advancement of biochemical techniques drove the chaperonin field, with new molecular biology methods allowing the creation of GroEL mutants and new biophysical methods allowing the examination of different aspects of protein folding. Although electron microscopy had previously revealed the crude structure of the chaperone, the crystal structures of GroEL, both apo and bound to GroES, provided key insights into the binding of ligands and the functioning of the chaperonin [152-154]. These structures allowed the creation of site-directed mutants of GroEL that could be used to determine specific details about its function.

With a greater understanding of the role and structure of GroEL, mechanistic details were becoming clearer. One component of GroEL function is the reaction cycle: the binding and release of substrate protein, ATP and GroES in an ordered manner, timed by the hydrolysis of nucleotide (Figure 1.1). *In vivo*, each GroEL tetradecamer is predominantly found in complex with one GroES heptamer, forming a closed cavity on one ring of GroEL [155-157]. This GroES-bound ring is referred to as the *cis* ring or *cis* cavity [154]. The entire complex is referred to as a GroEL-ES ATP bullet, named for the shape that the complex adopts and the fact that each GroEL monomer in the ring is bound to ATP. The open, or *trans*, ring of this bullet contains ADP, a product of the ATP hydrolysis from the previous reaction cycle [158-160]. This *trans* ring binds to a substrate protein, triggering the release of the ADP from the same ring [159]. With the

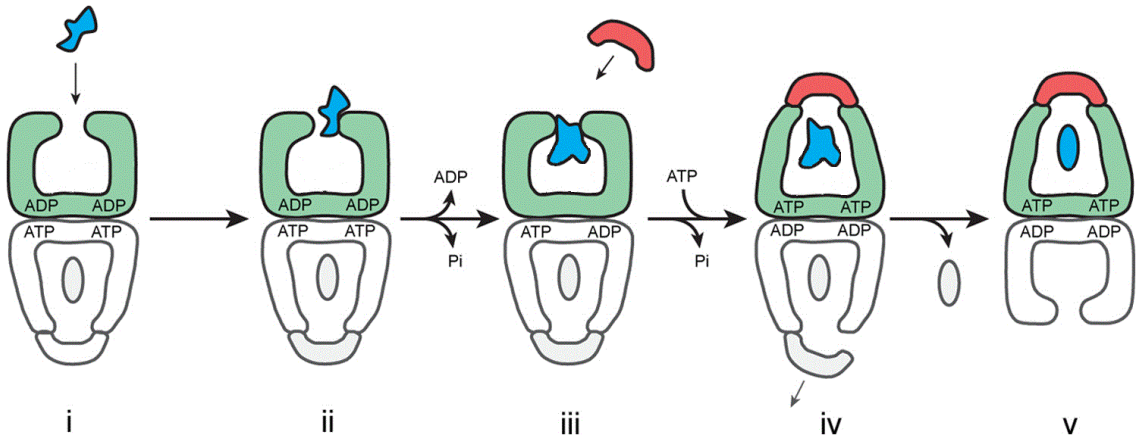


Figure 1.1: The GroEL Reaction Cycle. At the beginning of the reaction cycle, an ADP-bound *cis* GroEL ring (green ring, i) is unoccupied and can bind to a substrate protein. The binding of a substrate protein to an open *cis* ring leads to the dissociation of ADP from that ring (ii). At the same time, ATP in the *trans* ring (white ring) hydrolyses (ii and iii). Once the nucleotide binding sites in the *cis* ring are unoccupied, ATP binds (iii). ATP binding leads to the release of the substrate protein from the GroEL cavity wall. Immediately after ATP binding, the *cis* ring then binds GroES (iii). GroES binding creates a closed cavity in which the substrate protein is encapsulated (iv). ATP binding in the *cis* ring also leads to the dissociation of GroES and any encapsulated protein in the *trans* ring (iv). Inside the newly-formed GroEL-ES cavity, the substrate protein can fold (v). This complex (v) is identical to the starting complex (i), thus resetting the reaction cycle.

nucleotide binding sites open, ATP binds to this open ring, leading to major conformational changes in both rings due to the effects of both intra- and inter-ring allostery [154,159,161,162]. In the *cis* ring, the GroES dissociates, opening the cavity and facilitating the release of the enclosed substrate protein [163,164]. In the *trans* ring, large structural changes in the apical domains lead to several outcomes (Figure 1.2). First, the size of the cavity is expanded from ~80,000 $Å^3$ to ~175,000 $Å^3$ (Figure 1.2A and 1.2B) [152,154], which provides sufficient volume for a folding protein with a molecular weight of ~60 kDa. Second, the expansion of the cavity simultaneously expands the conformation of the bound substrate protein. This expansion is concomittal due to the strong interaction between specific hydrophobic residues of GroEL and the exposed hydrophobic residues of a substrate protein [165,166]. The physical extension of the bound substrate protein disrupts residual secondary and tertiary structure, leading to a more unfolded state [166-169]. Third, the binding sites on GroEL for the substrate protein and GroES are concurrently altered [154]. The region of the apical domains that was previously bound to the substrate protein has a changed orientation and position, diminishing the hydrophobicity of the surface (Figure 1.2C and 1.2D). Interactions between these GroEL residues and the substrate protein provide for the high-affinity binding [170], so upon the conformational change in GroEL, the chaperone's affinity for substrate proteins is greatly diminished. This transition yields the release of the bound substrate protein from the GroEL cavity wall [158,171,172]. At the same time, regions of the apical domain are reoriented to the surface of the protein, exposing the binding sites for GroES and allowing the cochaperone to form a closed cavity with GroEL

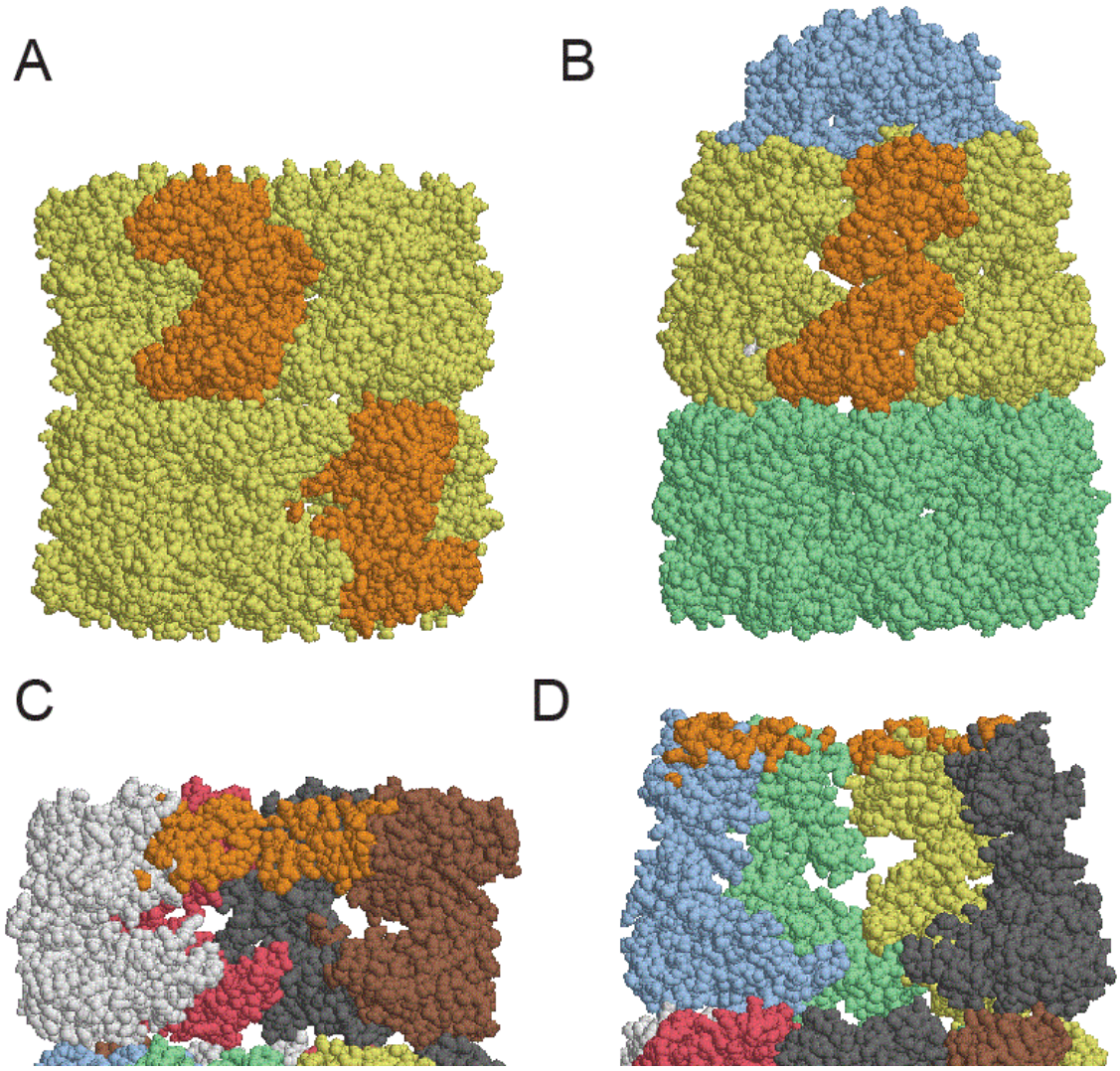


Figure 1.2: Structures of GroEL before and after ATP and GroES Binding. (A) The structure of the GroEL tetradecamer without ATP bound (PDB # 1SS8). One of the seven monomers of each ring is shown in orange. (B) The GroEL-GroES7 complex (PDB # 1AON) after nucleotide binding. The *cis* ring (yellow, with one orange monomer) has expanded in size relative to the apo ring (A) and can now bind GroES (blue). (C) The substrate protein binding sites (orange) account for a large surface of the apical domain of each monomer (different colors) in apo GroEL. (D) After nucleotide binding, the same residues (orange) that comprise the substrate binding site in (C) have translated upwards and outwards. These residues are now involved in binding GroES (GroES not shown).

[154,171,172]. Once this process is complete, GroEL has undergone a half cycle of the reaction. Since the *trans* ring is now bound to GroES, it has become the *cis* ring, and vice-versa. While a substrate protein folds in the new *cis* cavity, the new *trans* ring is competent to bind another substrate protein and continue the cycle.

Despite clarifying the complicated role of substrate binding behind the conformational changes of the chaperonin, these details do not explain the mechanism by which GroEL aids in the folding of substrate proteins. Early in the study of the chaperone reaction cycle, when protein folding was emerging as the likely role for GroEL, a debate arose concerning the release of substrate proteins into bulk solution during the folding process [127]. Although it would seem counterintuitive, some believed that unfolded proteins were released by the chaperonin into solution, where these proteins could then attempt to fold independently [164,173]. This theory became known as kinetic partitioning, because a substrate protein released from GroEL would fold productively with some probability and, depending on the outcome, would then partition into free solution or back to the chaperone. In this model, GroEL only served to unfold kinetically-trapped, non-native conformations of proteins, and the folding of the protein was entirely spontaneous [173-175]. Others contended that GroEL maintained contact with substrate proteins until productive folding was accomplished [155,176]. This concept emphasized the role of the GroEL-ES cavity in altering the energy landscape of a protein during the folding process. By reducing local energy minima, the chaperone could prevent or reverse substrate protein misfolding.

Although further work demonstrated that GroEL does release non-native proteins during the reaction cycle, the kinetic partitioning model was refined to account for the timing of the release of the substrate protein into solution [177,178]. After the encapsulation of a substrate protein inside the GroEL-ES cavity, a certain amount of time passes before the cavity is resolved and the substrate protein released [157,159,163]. There are two possible outcomes for the released substrate protein: it folded to a conformation that can only progress productively along the folding pathway, or it folded incorrectly or incompletely. In the first scenario, the substrate protein has reached a committed state in which the protein no longer associates with GroEL and from which correct folding can be accomplished independently. In the second scenario, the substrate protein remains in a conformation that is competent for GroEL-binding, allowing the substrate protein to partition back to the chaperone.

The development of the kinetic partitioning model required an alteration in the proposed mechanisms of chaperonin function. Besides acknowledging that substrate proteins are released into solution during the reaction cycle, the model that proposed a role for GroEL in altering how a protein folds remained mostly unchanged, stressing that interactions between the chaperone and the substrate protein prevent misfolding. This is known as the cavity confinement model, as the interactions necessary for promoting folding occur within the GroEL-ES cavity [179,180]. The original supporters of the kinetic partitioning model proposed two new theories. First, under the assumption that all proteins can fold independently, a model was developed asserting that the only role of the GroEL cavity is to block intermolecular interactions, particularly aggregation [181-

183]. This became known as the passive- or Anfinsen-cage model, so named for Christian Anfinsen, who originally postulated that the primary structure of a protein alone was sufficient determining the native conformation [62].

The second theory derived from kinetic partitioning maintained the significant role of substrate protein unfolding by the chaperonin [184-188]. In this model, the primary role of GroEL is to unfold kinetically-trapped conformations, resetting the substrate protein to a high-energy state from which the protein can fold. Increasing the energy in the substrate protein is accomplished by two GroEL-mediated processes: the annealing of the substrate protein to the GroEL apical domain and the expansion of the substrate protein conformation upon the binding of ATP to GroEL [166,168,185,187,188]. This model became known as the iterative annealing model, because substrate proteins undergo multiple rounds of binding and unfolding with the chaperone before productive folding is achieved. As the chaperonin field grew, the cavity confinement and iterative annealing models eventually fell under one umbrella, termed the active model. The active model attributes the ability of GroEL to stimulate folding to the changing of the energy landscape of the substrate protein during the folding process through specific interactions between the chaperone and the substrate protein [189,190].

Each of these three mechanisms attributes the ability of GroEL to assist folding to some component of the chaperone structure. The Anfinsen-cage model relies on the GroEL-ES cavity separating the substrate protein from the cellular environment; the cavity confinement model depends upon interactions between the substrate protein and

the GroEL cavity wall; and the iterative annealing model focuses on the expansion of the apical domain upon nucleotide binding. These distinctions may be clear, but the data that supports each of the models is not so distinct.

The primary argument for the Anfinsen-cage model is that many GroEL-dependent substrate proteins have a high propensity for aggregation [90,191-194]. Additionally, proteins are more prone to aggregation at higher temperatures, at times of cell stress, and when expressed at higher levels [195]. Many GroEL-dependent substrate proteins are highly abundant in the cell [89,90,196], so any perturbation in cellular homeostasis could result in a shift in the protein population towards an aggregated state. Although other chaperones, such as DnaK, can block aggregation, the clamp-like contact between those chaperones and a substrate protein can also inhibit productive folding during the lifetime of the interaction [197]. Because GroEL is capable of physically excluding a folding protein from the interactions of the cellular environment without maintaining a physical hold on the substrate protein [193,198], Ockham's razor suggests that GroEL exists to prevent this aggregation problem during protein folding. Besides this argument, biochemical data also supports the concept that GroEL functions as an Anfinsen cage. The folding rate of some proteins is not enhanced by the chaperone, indicating that GroEL cannot accelerate folding beyond the limit imposed by the primary structure [199-201]. The yield of folded protein is often a function of the concentration of the substrate protein in solution, with higher concentrations of protein leading to an increased dependence on the chaperone [149,195,201,202]. This concentration dependence indicates that the complication in the folding of these proteins is a

multimolecular process. Besides the assembly of native monomers into oligomers, any multimolecular process in protein folding would constitute aggregation, an unwanted, off-pathway product. In this passive model, the simple isolation of substrate proteins inside of the GroEL-ES cavity prevents the formation of aggregates and allows folding to occur as dictated by the primary sequence of the substrate protein.

Contrary to the notion that all proteins can fold independently, the active models define a role for GroEL in altering the folding landscape of the interacting substrate protein. The first active model, the cavity confinement model, is supported by several observations, the most important of which is that the folding rate of proteins can be stimulated by intra-cavity folding in the GroEL SR1 mutant [88,199,203-205]. The SR1 mutant is comprised of only a single, tetradecameric ring and forms complexes with GroES that are very long-lived [177]. In the caviy of this $GroEL_7$:$GroES_7$ complex, a substrate protein can fold in the absence of iterative annealing. Moreover, the folding rate of many proteins is the same in this non-cycling, single-ring mutant and in the cycling, wild-type chaperone [163,206]. These observations suggest that the unfolding action of the chaperone is not necessary for folding. The ability to fold in SR1, however, could easily be attributed to blocking intermolecular interactions that lead to aggregation [193]. To counter this notion, the folding of proteins that have no propensity for aggregation has been examined and found to be stimulated as well [88,205]. Beyond enhancing the ability to fold, the chaperone can induce changes in the folding pathway. A common observation that supports the cavity confinement model is that proteins undergo a compaction event inside the cavity [166,168], which is generally attributed to

a hydrophobic collapse that is thought to be common to the folding of most proteins [77,81]. The compaction of a substrate protein is taken as the result of unfavorable interactions between the charged residues on the surface of the GroEL cavity wall and the hydrophobic residues that often comprise the core of a folded substrate protein (Figure 1.3) [179,199,207,208]. Supporting this theory, alteration of these charged residues on the inner surface of the GroEL cavity reduces the stimulation of substrate protein folding [203,204]. The folding of some proteins may not be driven by a hydrophobic collapse [80,81], so if GroEL promotes the compaction of a substrate protein, it may affect some components of protein structure or folding more than others. This unequal stimulation was observed in a recent study that showed the folding of a core component of the protein structure to be vastly accelerated, while the folding of other components was not as stimulated [88]. Together, biochemical and biophysical data suggests a role for confining substrate proteins inside the GroEL-ES cavity as a means of driving protein folding.

Many of the observations used to support the cavity confinement model can also be used to support the iterative annealing model. Although folding in SR1 and wild-type GroEL is often reported as occurring at the same rate, these experiments were done at (or near) 25˚C [159,163,192,199,203,204,209]. A key prediction of the iterative annealing model is that any change that enhances the rate or magnitude of substrate protein unfolding results in an increased stimulation of folding [188]. Until structural stability is compromised, the activity of enzymes increases as a function of temperature [210]. Although the folding action of GroEL is not a chemical reaction, the ATPase

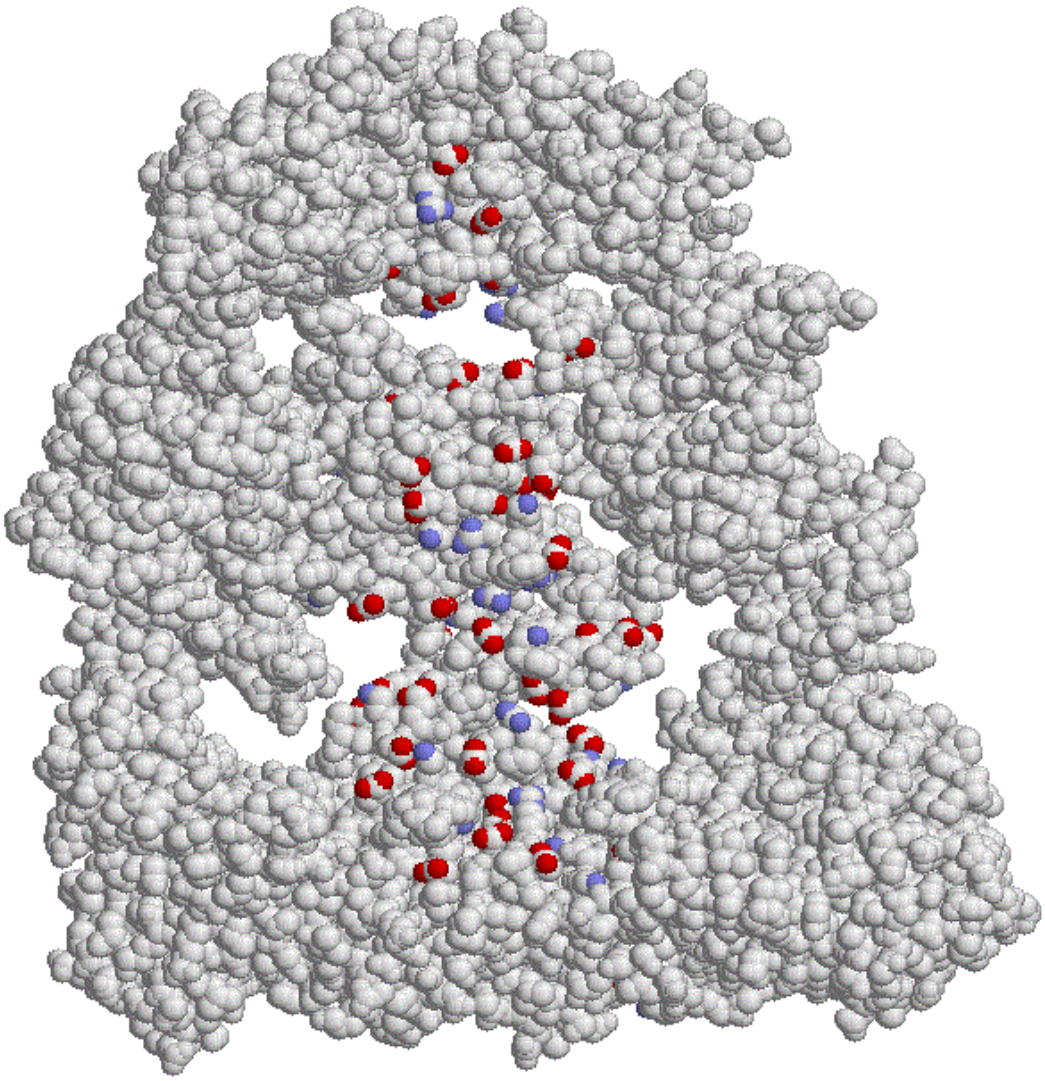


Figure 1.3: Charged Residues on the Surface of the GroEL-ES Cavity. The presence of hydrophilic residues, and specifically charged residues, is important for the cavity confinement model of GroEL function. The oxygen atoms of the sidechains of aspartic acids and glutamic acids are shown in red. The nitrogen atoms of the sidechains of lysines and arginines are shown in blue. These residues are only colored for one monomer of GroEL and one monomer of GroES.

activity of the chaperone qualifies. As the cycling rate of GroEL is dependent on the ATPase rate, increasing the temperature would result in an increase in cavity turnover. This prediction was confirmed experimentally: an increase in temperature resulted in both an increased turnover rate and an increased folding rate [159]. An increase in temperature had no effect on folding in the SR1 mutant [159], however, indicating that the stimulation of folding by cavity confinement may be limited by some other variable. This variable may be the binding-driven unfolding of substrate proteins that occurs on the GroEL apical surface before encapsulation within the GroEL-ES cavity. After a substrate protein is bound on the apical domains of a *trans* ring before encapsulation, it adopts a more unfolded conformation over time [166]. When this substrate protein is eventually encapsulated, the fraction that folds immediately (within the first ten seconds) increases as a function of the pre-encapsulation incubation time [166]. Therefore, increasing the extent of unfolding prior to encapsulation increases the ability of substrate proteins to fold on the physiologically-relevant timescale of the cycling chaperone. Despite the importance of iterative annealing in GroEL-mediated folding, this stimulation requires the presence of GroES [150,173,176,211,212]. This suggests that the benefits of unfolding exist only when coupled with confinement inside the cavity, indicating the requirement of at least two of the proposed mechanisms for the enhancement of folding.

A small selection of proteins, chosen for their physiological relevance and availability, was utilized in the testing of these models, although few were proteins with which GroEL naturally interacts in the *E. coli* cytoplasm. To investigate the magnitude

of chaperone interactions inside the cell, a proteomics survey was undertaken [89]. By trapping substrate proteins inside the GroEL-ES cavity and purifying these complexes, it was possible to identify ~250 *E. coli* proteins that interact with the chaperonin. Acquiring data on the entire proteome allowed for a comparison of proteins that interact with the chaperone, leading to the categorization of proteins into three classes. Class I proteins bind GroEL but do not need the chaperone to fold. Class II proteins do not have an absolute requirement for GroEL but greatly benefit from the interaction. Class III proteins are incapable of folding without the assistance of the chaperone. This third group consists of ~85 proteins, 13 of which are essential for cell viability.

This proteomics study in the lab of Ulrich Hartl [89], and a second, independent proteomic survey [90], demonstrated the broad specificity of the chaperone and explained that the necessity of GroEL for cell viability stems from the essential roles of some of its substrate proteins. It was found that although GroEL interacts with ~250 proteins, the ~85 class III proteins populate the chaperone ~80% of the time. The explanation for this discrepancy may be found in the folding of an obligate substrate protein, RuBisCO, which does not come from *E. coli* but rather from the bacterium *Rhodospirillum rubrum*. This RuBisCO, like all RuBisCO homologs from bacteria to plants, requires GroEL to fold [87,91,148], but the fraction that is folded after a single turnover in the reaction cycle is not 100% [87,148,163,192,209]. Rather, RuBisCO requires multiple rounds of folding with the chaperone machinery to reach a maximal yield [159,163,192]. If this is true of many obligate substrate proteins, then the proteins

that require GroEL can only rebind to the chaperone after a failed folding attempt, whereas non-obligate substrate proteins are capable of folding independently.

Despite the importance of GroEL in *E. coli*, in phage growth, and in the mitochondria and chloroplasts of eukaryotes, GroEL is not universally conserved. In a small number of genera of the bacterial family Mycoplasmataceae, a GroEL homolog is not found in the genome [213]. Importantly, although these organisms maintain smaller genomes, they also maintain many of the essential proteins that are obligate GroEL substrate proteins in *E. coli* [214,215]. There are two explanations for this phenomenon. First, in organisms such as *Ureaplasma urealyticum*, the concentration of the osmolyte urea under normal growth conditions is comparatively very high [215,216], and this osmolyte has been shown to influence the thermodynamics of protein folding [217-219]. The second group of organisms is the bacteria in the *Mycoplasma* genus. These bacteria are minimalist parasites that survive by importing every possible nutrient from the host organism [220]. A component of this minimalist nature is a reduced genome consisting of ~500 kbp [214]. While a decreased production of protein can reduce the stress on the folding proteome [221,222], other changes may have occurred in the evolution of these bacteria.

The *Mycoplasma* homolog of DapA, which is GroEL-dependent in *E. coli*, can fold spontaneously under conditions incompatible with *E. coli* DapA folding [88]. Examination of the folding of *E. coli* DapA, both spontaneously and when chaperone-mediated, revealed that the folding of specific segments of DapA was accelerated when folding with GroEL [88]. In the spontaneous folding of the *Mycoplasma* homolog, the

folding of these same segments was also accelerated. This suggests that the evolution of this *Mycoplasma* protein has altered its folding pathway to adopt changes similar to those effected by the chaperone. The similarities in the chaperone-mediated folding of *E. coli* DapA and the spontaneous folding of *Mycoplasma* demonstrate that the utilization of alternative intermediate conformations is an effective strategy for the improvement of folding, regardless of whether this change in folding pathway is the effect of mutation or chaperone involvement.

The Specialization of GroEL Function

Due to the requirement of GroEL to aid the folding of hundreds of different proteins, the chaperonin has maintained broad specificity over billions of years of evolution. However, this specificity is not cast in stone and can be adjusted by mutations in either GroEL or GroES. For example, the ability of GroEL to propagate phage can be abolished by single nucleotide polymorphisms without abolishing the chaperone's capability to fold essential bacterial proteins [223]. Alternatively, GroEL can be mutated to specifically enhance the folding of one protein [224]. This enhancement, however, creates deficiencies in the chaperone-mediated folding of other proteins. Throughout evolution, GroEL has likely mutated to strike the correct balance among all of its substrate proteins to assure the greatest benefit for cellular fitness.

In some organisms, gene duplication events have led to multiple copies of the chaperonin gene in the genome. The number of *groEL* genes in a single organism has been observed to be as high as eight [213]. In these cases, one copy of the gene is

maintained with broad specificity for the housekeeping function of folding many proteins, and other copies of the gene acquire mutations for a specialized function. This function is often the folding of one or more specific proteins that are essential for a specific cellular behavior or function [213]. Despite having conserved sequences and maintaining the structure of the housekeeping chaperonin, these specialized variants of groEL are often incapable of complementing the housekeeping gene of either the original organism or of *E. coli* [213,225,226]. The benefits of specifically enhancing the folding of a small subset of substrate proteins must outweigh the energetic costs of producing multiple, similar chaperones.

The utilization of multiple chaperone components is not limited to bacteria, as the genomes of a number of bacteriophage are known to include similar genes. Perhaps the most studied example of this is bacteriophage T4, a phage that infects *E. coli* and carries a gene encoding a GroES homolog [227]. A major capsid protein of T4 (gp23) requires the assistance of the bacterial GroEL and the T4-encoded GroES homolog (gp31) for productive folding [134,136,227]. The bacterial GroES is not capable of complementing gp31 for the folding of gp23; however, gp31 can complement GroES for the folding of the *E. coli* proteome [227,228]. In this case, the alteration of the co-chaperone gene required for altered function did not drastically alter the capability or specificity of the chaperone machinery. This is likely due to the presence of the substrate protein binding sites on GroEL and the lack of such sites on GroES.

While bacteriophage T4 carries a genes for a GroES homolog, several phage are known or predicted to carry genes for homologs of GroEL in their genomes [229-233].

One such virus is bacteriophage EL, which infects *Pseudomonas aeruginosa*, a Gram-negative bacterium that is in the same phylogenetic class as *E. coli* [233]. Gene product 146 (gp146) of bacteriophage EL is a homolog of GroEL [233,234]. Despite low sequence similarity with the bacterial chaperonin (Figure 1.4**)**, gp146 forms ring-shaped tetradecamers. This phage protein was demonstrated to specifically interact with another phage protein, gp188, which is the monomeric endolysin required for lysis of the host cell [234]. No other proteins were reported to interact with gp146, indicating a strong specificity in substrate protein binding.

Production of gp146 begins ~15 minutes after the initial infection of the cell; production of gp188 begins ~60 minutes after infection [234]. This timing supports the concept that gp146 is required for the productive folding of gp146. Unfortunately, the folding of this endolysin, either spontaneously or in the presence of a chaperone system, has not been examined. However, at elevated temperatures, gp188 has been shown to aggregate, and this aggregation is mitigated by the presence of the bacteriophage GroEL homolog [234,235]. In the absence of chaperonin, a decrease in endolysin activity is also observed over time at 37˚C, the optimal temperature for *P. aeruginosa* growth and replication [236].

The maintenance of endolysin by the chaperonin required the presence of ATP, but not of any co-chaperone. This suggests several possible roles for gp146 in regards to gp188. First, the native conformation of endolysin may be susceptible to converting to a misfolded conformation that is aggregation prone, and gp146 serves to unfold this substrate protein to allow for spontaneous refolding. Another possibility is that the

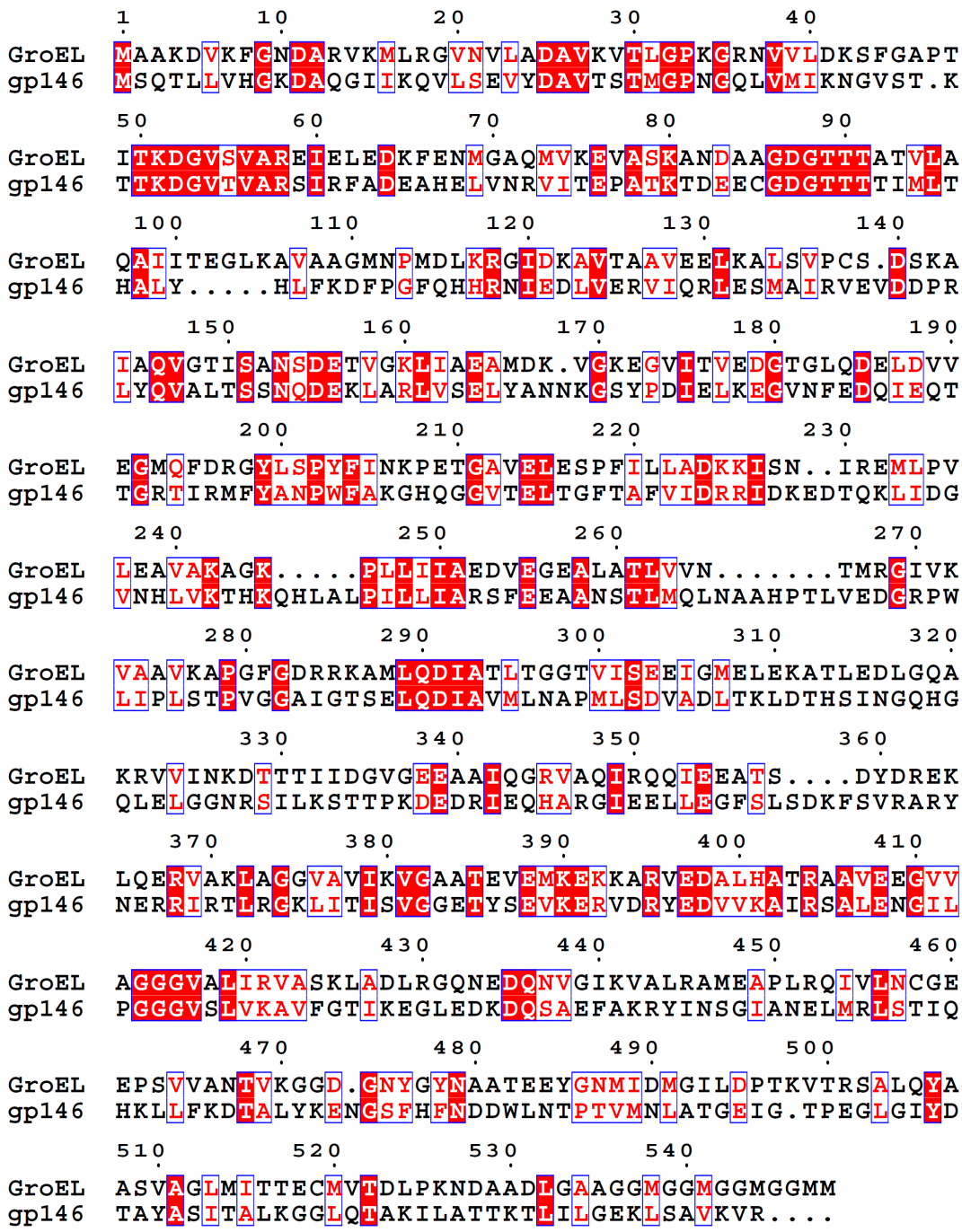


Figure 1.4: Sequence Alignment of *E. coli* GroEL and *Pseudomonas* Phage EL gp146. Alignment of GroEL (accession number P0A6F5) and gp146 (Q2Z0T5) was performed using CLUSTALW [237] and graphically organized with ESPript [238]. Completely conserved residues are highlighted in red and highly conserved residues or regions are boxed and shown in red text. Numbering shown is for *E. coli* GroEL.

chaperonin functions as a reservoir, holding the protein in a pro-functional state. This may allow the effective concentration of endolysin to surpass its critical concentration for aggregation, which would increase the subsequent rate of peptidoglycan degradation during host cell lysis. ATP would be required for this function to partition the substrate protein off of the chaperonin ring. Differentiating these models or defining other models will require more biochemical and biophysical data. Many of the traditional experiments used in the examination of GroEL substrate proteins have not yet been done, leaving open questions concerning the role of the GroEL-GroES cavity in the folding or prevention of aggregation of gp188.

Although much remains unknown about how gp146 can aid in the folding or maintenance of gp188, or about the interactions that lead to the altered specificity of gp146, mutations in the chaperonin have changed the array of interactions that gp146 can make with substrate proteins, allowing for a new, highly specialized function. Although the host chaperonin could facilitate both roles in unfolding and in increasing the allowable concentration of endolysin in the cell, the availability of a specific chaperonin would provides a key advantage. During the phage infection, the host GroEL would still be functioning in the folding of host proteins. Competing against the host proteome would lead to reduced accessibility to the host chaperonin for gp188. Depending on the mechanism behind the aggregation of gp188, either the bulk concentration or the concentration of a potentially misfolded sate would be increased in this scenario, leading to the loss of functional protein through aggregation.

Another challenge posed by utilizing the host machinery is that gp188 would displace the proteins that would otherwise be interacting with the host GroEL. These proteins would partition onto other chaperones, leading many proteins to degradation pathways [123,239,240]. The stress response induced to cope with widespread misfolding and aggregation would likely temper the production of phage proteins. The production of a specific chaperonin for the maintenance of the endolysin monomers would alleviate both challenges associated with utilizing the host machinery. The enhanced specificity for gp188 would also reduce interactions with bacterial proteins, which would reduce the amount of gp146 required to associate with gp188, decreasing the amount of resources needed for this process.

Specific interactions between the chaperone and substrate proteins have been shown to be required for the productive folding and assembly of certain proteins in organisms with multiple chaperonins [213,225,241-243]. This same strategy is being utilized by pseudomonas phage EL for the production of a lysis protein. Although gp146 has a role in preventing gp188 aggregation, it is also possible that the phage chaperone also has a role in the folding of the endolysin. Much is still unknown about gp146; however, the examination of its sequence and preliminary biochemical finding suggest that this chaperonin evolved for a specific function, and this function was important enough to be retained through billions of years of phage evolution.

The Importance of the GroEL C-terminus

Considerable work has been done in the study of the various domains of GroEL, but the last 23 amino acids of the chaperonin escaped much of this scrutiny devoted to other aspects of GroEL function [244]. These residues, K-N-D-A-A-D-L-G-A-A-G-G-M-G-G-M-G-G-M-G-G-M-M, follow proline-525, which is the last visualized residue of GroEL in every crystal structure of the chaperone [152-154,227,245]. Lacking a resolvable structure and containing nine glycines, these residues were believed to adopt a disordered conformation. Removal of the C-terminus *in vivo* by mutation of the *E. coli* genome was not lethal for the bacteria under various common growth conditions, showing that these residues are not essential for chaperone function [246-248]. Although the studies using genomic mutations of GroEL did not investigate the bacterial genome for second-site suppressors, the conclusion has been widely accepted [198,244,249]. Without a structure or any observed interactions, the role of the C-terminal tail was examined in the context of non-specific functions, including the occupation of space inside the cavity and altering the timing of the GroEL reaction cycle.

One potential impact of the GroEL C-terminus that was investigated was a role in establishing the available volume of the GroEL-ES cavity. The charged residues in the first six of the 23 C-terminal amino acids creates a repulsion that prevents these appendages from coalescing [250], causing these residues to extend upwards from the bottom of the cavity. Because each GroEL ring contains seven monomers, a total of 161 residues with a combined molecular weight of 14.4 kDa occupy the bottom of the cavity.

Increasing or decreasing the length of this C-terminus would alter the available volume in the cavity for a substrate protein to fold. If the evolution of the chaperonin selected for this length for the C-terminus, then one would predict that an increase or decrease in the number of residues after proline-525 should be detrimental to folding. This theory was tested through the use of various C-terminal truncation and extension mutants [203,204]. These studies showed that none of the alterations in the length of the C-terminal tail were beneficial for the folding of several substrate proteins. Although this result suggests that the length of the GroEL C-terminus was under selective pressure during the evolution of the chaperonin, it could not exclude roles for the C-terminal tail other than cavity volume adjustment.

A second role for the GroEL C-terminus was examined concurrently with the cavity volume role. Altering the length of the C-terminus impacted the rate of ATP hydrolysis by the chaperone, which drives the turnover of this molecular machine [157,163]. Increasing the length of the C-terminal tail increases the ATP hydrolysis rate, while a truncation decreases this rate [209,244]. The focus of these studies was the effect of these changes on the lifetime of the GroEL-ES cavity. As in the other studies, detrimental effects were observed with both extensions and truncations of the C-terminal residues [209]. In the context of complex turnover, the conclusion was that decreasing the lifetime of the cavity provided less opportunity for the substrate proteins to fold, therefore decreasing the folding rate. However, these authors provided no explanation for why the folding rate was decreased in the absence of the C-terminal tail.

The results of these studies were fundamentally the same, though each was interpreted in the context of a different model for GroEL function. The results of both studies were complicated by other impacts of the C-termini, including the roles described by the competing studies [203,204,209], the decrease in substrate protein encapsulation [209], the possible aggregation of the C-terminal residues [250], and other aspects of GroEL function that were not considered. An important result of both studies was that alterations of the GroEL C-terminus did not impact every substrate protein equally. The GroEL cavity evolved under the selective pressure of aiding multiple essential proteins [89,90,196], and changes in the cavity that benefit some proteins are generally detrimental to other proteins [224,251]. The same is likely true of the C-terminal tail, such that the role of these residues strikes a balance for the optimal effect on the entire proteome.

Although research has primarily focused on the *E. coli* chaperonin, the composition and length of the C-terminal tail of the mitochondrial chaperonins has remained mostly unchanged from the C-terminus of *E. coli* GroEL (Figure 1.5) [252-260]. For example, the *E. coli* C-terminal tail contains four copies of a Gly-Gly-Met motif, while the human mitochondrial Hsp60 ends with three repeats of a Gly-Gly-Gly-Met motif [253,255,256]. The conservation is striking, given that the proteomes of an *E. coli* cell and human mitochondria are of significantly different size and composition [255,256,261,262]. The conservation from prokaryotes to eukaryotes implies an importance for the mild hydrophobicity and flexibility of these residues. These attributes would not be necessary if the role of these residues was simply to occupy

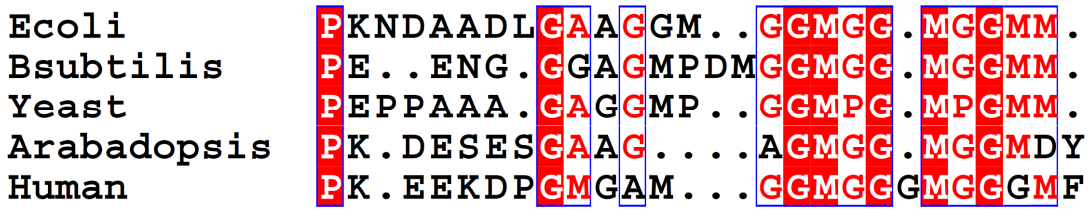


Figure 1.5: Sequence Alignment of Conserved C-Terminal Sequences in GroEL Homologs. Alignment of the *E. coli* GroEL (top, accession number P0A6F5) C-terminal residues (starting with proline-525) and the C-terminal sequences of GroEL homologs was performed using CLUSTALW [237] and graphically organized with ESPript [238]. Completely conserved residues are highlighted in red and highly conserved residues or regions are boxed and shown in red text. From the top, sequences shown are from: *E. coli* GroEL (accession number P0A6F5), *Bacillus subtilis* GroEL (P28598), *Saccharomyces cerevisiae* mitochondrial Hsp60 (P19882), *Arabidopsis thaliana* mitochondrial Cpn60 (P29197), and *Homo sapiens* mitochondrial Hsp60 (P10809).

space or to modulate the rate of ATP hydrolysis by the chaperone. This conservation suggests that there are other roles for these residues.

One possible role for the C-terminal tail of GroEL is suggested by the similarity of the C-terminal sequence with sequences found in other chaperones. Hsp40s are chaperone proteins that bind to non-native proteins and function in concert with Hsp70s to perform a number of roles inside the cell, such as protein disaggregation, complex disassembly, and protein folding [124,197,263,264]. In Sis1, an Hsp40 found in *S. cerevisiae*, the linker region between the two domains of the protein has been implicated in having some role in responding to substrate protein binding [265]. This region of the protein contains multiple repeats of the Gly-Gly-Met motif [266,267], and has been identified in other Hsp40s [264,265,268-270]. Although not identical to the Gly-Gly-Met motif, the same region in another *S. cerevisiae* Hsp40, Ydj1, contains Gly-Gly-X repeats, with X being alanine or phenylalanine [267,271]. The *E. coli* Hsp40, DnaJ, also contains multiple Gly-Gly-X repeats, with X being methionine or phenylalanine [255,272]. Not only do many Hsp40s contain this motif, but it is also found in many Hsp70s [246]. The presence of this motif in chaperone proteins that are responsible for binding to non-native proteins suggests a similar role in substrate protein binding for the C-terminal tail of GroEL.

The Specialization of the GroEL C-terminus

The C-terminal tail of *E. coli* GroEL may have evolved for broad specificity, but the same is not true of the C-terminus of the chaperonin in every bacteria. Duplication

events in the genomes of some bacteria have created second copies of *groEL* in those organisms [213,273,274]. With the original GroEL providing the essential function of broad-spectrum protein folding, the additional genes have not faced strict selective pressure for viability [275]. This has led to mutations arising in the additional chaperonin genes, specifically in the region encoding the C-terminal tail, in organisms such as actinobacteria, cyanobacteria, and α-proteobacteria (Figure 1.6) [213,226,276-278]. These specialized chaperonins have developed altered specificities in support of particular functions in these bacteria.

Although many actinobacteria have been identified with mutations in the C-terminal region of the chaperonin, the most well-known genus of actinobacteria is *Mycobacterium*. *Mycobacterium* includes several pathogenic organisms [276,279], the most well-known of which is *Mycobacterium tuberculosis*. Mycobacteria, including *M. tuberculosis*, contain at least one additional copy of *groEL* [213]. This gene, *cpn60.2*, is located distant from the *cpn60.1* operon in the genome and does not have an adjacent copy of *groES* [280]. Interestingly, *cpn60.2* is the essential gene and is capable of complementing Δ*groEL* in *E. coli*, while *cpn60.1* is not essential and cannot complement *E. coli* Δ*groEL* [225,275,281]. Although both Cpn60.1 and Cpn60.2 form tetradecamers, bind nucleotide, bind GroES and are capable of folding proteins [213,225,273,278,281,282], two significant differences exist in the sequences of these paralogs. First, the sequence of the Cpn60.1 N-terminus, a region of the chaperonin that is involved in oligomerization, is highly altered [280,283]. This difference causes the chaperone to be less stable, requiring the presence of nucleotide or high concentrations

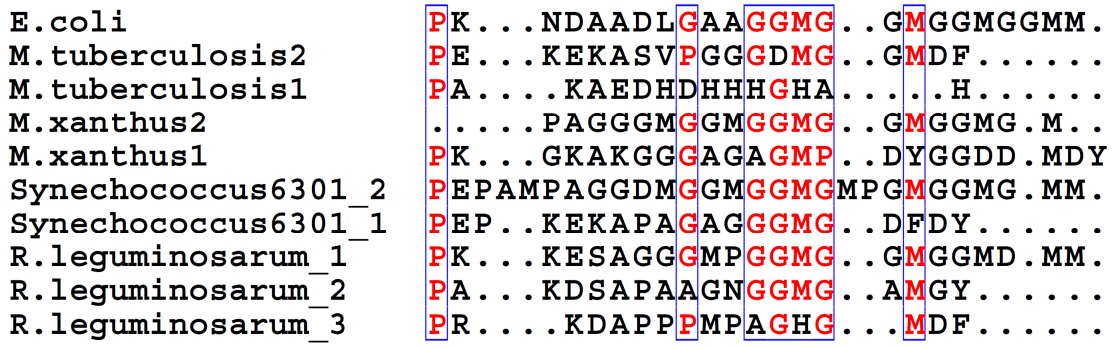


Figure 1.6: Sequence Alignment of Divergent C-Terminal Sequences in GroEL Homologs. Alignment of the *E. coli* GroEL (top, accession number P0A6F5) C-terminal residues (starting with proline-525) and the C-terminal sequences of GroEL homologs was performed using CLUSTALW [237] and graphically organized with ESPript [238]. Completely conserved residues are highlighted in red and highly conserved residues or regions are boxed and shown in red text. From the top, sequences shown are from: *E. coli* GroEL (accession number P0A6F5), *Mycobacterium tuberculosis* Cpn60.2 (P9WPE7), *M. tuberculosis* Cpn60.1 (P9WPE9), *Myxococcus Xanthus* Cpn60.2 (Q1D2S1), *M. Xanthus* Cpn60.1 (Q1D3Y5), *Synechococcus sp*. (PCC 6301) GroEL2 (Q5N3T6), *Synechococcus sp*. (PCC 6301) GroEL1 (P12834), *Rhizobium leguminosarum* GroEL1 (Q1MKX4), *R. leguminosarum* GroEL2 (Q1MJF2), and *R. leguminosarum* GroEL3 (Q1M3H2).

of salt to remain oligomerized [278,283,284]. Although this factor may influence the initial formation of the tetradecamer, most models of the GroEL reaction cycle show the chaperone as constantly bound to nucleotide [157,163,206], which may solve this stability problem. Importantly, while the reaction cycle has been studied extensively using *E. coli* GroEL, it has not yet been demonstrated that the mycobacterial protein utilizes the same cycle.

The second region of the mycobacterial Cpn60.1 that differs from Cpn60.2 or *E. coli* GroEL is the C-terminal tail [225,280,283]. After the conserved proline residue in *M. tuberculosis*, the sequence of the Cpn60.1 is only 14 amino acids long (A-K-A-E-D-H-D-H-H-H-G-H-A-H). Although this sequence retains the presence of charged residues distal to the proline residue, the subsequent Gly-Gly-Met repeats have been replaced with a shorter, less flexible, more polar sequence. The presence of multiple histidines in the C-terminal tail of GroEL paralogs is unique to actinobacteria [225]. The role of this region of Cpn60.1 was revealed by its role in phage development. *Mycobacterium smegmatis* can be infected by the mycobacteriophage Bxb1, which integrates into the bacterial genome at the 3'-end of *cpn60.1* [285]. This in-frame insertion of phage DNA changes the sequence and shortens the C-terminal tail of Cpn60.1 [285]. This mutation in *cpn60.1*, or the deletion of the entire gene, has no effect on normal growth of the bacteria, but it does render the bacteria incapable of forming biofilms [225,285]. This effect is due to the role of Cpn60.1 in the maturation of two proteins, KasA and KasB [225]. These two proteins function in the synthesis of mycolic acid, a critical component of the mycobacterial cell wall during biofilm

maturation [286]. Although Cpn60.1 physically interacts with KasA utilizing the same binding surface as classical chaperonins [225,282], no evidence yet exists to support a role for Cpn60.1 in folding KasA. Whether Cpn60.1 stimulates the folding of KasA or performs some other function, such as trafficking KasA inside the cell, the unique C-terminal sequence of this paralog is necessary for both the maturation of this protein and the maturation of the bacterial biofilm.

In Gram-negative bacteria, one example of an organism containing multiple, differentiated chaperonin genes is *Myxococcus xanthus*. The genome of *M. xanthus* is ~9.1 Mbp and contains many duplicated genes, including two copies of *groEL* [287,288]. As in the *M. tuberculosis* Cpn60s, the C-terminal tails of the GroEL paralogs of *M. xanthus* are starkly different. While GroEL2 retains multiple repeats of the Gly-Gly-Met motif, the glycine-rich C-terminus of GroEL1 includes four aspartic acid residues in the final nine positions of the protein [287]. Unlike the mycobacterial gene duplication, the differentiation of GroEL1 and GroEL2 in *M. xanthus* did not lead to a loss of the house-keeping capability for either chaperone, since either chaperone is independently sufficient for cell viability [241]. While maintaining this house-keeping role, both chaperonins also gained specialized functions. It was reported that these chaperonins are differentially expressed under normal growth conditions and differentially induced by heat-shock [288]. Although the basal abundance of GroEL1 is roughly 5-10 fold lower than GroEL2, GroEL1 expression was observed to quickly surpasses GroEL2 upon heat-shock and achieves a higher maximal expression level. Although GroEL2 was up-regulated approximately fivefold under heat-shock [288], only

GroEL1 was capable of maintaining cell viability at higher temperatures [241,242]. Maintaining the cellular proteome under stress conditions is a role for GroEL in most organisms [121,289,290], but this differential regulation suggested that these proteins may serve different functions.

Myxobacteria are characterized by a number of features that are uncommon in other bacteria, such as colonial predation, intercellular spore formation processes, and the presence of eukaryotic-like gene regulation [291-293]. In these functions a differentiated role for *M. xanthus* GroEL1 and GroEL2 can be found. Although *M. xanthus* can survive on normal laboratory growth medium, these bacteria can also utilize other bacteria (such as *E. coli*) as a food source [291]. The ability of these myxobacteria to swarm and prey upon other microbes is strongly dependent on the presence of GroEL2 [241]. Conversely, in times of starvation, these bacteria form multicellular fruiting-bodies that develop into spores [291]; these processes in development are severely deficient in the absence of GroEL1 [241]. As chaperone proteins, neither GroEL1 nor GroEL2 would have a direct function in either predation or development, but substrate proteins that require the assistance of these chaperonins could have roles in these processes. An investigation of the interaction proteomes of GroEL1 and GroEL2 identified several dozen substrate proteins that interact specifically with one chaperonin or the other [242].

Although many proteins in the *M. xanthus* proteome remain uncharacterized, several putative substrate proteins, annotated with functions in polyketide or non-ribosomal peptide synthesis, were identified as specifically using either GroEL1 or

GroEL2 [242]. Bacteria use polyketide and other non-ribosomal mechanisms for the biosynthesis of small molecules involved in the development and the defense of the manufacturing organism [294-296]. In *M. xanthus* predation, the secondary metabolite myxovirescin serves as an antibiotic to kill potential food sources [297]; the production of this compound requires the GroEL2 gene [243]. In examining the role of the GroEL paralogs, not only for the production of myxovirescin, but also for other aspects of predation and for the development of fruiting-bodies and spores, the C-termini of the chaperonins were interrogated by site-directed mutagenesis [242,243]. For both GroEL1 and GroEL2, removal of the C-terminal sequence or exchanging the sequences of the two paralogs results in a loss of the observed phenotypes associated with each chaperonin [242,243]. The deletion of the C-teminus from either chaperonin could not be complemented by a variant of the other chaperonin featuring a mutated C-terminus that matched the deleted sequence. This suggests that these C-terminal tails function along with other components of the GroEL structure in the folding of these specific substrate proteins. However, the details of this mechanism and the details about the interactions between the C-termini and the substrate proteins that promote their folding remain unresolved.

The development of functional specialization for additional chaperonins has also been identified in the chloroplasts of some plants [298-302]. Two different types of subunits (alpha (α) and beta (β)) comprise the chloroplastic chaperonin of *Arabidopsis thaliana*, and each subunit is encoded in multiple genes [303,304]. Each of these genes is expressed at different levels, but not all genes are required for viability [305,306].

These subunits form mixed tetradecamers that are generally composed of seven alpha and seven beta subunits [303]. Three of the four beta subunits expressed in *A. thaliana* are ~90% identical with each other; however, Cpn60β4 is only ~60% identical to the other three subunits [304]. Deletion of the Cpn60β4 subunit does not lead to a loss of plant viability; the only observed effect was a loss of NADH dehydrogenase-like complex (NDH) activity [307]. Although the NDH protein was found to bind to Cpn60 tetradecamers that lacked the β4 subunit, tetradecamers lacking this subunit could not compensate for the loss of the Cpn60β4 subunit [307]. The Cpn60β4 subunit shares only ~60% identity with the other beta subunits; the C-termini of these subunits are notably different. The C-terminal tail of the Cpn60β4 subunit is more positively charged and at least 26 amino acids longer than the C-termini of the other Cpn60β subunits in *A. thaliana* [307]. Extension or deletion of this β4 C-terminal tail drastically reduces NDH activity, although not to the same extent as a deletion of the entire subunit [307]. This suggests that the C-terminus has a critical, but not essential, role in the folding of NDH by Cpn60β4. Despite the importance of this subunit, no research has been conducted into the actual mechanism used by Cpn60β4 to support the folding of NDH.

Even after multiple studies investigating the role of the C-terminus, its role in chaperonin function has not been definitely established. In general, the GroEL C-terminus is highly conserved in many organisms. However, when this sequence has changed, the new GroEL paralog supports a secondary cellular function. Sequence similarity of the GroEL C-terminus to regions of the Hsp40 and Hsp70 proteins suggests a role in substrate protein binding; this concept is supported by the observed specificities

of the diverged C-terminal sequences of the myxobacterial chaperonin paralogs. Whether this potential binding activity is required for the initial capture of the substrate protein on the apical domain of GroEL, for the subsequent encapsulation of the substrate protein, or for intracavity folding remains unknown. Only through extensive, focused biochemical and biophysical characterization will an understanding of the mechanistic role of the GroEL C-terminal tails be gained.

CHAPTER II

VISUALIZING GROEL/ES IN THE ACT OF ENCAPSULATING A FOLDING PROTEIN*

Many essential proteins fold only when assisted by ATP-powered machines known as molecular chaperones [239]. The GroEL/ES system of E. coli is a well-studied example of the chaperonin class of molecular chaperones [190,198]. GroEL is a tetradecamer of 57 kDa subunits, arranged as two stacked, seven-membered rings, each containing a large, solvent-filled cavity [152]. The cavity-facing surface of the apical domain of each subunit is lined with hydrophobic amino acids that tightly bind substrates that are neither random coil, nor natively folded proteins (so-called non-native proteins; [165]). Efficient folding of proteins that strictly depend on GroEL (so-called stringent substrate proteins) requires encapsulation of the non-native substrate protein within a cavity formed by GroEL plus the smaller, ring-shaped co-chaperonin GroES [163,177,206,308]. Encapsulation seals the GroEL cavity and results in the release of the substrate protein into an enlarged GroEL-GroES chamber (a cis complex). Upon release, folding is initiated and continues for a brief period, until the cavity is disassembled and the protein, folded or not, is ejected back into free solution [164,173,177,206,308].

Encapsulation and initiation of folding ultimately depend upon ATP-driven structural rearrangements [157,161,162]. ATP binding to GroEL equatorial domains

*Reproduced, with modifications, with permission from Chen, et al. (2013) Cell 153: 1354-1365. Copyright 2013 Elsevier Inc. For the original publication: JW designed and performed biochemical experiments, and analyzed the corresponding data. JW did not perform or analyze any of the microscopy experiments, which were done by DHC.

results in large, cooperative rearrangements, which elevate and rotate the apical domains [154,309,310]. Exposed sites bind GroES, which results in a switch of the apical-domain surfaces from hydrophobic to polar [154], a switch believed to be essential for releasing substrate and triggering folding. While protein folding is initiated inside the GroEL-GroES cavity, the relatively short lifetime of this complex limits the amount of time a protein has to fold [157,164,173,309]. The timer for complex disassembly is set by the rate of ATP hydrolysis within the cis cavity, ranging from 4-20 sec, depending on temperature and the concentration of non-native substrate protein [157,160,309].

Recent structural work provides insight into how a non-native substrate protein is bound to an open GroEL ring and a view of a fully folded protein inside the GroEL cavity [311-314], but structural information about non-native proteins during and immediately following encapsulation, the point at which folding is initiated, remains limited. In fact, the cooperative structural transitions of the GroEL ring that occur in response to ATP binding appear to create a paradox. Given that non-native substrate proteins and GroES are thought to bind to overlapping sites on the GroEL apical domains, how is it possible for ATP binding to drive a GroEL ring into a state with high affinity for GroES without causing premature release of the folding intermediate outside the chaperonin?

Binding competition between GroES and a substrate protein to the apical domains could be avoided if the ATP-bound GroEL ring populates an intermediate conformation that transiently binds both substrate protein and GroES [171]. Entry into and exit from such a state would require an orderly allosteric cascade designed to

enforce a specific ligand binding sequence [158,171,172,315]. The sequence begins with the non-native substrate protein binding to the open trans ring of an asymmetric GroEL-GroES complex with high affinity for the substrate protein, but without significant affinity for GroES (Figure 2.1A; [166]). Cooperative binding of ATP to the same GroEL ring is then thought to initiate encapsulation through a series of conformational states, which sequentially weaken the interaction between GroEL and the substrate protein, while simultaneously strengthening the interaction with GroES (Figure 2.1B-1F). While functional and kinetic studies strongly suggested the existence of such an allosteric cascade, because they are only transiently populated, the structural nature of these key intermediate states has remained poorly understood. A recent electron cryo-microscopy (cryo-EM) study of GroEL in the presence of ATP has begun to fill in some of these missing details, by successfully classifying several intermediate conformations of the GroEL apical domains [315]. This study provides structural evidence for a sequential allosteric cascade, as well as insight into the intermediates populated by an ATP-bound GroEL ring prior to encapsulation. However, the absence of GroES and non-native substrate protein in these studies leaves unresolved the key structural transitions that lead to substrate protein encapsulation, release and folding.

Here we show a non-native substrate protein trapped inside the GroEL-GroES cavity during encapsulation. We used cryo-EM and single-particle three-dimensional (3D) reconstruction to determine the structure of a chemically modified GroEL mutant, which stalls in an allosteric state just prior to substrate protein release (the R2 state; [158]).

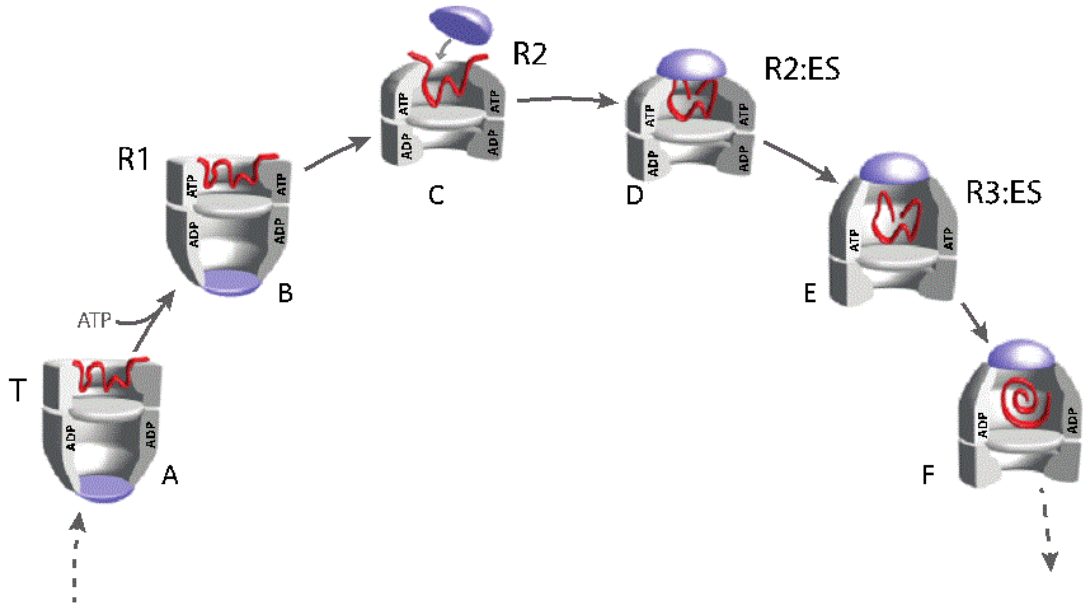


Figure 2.1: The GroEL Protein Folding Cycle Involves a Series of Allosteric Transitions within the Chaperonin Complex. Non-native substrate proteins enter the GroEL reaction cycle by binding to the open trans ring of an asymmetric GroEL-GroES complex, pulling the trans ring into the high-affinity "T" state (A; for cycle details, see [171,190,198]). Protein encapsulation is initiated by highly cooperative binding of ATP to the *trans* ring, populating the R1 state (B), a conformational state of the GroEL ring with high affinity for the non-native protein but not yet for GroES. The R2 state (C) retains substantial, though weakened, affinity for the non-native protein, binds GroES and encapsulates the substrate protein (D). Transitions into or between the R1 and R2 states are also linked to disassembly of the GroEL-GroES complex on the opposite ring. The ATP-bound GroEL-GroES complex has a high affinity for GroES in the R3 state (E), which releases the non-native substrate protein into the enclosed *cis* cavity, to initiate folding. Hydrolysis of ATP within the *cis* ring triggers a transition of the complex to at least one additional conformational state (F).

Experimental Procedures

Proteins. Wild-type GroEL, EL398A, EL43C and EL43C398A, were expressed and purified as previously described [166]. ELΔ526 was purified in the same manner as wild-type GroEL. GroES, RuBisCO and GFP were also expressed and purified as previously described [157,163,166,190]. Bovine rhodanese was purchased from Sigma and purified as previously described [158,173].

Labeling of proteins with fluorescent dyes. EL43C and EL43C398A were specifically labeled at Cys43 with the thiol-reactive dye N-1-pyrene maleimide (PM) to generate EL43Py and EL43Py398A, as previously described [158]. Wild-type RuBisCO was fluorescently labeled at Cys58 with the thiol-reactive dye 5-iodoacetamidofluorescein (5IAF) to create 58F-RuBisCO as previously described [157,166,190]. Both PM and 5IAF were obtained from Invitrogen (Molecular Probes; Eugene, OR) and were prepared fresh from dry powder in anhydrous DMF immediately prior to use. The extent and specificity of dye conjugation was confirmed as previously described [166,168,316]. For 58F-RuBisCO, EL43Py and EL43Py398A, the labeling efficiency was confirmed to be 98-100% by at least two different methods of analysis [158,316].

Refolding, enzymatic and encapsulation assays. Rhodanese folding was assayed essentially as previously described [206]. Native rhodanese was first denatured in acid urea buffer (25 mM glycine-phosphate, pH 2.0, 8 M urea), and then diluted 50 fold into refolding buffer containing a given GroEL variant. Folding was then initiated by addition of GroES and ATP. For single turnover experiments, GroEL cycling was

prevented by the addition of hexokinase and glucose within 5 s of the initial ATP addition, as previously described [158,166].

All encapsulation experiments were conducted in sample buffer: 50mM HEPES, pH 7.6, 15 mM $Mg(OAc)_2$, 100 mM KOAc and 2 mM DTT. In all cases, the substrate protein (58F-RuBisCO, GFP or rhodanese) was denatured in acid urea buffer for 30-60 min at room temperature (in the dark) to yield working stocks of dRub, dGFP, dRho. ADP bullets complexes were created essentially as previously described, by adding GroEL (WT or Δ526), GroES and ATP to final concentrations of 7 μM, 14 μM and 150 μM respectively, followed by a 10 min incubation at room temperature [166]. ADP bullet complexes with either non-native GFP or rhodanese bound to the trans ring were formed by adding dGFP or dRho (250 nM final) to cold buffer containing 250 nM ADP bullets. After a five min incubation at room temperature, ATP was added to 1mM to permit GroES binding and protein encapsulation. Following a 10 s incubation, the ambient ATP was quenched by the addition of hexokinase (50 U/ml) and glucose (10 mM). In the case of GFP, the sample was then incubated for an additional 10 min at room temperature to allow GFP folding, followed by injection of the sample onto a Superose 6 gel filtration column, equilibrated in sample buffer and supplemented with 50 μM ADP, connected to an in-line fluorescence detector. For experiments with rhodanese, unencapsulated protein was removed by treatment with 0.25 mg/ml Proteinase K for 10 min. The protease was inactivated at room temperature by addition of 1 mM PMSF prior to injection of the sample onto the gel filtration column. The GroEL-GroES peak was collected and total protein was precipitated with cold TCA. The

concentrated protein sample was then loaded onto a 10% SDS-PAGE gel to separate rhodanese from GroEL and GroES. Following Coomassie staining, the intensity of the rhodanese band was quantified by densitometry. Encapsulation experiments with 58F-RuBisCO were conducted in a similar manner, except that ATP was added 3 s after the non-native RuBisCO was bound to the ADP bullet trans ring. Unencapsulated RuBisCO was removed prior to injection on the column by treatment of the sample with Proteinase K, as described above. As with GFP, the extent of 58F-RuBisCO encapsulation was determined by direct fluorescence measurement of the GroEL-GroES peak.

TCA Precipitation of GroEL-GroES Samples. 1.3 ml of each sample was added to 200 ml of 100% TCA in a 1.5 ml microcentrifuge tube. These were incubated overnight at 4 C. Samples were centrifuged at 13,000 rpm at 4 C for 15 min. The supernatant was removed and the pellets were washed with acid acetone. Samples were centrifuged at 13,000 rpm at 4 C for 15 min. The supernatant was removed and the pellets were allowed to air-dry. Pellets were then resuspended for gel electrophoresis (2:1 solution of 50mM NaHCO3, pH 10.5 and gel loading buffer).

Preparation of Chaperonin Complexes for Cryo-EM. Preparation of the EL43Py398A-GroES-ATP complex was performed by mixing EL43Py398A (8 μM), GroES (7.2 μM) and ATP (100 μM) in 50 mM HEPES (pH 7.6), 5 mM KOAc, 10 mM $Mg(OAc)_2$, 2 mM DTT. The reaction products were immediately deposited to the grids for freezing. The EL43Py398A-RuBisCO-GroES-ATP and EL398A-RuBisCO-GroES-ATP complexes were prepared in a similar manner, as follows: 0.1 μM denatured RuBisCO was quickly diluted into 1 ml of cold, high K+ refolding buffer (same

composition as above, but with the KOAc concentration increased to 50 mM) containing 0.1 μM EL43Py398A or EL398A. The solution was warmed to room temperature for 5 min to allow RuBisCO binding to the chaperonin, before 0.1 μM GroES and 1 mM ATP were added. Within 10 s, 600 ml of solution was transferred to a Vivaspin 500 ultrafiltration unit (100K MWCO) and centrifuged at 12,000 x g in a swinging bucket rotor for 6 min in an Eppendorf Microcentrifuge 5417 R at 4 C. Approximately 20 ml of this concentrated solution was recovered, from which a sample was immediately applied to the holey grids, blotted and plunged into liquid ethane.

Cryo-EM. Sample preparation and freezing for all experiments was complete within 10 min. Using freshly glow-discharged Quantifoil grids (R1.2/1.3, Quantifoil Micro Tools GmbH, Jena, Germany), 3 ml of the sample solution was applied onto the holey film side of each grid and then blotted at room temperature for 1.0 s at 95%–100% humidity prior to plunging into liquid ethane cooled by liquid nitrogen using an automated vitrification device, Vitrobot (FEI Company, http://www.fei.com/Vitrobot). Frozen grids were stored in liquid nitrogen until examined in an electron cryo-microscope. Images for the EL43Py398A-GroES-ATP complex were collected on Kodak SO-163 photographic film at an effective magnification of 61,060x with the specimen temperature at 4 K, a 2 s exposure time, and a dose of 25-36 e/Å 2 per image from a 300-kV JEOL JEM-3000SFF liquid-helium-cooled electron microscope. Images for the EL43Py398A-RuBisCO-GroES-ATP complex were collected on a Gatan Ultrascan 4k x 4k CCD at a detector magnification of 73,530x with a specimen temperature of 50 K, a 2 s exposure time, and a dose of 30 e/Å 2 per image on a JEM-

3000SFF. Images for the EL398A-RuBisCO-GroES-ATP complex were collected using a Gatan Ultrascan 4k 3 4k CCD at a detector magnification of 70,760x using a 300-kV JEOL JEM-3200FSC with a specimen temperature of 101 K, a 1 s exposure time, a dose of 20 e/Å 2 per image and an energy slit of 20 eV in the in-column energy filter. The pixel size for the CCD images of the EL43Py398A-RuBisCOGroES-ATP and the EL398A-RuBisCO-GroES-ATP complexes were 2.04 Å and 2.12 Å at the specimen level, respectively. The films for recorded images of the EL43Py398A-GroES-ATP complex were developed for 12 min in Kodak D19 developer, fixed for 10 min in Kodak fixer at 20 C and then scanned using a Nikon Super Coolscan 9000 scanner at a 6.35 mm step size, yielding a pixel size of 1.04 Å at the specimen level. The scanned EL43Py398A-GroES-ATP images were then averaged twice to produce data at 2.08 Å /pixel. The digital images were pre-processed by boxing particle images and determining contrast transfer function parameters for each micrograph using EMAN1 program [317].

Image Processing. Particle selection for each complex was performed semi-automatically using the EMAN1 [317] program, boxer, followed by manual screening-out of large contaminants and aggregates. The contrast transfer function fitting for each photographic film or CCD frame was performed automatically using the program, fitctf.py [318], and then fine-tuned manually using the EMAN1 program, ctfit. The methodology of EMAN1 multiple-model refinement (EMAN1 program multirefine) for compositionally and conformationally heterogeneous complex analysis, which has been previously described [319], was used to sort different particle sub-populations from all

the data sets (Figure 2.2A). Briefly, several consecutive multiple-model refinements were applied to ‘‘computationally purify’’ a sub-population of the relatively homogeneous bullet-shaped particle images of greatest interest for each of three complexes described in this study. Each sub-population with ‘‘purified’’ particle images was then subjected to a single-model refinement (EMAN1 program refine) to obtain a final converged bullet-shaped 3D reconstruction. In the case of the EL43Py398A + GroES + ATP sample (71,200 particle images in total), an initial model was generated from the crystal structure of the GroEL-GroES-ADP complex (PDB ID code 1AON) with C7 symmetry by first low-pass filtering this structure to 40 Å resolution. Flatband Gaussian noise was then added to this 40-Å map, yielding three structurally identical initial models that possessed distinct Gaussian noise. These three initial models were used for the first-round multiple-model refinement. Tens of cycles for this first multiple-model refinement were performed until a reasonable convergence was achieved for each sub-population. The number of refinement cycles was determined when the number of particle images was stabilized in successive iterations for each sub-population data. This initial multiple-model refinement divided the data set into three sub-populations: free GroEL tetradecamer (no GroES bound, 12,490 particle images), bullet-shaped GroEL-GroES complexes (with GroES bound to only one end of GroEL tetradecamer, 35,170 particle images) and football-shaped GroEL-GroES2 complexes (with GroES bound to both ends of the GroEL tetradecamer, 23,540 particle images).

Because of substantial heterogeneity in the EL43Py398A + GroES + ATP sample, single-round multiple-model refinement based on three initial models, which

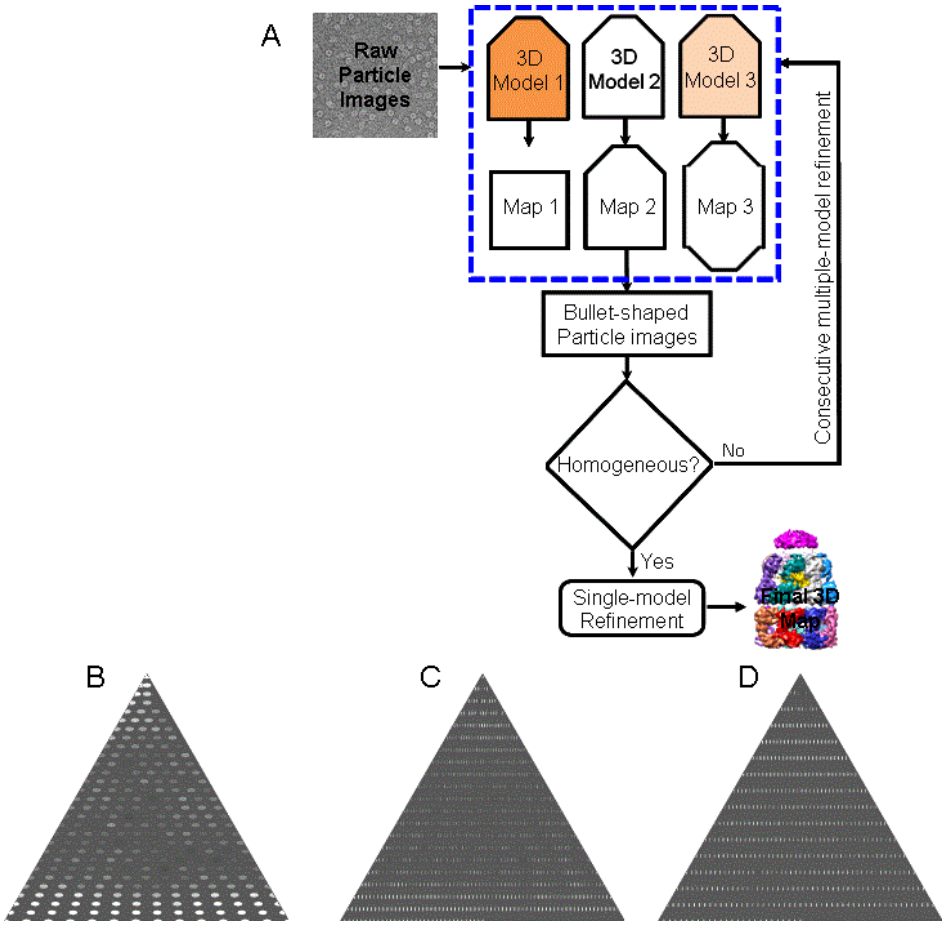


Figure 2.2: Data Processing Flow Chart and Euler Angle Distributions of Bullet-Shaped Complexes. (A) A general flow chart for our data processing strategy using the consecutive multiple-model refinement is shown. The first-round multiple-model refinement was seeded with three initial models. These initial models were derived from the GroEL-GroES-ADP (PDB ID code 1AON) X-ray structure after it was low-pass filtered to 40 Å and different Flatband Gaussian noises were added to each of the three initial models. Tens of cycles for this first-round multiple-model refinement were performed until a reasonable convergence was obtained for each sub-population. The bullet-shaped particle images were extracted for a few more multiplemodel refinements until a relatively homogeneous data set for this bullet-shaped particle images was sorted out. This final highly computationally selected bullet shaped particle images were subjected to the single-model refinement to generate a converged 3D reconstruction. (B) The Euler angle distribution for the C7 structure of EL43Py398A-GroES-ATP shown in Figure 2.3A. (C) The Euler angle distribution for the symmetry-free structure of EL43Py398A-RuBisCO-GroES-ATP shown in Figure 2.4A. (D) The Euler angle distribution for the symmetry-free structure of EL398A-RuBisCO-GroES-ATP shown in Figure 2.5A. The top of each triangle represents top views and the two bottom corners of each triangle represent side views.

sorted all the particle images into only three sub-populations, could not reliably sort the entire population. In order to further differentiate the sub-population of bullet-shaped GroEL-GroES complexes from the EL43Py398A + GroES + ATP sample, several additional rounds of consecutive multiple-model refinement were applied to the sub-population of bullet-shaped GroEL-GroES complexes obtained from the first-round multiple-model refinement. The second-round multiplemodel refinement sorted the bullet-shaped GroEL-GroES complexes (35,170 particle images) into three sub-populations: 5,370 particle images without GroES bound, 17,950 particle images with bullet shape and 11,850 particle images with football shape. The third-round multiple-model refinement further purified the bullet-shaped GroEL-GroES complexes (17,950 particle images) into a relatively homogeneous sub-population of bullet-shaped particle images (8,372 particle images). This final bullet-shaped sub-population was further processed using a single-model refinement with imposed C7 symmetry for a map at 8.9 Å resolution.

The strategies employed for heterogeneity sorting for EL43Py398A + RuBisCO + GroES + ATP sample and EL398A + RuBisCO + GroES + ATP sample were similar. In the case of the EL43Py398A + RuBisCO + GroES + ATP sample (235,000 particle images), the three initial models was also generated from the crystal structure of the GroEL-GroES-ADP complex (PDB ID code 1AON) with C7 symmetry by first low-pass filtering this structure to 40 Å resolution and adding flatband Gaussian noise. The first several consecutive multiple-model refinements reduced the EL43Py398A + RuBisCO + GroES + ATP sample to 62,350 particle images with the bullet-shape. The final

multiple-model refinement sorted this bullet-shaped data set to two major sub-populations: one without substrate RuBisCO (24,580 particle images) and one with substrate RuBisCO (17,610 particle images). The final sub-population with RuBisCO inside the cis cavity for the EL43Py398A + RuBisCO + GroES + ATP sample was processed using a single-model refinement for a converged bullet-shaped structure without imposing symmetry at 9.2 Å resolution.

In the case of the EL398A + RuBisCO + GroES + ATP sample (284,070 particle images), the first multiple-model refinement sorted the heterogeneous sample for the bullet-shaped complex with 108,870 particle images. After several consecutive multiple-model refinements, the final refinement round sorted this bullet-shaped data set into the sub-population of 8,189 particle images containing RuBisCO inside the cavity. The final sub-population with RuBisCO inside the cis cavity for the EL398A-RuBisCO-GroESATP sample was processed using a single-model refinement for a converged bullet-shaped map without imposed symmetry at 15.9 Å resolution.

In the initial round of multiple-model refinement for heterogeneity sorting, C7 symmetry was applied. However, subsequent rounds of multiple-model refinement and each final single-model refinement did not impose any symmetry on the data from the EL43Py398A + RuBisCO + GroES + ATP sample or the EL398A + RuBisCO + GroES + ATP sample. However, a C7 symmetry restraint was imposed on all multiple-model and single-model refinements for the EL43Py398A + GroES + ATP sample.

The Euler angular distributions of particle images corresponding to each of the reconstructions (Figure 2.3A, 2.4A and 2.5A) are shown in Figures 2.2B-D, respectively.

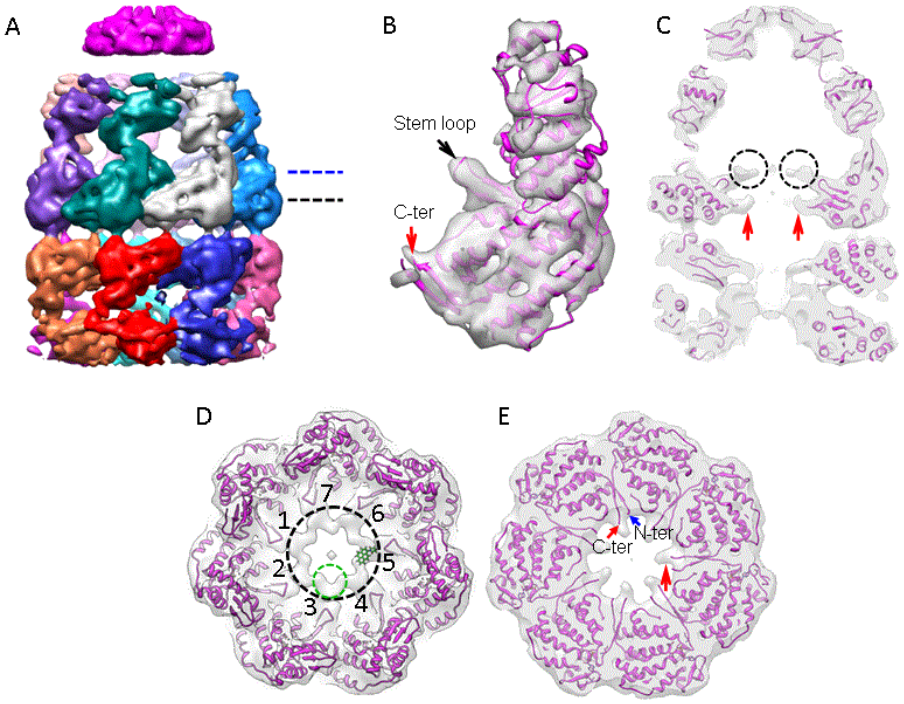


Figure 2.3: The Structure of the EL43Py398A-GroES-ATP Complex Determined at 8.9 Å Resolution by Cryo-EM with C7 Symmetry Imposed. (A) Side view of the EL43Py398A-GroES-ATP density map displayed at a contour level of 1.3 σ. Individual GroEL subunits are shown in different colors; GroES is magenta. All other density maps shown in this study are displayed at a contour level of 1.0 σ (unless otherwise noted). (B) Close-up view of a single EL43Py398A cis-ring subunit (contour level of 1.5 σ) overlapped with a rigid-body, flexibly refined fit of the GroEL-GroES-ADP crystal structure (PDB ID: 1AON; magenta) using the program *DireX* [320]. The stem loop containing Cys 43 and the GroEL C-terminus are labeled with arrows. (C) A medial slice of the density map shown in (A), with the density rendered transparent and superimposed on a rigid-body, flexibly refined fit of the GroEL-GroES-ADP crystal structure. In (C), extra density is visible at the tips of the equatorial stem loops of each GroEL subunit (amino acids 34-52; black dashed circles). The observed densities beyond amino acid 525 in the C-terminal tails are indicated by red arrows. (D) The additional stem-loop density for each subunit is shown (inside of black dashed circle), viewed from above, at a slice level indicated by the dashed blue line in (A). The seven stem loops are labeled 1-7, respectively. A single N-1-pyrene maleimide dye molecule (green) was rigid-body fit into the density at the tip of one stem loop in the EL43Py398A-GroES-ATP complex using Chimera. (E) View of the cis-ring equatorial domain near the subunit C-termini, viewed from above, at the slice level indicated by the black dashed line in (A). Substantial density (large red arrow) is visible in the region of the subunit C-termini, well beyond the last crystallographically resolved residue (small red arrow). The position of the GroEL subunit N-terminus is indicated by the blue arrow.

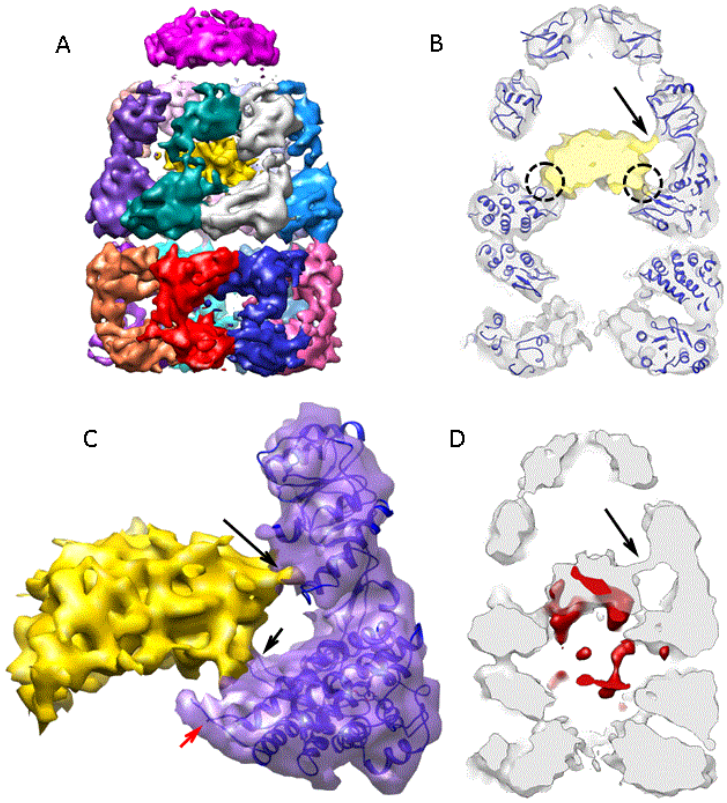


Figure 2.4: The Structure of the EL43Py398A-GroES-ATP Complex Containing Non-Native RuBisCO within the Cis Cavity Determined at 9.2 Å by Cryo-EM without Imposed Symmetry. (A) A side view of the density map of EL43Py398A-RuBisCO-GroES-ATP complex (contour level 1.23 σ) shown colored as in Figure 2.3, with density from the encapsulated, non-native RuBisCO monomer shown in gold. (B) A medial slice of the EL43Py398A-RuBisCO-GroES-ATP complex with the density rendered transparent and overlapped with a rigid-body, flexibly refined fit of the GroEL-ADP-GroES crystal structure (PDB ID: 1AON) to the cryo-EM map. Additional density around the GroEL equatorial domain stem loops makes direct contact with the non-native RuBisCO monomer (dashed black circles). The RuBisCO is also in contact with the lower region of the apical domain of one cis-ring GroEL subunit in the region of F281 (black arrow). (C) A close-up view of one cis-ring GroEL subunit in direct contact (long black arrow) with the non-native RuBisCO monomer (gold; contour level of 1.05 σ). The GroEL subunit stem loop (short black arrow) and C-terminus (red arrow) are indicated. (D) A medial slice of the variance map derived for the EL43Py398A-RuBisCO-GroES-ATP complex (red; see Experimental Procedures) is shown overlapped with the average map of the complex (gray; orientation as in panel B), calculated from 100 3D reconstructions of the complex computed during the variance calculations. The largest variations in the density map are from the non-native RuBisCO monomer and cavity-facing regions of the GroEL equatorial domains, most likely the C-termini of the cis and trans rings.

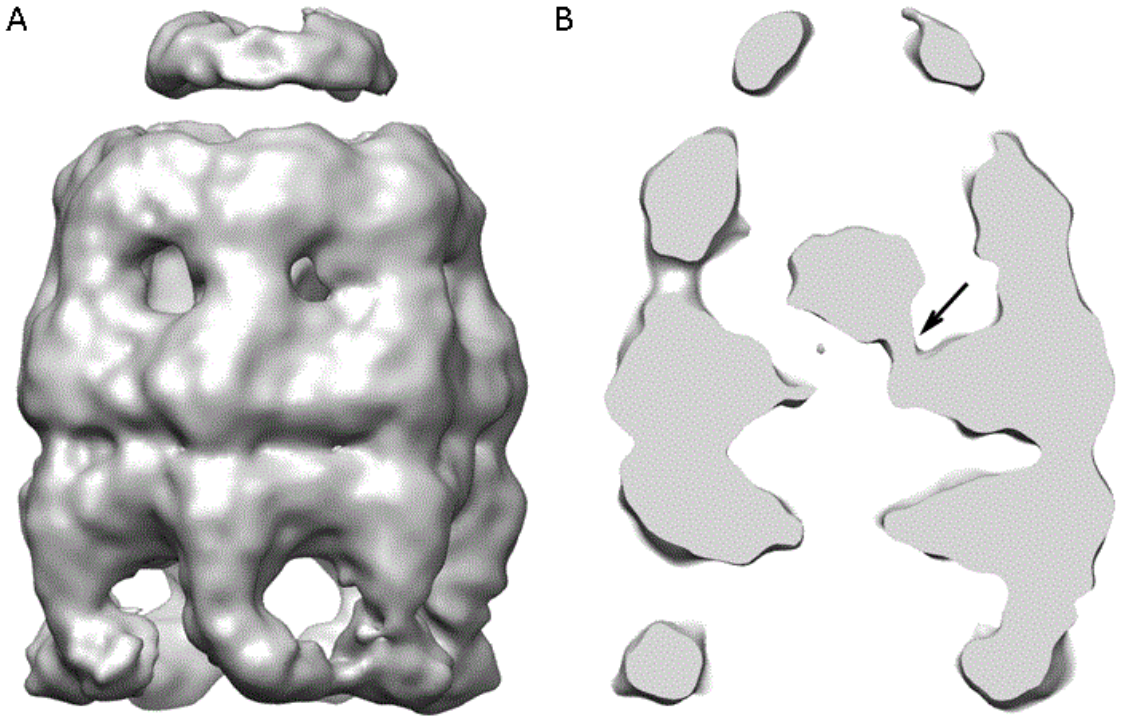


Figure 2.5: The Structure of the EL398A-GroES-ATP Complex Containing RuBisCO within the Cis Cavity Determined at 15.9 Å by Cryo-EM Reconstruction without Imposed Symmetry. (A) A side view of the EL398A-RuBisCO-GroES-ATP density map. (B) Medial slice of the EL398A-RuBisCO-GroES-ATP complex indicates direct contact between the folding RuBisCO monomer and the C-terminal and stem-loop region of one cis-ring GroEL subunit (black arrow).

The final resolutions for the refined structures were assessed using the gold-standard criterion of Fourier Shell Correlation (FSC) cut-off at 0.143 from two independent half-sets of data [321] after the particle images were highly purified from our consecutive multiple-model refinement procedures. Chimera [322] was used for the surface representations of all the cryo-EM density maps. The 3D variance map of the EL43Py398A-RuBisCO-GroES-ATP complex was calculated using the EMAN1 program calculateMapVariance.py with the bootstrap technique implemented [323,324].

Resolution Assessment of Reconstruction from Computationally Purified Particle Images. To further confirm our resolution for each finally sorted subpopulation, we performed the new gold standard resolution test [321,325], which was recently adopted by the cryo-EM community as the most rigorous criterion. First, the data set was separated into two half sets of the finally sorted particle images for completely independent reconstructions. Second, initial model for each of the two half data sets was built from scratch based on the reference-free class-averages from EMAN1's program refine2d.py. Third, each data set was refined to a converged map. Fourth, two final maps were aligned using EMAN20 s program e2align3d.py. Lastly, the FSC between the two independently reconstructed maps was computed. According to the gold standard FSC = 0.143 criterion, the resolutions are: 8.9Å, 9.2Å and 15.9Å, for EL43Py398A-GroES-ATP (in C7), EL43Py398A-RuBisCO-GroES-ATP (in C1) and EL398A-RuBisCO-GroES-ATP (in C1) respectively (Figures 2.6C, 2.7A, and 2.8B).

Variance Map Calculations. Briefly, 100 3D reconstructions were obtained from randomly sampled subsets of particle images with determined orientations from which

Figure 2.6: The EL43Py398A-GroES-ATP Complex. (A) A hydrolysis deficient variant of EL43Py (EL43Py398A) stalls at the same point in the GroEL-GroES folding cycle as EL43Py. Assisted folding by EL43Py and EL43Py398A was examined using a single turnover folding assay. So that the extent of folding could be assessed inside the GroEL-GroES cavity, without the need to disrupt the complex and release the protein for assembly, the monomeric enzyme Rhodanese (rho) was employed. Native rho was denatured using acid-urea and was then mixed with either of four different tetradecamer GroEL variants to form binary complexes: (1) wild-type GroEL (wt EL), (2) EL43C/D398A, the same variant of GroEL from which EL43Py398A is derived, but with no dye attached to position 43 (EL43C398A), (3) EL43Py and (4) EL43Py398A. In each case, 100 nM denatured rho was mixed with 250 nM GroEL tetradecamer and the samples were then incubated at room temperature for 10-15 min. Each complex was then rapidly mixed with 500 nM GroES and 0.1 mM ATP, incubated for 5 s, followed by rapid depletion of excess ATP with hexokinase and glucose to prevent cycling. Because the added GroES can bind to either the substrate-occupied or empty rings of each tetradecamer (but not, under these conditions, both rings), the maximum theoretical folding yield for wild-type GroEL should be 50%. Both wt GroEL and the hydrolysis deficient, but dye free, EL43C398A variant display essentially identical rho refolding curves, with a plateau at 55%–60%. The EL43Py398A variant, despite being unable to hydrolyze its bound ATP, displays folding behavior comparable to EL43Py. Error bars are one SD. (B) A typical cryo-EM image for the EL43Py398A- GroES-ATP complex at a defocus of 2.88 mm is shown. (C) Resolution assessment EL43Py398A-GroES-ATP complex map using the gold standard 0.143 FSC cut-off criterion yielded a resolution of 8.9 Å. The subpopulation of the asymmetric, bullet- shaped EL43Py398A-GroES-ATP complex was sorted from the heterogeneous mixture of several other complexes (e.g., uncomplexed and symmetric EL43Py398A-GroES2- ATP; see Experimental Procedures) and reconstructed with C7 symmetry imposed. (D) The crystal structure of the GroEL-GroES-ADP complex does not display extra density around the equatorial stem loops and C-termini. The X-ray structure of the GroEL- GroES-ADP complex (PDB ID code 1AON; green) is superimposed on its transparent density map, low-pass filtered to 8 Å (gray). A medial slice of the structure along the 7- fold symmetry axis is shown. Note that no extra density exists beyond the stem loops
(dashed circles) and C-termini, compared to Figure 2.3C. (E) Association of the GroEL C-terminal tails with the engineered pyrene dyes in the EL43Py398A-GroES-ATP complex. A slice of the EL43Py398A-GroES-ATP complex along the 7-fold symmetry axis is shown at a slightly lower contour level (0.9 s) than in Figure 2.3C and is overlaid with a model obtained from flexibly refined, rigid-body fitting (magenta). The density for the EL43Py398A-GroES-ATP is rendered transparent. The C-termini from the cis (upward pointing red arrows) and trans (downward pointing red arrows) rings curve toward, and appear to make intimate contact with, the engineered pyrene dyes on the equatorial stem loops.

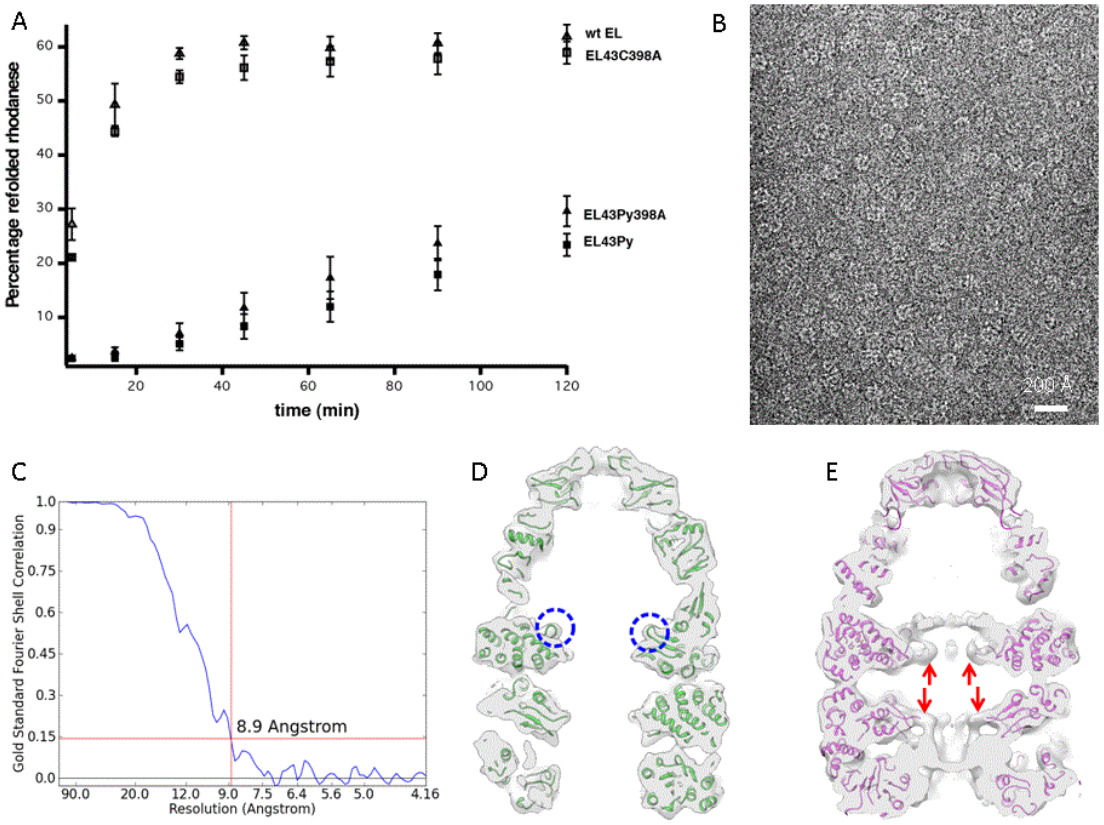
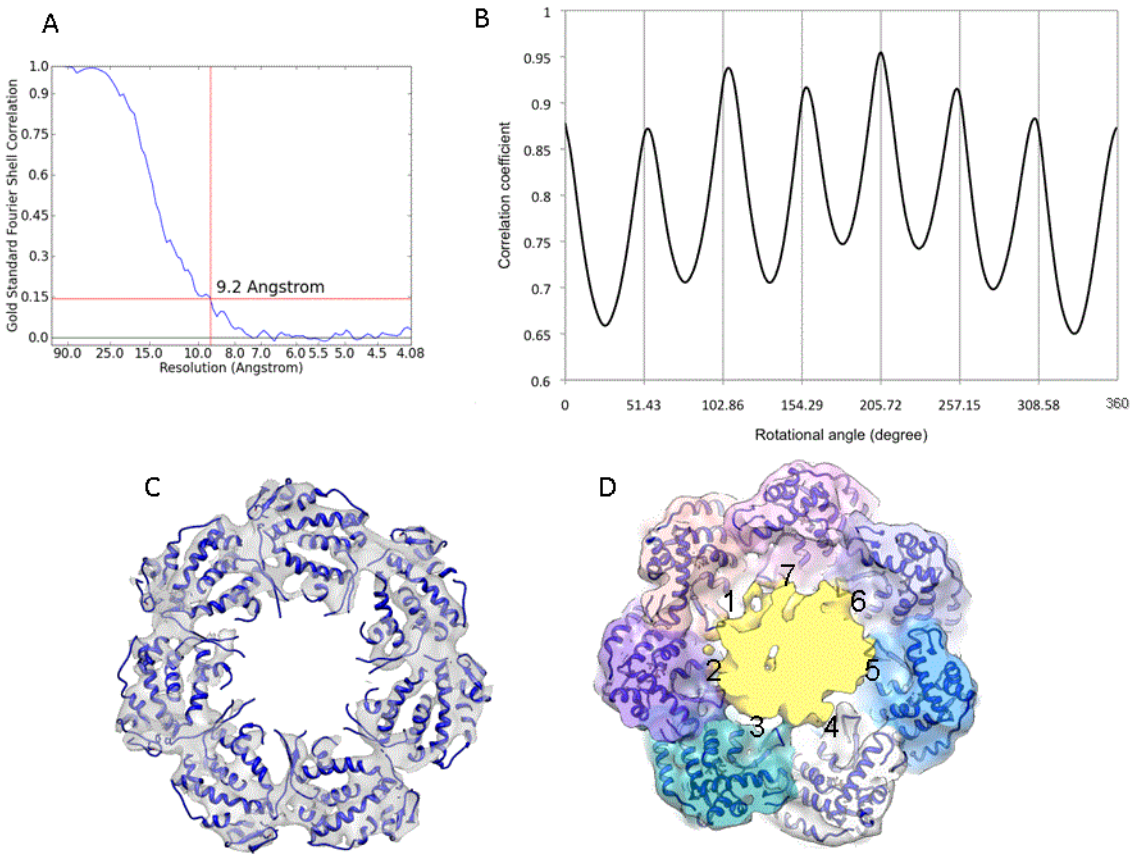

A
Percentage refolded rhodanese
time (min)
wt EL
EL43C398A
EL43Py398A
EL43Py
B
200 Å
C
Gold Standard Fourier Shell Correlation
Resolution (Angstrom)
8.9 Angstrom
D
E

Figure 2.7: The EL43Py398A-RuBisCO-GroES-ATP Complex. (Related to Figure 2.4) (A) The FSC curve for the EL43Py398A-RuBisCO-GroES-ATP structure is shown. The resolution of the structure is 9.2 Å by the gold standard 0.143 FSC cut-off criterion. (B) The C7 symmetry of the GroEL trans ring is slightly broken in the EL43Py398A-RuBisCO-GroES-ATP. A plot of the cross correlation coefficient between the trans ring apical domains of the EL43Py398A-RuBisCO-GroES-ATP complex (shown in Figure 2.4A) and one subunit from the C7-symmetrized map of the same trans-ring apical domains, rotated along the applied C7 symmetry axis, is shown. The peaks in each curve have different heights and spacing, indicating the C7 symmetry of the canonical GroEL rings is broken on the trans ring, as well as the cis ring (Figure 2.6B). (C) A slice across the cis ring equatorial domain is shown at a contour level of 1.35 s, illustrating the well resolved equatorial helices and highlighting the subnanometer resolution of the reconstruction. (D) Interaction of the non-native RuBisCO monomer with density near the pyrene dye attachment points on the cis-ring equatorial stem loops. A transverse slice across the cis ring equatorial domain of the EL43Py398A-RuBisCO-GroES-ATP structure, overlapped with the model obtained from strongly restrained flexible fitting (blue) with the density rendered transparent. This slice is slightly above the stem loops and is viewed from the cis ring toward the trans ring. The nonnative RuBisCO monomer is rendered in gold and each of the seven GroEL subunits are shown in a different color and labeled (1-7). At this contour level (1.0 s) the stem loop can be seen to make direct connect with the RuBisCO monomer, though with different apparent strengths; the stem-loop region from subunit 4 shows no apparent connection to the RuBisCO monomer.


A
Gold Standard Fourier Shell Correlation
1.0
0.9
0.75
0.6
0.45
0.3
0.15
0.0
9.2 Angstrom
90.0 25.0 15.0 10.0 8.0 7.0 6.0 5.5 5.0 4.5 4.08
Resolution (Angstrom)
B
Correlation coefficient
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0 51.43 102.86 154.29 205.72 257.15 308.58 360
Rotational angle (degree)
C
D
1
2
3
4
5
6
7

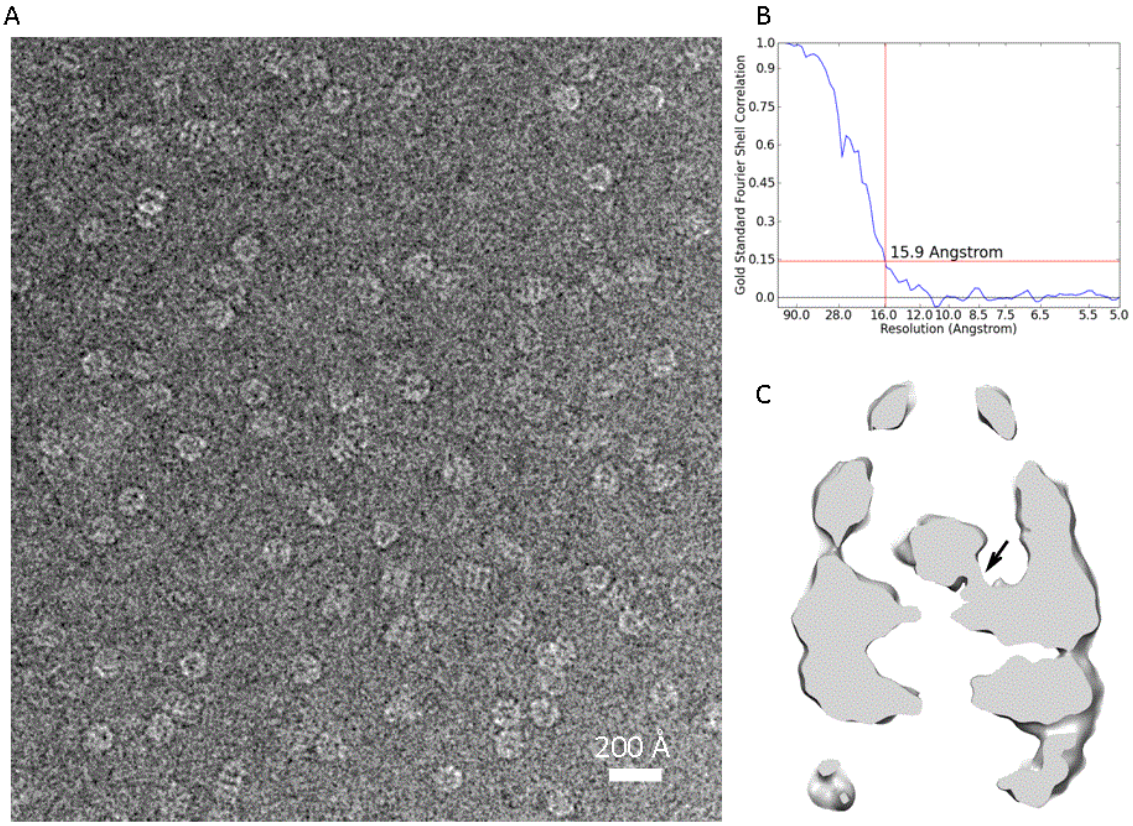


Figure 2.8: The EL398A-RuBisCO-GroES-ATP Complex. (A) A typical cryo-EM image for the complex at a defocus of 2.98 mm. (B) FSC curve for the structure, indicating a resolution of 15.9 Å by the gold standard 0.143 FSC cut-off criterion. (C) A medial slice of the refined structure is shown, using a different initial model low-pass filtered to 40 Å from the GroEL-GroES-ADP crystal structure (PDB ID code 1AON). Note that this starting model contains no density inside the cavity of GroEL/ES. Even though refined from a different starting model, the final structure still shows the association (black arrow) between the RuBisCO and one of C-terminal/stem-loop regions. In our gold standard resolution test, in which the initial models were built from scratch, the connection density between the substrate protein and the region of the GroEL C-termini and stem loops still exists in both independent maps from two separate half data sets (data not shown).

the 9.2 Å density map in Figure 2.4A was reconstructed. A 3D variance map and a 3D average map were calculated using these 100 reconstructions. The final variance map was low-pass filtered to 15 Å and the average map was low-pass filtered to 9.0 Å.

Flexible Model Fitting to Cryo-EM Maps. The GroEL-GroES complexes examined here all contain the hydrolysis deficient D398A mutation and, therefore, have ATP bound within the cis complex. In order to determine whether the results of the fitting analysis were affected by the starting X-ray structure, the EL43Py398A-RuBisCO-GroES-ATP complex was re-analyzed using the X-ray structure of an ATP-bound GroEL-GroES complex (PDB ID code 1PCQ). The differences between the fitted models derived from a starting ADP-bound X-ray structure (PDB ID code 1AON) and a starting ATP-bound X-ray structure were negligible, with a per subunit RMSD for the Ca atoms of 0.6-0.7 Å. This is less than the per-subunit Ca RMSD between the X-ray structures themselves (0.7-1.0 Å), and less than the difference between the starting 1PCQ model and the fitted model derived from this X-ray structure (0.9 - 1.3 Å). These results demonstrate that the final models obtained from the fitting analysis outlined above are not biased by the use of the ADP-bound GroEL-GroES-ADP structure (PDB ID code 1AON) as the starting model.

A network of elastic restraints was used, with in total 117,388 (two times the number of atoms) harmonic distance restraints which were randomly chosen from the list of atom pairs that are initially within a distance of 3 to 12 Å. This choice of restraints was used for all model refinements expected for the lower resolution EL398A-RuBisCO-GroES-ATP map, for which 176,610 distance restraints (three times the

number of atoms) were chosen from a distance interval of 3-15 Å , leading to overall stronger restraints (almost equivalent to rigid-body fitting). With these strong restraints, the structure of each subunit remained very close to the starting crystal structure, with an RMSD value of only 0.3 Å for the cis-ring subunits. In general, no restraints between the different monomers were used. It should be noted that the elastic network was chosen not to be deformable, i.e., the gamma value was set to zero. In total, 100 steps of refinement with DireX were performed. A cross-validation approach was used to identify the optimally fitted model and to prevent overfitting. In brief, only Fourier components of the cryo-EM density map lower than 9 Å were used for fitting, while Fourier components from the so called 'free' interval 7–9 Å were used for validation only. For this, the cross-correlation coefficient, Cfree, is calculated between the model density map and the cryo-EM density map, both of which were band-pass filtered using the free interval, thus containing information that has not been used for fitting. Both the choice of the restraints and the choice of which of the 100 models generated during fitting were picked for final analysis were based on maximizing the Cfree value. For the EL398A-RuBisCOGroES-ATP case a free interval of 9–12 Å was used, i.e., only Fourier components with a resolution lower than 12 Å were employed for fitting. Elastic restraints within secondary structure elements were weighted twice as strong as in loop regions.

Fitting of X-ray structure into cryo-EM density maps. To standardize the pixel size of the substrate occupied and empty density maps of the GroEL-GroES complexes, a number of density maps were generated with different pixel sizes between 2.0 and 2.2

Å because different electron microscopes and recording media were used. The X-ray structure of the GroEL-GroES-ADP complex (PDB ID: 1AON) was then refined against each of these density maps using *DireX* [320]. As the orientation of the equatorial domains in the density maps was very similar to the X-ray structure, we used those domains as reference regions for magnification calibration for each map. The RMSD between all equatorial domains of the fitted models and the X-ray structure of the GroEL-GroES-ADP complex was calculated and the optimal pixel size was chosen as the one that leads to the smallest RMSD value. The optimized pixel sizes for the three density maps of the EL43Py398A-GroES-ATP, EL43Py398A-RuBisCO-GroES-ATP and EL398A-RuBisCO-GroES-ATP complexes were 2.08, 2.04 and 2.12 Å, respectively.

The GroEL-GroES-ADP crystal structure (PDB ID: 1AON) was fitted as a rigid body into the density maps of the EL43Py398A-GroES-ATP, EL43Py398A-RuBisCO-GroES-ATP, EL398A-RuBisCO-GroES-ATP and EL398A-GroES-ATP complexes. This rigidly docked structure served as the starting point for our strongly restrained flexible fitting using the program DireX [320].

Accession Numbers. The cryo-EM density maps for EL43Py398A-GroES-ATP complex, EL43Py398A-RuBisCO-GroES-ATP complex and EL398A-RuBisCO-GroES-ATP complex are deposited to the EMDB with accession numbers EMD-2325, EMD-2326 and EMD-2327, respectively. The fitted models are deposited in the PDB with accession numbers 3zpz, 3zq0 and 3zq1, respectively.

Results

Cryo-EM of a functionally trapped GroEL (EL43Py)-GroES complex. Previous work showed that the GroEL variant EL43Py is a potent tool for examining the linkage between substrate protein encapsulation, release and folding [158]. EL43Py was created through homogeneous N-1-pyrene maleimide alkylation of a surface-exposed Cys residue engineered into a stem-loop at the bottom of the GroEL cavity. EL43Py encapsulates non-native substrate proteins beneath GroES, but only very slowly releases them into the GroEL-GroES cavity to initiate folding. The EL43Py variant thus provides an excellent opportunity to trap and structurally characterize a key conformation of the GroEL-GroES complex that is essential for substrate protein encapsulation, but which is normally highly transient (the GroES-bound R2 state of the GroEL ring; Figure 2.1D). In order to facilitate this study, we incorporated one additional modification into the EL43Py background, introducing a well-established mutation (D398A) that prevents ATP hydrolysis by GroEL, without affecting ATP or GroES binding [163]. EL43Py398A stalls at the same point in the allosteric cycle as EL43Py (Figure 2.6A), but cannot hydrolyze ATP (data not shown). Using EL43Py398A and limiting amounts of ATP and GroES, we were thus able to create a chaperonin sample enriched in asymmetric EL43Py398A-GroES-ATP complexes (a so-called ATP bullet complex) with the cis cavity trapped in the R2 configuration. Cryo-EM was used to image this sample (Figure 2.6B), which contains multiple molecular species even when using an optimized mixing protocol, because the assembly reaction

can never be driven to completion to yield a single, unique EL43Py398A-GroES-ATP bullet complex.

We applied a consecutive multiple-model refinement strategy, which was used to successfully analyze images of chaperonins with mixed conformations and compositions [319,323,326]. The first round of processing of 71,200 particle images yielded three sub-populations of images that resulted in free GroEL tetradecamer (no GroES bound), bullet-shaped GroEL-GroES complexes (with GroES bound to only one end of the GroEL tetradecamer) and football-shaped GroEL-$GroES_2$ complexes (with GroES bound to both ends of the GroEL tetradecamer). Because we were only interested in the structure of the bullet-shaped complex in the present study, we did not pursue a structural determination of the other sub-populations.

The subset of images corresponding to the bullet-shaped complex was subjected to several additional rounds of multiple-model refinement to yield a final homogeneous dataset of 8,372 bullet-shaped particle images. This final data set was split into two halves for two completely independent reconstructions for a gold standard resolution assessment [321]. A 3D structure of the bullet complex at ~ 8.9 Å was reconstructed using C7 symmetry (Figure 2.3A and Figure 2.6C) from all 8,372 highly selected particle images. A symmetry-free reconstruction from the same set of 8,372 highly selected particle images was also obtained at 13.9 Å resolution. Without a symmetry imposition, the subunits are not perfectly symmetrically arranged, but do not deviate far from 7-fold symmetry (data not shown). Because the symmetry-free map is close to C7

symmetry, and the symmetry imposition generated a higher resolution map, we employed the symmetry-imposed map for subsequent structural analysis.

The structures of the wild type GroEL-GroES-ADP complex and the EL43Py398A-GroES-ATP complex bear substantial similarities at ~ 9 Å resolution, except in three locations. First, the position of the GroEL apical domains and the orientation of GroES, are slightly shifted in the EL43Py398A-GroES-ATP complex (an R2-ES complex; see below). Secondly, substantial additional density protrudes into the chaperonin cavity from the end of a stem loop (amino acids 34 - 52) at the base of the EL43Py398A-GroES cis cavity (Figure 2.3B-2D and 2.6D). This density emanates in part from the expected attachment point of the pyrene dyes at position 43, which resides at the tip of the stem loop (Figure 2.3B). However, the density near this position is larger than can be accounted for by the dye molecule alone (Figure 2.3D). Thirdly, significant density rises up from the bottom of each GroEL subunit (large red arrow in Figures 2.3C, 2.3E and 2.6E), beyond the last crystallographically resolved residue at position 525, toward the dye attachment position. The location of this additional density is consistent with the normally flexible C-terminal tails of the GroEL subunits, which extend from residue 526 to the C-terminus (a total of 23 amino acids), rising from the bottom of the GroEL subunits and interacting with the pyrene dyes attached to the protruding stem loop (Figure 2.3C and 2.3D, Figure 2.6E). Additional density is also apparent in the trans ring of the complex (Figure 2.6E), though the density in this ring is somewhat more complex than that observed in the cis ring, and may suggest that the

GroEL C-terminal tails in the trans ring make contact with both the pyrene dyes and the apical domains.

Visualizing an encapsulated non-native protein. We next examined the conformation of the EL43Py398A-GroES-ATP complex in the presence of a non-native substrate protein. To accomplish this goal, we modified our original preparation protocol to add the well-characterized GroEL-dependent substrate protein RuBisCO. In brief, EL43Py398A was first mixed with non-native RuBisCO to form a binary complex. The EL43Py398A-RuBisCO binary complex was then mixed with limiting ATP and GroES, which results in the formation of multiple species, including the bullet-shaped EL43Py398A-RuBisCO-GroES-ATP complex, both with and without non-native RuBisCO inside the cis chamber. After cryo-EM imaging (Figure 2.9A) and heterogeneity sorting of the particle images, we examined the first of two major sub-populations of bullet-shaped particle images, which were used to produce both symmetry-imposed and symmetry-free maps. The symmetry-imposed map for this RuBisCO-free sub-population, which is devoid of density within either cis or trans cavities (Figures 2.9B-2.9C), is very similar to that of the empty EL43Py398A-GroES-ATP complex (Figure 2.3A). Additionally, the unwrapped density in the symmetry-free reconstruction from the same set of RuBisCO-free particle images shows that the complex retains 7-fold symmetry (Figure 2.9D).

We examined the second major sub-population of bullet-shaped particle images to generate a 9.2 Å symmetry-free reconstruction of the bullet-shaped complex (Figure 2.4A and Figure 2.7A). Remarkably, this map displays strong density within the cis

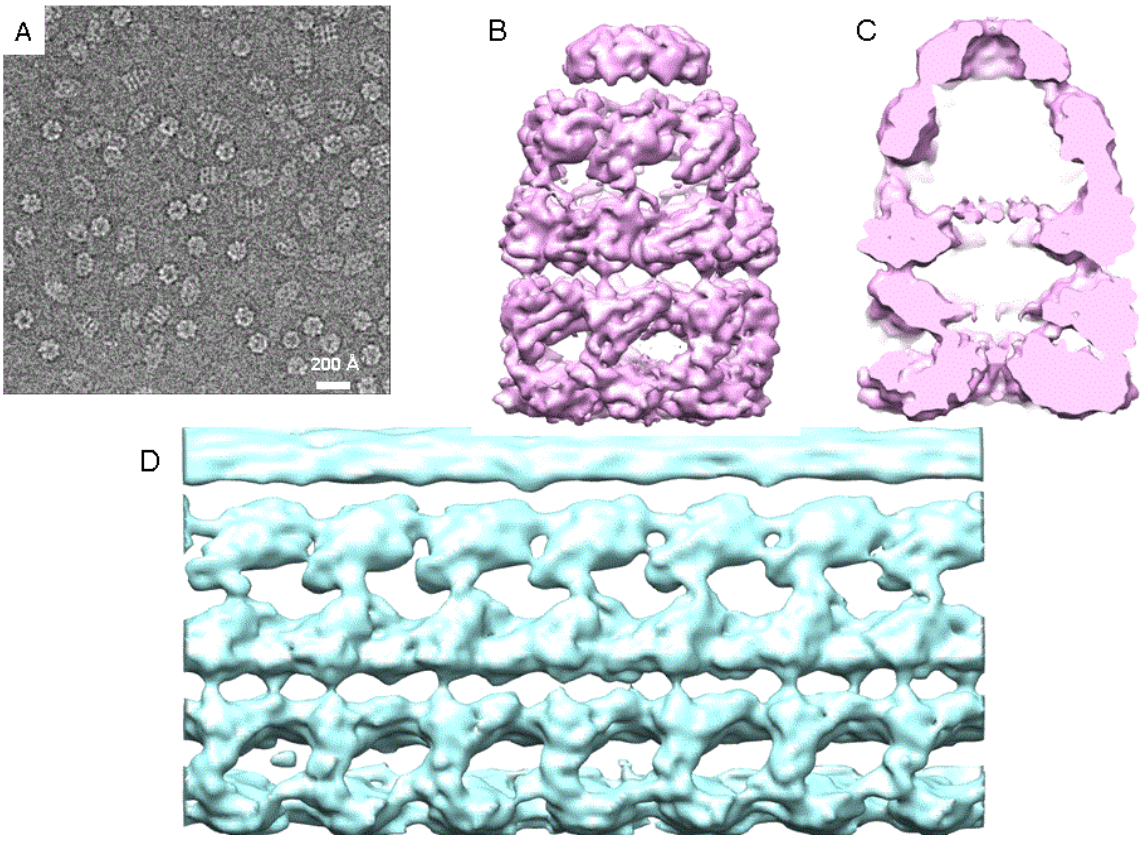
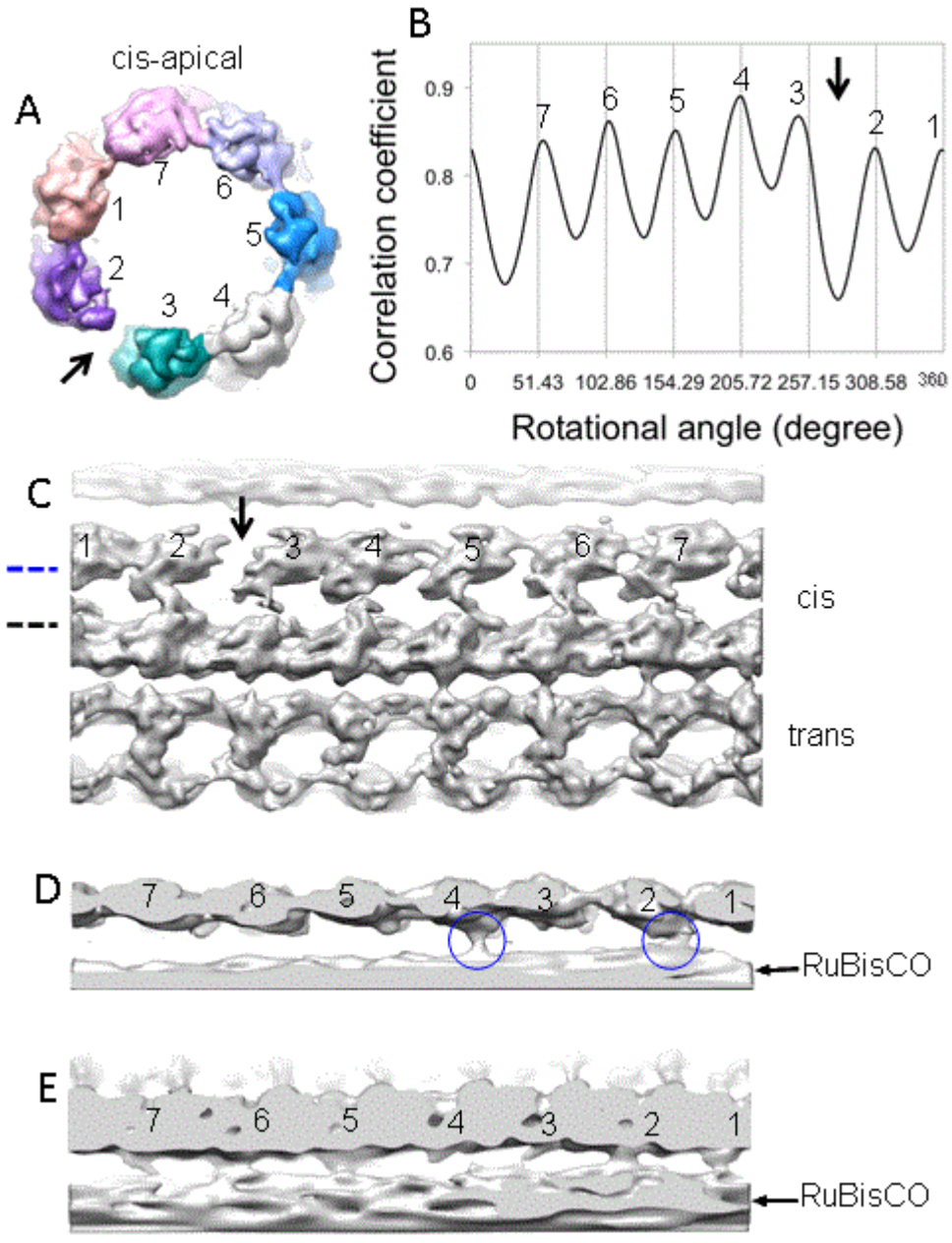


Figure 2.9: Reconstructions of a RuBisCO-Free, Bullet-Shaped Subpopulation from the EL43Py398A+RuBisCO+GroES+ATP Sample. (A) A typical cryo-EM image of the complex at a defocus of 3.1 mm. (B and C) C7 symmetry was imposed for these reconstructions to achieve subnanometer resolution. The isosurface thresholds for the whole map (B) and the cutaway map (C) were 1.2 s and 0.9 s, respectively. Note that (B) is very similar to Figure 2.3A and no substrate protein density was observed from (C). (D) Side view of the unwrapped map for a separate, symmetry-free reconstruction of the same RuBisCO-free, bullet-shaped sub-population at 10 Å resolution. The isosurface threshold was 1.1 s. Note that there is no apparent symmetry break-down among the subunits within the cis ring.

cavity, most likely from the non-native RuBisCO monomer trapped within the stalled R2 complex (gold in Figure 2.4A). The estimated mass of the visible RuBisCO monomer at a contour level of 1.0 σ is ~ 35 kDa, representing roughly 70% of the native RuBisCO monomer mass, assuming the central density comes from the RuBisCO alone. However, no regular secondary structural elements could be defined in the putative RuBisCO density either visually or quantitatively (based on SSEHunter; [327]), and no fragment of the RuBisCO crystal structure could be docked convincingly into the density.

The presence of the non-native RuBisCO within the R2 cavity also alters the structure of the GroEL-GroES complex itself. The rotational symmetry of the EL43Py398A apical domains, on both the cis and trans rings, is broken in the presence of non-native RuBisCO (Figure 2.10A-2.10C and Figure 2.7B), with a gap appearing between the apical domains of two cis-ring neighboring subunits (Figure 2.10A and 2.10C). Interestingly, the point at which the cis ring appears to break 7-fold rotational symmetry coincides with direct physical interaction between the non-native RuBisCO monomer and the lower segment of the apical domains of two cis ring subunits (Figure 2.4B-2.4D; Figure 2.10C and 2.10D). The reliability of this connecting density is substantiated by its low variance in the 3D variance analysis from 100 reconstructed maps with different subsets of particle images (red in Figure 2.4D). The cis-ring equatorial domains also deviate slightly from C7 symmetry, which can be observed as differences in the separation of the equatorial domain helices between different subunits (Figure 2.7C).

Figure 2.10: The C7 Symmetry of the GroEL Cis-Ring is Broken in the EL43Py398A-RuBisCO-GroES-ATP Complex near Points of Contact between the Non-Native RuBisCO and the GroEL Cavity Wall. (A) Cis-ring apical domains of the EL43Py398A-RuBisCO-GroES-ATP structure are shown as in Figure 2.4A, viewed from the top of the cis ring. A gap (black arrow) in the ring density is observed between subunit 2 (purple) and subunit 3 (dark cyan). (B) The cross correlation coefficient between the map of the cis-ring apical domains and a symmetric reference indicates subunit 4 is closer to subunit 3, which is approximately 9 degrees off its C7 symmetrical position, leaving a gap between subunits 3 and 2 (black arrow). (C) The gap (black arrow) between two neighboring GroEL subunits is shown in an unwrapped, planar display from the outside of the 9.2-Å density map of EL43Py398A-RuBisCO-GroES-ATP, as viewed from the side. (D) Top-view slice of the planar map, through the lower region of the cis ring apical domains (panel C, blue dashed line) shows interactions between the non-native RuBisCO monomer and the lower aspect of the GroEL apical domains of subunits 2 and 4 (blue circles). (E) Top- view slice of the planar map through the upper section of the equatorial domains (panel C, dashed black line) indicates contacts with the GroEL subunits near the stem-loop region of the equatorial domain. The isosurface threshold for (D and E) is 0.9 σ.


A
cis-apical
1
2
3
4
5
6
7
B
Correlation coefficient
0.9
0.8
0.7
0.6
0
51.43
102.86
154.29
205.72
257.15
308.58
360
Rotational angle (degree)
C
cis
trans
D
RuBisCO
E
RuBisCO

A role for the GroEL C-terminal tails in protein encapsulation. A second striking feature of the R2 state revealed by the EL43Py398A-RuBisCO-GroES-ATP complex is a direct, physical contact between the non-native RuBisCO monomer and density from the base of the GroEL cavity wall (Figure 2.4B-2.4C; Figure 2.10E; Figure 2.7D). This interaction site is principally in the region of the dye-modified stem loops near amino acid 43, and some of this interaction is probably due to contact between the pyrene dyes and the non-native substrate protein (Figure 2.7D). However, in the absence of RuBisCO substantial density from the GroEL C-termini is also present in this region (Figure 2.3C and 2E; Figure 2.6E). Several studies have suggested that the C-terminal tails of the GroEL subunits play important, though poorly defined, roles in protein folding and regulation of the GroEL ATPase cycle [203,209,246,250]. Additionally, 3D variance analysis of the EL43Py398A-RuBisCO-GroES-ATP structure suggests a direct and heterogeneous interaction between the non-native RuBisCO and the C-terminal region of the GroEL subunits (Figure 2.4D). By far some of the largest 3D variance in the GroEL subunits in the EL43Py398A-RuBisCO-GroES-ATP complex appears around the equatorial domains, in regions near or containing the C-terminal tails of the GroEL subunits. While high variance regions (red in Figure 2.4D) are observed on both the cis and trans rings, in the cis ring these regions appear to be in intimate contact with the non-native substrate protein (Figure 2.4D). The putative RuBisCO density also displays high variance, suggesting that the non-native protein remains conformationally heterogeneous at this stage of the GroEL reaction cycle.

We next considered whether the observed contacts between the C-terminal tails and non-native RuBisCO require the presence of the pyrene dyes. Using a mixing protocol similar to that described above, we created a population of asymmetric complexes using the GroEL variant D398A (EL398A) that does not contain the pyrene dye. Following imaging (Figure 2.8A) and heterogeneity sorting, the bullet-shaped structure of the EL398A-RuBisCO-GroES-ATP complex with RuBisCO within the cis cavity was solved to 15.9 Å without imposing a symmetry constraint (Figure 2.5 and Figure 2.8B). Because the EL398A-GroES complex does not stall in the R2 state, but productively releases the substrate into the cis cavity and initiates folding, though it cannot hydrolyze its bound ATP [163], the EL398A-RuBisCO-GroES-ATP complex constitutes a substrate-occupied R3 state of the cis ring (Figure 2.1E).

Once again, a substantial amount of density from the RuBisCO monomer is visible within the cis cavity (Figure 2.5B). The apparent mass of the RuBisCO monomer in this complex appears to be less than that in the EL43Py398A-RuBisCO-GroES-ATP structure (Figure 2.4B and Figure 2.7D). This is likely due to the fact that the R3 cavity of the EL398A-GroES complex is fully folding active (Figure 2.1E), unlike the R2 cavity of the EL43Py398A-GroES complex (Figure 2.1D). Even though the EL398A complex was rapidly processed for cryo-EM freezing to prevent complete folding of the encapsulated RuBisCO, the initiation of folding in this complex could not be blocked at a specific step, as it is with EL43Py398A. The enclosed RuBisCO monomer will thus be a highly heterogeneous mix of both folded and non-native states, resulting in a lower resolution reconstruction. The C-terminal tails of the GroEL

subunits are also not resolved, because the map represents a heterogeneous ensemble of interactions between the RuBisCO monomer and the C-terminal tails. Nonetheless, a significant contact between the base of the GroEL cavity wall, in the region of the C-terminal tails and stem loop and the RuBisCO monomer is apparent (Figure 2.5B). To further validate this structure, a completely independent reconstruction of this structure with a different initial model in which the cis cavity was empty (a low-pass filtered X-ray structure of GroEL-GroES-ADP complex; PDB ID: 1AON) still converged well and displayed a very similar contact between the encapsulated RuBisCO and the GroEL C-terminal region and stem loops (Figure 2.8C). As expected for a released and folding competent RuBisCO monomer, this contact is less substantial than observed in the EL43Py398A complex (Figure 2.10E, Figure 2.7D and Figure 2.5B) and, more importantly, does not depend upon the presence of the pyrene dyes.

The contact between the GroEL C-terminal tails and non-native RuBisCO suggests that the C-termini play a direct and important role in ensuring efficient substrate protein encapsulation beneath GroES. In order to test this hypothesis, we generated a GroEL variant with a C-terminal truncation at the last crystallographically resolved residue (ELΔ526). A similar truncation has been previously observed to display both perturbed ATPase activity and a reduced ability to support folding of several model substrate proteins [203,209,250]. We employed this GroEL variant in a gel filtration assay designed to score the efficiency of protein encapsulation (Figure 2.11A). When non-native GFP is bound to the trans ring of a wild-type GroEL-GroES-ADP bullet, approximately half of the initially bound protein is encapsulated inside a new cis cavity

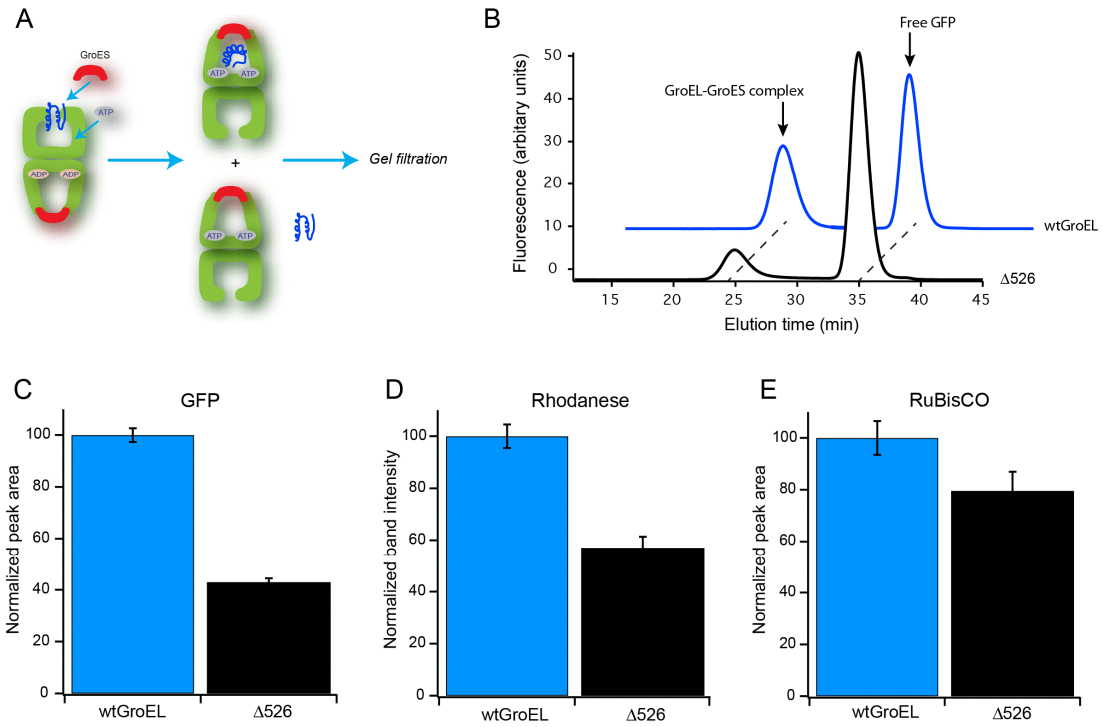


Figure 2.11: Removal of the GroEL C-Terminal Tails Results in Premature Substrate Protein Release and Reduced Encapsulation Efficiency. (A) Experimental schematic: non-native substrate protein (blue) is bound to the open trans ring of a GroEL ADP bullet complex in the presence of excess GroES. Encapsulation is initiated by the addition of ATP. ATP binding and turnover is limited to a single round by addition of hexokinase and glucose within 10 sec of ATP addition. Complexed and free substrate proteins are separated by gel filtration chromatography with an in-line fluorescence detector. (B) Example of an encapsulation experiment using GFP as the substrate protein. The positions of encapsulated GFP (GroEL-GroES complex) and released GFP (free GFP) are indicated with arrows, for both wild-type GroEL (wtGroEL) and the Δ526 truncation mutant (Δ526). Encapsulation is quantitated for three independent substrates: (C) GFP (normalized fluorescence peak area; n = 6), (D) rhodanese (normalized SDS-PAGE band intensity by densitometry; n = 4) and (E) RuBisCO (fluorescently labeled; n = 6). The reduction in encapsulation of non-native substrate protein by Δ526 GroEL relative to wtGroEL is robust: $P = 6.5 \times 10^{-9}$ for GFP, P = 0.0007 for rhodanese, and P = 0.0007 for RuBisCO (paired t-test; error bars are one standard deviation).

upon the addition of limiting ATP (Figure 2.11B). However, when the GroEL C-terminal tails are removed in the ELΔ526 variant, the efficiency of GFP encapsulation beneath GroES drops dramatically (Figure 2.11C). Similar results are obtained when the same experiment is conducted with both non-native rhodanese (Figure 2.11D) and RuBisCO (Figure 2.11E), though the drop in encapsulation efficiency is not as substantial.

Structural changes in the GroEL-GroES complex between the R2 and R3 states Following GroES binding and substrate protein encapsulation, the R2 state of the GroEL-GroES complex must then execute a shift to the R3 state, whereupon the substrate protein is released into the GroEL-GroES cavity and folding is triggered (Figure 2.1). In order to gain additional insight into this transition, we re-examined our cryo-EM data using strongly restrained flexible fitting of the atomic model of the GroEL-GroES-ADP complex (PDB ID: 1AON) into our cryo-EM maps (Figures 2.3A, 2.4A and 2.5A) with the program DireX [320]. The models generated from each flexible fitting analysis were then compared with each other in an attempt to isolate the conformational changes that lead from the R2 to the R3 state of the GroEL-GroES complex.

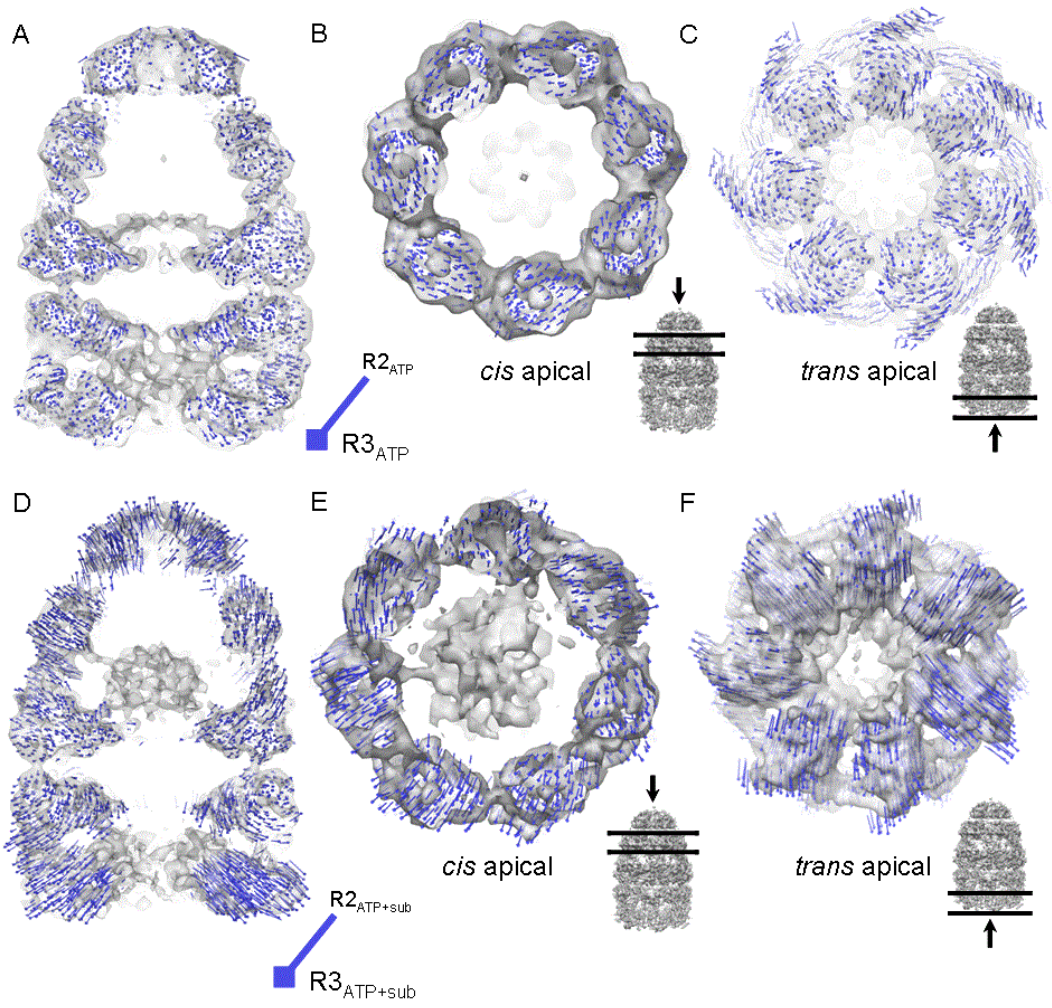
We first sought to identify structural differences between the R2 and R3 complexes that do not depend on the presence of the non-native substrate protein. This was accomplished by comparing the EL43Py398A-GroES-ATP complex, representing an empty R2 complex, to a previously described EL398A-GroES-ATP complex [328], representing, in principle, an empty R3 complex (Figure 2.1E). Surprisingly, the

conformational differences between these two complexes are relatively small. The cis ring apical domains appear to display slight counter-clockwise rotations (1-2 degrees) within the plane of the ring, but the overall position and elevation of the apical domains do not appear to change substantially (Figure 2.12A and 2.12B). Likewise, the equatorial domains show almost no movement at the current resolution (Figure 2.12A). The position and conformation of the GroES heptamer also appear mostly unchanged (Figure 2.13A). The lack of substantial conformational differences between these two complexes suggests that either (1) the conformation we observe for the EL43Py398A-GroES-ATP complex is further along the R2-to-R3 transition than expected or (2) that the detailed conformational properties of either the R2 or R3 state are not observable or stable in the absence of the non-native substrate protein.

To address these questions, we compared the empty R2 complex (EL43Py398A-GroES-ATP from Figure 2.3A) with the substrate protein occupied R2 complex (EL43Py398A-RuBisCO-GroES-ATP from Figure 2.4A). As shown in Figure 2.14, only slight differences between the two complexes are apparent, with the notable exception of the disruption of rotational symmetry and local structural shifts in the apical domains that make direct contact with the non-native substrate protein. These observations suggest that the global conformation of the R2 state of the EL43Py398A cis ring is stable in the presence of the substrate protein.

We next examined whether conformational differences between an R2 and R3 complex can be detected when the cis cavity is occupied by a substrate protein. Strikingly, the substrate-occupied R3 complex (EL398A-RuBisCO-GroES-ATP from

Figure 2.12: The Transition from the R2 to the R3 State in the Presence of RuBisCO Involves Large Structural Rearrangements of Both the Cis and Trans Rings.
(A-C) Atomic models of the GroEL-GroES complex (PDB ID: 1AON) were refined against density maps of the empty EL43Py398A-GroES-ATP ($R2_{ATP}$; Figure 2.3A) and EL398A-GroES-ATP ($R3_{ATP}$; [328]) complexes. (D-F) Atomic models of the GroEL-GroES complex (PDB ID: 1AON) were refined against density maps of the EL43Py398A-RuBisCO-GroES-ATP complex ($R2_{ATP + sub}$; Figure 2.4A) and the EL398A-RuBisCO-GroES-ATP ($R3_{ATP + sub}$; Figure 2.5A). (A) Side view of the EL43Py398A-GroES-ATP density map: structural shifts associated with movement from an empty R2 complex to an empty R3 complex are illustrated with a field of difference vectors (*blue lines and dots*) to indicate the change in Cα-positions from R2 (start) to R3 (end; *square*). Vector lengths are scaled by a factor of 2 to improve visibility. (B) View of structural changes in the *cis* apical domains and (C) *trans* ring apical domains. (D) Side view of the EL43Py398A-RuBisCO-GroES-ATP density map: structural shifts indicating the differences between the substrate-occupied R2 and R3 complexes. (E) View of structural changes in the *cis* apical domains, and (F) the *trans* ring apical domains. For (B), (C), (E) and (F) the viewing direction and selected slice density are indicated by the black arrow and horizontal lines on the GroEL-GroES density map shown in the inset, to the lower right. In all cases, strongly restrained flexible model refinement was carried out with *DireX*. The designations R2 and R3 reference the functional allosteric states of the GroEL ring, illustrated in Figure 2.1.


A
B
C
R2ATP
R3ATP
cis apical
trans apical
D
E
F
R2ATP+sub
R3ATP+sub
cis apical
trans apical

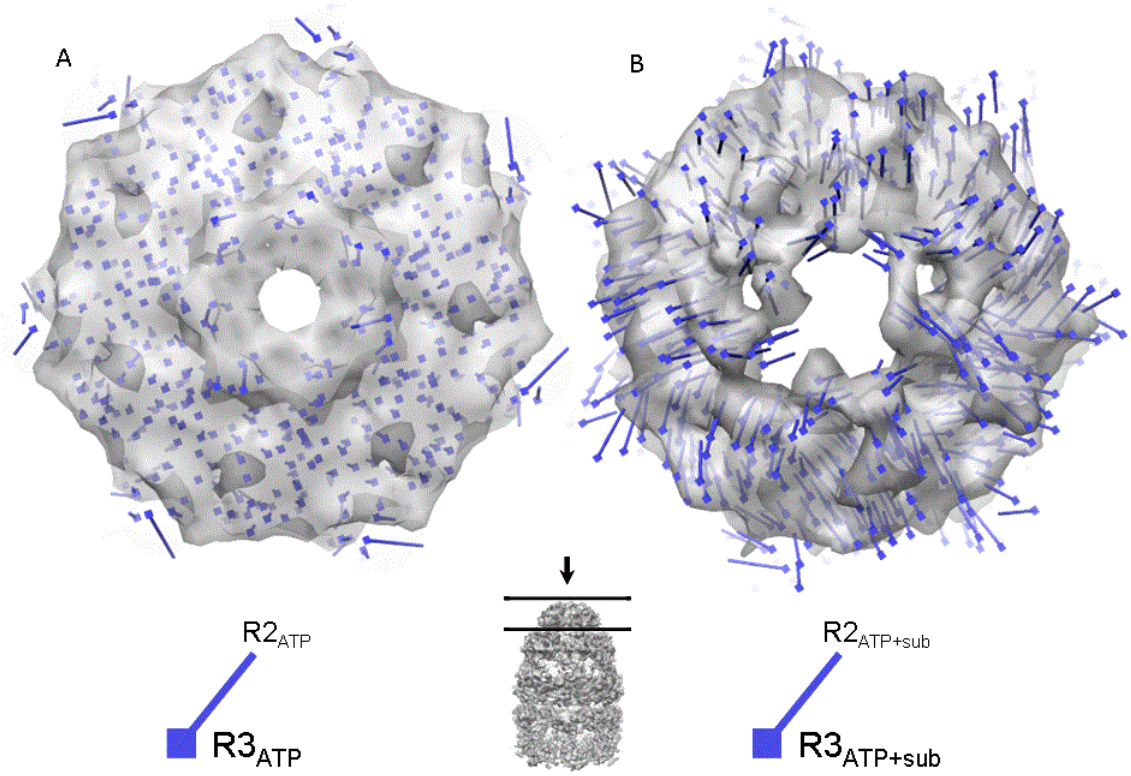


Figure 2.13: The Orientation of GroES Shifts as the Cis Ring Switches between Allosteric States. (A) The conformational differences of the GroES heptamer between an empty R2 and an empty R3 state are shown. The viewing direction and selected density region for each comparison are shown by the black arrow and horizontal lines on the GroEL-GroES density map in the lower middle panel. (B) The change in conformation of the GroES heptamer as the RuBisCO-occupied cis cavity transitions from the R2 to the R3 state. As in Figure 2.12, the conformational shift for each comparison is illustrated with a set of difference vectors (blue lines) illustrating the change in Ca-positions from the starting point to the ending point (square). The models used to generate each difference vector field, as well as the superimposed density maps shown, are the same as in Figure 2.12.

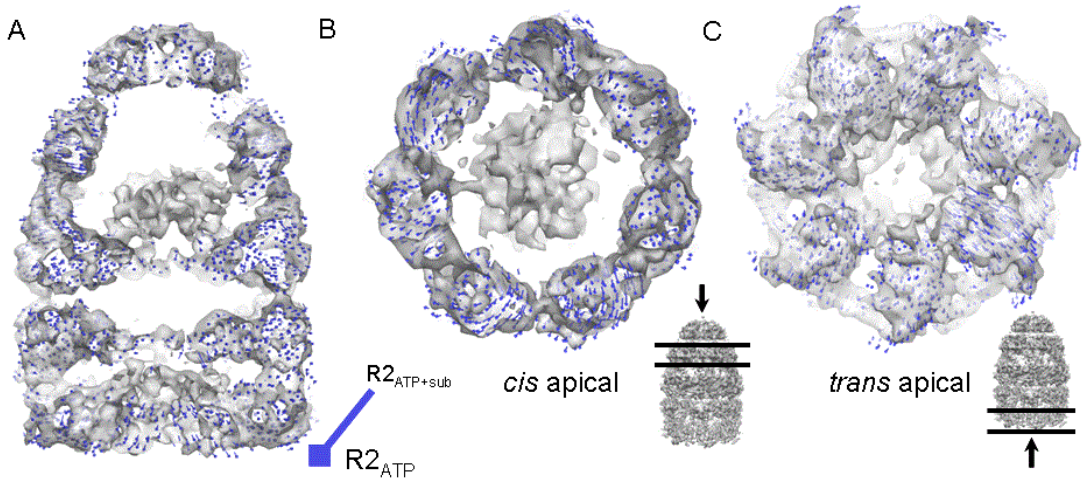


Figure 2.14: Nonnative RuBisCO Induces Modest, Asymmetric Changes in the Apical Domains of the R2 State of the GroEL-GroES Complex. (Related to Figure 2.12) (A–C) Atomic models of the GroEL-GroES complex (PDB ID code 1AON) were refined against density maps of the EL43Py398A-GroES-ATP (R2ATP) and EL43Py398A-RuBisCO-GroES-ATP (R2ATP + sub) complexes. Flexible model refinement was carried out with DireX. (A) A side view of the EL43Py398A-RuBisCO-GroES-ATP density map is displayed, superimposed on a vector field showing the difference between the empty (R2ATP) and substrate protein-occupied (R2ATP + sub) complexes. The conformational differences between the empty and occupied complexes are illustrated with a set of difference vectors (blue lines and dots) showing the change in Ca-positions from the starting point to the ending point (square). The vector lengths are scaled by a factor of 2 for improved visibility. (B) A view of structural changes in the cis apical domains is shown, with the viewing direction and selected slice density indicated by the black arrow and horizontal lines on the GroEL-GroES density map shown to the lower right. (C) A view of the structural changes in the trans-ring apical domains is shown, with the viewing direction and selected slice density indicated by the black arrow and horizontal lines on the GroEL-GroES density map shown to the lower right.

Figure 2.5) shows substantial rearrangements compared to the substrate-occupied R2 complex (EL43Py398A-RuBisCO-GroES-ATP from Figure 2.4; see Figure 2.12D-2.12F). The cis apical domains of the substrate-occupied R3 ring display sizable outward tilts and elevations (Figure 2.12D-2.12E), increasing the cavity volume compared to the R2 ring, as well as shifting the position of the bound GroES heptamer upward (Figure 2.12D). This tilt and elevation are also associated with a small clockwise rotation of the apical domains within the plane of the ring (Figure 2.12E). The conformation of the bound GroES heptamer also changes, with the average position of the GroES subunits shifting outward in concert with the apical domains, resulting in a somewhat larger opening in the GroES dome orifice (Figure 2.13B). The equatorial domains display only small movements, which are most notable as an outward shift in the region near the C-termini (Figure 2.12D). The collective movements of the R3 cis ring thus appear poised to peel away the remaining contacts between the non-native substrate protein and the apical domains and C-terminal tails. Notably, both release events occur once GroES is already bound. The final elevation and rotation of the apical domains in the R3 state are likely responsible for locking GroES into its highest affinity state for the ATP-bound GroEL ring [163,171].

Discussion

Conformational properties of an encapsulated folding intermediate. The structure of the EL43Py398A-RuBisCO-GroES-ATP complex provides the first view of a protein folding intermediate inside the GroEL-GroES cavity (Figure 2.4). While

earlier studies have visualized non-native proteins bound to an open GroEL ring [312,313,315], as well as fully folded proteins inside the GroEL-GroES cavity [311,314], the RuBisCO folding intermediate we describe here exists in transition between the two. Earlier work demonstrated that the RuBisCO monomer upon which GroEL operates is likely a middle- to late-stage protein folding intermediate [166,168,329], where the polypeptide chain has collapsed but is not as compact as the native state, and which possesses significant secondary structure, but poorly organized and highly heterogeneous tertiary structure. We examined the RuBisCO density in the EL43Py398A-RuBisCO-GroES-ATP complex for recognizable structural elements of native RuBisCO. Secondary structural elements should theoretically be identifiable in this structure, given the sub-nanometer resolution of the entire reconstruction. For example, helices in the equatorial domain of the GroEL subunits can be readily assigned (Figure 2.7C). However, our analysis failed to objectively identify any native state secondary structural elements. The lack of identifiable secondary structural elements in the putative RuBisCO region is consistent with the RuBisCO monomer populating a heterogeneous ensemble of collapsed and partially organized states (Figure 2.4A).

Coordinated action of the GroEL apical domains and C-termini. Our results suggest that direct contact between non-native substrate proteins and the C-terminal tails of the GroEL subunits helps prevent premature substrate protein escape during encapsulation beneath GroES. However, this interaction alone cannot fully explain efficient encapsulation on an R2 ring. Indeed, most of the non-native RuBisCO and rhodanese are still correctly captured in the absence of the C-termini (Figure 2.11D-

2.11E), suggesting that contacts between the GroEL apical domains and non-native RuBisCO must also play an important role. The structure of the R2 cavity in the EL43Py398A-RuBisCO-GroES-ATP complex provides strong evidence for this mechanism. As shown in Figure 2.4 and Figure 2.10, the non-native RuBisCO makes direct, physical contact with the lower section of two cis apical domains in the region of Phe 281, a segment of the inner apical domain previously identified as important for substrate protein encapsulation and folding [165]. Simultaneous binding of the non-native substrate protein by both the C-terminal tails and the lower segment of the cis apical domains could thus provide a mechanism for retaining the non-native substrate protein while the GroEL ring shifts into the R2 state to permit loading of GroES.

How GroES makes initial contact with a GroEL ring already occupied by a large and bulky non-native substrate protein remains unclear. The earliest stages of the encapsulation reaction undoubtedly follow an ATP-driven elevation and movement of the GroEL apical domains, structural shifts that are capable of mechanically unfolding the bound substrate protein [166,315]. However, only a subset of the apical domains must maintain contact with the non-native protein during the encapsulation reaction [330]. This observation suggests that, in the earliest stages of contact between GroES and a substrate-occupied GroEL ring, apical domains not in direct contact with the substrate protein are the ones employed to initially capture GroES. Such a loading mechanism would likely require that the cooperative interactions between the apical domains in the R2 cis ring be relaxed or partially uncoupled, in order for different apical domains to bind to two distinct ligands in different positions. In support of this idea, we

find that the substrate-occupied EL43Py398A-GroES complex breaks C7 rotational symmetry (Figure 2.10A-2.10C).

Effects of non-native substrate protein on inter ring allostery. While changes in the cis ring complex are essential for the progression of the GroEL folding cycle, the trans ring also plays a central role. Substrate proteins first enter the GroEL reaction cycle on the open trans ring of the asymmetric GroEL-GroES complex, and the nucleotide state of each ring directly influences the functional state of the other ring [163,166,331,332]. For example, the presence of ATP on one ring inhibits ATP binding to the other ring (negative cooperativity) and the presence of ADP on one ring, while permitting ATP to bind to the second ring, nonetheless non-competitively inhibits ATP hydrolysis on the other ring [156,309,310]. These trans ring effects are thought to be essential for imposing the ring-ring asymmetry needed for the GroEL-GroES machine to function as a two-stroke motor [156,157,309,333,334]. While the structural nature of this ring-ring allostery remains incompletely understood, our flexible fitting analysis of different GroEL-GroES complexes provides insight into structural changes imposed on the trans ring by the ligand status of the cis ring.

The occupancy of a cis ring R2 cavity by non-native RuBisCO appears to be communicated to the trans ring through substantial and asymmetric displacements of the trans ring apical domains (Figure 2.12C, 2.12F and Figure 2.7B). The apical domains of the R2 complex trans ring appear to be drawn inward, resulting in a smaller ring opening (Figure 2.12C and 2.12F). This change involves both counter-clockwise rotations and outward tilting of the trans ring apical domains (Figure 2.12C and 2.12F), resulting in a

reordering the cavity-facing apical surface. Interestingly, the conformational shift of the trans ring is different in detail when RuBisCO is present in the cis cavity, with the magnitude of the apical domain movement in the trans ring being considerably larger, and the extent of domain rotation being much smaller (Figure 2.12C and 2.12F). The observed closing down of the trans ring opening in the R2 complex, both with and without non-native protein in the cis cavity, could provide a mechanism to prevent non-native substrate proteins from binding to the trans ring until the substrate protein inside the cis complex is committed to release and folding.

A model for substrate protein encapsulation, release and folding. The observations described here thus suggest a multi-step model for substrate protein encapsulation, release and folding. Initial capture of a non-native substrate protein on the apical face of a GroEL ring is accompanied by additional binding contacts between the substrate protein and the C-terminal tails of the GroEL subunits. Subsequent binding of ATP to the GroEL ring initiates the movement of the GroEL apical domains, weakening the interaction between the non-native substrate protein and the apical domains [151,176,335]. Binding contacts between the non-native substrate protein and the C-terminal tails at the base of the cavity serve to reduce the probability of premature substrate protein escape as the apical domains move to accommodate GroES. Our structural analysis of the EL43Py398A complexes further suggests that population of the GroES acceptor state (the R2 state) requires an intermediate arrangement of the GroEL apical domains (Figure 2.12). The consequence of this altered apical position involves a shift in the binding position of the GroES heptamer and the simultaneous exposure of a

partial binding surface for the non-native substrate protein at the bottom of the apical domains (Figure 2.12). A subsequent allosteric transition of the GroEL-GroES cavity to the R3 state of the ring then results in a shift of the apical domains to their high-affinity state for GroES, fully occluding the apical binding surface and ejecting the non-native substrate protein from the apical face [158,171,172]. Coordinated movements in the C-terminal regions of the equatorial domains serve to draw the C-terminal tails away from the substrate protein, resulting in full release of the non-native substrate protein and the initiation of folding. However, this release from the C-termini does not appear to be total, as the C-terminal tails continue to make ongoing, though reduced, physical contacts with the substrate protein following release and the initiation of folding (Figure 2.5). Whether these ongoing contacts directly influence the folding of a substrate protein remains to be determined.

CHAPTER III

THE C-TERMINAL TAILS OF THE BACTERIAL CHAPERONIN GROEL STIMULATE PROTEIN FOLDING BY DIRECTLY ALTERING THE CONFORMATION OF A SUBSTRATE PROTEIN*

To function, most proteins must fold into specific three-dimensional structures. Although the native conformation of a protein is ultimately governed by the thermodynamics of its amino acid sequence in aqueous solution, protein folding is often prone to errors [62,82]. Side reactions, like misfolding and aggregation, frequently occur and can be especially serious for large and topologically complex proteins inside the concentrated interior of a living cell [191,336]. Fundamentally, the twin problems of misfolding and aggregation are kinetic in nature and biologically solved by the early evolution of several families of specialized machines known as molecular chaperones [337]. In general, molecular chaperones prevent or correct folding and assembly errors and thereby permit proteins to attain the native states thermodynamically encoded in their sequences [239]. Molecular chaperones are thus kinetic editors of protein folding reactions.

Within the network of molecular chaperones that maintain cellular protein homeostasis, the Hsp60s or chaperonins occupy a central and essential hub

*Reproduced, with modifications, with permission from Weaver J, Rye HS (2014) J Biol Chem 289: 23219-23232. Copyright 2014 American Society for Biochemistry and Molecular Biology, Inc. For the original publication: JW conceived and designed experiments, performed experiments, analyzed data, and contributed to the writing of the manuscript.

[123,198,239]. The Hsp60s are ancient and widespread and are present in virtually every organism currently known. The chaperonin system of Escherichia coli, GroEL-GroES, is perhaps the best studied example of this molecular chaperone family [190,198,338]. GroEL is a homotetradecamer composed of 57-kDa monomers arranged in two stacked, heptameric rings [152]. Each GroEL ring contains a large solvent-filled cavity and the upper cavity-facing surface of each ring is lined with hydrophobic amino acids that capture incompletely folded substrate proteins (non-native proteins) [152,165]. Proteomic surveys suggest that ~80–100 E. coli proteins possess an obligate dependence on GroEL for folding, with an additional larger number of proteins gaining an intermediate level of assistance from the GroEL-GroES system [89,90].

Following capture, most substrate proteins are encapsulated within a sealed cavity formed between the GroEL ring and the smaller lid-like GroES co-chaperonin, a heptamer of 10-kDa subunits [154,163,177,206,308]. The assembly of the GroEL-GroES folding cavity results in the initiation of protein folding by release and confinement of the substrate protein inside the privileged volume of the GroEL-GroES chamber [163]. The formation of the GroEL-GroES folding cavity is a highly ordered process, in which binding of the non-native protein on the open trans ring of a GroEL-GroES complex is followed by the obligate binding of ATP and then GroES to the same ring [157-159,163,166]. The encapsulated protein can persist and fold within the GroEL-GroES cavity for a brief period of ~5–25 s, depending on conditions [157,159,309]. Hydrolysis of the ATP inside the GroEL-GroES cavity (the cis cavity) prepares the complex for disassembly upon binding of substrate protein and ATP to the

second trans ring [157]. Disassembly of the GroEL-GroES cavity results in the release of the full population of enclosed substrate protein, folded or not [164,173,339]. Folding intermediates that do not commit to their native states within the lifetime of the GroEL-GroES cis cavity and that cannot complete folding in free solution must be recaptured for another round of processing. Thus, under the biologically relevant conditions of steady-state ATP turnover, the GroEL-GroES machine proceeds through a highly dynamic reaction cycle, the timing of which is ultimately set by the rate of ATP hydrolysis [157,159,309,340].

Despite over 2 decades of effort, the precise manner in which GroEL facilitates protein folding remains controversial. Several mechanisms have been proposed, which cluster into two classes based upon whether GroEL is postulated to act passively or actively [190,198,239,338]. The prevailing passive mechanism, referred to as the Anfinsen cage model, assumes that the folding of GroEL-dependent proteins is, in general, only limited by the tendency of on-pathway folding intermediates to aggregate. In this view, GroEL facilitates folding by simply binding and sequestering aggregation-prone intermediates, blocking aggregation, and thereby allowing the inherent thermodynamic drive programmed into the protein sequence to express itself unencumbered [192,198,338]. Active mechanisms, by contrast, accept the possibility that GroEL-dependent substrate proteins populate off-pathway states that have no direct access to the native state. Such misfolded conformations are also likely to be highly aggregation-prone but, because they cannot be rescued by simple sequestration, require an additional corrective action by the chaperonin [190,239]. The nature of this corrective

mechanism remains poorly understood but has been suggested to come from either (i) repetitive unfolding and iterative annealing [187,341] or (ii) smoothing of a substrate protein's free energy landscape as a result of confinement inside the GroEL-GroES cavity, where either spatial constraints or interactions between the substrate protein and the chaperonin cavity alter the ensemble of folding intermediates, eliminating inhibitory states or favoring productive ones [190,239,338].

One reason a coherent picture of GroEL-mediated protein folding has yet to emerge stems from the broad range of proteins upon which GroEL operates. Several proteins have been shown to satisfy the conditions required for a purely passive folding mechanism [200,342-344]. However, other proteins appear to require more active participation of the GroEL-GroES system to fold [168,199,205]. In examining active mechanisms of GroEL-mediated protein folding, we have focused on the CO2-fixing enzyme from *Rhodospirillum rubrum*, RuBisCO, one of the most highly GroEL-dependent substrate proteins known. We previously demonstrated that RuBisCO populates a kinetically trapped, misfolded monomer that is efficiently rescued by GroEL in the absence of aggregation [168]. We also showed that GroEL assists RuBisCO folding, at least in part, through two phases of multiple axis unfolding of the misfolded RuBisCO monomer [159,166,173]. In addition, we observed an encapsulation-dependent compaction of the RuBisCO folding intermediate, hinting at the possibility of an active role for confinement inside the GroEL-GroES chamber [159,166,173]. Work on RuBisCO by others [199,203] also suggested that confinement of the RuBisCO monomer alters the folding landscape of the protein. In this case,

confinement was suggested to be the dominant mechanism of folding assistance, and the nature of the effect was assigned to either the restricted volume or charge character of the GroEL-GroES cavity [199,203,204,345]. More recent observations have challenged some of these conclusions, however, calling into question confinement-based active mechanisms [192,209,346].

To probe the relative contribution of unfolding versus confinement, we have re-examined the folding of RuBisCO using a variant of GroEL in which the character and volume of the GroEL cavity have been altered by removal of the unstructured C-terminal 23 amino acids of the GroEL subunits. Deletion of these C-terminal tails, which project upward from the equatorial plane of the GroEL ring into the bottom of the cavity, has no effect on the stability of the GroEL tetradecamer but has a notable impact of the folding of several proteins, as well as a modest in vivo phenotype [203,204,209,246,250]. Additionally, although the C-terminal tails are not resolved in the x-ray crystal structures of the chaperonin, using high resolution cryo-EM we recently demonstrated that the C-terminal tails interact directly with the non-native RuBisCO folding intermediate during and immediately after encapsulation by GroES [347]. At the same time, other studies have suggested that the amphipathic character of the C-terminal tails is essential for efficient folding of several proteins [250]. Still other work has suggested that the combined mass of the seven C-termini at the base of the GroEL-GroES cavity is important for restricting the volume of the cavity [203,204].

Consistent with previous observations, we find that removal of the C-terminal tails results in a several fold drop in RuBisCO folding with actively cycling

tetradecamer GroEL. However, when RuBisCO folding is confined to the GroEL-GroES cavity by using a noncycling, single ring version of GroEL, removal of the C-terminal tails displays a much smaller impact on RuBisCO folding. We show that, in part, the larger impact of tail removal on the cycling tetradecamer is due to a slowing of the GroEL reaction cycle and a large shift in a key allosteric transition of the GroEL machine. Removal of the tails also results in a modest slowing of several kinetically distinct RuBisCO folding transitions inside the GroEL-GroES chamber. Surprisingly, however, the more substantial consequence of C-terminal tail removal is a large reduction in unfolding of the RuBisCO monomer at the early stages of the GroEL reaction cycle. Reduced unfolding results in a decrease in the fraction of the RuBisCO monomers that rapidly commits to the native state. The GroEL C-terminal tails thus not only assist in substrate protein encapsulation, but directly participate in protein unfolding, a process that is required for maximally efficient folding of the highly stringent substrate protein RuBisCO.

Experimental Procedures

Proteins. Wild-type and variants of GroEL (SR1, single-cysteine mutants and C-terminal truncation mutants), wild-type and E98C GroES, and wild-type and cysteine mutants of *R. rubrum* RuBisCO were all expressed and purified as described previously [158,159,166,168].

Labeling of Proteins for FRET. RuBisCO variants were labeled as described previously [158,159,166,168], using either 5-iodoacetamidofluorescein (fluorescein, F)

and/or 5-(2-acetamidoethyl) aminonaphthalene 1-sulfonate (EDANS, ED). GroEL variants were labeled with fluorescein, and GroES-98C was labeled with EDANS as described [157,316]. All dyes were purchased from Molecular Probes (Eugene, OR) and prepared fresh in anhydrous N, N-dimethylformamide immediately before use. The extent of labeling was determined by protein quantification by the Bradford assay (Bio-Rad) and dye quantification under denaturing conditions using known molar extinction coefficients [157,316]. For RuBisCO variants, site-specific labeling was verified through denaturing anion-exchange chromatography and analysis of proteolytic fragments [316]. GroES-98ED was re-purified via anion exchange, and only protein labeled with 2 dyes/heptamer was used. All GroEL variants were labeled with 2–3 dyes/ring.

Stopped-flow Fluorescence. Stopped-flow experiments were conducted as described previously [166,168,316], using an SFM-400 rapid mixing unit (BioLogic) equipped with a custom-designed two-channel fluorescence detection system. Mixing was done using two syringes, one containing GroEL-RuBisCO complexes and one containing GroES and ATP. The donor-side FRET efficiency was calculated from matched sets of donor-only and donor-acceptor experiments as described previously [157,316].

Enzymatic Refolding. The refolding of wild-type RuBisCO was assayed by enzymatic activity as described previously [157,163]. Acid-urea-denatured RuBisCO was bound to an excess of full-length or truncated SR1. After a 10-min incubation at

25°C, GroES and ATP were added. Intra-cavity folding was quenched by addition of EDTA and incubation at 4°C, as described [159].

Steady-state and Time-resolved FRET. Steady-state fluorescence measurements were conducted with a PTI fluorometer, with temperature regulation through a jacketed cuvette holder (Neslab). Fluorescence lifetimes were measured in the time domain using a TimeMaster Fluorescence Lifetime Spectrometer (PTI), with sample excitation from a pulsed nitrogen laser coupled to an optical boxcar detector [348]. The average distance between donor and acceptor probes was calculated from donor-side FRET efficiencies, extracted from both steady-state and time-resolved data, as described previously [166]. The Förster distance, R_0, was calculated independently for each donor-acceptor pair in each complex examined. The value of κ^2 was assumed to be ⅔ for all distance measurements. In all cases, the average anisotropy of the donor and acceptor probes, at each labeling position, was the same when non-native RuBisCO was bound to either full-length or truncated GroEL (data not shown). All protein complexes were equilibrated with a 10-min incubation at 25°C prior to measurement. Note that in the absence of the C-terminal tails, and under active cycling conditions, 10-20% of the RuBisCO initially bound to a GroEL ring can escape prior to GroES binding [347]. However, under the experimental conditions employed here, in which RuBisCO-GroEL binary complexes are incubated for several minutes prior to the addition of GroES and ATP, the amount of RuBisCO that escapes from a Δ526 ring prior to encapsulation by GroES, compared with a wild-type GroEL ring, is negligible (<2%, data not shown).

Protease Protection. The protease sensitivity of non-native RuBisCO bound to a GroEL ring was conducted as described previously [158,166]. Briefly, 58F-RuBisCO was bound to asymmetric GroEL-GroES ADP bullets. Trypsin was added, and time points were taken, with the reaction stopped with PMSF. Samples were run on 10% SDS-PAGE and imaged using a Typhoon Trio (GE Healthcare).

Results

Deletion of GroEL C-Termini Slows RuBisCO Folding and Perturbs the GroEL Reaction Cycle. To examine how a change in cavity character impacts the folding of a stringent substrate protein, we created a GroEL variant in which the C-terminal 23 amino acids of the GroEL subunits have been removed. This variant, GroELΔ526–548 (hereafter Δ526), displays no detectable alteration in tetradecamer assembly or stability at room temperature. The behavior of the Δ526 variant during expression, purification, and on gel filtration chromatography is identical to wild-type GroEL (data not shown). The Δ526 variant also supports a functional chaperonin cycle, in which GroES is bound and released in an ATP hydrolysis-dependent manner. However, key properties of the GroEL hydrolytic cycle change upon removal of the C-termini. When folding of the stringent substrate protein ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) from *Rhodospirillum rubrum* is examined by tryptophan fluorescence, we observe a greater than 2-fold decrease in the apparent folding rate constant with the Δ526 variant, compared with full-length GroEL (Figure 3.1A). Notably, the yield of folded RuBisCO is not significantly different between full-length GroEL and Δ526 (Figure 3.1A, inset),

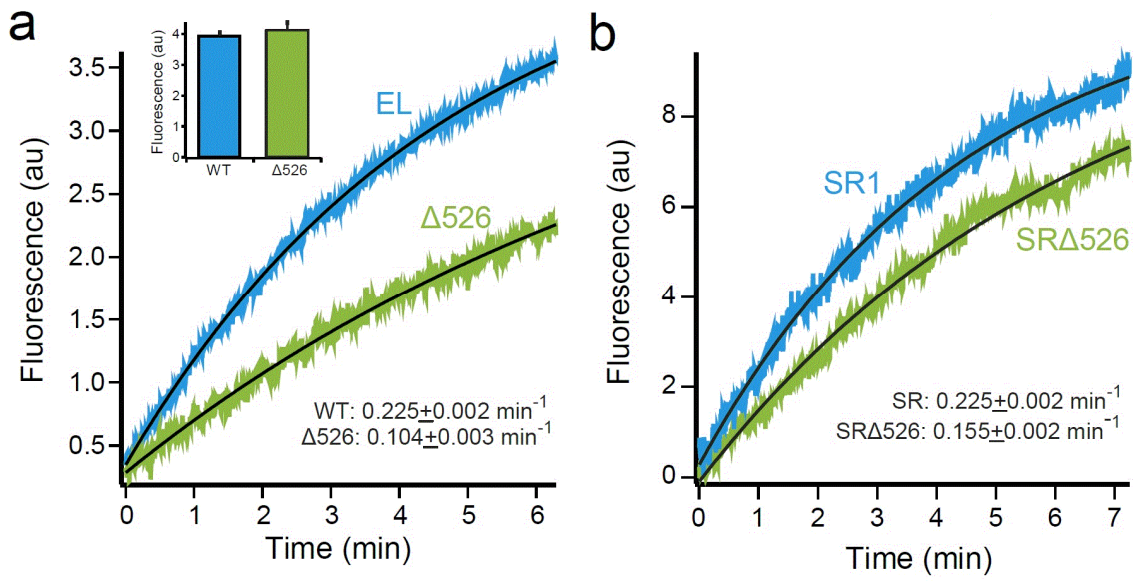


Figure 3.1: Presence of the GroEL C-Terminal Tails Enhances Protein Folding. (A) folding of RuBisCO by cycling GroEL-GroES was monitored by an increase in tryptophan fluorescence. Chemically denatured, wild-type RuBisCO (100 nM) was bound to GroEL (200 nM), wild-type (EL; blue), or C-terminal deletion (Δ526; green) and rapidly mixed with an equal volume of excess GroES (400 nM) and ATP (2 mM) in a stopped-flow apparatus. Curves were fit to a single-exponential rate law (black line), resulting in observed rate constants of 0.225 ± 0.002 min^{-1} for GroEL and 0.104 ± 0.003 min^{-1} for Δ526. In each case, the overall folding yield was examined by allowing each folding reaction to run to completion (~30 min), followed by a measurement of the total tryptophan fluorescence (inset). (B) folding of RuBisCO inside stable single ring-GroES complexes monitored by an increase in tryptophan fluorescence. Chemically denatured wild-type RuBisCO (100 nM) was bound to full length SR1 (SR, blue) or Δ526 (SRΔ526, green) single-ring variants of GroEL (300 nM) and rapidly mixed with an equal volume of excess GroES (600 nM) and ATP (2 mM) in a stopped-flow apparatus. Curves were fit to single-exponential rate laws, yielding observed rate constants of 0.225 ± 0.002 min^{-1} for SR1 and 0.155 ± 0.002 min^{-1} for SRΔ526. n = 10 replicates, with a 100-ms sampling time. AU, arbitrary units.

indicating that the observed decrease in apparent folding rate is not due to an increase in aggregation in the presence of cycling Δ526. This result is consistent with observations made using other C-terminally perturbed tetradecameric GroELs, using different assay methods [203,204,209,250].

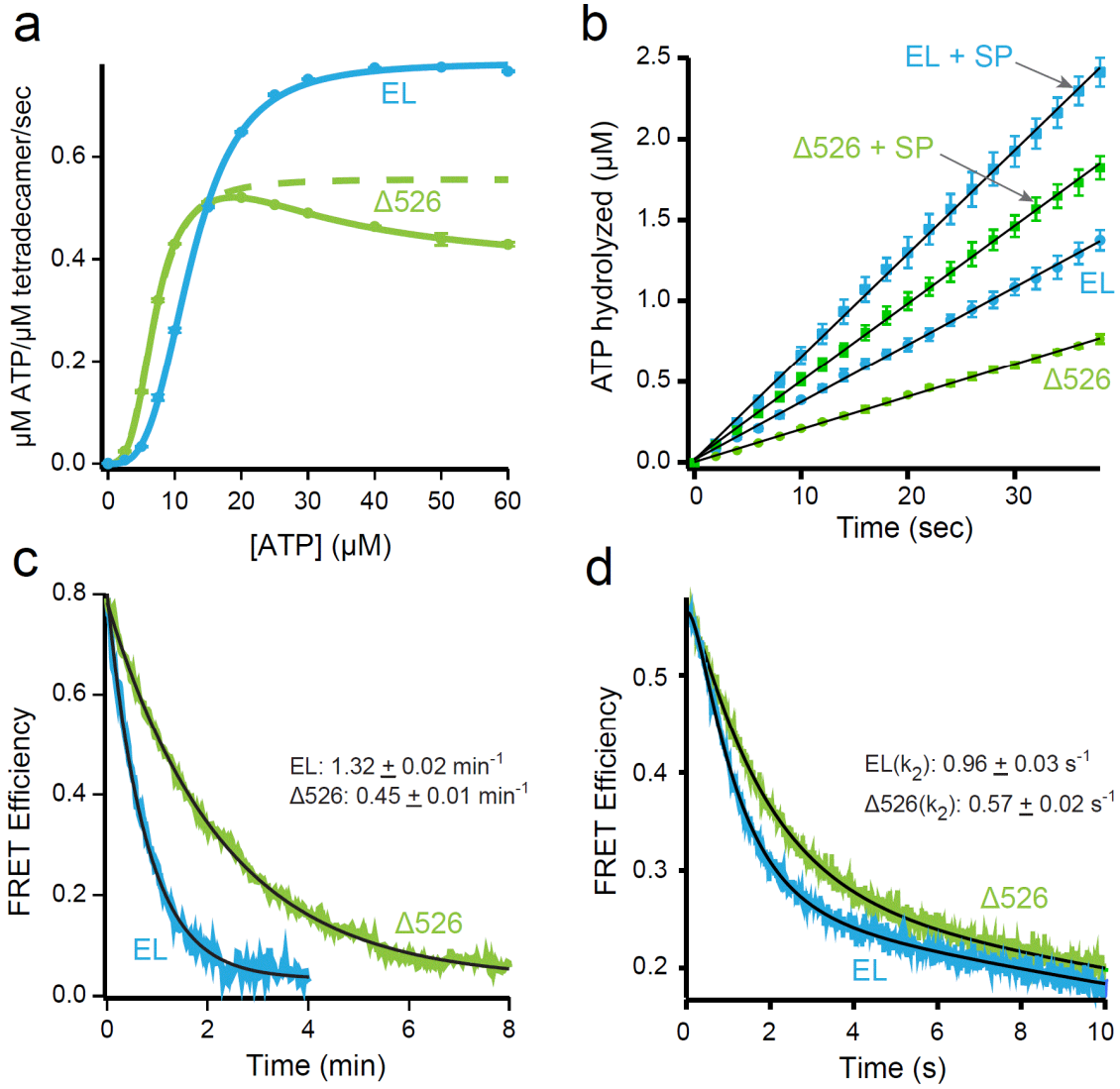
We next considered whether slower RuBisCO folding is due to a change in intra-cavity folding in the absence of the GroEL C-termini. To prevent cycling and confine RuBisCO folding to the interior of the GroEL-GroES cavity, we employed a previously described single-ring variant of GroEL, SR1 [177]. For comparison, we constructed a Δ526 variant of SR1 (hereafter SRΔ526) that is the single ring analog of tetradecameric Δ526 [203,209,250]. Binding of RuBisCO and GroES to SRΔ526 was unchanged from SR1 (data not shown). However, RuBisCO folding within the SRΔ526-ES cavity is 50–70% slower than folding inside the SR1-ES cavity (Figure 3.1B). Thus, although we do observe a drop in intra-cavity folding rate, the magnitude of this decrease cannot fully explain the greater than 2-fold drop in folding rate observed with tetradecameric Δ526.

Because progression of the GroEL ATPase cycle is essential for facilitated folding, we next examined the hydrolytic cycle of Δ526 in greater detail. Previous studies demonstrated that C-terminal GroEL truncations like Δ526 can result in a reduction of the steady-state ATP hydrolysis rate [204,209,244,250]. However, the source of this reduced turnover, as well as the impact on the overall GroEL reaction cycle, has not been assessed. The decrease in ATPase rate caused by C-terminal truncations, as well as our previous observation of accelerated ATP turnover caused by dye-induced perturbations of the C-termini [158], suggested that the C-terminal tails are

coupled to the allosteric transitions of the GroEL machine. To test this, we examined the rate of ATP hydrolysis by both GroEL and Δ526 at different ATP concentrations. With full-length GroEL, we observe the well described, positively cooperative binding of ATP to the first GroEL ring (Figure 3.2A) [174,310,349]. Fitting of this transition to the Hill equation results in values for the turnover (k_{cat} = 0.12 s^{-1} per active subunit), half-saturation point ($K_{½}$ = 12.4 μM), and Hill coefficient (n_H = 3.3) that are consistent with previous measurements. By contrast, ATP hydrolysis by Δ526 displays a substantially different response. Although Δ526 shows an initial, positively cooperative transition similar to full-length GroEL, the maximal ATPase rate attained is ~40% lower than full-length (Figure 3.2A).

Strikingly, Δ526 has a marked roll-off in ATP turnover as the ATP concentration exceeds ~20 μM. This decrease in ATP turnover is highly reminiscent of the negative cooperativity roll-off displayed by full-length GroEL, except in the case of Δ526, it occurs at a far lower ATP concentration. Positive cooperativity in ATP binding to the first GroEL ring, followed by negative cooperativity between the rings as the second ring fills with ATP is a defining feature of the nested cooperativity model of GroEL allostery [310]. The ATPase response curve of Δ526 is well described by the nested cooperativity model, with parameters for the initial, positively cooperative event (n_H = 3.8; $L_1 = 2.5 \times 10^{-3}$, where L_1 is the apparent allosteric constant for the first transition) shifted to a lower turnover rate (k_{cat} = 0.09 per active subunit) and lower half-saturation concentration ($K_{½}$ = 1.5 μM). Additionally, Δ526 displays a dramatic shift in

Figure 3.2: Stimulated Folding of RuBisCO is Not the Product of Extended GroEL-ES Cavity Lifetime. (A) full-length (EL; blue) or C-terminal deletion (Δ526; green) GroEL (200 nM) was mixed with various concentrations of ATP, and the steady-state rate of hydrolysis was measured. For GroEL, the observed change in initial rate was well fit by the Hill equation, yielding $k_{cat} = 0.12 \pm <0.01$ s^{-1} per active subunit, $K_{1/2} = 12.4 \pm 0.1$ μM, $n_H = 3.3 \pm 0.1$. The data for Δ526 at low ATP concentrations was also well fit to a two-state Hill model (green dashed line), yielding $k_{cat} = 0.08 \pm <0.01$ s^{-1} per active subunit, $K_{1/2} = 6.9 \pm 0.1$ μM, $n_H = 3.2 \pm 0.1$. The full Δ526 data set was also fit to the nested cooperativity model [310], yielding $k_{cat,1} = 0.09 \pm <0.01$ s^{-1} per active subunit; $k_{cat,2} = 0.05 \pm <0.01$ s^{-1} per active subunit, $K_{1/2} = 1.5 \pm 0.1$ μM, $n_H = 3.8 \pm 0.1$, $L_1 = 2.5 \times 10^{-3} \pm 0.1 \times 10^{-3}$, $L_2 = 9.3 \times 10^{-5} \pm 2.7 \times 10^{-5}$. (B) steady-state ATP hydrolysis by GroEL (125 nM) and Δ526 (125 nM) was measured in the presence of excess GroES (250 nM), with and without non-native denatured RuBisCO (dRub, 100 nM; SP). Addition of dRub to the GroEL-GroES system results in a hydrolysis rate enhancement of 1.8-fold (2.1 μM/min to 3.8 μM/min), consistent with previous observations using saturating levels of the substrate protein malate dehydrogenase [157]. Addition of dRub to the Δ526-GroES system results in a rate enhancement of 2.4-fold (1.2 μM/min to 2.9 μM/min). (C) lifetime of the GroEL-GroES complex in the absence of substrate protein was examined using a previously described FRET assay [157]. ATP (2mM) was added to GroES-98ED (100 nM) and either 315F- labeled full-length (EL; blue), or C-terminal deletion (Δ526; green) GroEL (125 nM). After a one min incubation, unlabeled GroES (2 μM) was introduced as a competitor. The change in FRET was calculated from matched donor-only and donor-acceptor traces. The curves were fit to a single-exponential rate law, yielding rates of 1.32 ± 0.02 min^{-1} (EL) and 0.45 ± 0.01 min^{-1} (Δ526). The average of three experiments is shown. (D) lifetime of the GroEL-GroES complex upon addition of a stoichiometric amount of non-native RuBisCO was examined using the same FRET assay as in (C) [157]. Experiential conditions are similar; except that a stopped-flow apparatus was employed, and 100 nM denatured RuBisCO was mixed with the cycling GroEL-GroES system simultaneously with excess, unlabeled GroES. The average of 10 experiments is shown. As reported previously, the observed change in FRET efficiency in the presence of non- native substrate protein requires a triple-exponential rate law for a good fit [157]. The rate of the dominant decay component (k_2), previously shown to reflect the substrate protein-induced acceleration of GroES release [157], is ~1 s^{-1} for GroEL and ~0.6 s^{-1} for Δ526.


a
μM ATP/μM tetradecamer/sec
EL
Δ526
[ATP] (μM)
b
ATP hydrolyzed (μM)
EL + SP
Δ526 + SP
EL
Δ526
Time (sec)
c
FRET Efficiency
EL: 1.32 ± 0.02 min-1
Δ526: 0.45 ± 0.01 min-1
Δ526
EL
Time (min)
d
FRET Efficiency
EL(k2): 0.96 ± 0.03 s-1
Δ526(k2): 0.57 ± 0.02 s-1
Δ526
EL
Time (s)

the value of the second allosteric coupling parameter, which describes ATP binding to the second ring (L_2= 9.3 × 10^{-5}), compared with full-length GroEL (L_2 = 6.0 × 10^{-9}, where L_2 is the apparent allosteric constant for the second transition). These observations strongly suggest that both intra- and inter-ring allostery are altered by removal of the C-termini in Δ526.

Because non-native substrate protein can accelerate the GroEL reaction cycle [157] and may also induce a shift from a cycle dominated by asymmetric bullet-shaped complexes to one dominated by symmetric football-shaped complexes [340,350-352], we examined the effect of non-native substrate protein on both ATP turnover and GroES release. As with GroEL, non-native substrate protein accelerates the rate of ATP hydrolysis by Δ526 in the presence of GroES (Figure 3.2B). However, although the magnitude of substrate-stimulated turnover is larger for Δ526 (1.8-fold for GroEL versus 2.6-fold for Δ526), the maximal turnover rate for Δ526 is still ~60% slower than GroEL. These observations are consistent with steady-state ATPase stimulation seen with GroEL and a C-terminal truncation variant using other substrate proteins [244].

The changes in ATP turnover and allostery exhibited by Δ526 suggested that this C-terminal truncation variant progresses through the canonical GroEL reaction cycle more slowly than wild-type GroEL. If true, the lifetime of the Δ526-GroES complex should lengthen compared with full-length GroEL. To test this possibility, we exploited a previously described fluorescence resonance energy transfer assay (FRET) capable of tracking the dynamics of the GroEL-GroES complex [157]. In this assay, an acceptor

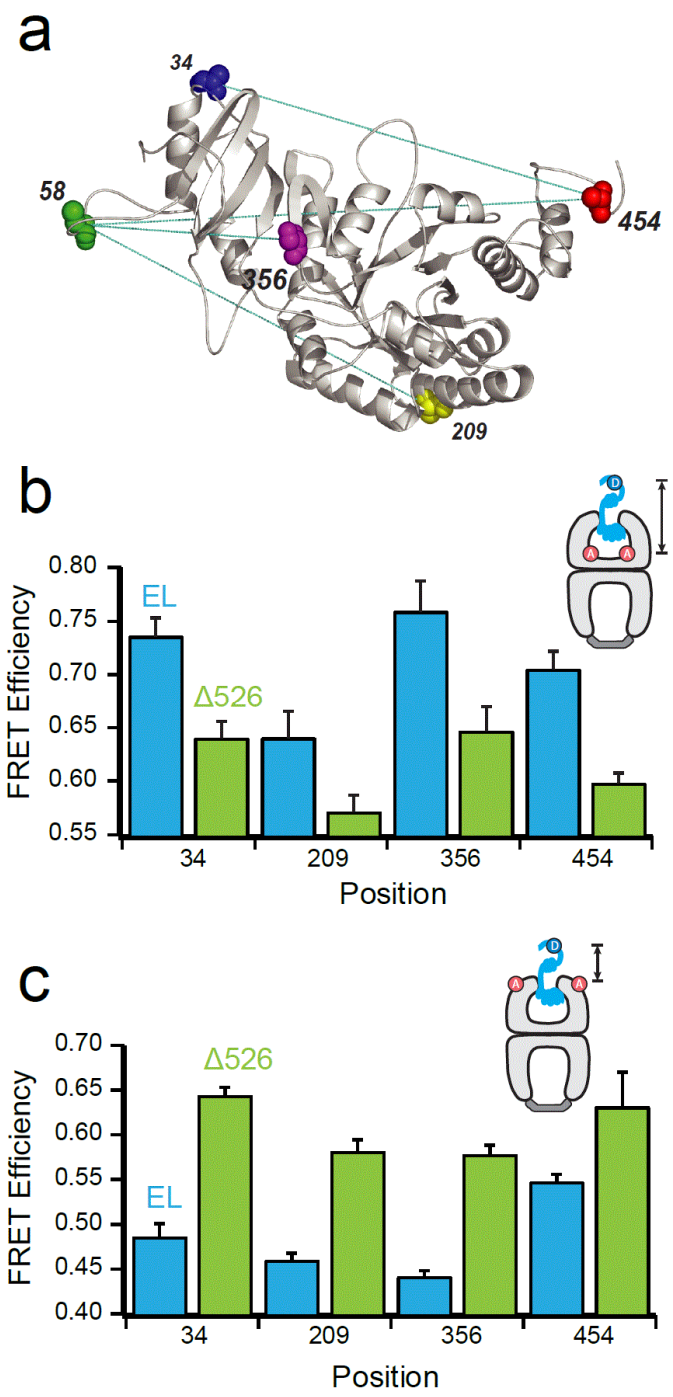
fluorescent probe was covalently attached to a nonperturbative cysteine substitution on the backside of the apical domain of either full-length GroEL or Δ526. A donor probe was attached to GroES, and the extent of complex formation between the GroEL and GroES was then observed by FRET. As shown in Figure 3.2C, the lifetime of the Δ526-GroES complex in the absence of substrate protein is substantially longer than the GroEL-GroES complex (~2.5 times) under conditions of steady-state ATP turnover. The rate of GroES release from a Δ526-GroES complex is also slower in the presence of non-native substrate protein, although the difference is not as dramatic (Figure 3.2D). Interestingly, the differences between GroEL and Δ526 in substrate protein-stimulated GroES release and ATP turnover were essentially identical (Δ526 is 60% slower). This observation suggests that the Δ526 reaction cycle is limited by the same substrate-driven release of ADP from the trans ring as is GroEL [158,160,340], but in the absence of the C-terminal tails, the system responds more slowly. In combination, these results show that the progression of the GroEL reaction cycle is coupled to the dynamics of the GroEL C-terminal tails, such that removing the tails slows down the functional GroEL cycle.

It is important to note that the stimulated ATPase rate observed with near saturating levels of non-native substrate proteins (Figure 3.2B) is unlikely to apply to the RuBisCO folding experiments described here. To minimize aggregation, the GroEL-GroES or Δ526-GroES systems are typically present in a several fold molar excess over the non-native RuBisCO in refolding experiments. Consequently, the overall rate of ATP turnover will only be slightly affected by this amount of substrate protein, and the

limiting cycling rate for both GroEL and Δ526 will approach the rate observed in the absence of substrate protein (Figure 3.2, B and C).

GroEL C-Termini Alter the Conformation of a RuBisCO Folding Intermediate. We next considered whether, at different points in the GroEL reaction cycle, the conformation of the RuBisCO monomer is altered by the presence of the C-terminal tails. We first examined whether the RuBisCO monomer is bound in the same average position on a GroEL ring in the presence and absence of the C-termini. For these FRET experiments, the donor was attached to the RuBisCO monomer (Figure 3.3A), and the acceptor was positioned either near the outer and upper edge of the ring (EL315-F) or near the bottom of the cavity (EL43-F). In all cases, the RuBisCO labeling positions were homogeneously derivatized with the donor dye, although the GroEL variants were lightly modified at ~2–3 dyes per ring. Acceptor-labeled, ADP-bullet complexes of both full-length GroEL and Δ526 were prepared and then mixed with one of four different, denatured, and donor-labeled RuBisCO variants. Matched donor-only experiments were also performed with unlabeled full-length GroEL and Δ526. Following RuBisCO binding, the extent of FRET between four different locations on the folding intermediate, relative to two positions on the GroEL tetradecamer, was determined. For all donor positions and both acceptor locations, a strong FRET signal is observed, ranging in efficiency from 0.42 to 0.75 (Figure 3.3, B and C). Robust differences between the different donor-labeled positions suggest that the RuBisCO monomer is asymmetrically bound to both the full-length GroEL and Δ526 trans rings. More striking, however, is the distinct pattern of differences that emerge when full-length GroEL and Δ526 are

Figure 3.3: Contact with the C-Termini Promotes Deeper Initial RuBisCO Binding within the GroEL Cavity. (A) structure of one monomer of the native RuBisCO dimer (PDB ID: 9RUB) is shown. Positions of five amino acid locations employed for the attachment of exogenous fluorescent probes are indicated. Except for position 58, which is a naturally occurring surface-exposed Cys residue, these positions were mutated to encode Cys and labeled with thiol-alkylating fluorescent dyes as described previously [157,316]. Sites successfully paired for intramolecular FRET assays, in which donor and acceptor dyes are attached to the same RuBisCO monomer, are indicated by dotted lines. (B) steady-state FRET measurements between chemically denatured, donor-labeled RuBisCO (100 nM) bound to the trans ring of acceptor-labeled, full length (EL; blue) or C-terminal deletion (Δ526; green) GroEL-GroES ADP bullets (120 nM). For these measurements, RuBisCO was labeled with EDANS at each of four sites (amino acids 34, 209, 356, and 454), and GroEL was labeled with fluorescein near the bottom of the cavity through a unique, introduced Cys at position 43 [353]. Error bars represent the standard deviation for n = 3 experiments. (C) steady-state FRET measurements as in (B), but using GroEL labeled with fluorescein near the top of the cavity, through a unique introduced Cys at position 315 [157].


a
34
58
454
356
209
b
EL
Δ526
FRET Efficiency
0.80
0.75
0.70
0.65
0.60
0.55
34
209
356
454
Position
c
Δ526
EL
FRET Efficiency
0.70
0.65
0.60
0.55
0.50
0.45
0.40
34
209
356
454
Position

compared.

For all RuBisCO donor positions, full-length GroEL shows a consistently higher FRET efficiency than Δ526 when the acceptor dye is located at the base of the GroEL ring (Figure 3.3B). By contrast, this pattern reverses when the acceptor dye is moved to the upper edge of the GroEL cavity, with Δ526 showing a consistently higher FRET efficiency (Figure 3.3C). These observations suggest that the RuBisCO monomer is, on average, shifted higher up the GroEL cavity when the C-termini were removed.

To investigate the effect of the C-terminal tails on RuBisCO folding in greater detail, we next employed an intramolecular FRET assay. The assay is designed to follow the conformation of a RuBisCO monomer as it transits the GroEL reaction cycle [159,166,168]. A series of RuBisCO variants were used, in which a donor probe was attached to one position and an acceptor probe attached to another (Figure 3.3A). We used four distinct FRET pairs to measure the intramolecular FRET efficiency of the RuBisCO monomer bound to the trans ring of full-length GroEL and Δ526 ADP bullets. In all cases, the donor-side FRET efficiency was determined by both steady-state and time-domain lifetime measurements. An example of experimental data from one FRET pair (209ED-58F) is shown in Figure 3.4A and 3.4B, and the FRET efficiencies and calculated distances for all four pairs are shown in Table 3.1. For each pair of sites, the measured intramolecular distance was shorter when the RuBisCO monomer is bound to a Δ526 ring, compared with a full-length GroEL ring. These measurements suggest that the non-native RuBisCO monomer is more compact and less unfolded when it is bound to a Δ526 ring. If true, this predicts that RuBisCO bound to a Δ526 ring should be more

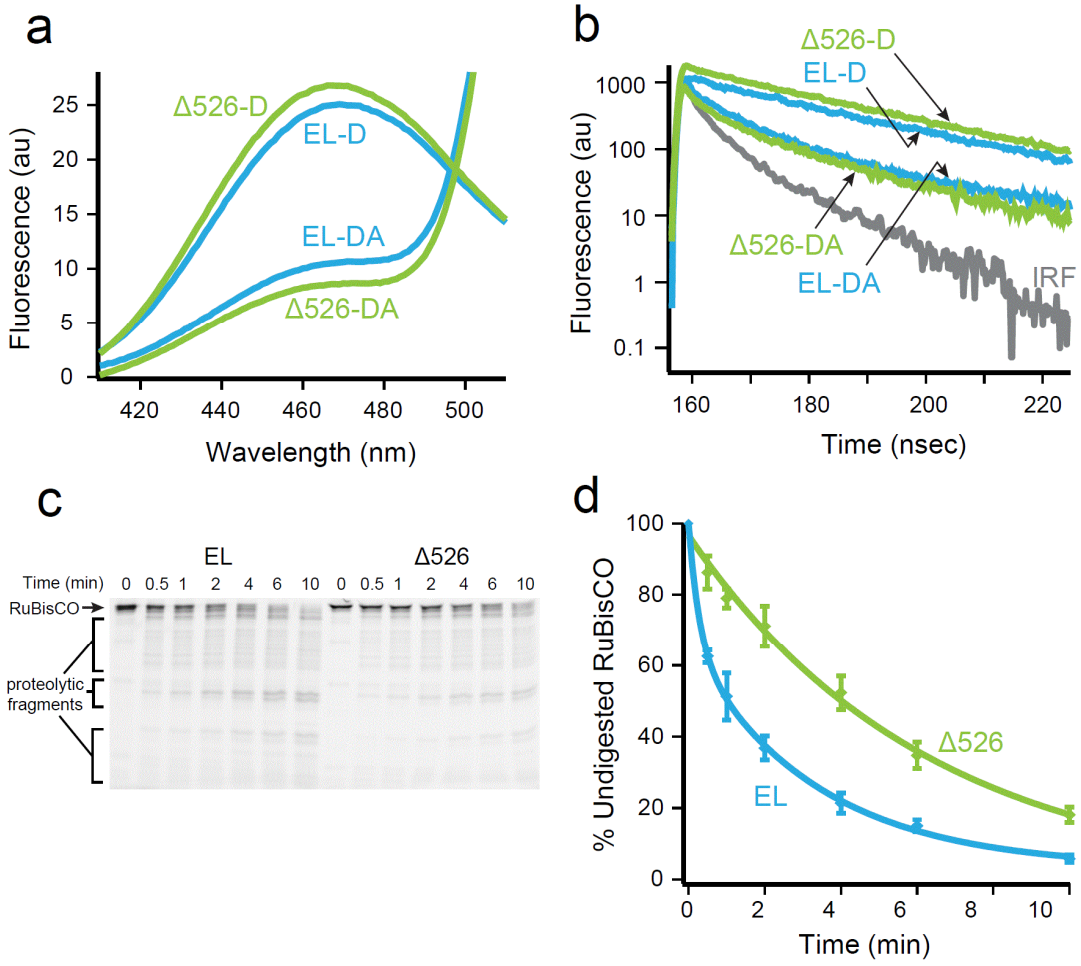


Figure 3.4: RuBisCO Adopts a More Unfolded Conformation on a GroEL Ring in the Presence of the C-Termini. (A) steady-state fluorescence of chemically denatured, donor-only (209ED) or donor-acceptor (209ED-58F) labeled RuBisCO (100 nM) bound to the trans ring of full-length (EL, blue) or C-terminal deletion (Δ526; green) GroEL-GroES ADP bullets (120 nM). The spectra shown are the average of n = 3 experiments. (B) time-resolved fluorescence decay measurements of the same complexes in (A). The instrument response function is also shown (IRF). (C) fluorescently labeled RuBisCO (58-F; 100 nM) was chemically denatured and bound to the trans ring of either GroEL or Δ526 ADP bullets (120 nM) and then treated with trypsin for the indicated times before quenching with PMSF (1mM). The samples were analyzed by SDS-PAGE and laser-excited fluorescence gel scanning. An arrow shows the migration position of full-length RuBisCO, and brackets indicate the position of three dominant groups of proteolytic fragments. (D) amount of full-length RuBisCO remaining at each time point in (C) was quantified and plotted as a function of time. The average half-time for the digestion of RuBisCO bound to full-length ADP bullets was ~1.5 min (EL, blue) and ~4 min for C-terminal deletion ADP bullets (Δ526, green). AU, arbitrary units. Error bars represent the standard deviation of n = 3 experiments.

Table 3.1: Intramolecular FRET Measurements of RuBisCO Bound to Wild-Type and Δ526 GroEL

	FRET pair	Distance[a]	
		WT[b]	Δ526[c]
		Å	
Steady state	34ED[d]-454F[e]	44.1	42.0
	58F-209ED	46.9	41.2
	58F-356ED	43.9	40.1
	58F-454ED	46.2	42.0
Time-resolved	34ED-454F	46.1	43.5
	58F-209ED	49.1	41.7
	58F-356ED	43.2	39.3
	58F-454ED	44.5	42.3

[a] Steady-state and time-resolved intramolecular FRET measurements between different positions on chemically denatured, doubly-labeled RuBisCO bound to the *trans* ring of Δ526 GroEL/GroES ADP bullets.
[b] Full-length GroEL/GroES ADP bullets.
[c] Δ526 GroEL/GroES ADP bullets.
[d] Rubisco labeled with EDANS at amino acid 34.
[e] Rubisco labeled with fluorescein at amino acid 454.

resistant to protease digestion than a monomer bound to a full-length GroEL ring. As shown in Figure 3.4C and 3.4D, RuBisCO bound to a full-length GroEL ring is digested by trypsin more rapidly (~3 times) than RuBisCO bound to a Δ526 ring.

We previously demonstrated that the C-terminal tails help prevent premature substrate-protein escape during GroES binding [347]. This finding suggests that the C-termini maintain contact with the folding intermediate even as the apical domains, which bind different parts of the substrate protein, re-arrange in response to ATP binding [347]. As the C-terminal tails appear to contribute to binding-induced unfolding, we reasoned they may also impact a later step, ATP-driven forced unfolding [159,166]. To test this possibility, we examined the conformation of the GroEL-bound RuBisCO monomer during ATP and GroES binding. Using the same set of intramolecular FRET pairs described above, in combination with stopped-flow rapid mixing, we examined the time-resolved changes in conformation of a trans ring-bound RuBisCO monomer as ATP and GroES bind. An example of the changes for two FRET pairs (34ED-454F and 209ED-58F) is shown in Figure 3.5A and 3.5B. All four pairs displayed similar overall behavior.

The kinetic behavior observed with full-length GroEL is consistent with our previous observations of forced unfolding upon ATP binding, followed by compaction of the RuBisCO monomer as GroES binds [159,166,168]. We observe a very rapid drop in FRET efficiency over the first 200–500 ms, which is followed by a slower increase in FRET efficiency (Figure 3.5A and 3.5B). With Δ526, we also observe the early and rapid drop in the FRET efficiency with all four site pairs, although the amplitude of the rapid FRET decrease is smaller with Δ526 compared with full-length GroEL (Figure

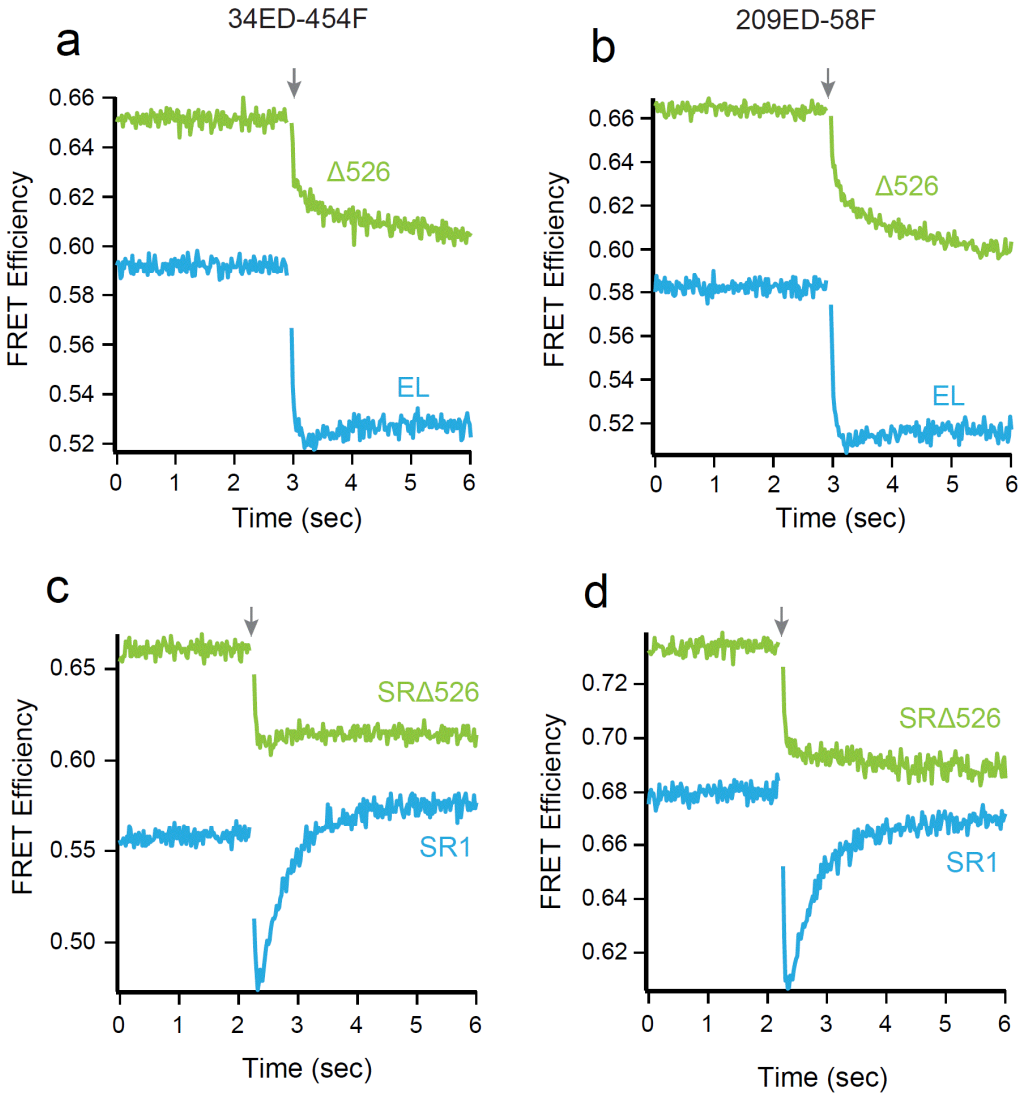


Figure 3.5: Removal of the GroEL C-Termini Diminishes Both Forced Unfolding and Compaction of RuBisCO. (A) and (B), change in FRET efficiency of labeled RuBisCO (100 nM), bound to the trans ring of either a full-length GroEL-GroES (blue) or ELΔ526-GroES (green) complex (120 nM), during encapsulation by GroES. The change in FRET efficiency for 34ED-454F is shown in (A) and for 209ED-58F is shown in (B). (C) and (D), change in FRET efficiency of labeled RuBisCO (100 nM) bound to the single ring GroEL variants SR1 (blue) or SRΔ526 (green) (500 nM), during encapsulation by GroES (1 μM). The change in FRET efficiency for 34ED-454F labeled RuBisCO is shown in (C) and 209ED-58F is shown in (D). In each case, the starting FRET efficiency of non-native RuBisCO bound to each GroEL variant is shown, beginning at t = 0. At t = 2.25 s (arrow), the complex was rapidly mixed with ATP and GroES; n = 21 replicates, with sampling times of 20 ms. The small gap in each plot is a consequence of the removal of data points collected during the mixing dead time.

3.5A and 3.5B). More striking is the change in the GroES-dependent compaction phase, which is absent when the RuBisCO intermediate is encapsulated on a Δ526 trans ring.

The change in compaction of the RuBisCO monomer was also observed using SR1 and SRΔ526. With SR1, we readily observed forced expansion of the RuBisCO monomer, followed by compaction of the folding intermediate. Because each FRET pair again displayed similar overall behavior, we only show the data for the 34ED-454F and 209ED-58F pairs in Figure 3.5C and 3.5D. Interestingly, we observe a larger amplitude compaction phase upon GroES binding to SR1 than is observed with the trans ring of the GroEL-GroES complex [159]. The reason for this difference is not known, but is likely due, as least in part, to the fact that the SR1 sample does not cycle. By contrast, for tetradecamer GroEL, cycling was initiated within the first several seconds of the experiment, resulting in asynchronous reaction phases for the RuBisCO monomers and an apparent difference in average FRET efficiency. When RuBisCO is bound to SRΔ526, we observed a reduced forced expansion phase, as well as the complete disappearance of the GroES-dependent compaction phase. The extent of GroES binding and RuBisCO encapsulation was similar for all of the GroEL variants employed in these experiments (both Δ526 and full-length) under the conditions used, ruling out trivial explanations for this change. Taken together, these observations indicate that the C-termini directly contribute to unfolding by binding and holding segments of a folding intermediate near the bottom of the GroEL cavity.

GroEL C-Termini Enhance Productive RuBisCO Folding. In addition to their involvement in protein unfolding, the C-terminal tails appear to affect protein folding

within the GroEL-GroES cavity. We and others observe a modest but clear difference in intra-cavity folding rate in the presence and absence of the C-termini (Figure 3.1) [204,209]. To gain greater insight into the impact of the C-termini on intra-cavity folding, we examined the changes in intramolecular FRET efficiency of labeled RuBisCO monomers during folding inside the SR-GroES and SRΔ526-GroES cavities. Because the four FRET vectors span several different regions and length scales of the RuBisCO monomer, we anticipated that this assay would provide a higher resolution examination of intra-cavity folding and potentially illuminate how the C-termini enhance the process.

All four RuBisCO FRET pairs demonstrate substantial decreases in FRET efficiency upon the initiation of folding inside both SR1-GroES and SRΔ526-GroES cavities (Figure 3.6). The observed decrease in FRET efficiency shows that as the monomer folds, the labeled sites move apart, indicating that, on average, the kinetically trapped RuBisCO folding intermediate is more structurally collapsed than the committed monomer. In most cases, the observed FRET transients were not well fit by single exponential rate laws and required double exponentials for good fits (Table 3.2). The precise reason for this kinetic complexity is unknown. However, one plausible explanation is the appearance of two kinetically resolvable RuBisCO subpopulations in the GroEL-GroES cavity that possess distinct conformational transition rates in the regions probed by the different FRET pair. Despite this kinetic complexity, the four FRET pairs detect three average temporal regimes in the folding process (Table 3.2). Although the 356ED-58F pair appears to track the same committed step in folding that is

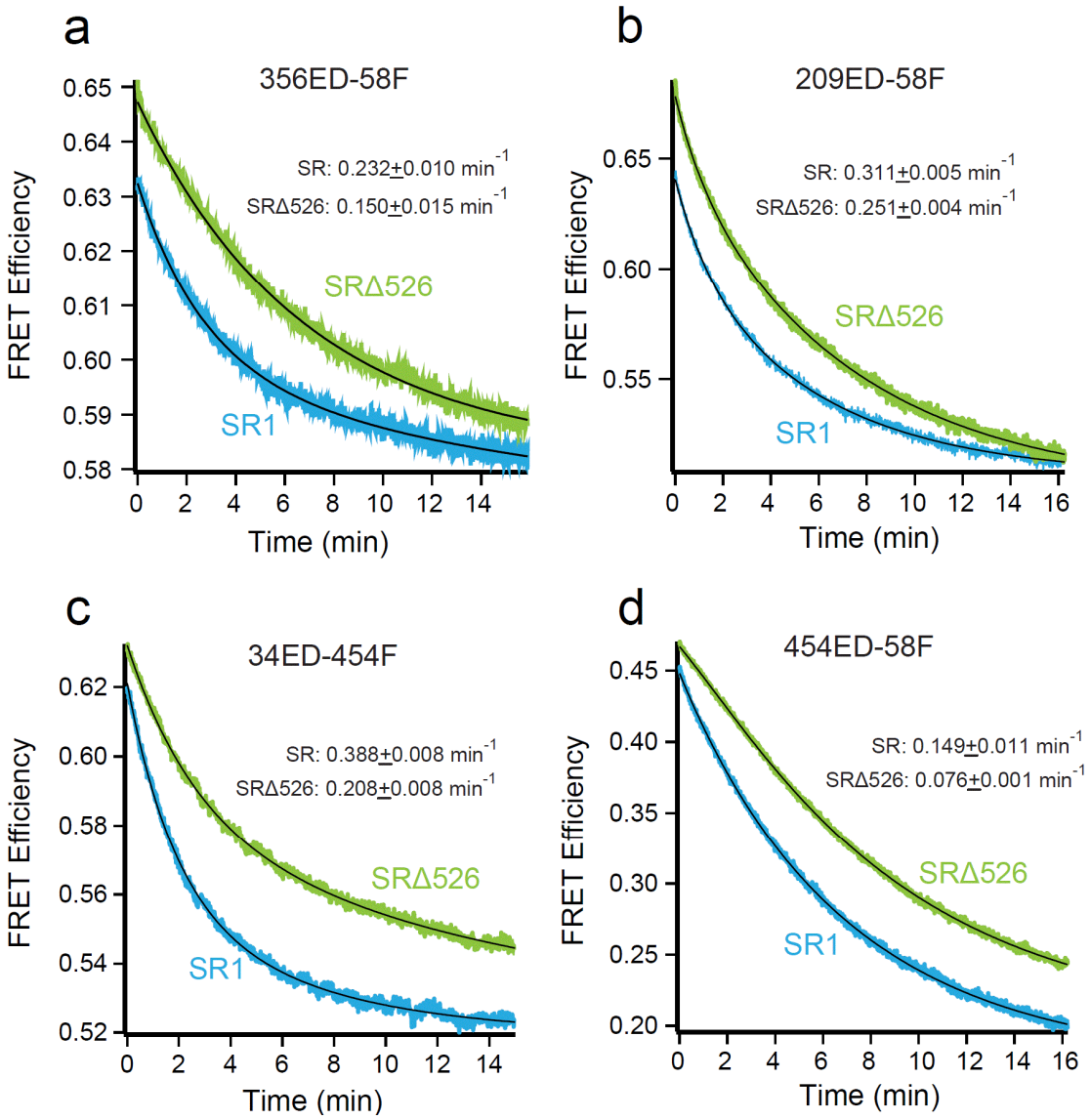


Figure 3.6: Intra-Cavity Folding of RuBisCO with and without the GroEL C-Termini Monitored by Intramolecular FRET. RuBisCO folding inside full-length and truncated single ring GroEL-GroES complexes monitored by intramolecular FRET with four different site pairs as follows: (A) 356ED-58F; (B) 209ED-58F; (C) 34ED-454F; and (D) 454ED-58F. Chemically denatured, fluorescently labeled RuBisCO (100 nM) was bound to full-length SR1 or SRΔ526 (500 nM). This complex was rapidly mixed with an equal volume of excess GroES (1 μM) and ATP (2 mM) in a stopped-flow apparatus. Shown is the average of n = 12 replicates of matched experimental pairs, calculated from donor-only and donor-acceptor samples. Sampling time was 150 ms. In all cases, the change in FRET efficiency was well fit by a bi-exponential rate law (black line), and the average folding rate constant is shown for each case. Table 3.2 contains the rate constants for each fit.

Table 3.2: Intra-Cavity Folding Rates of RuBisCO with and without the GroEL C-Termini

Complex	FRET Pair							
	58F[c]-454ED[d]		58F-356ED		58F-209ED		34ED-454F	
	A_1[e]	k_1[f]	A_1	k_1	A_1	k_1	A_1	k_1
SR1[a]	0.018	0.307	0.029	0.060	0.099	0.169	0.040	0.155
SRΔ526[b]	-0.035	0.466	0.002	0.399	0.150	0.139	0.058	0.066
	A_2[e]	k_2[f]	A_2	k_2	A_2	k_2	A_2	k_2
SR1	0.257	0.138	0.032	0.388	0.035	0.712	0.062	0.539
SRΔ526	0.301	0.121	0.063	0.142	0.028	0.850	0.051	0.371
	k_{AVG}[g]		k_{AVG}		k_{AVG}		k_{AVG}	
SR1	0.149		0.232		0.311		0.388	
SRΔ526	0.076		0.150		0.251		0.209	
SR1/SRΔ526[h]	1.96		1.55		1.24		1.86	

[a] Folding was inside full-length SR1-GroES cavity.
[b] Folding was inside truncated SRΔ526-GroES cavity.
[c] Rubisco was labeled with fluorescein at amino acid 58.
[d] Rubisco was labeled with EDANS at amino acid 454.
[e] Pre-exponential amplitude terms were from double exponential fits of folding transients.
[f] Rate constant terms were from double exponential fits of folding transients.
[g] Calculated average rate constant was for the double exponential fit.
[h] Ratio of average folding rate constant was for wild-type and Δ526 GroEL.

observed by enzymatic activity, the other pairs report on steps both slower (454ED-58F) and faster (209ED-58F and 34ED-454F). Interestingly, each of the three kinetic phases identified during folding within the SR1-GroES cavity are also present when RuBisCO folds in the absence of the C-termini inside the SRΔ526-GroES cavity. However, the average rate of each kinetic phase is slower by 20–50% when folding proceeds in the SRΔ526-GroES cavity (Table 3.2).

As a complement to the stopped-flow FRET measurements, we employed changes in tryptophan fluorescence to report on the folding of the RuBisCO monomer at early times inside the SR1-GroES and SRΔ526-GroES cavities. The tryptophan emission spectra of non-native RuBisCO bound to SR1 and SRΔ526 are very similar (Figure 3.7A, inset), suggesting that the average exposure and environment of the RuBisCO tryptophan residues are also similar in both starting complexes. Consistent with our earlier measurements, a rapid decrease in fluorescence intensity was followed by a much slower rise in fluorescence that matches the rate of the committed folding step (Figure 3.7B), when ATP and GroES are added to the SR1 complex [163]. The early drop in tryptophan fluorescence intensity likely reports on the process of conformational expansion, GroES binding, and release of the RuBisCO monomer into the SR1-GroES cavity. However, the early folding behavior with SRΔ526 is quite different. Although we observe a rapid decrease in fluorescence intensity upon the addition of ATP and GroES to the SRΔ526 complex, the decrease is slower, biphasic, and shows a greater amplitude compared with SR1 (Figure 3.7A and 3.7B). In addition, the rising phase that reports on formation of the committed RuBisCO monomer is dramatically delayed, relative to SR1.

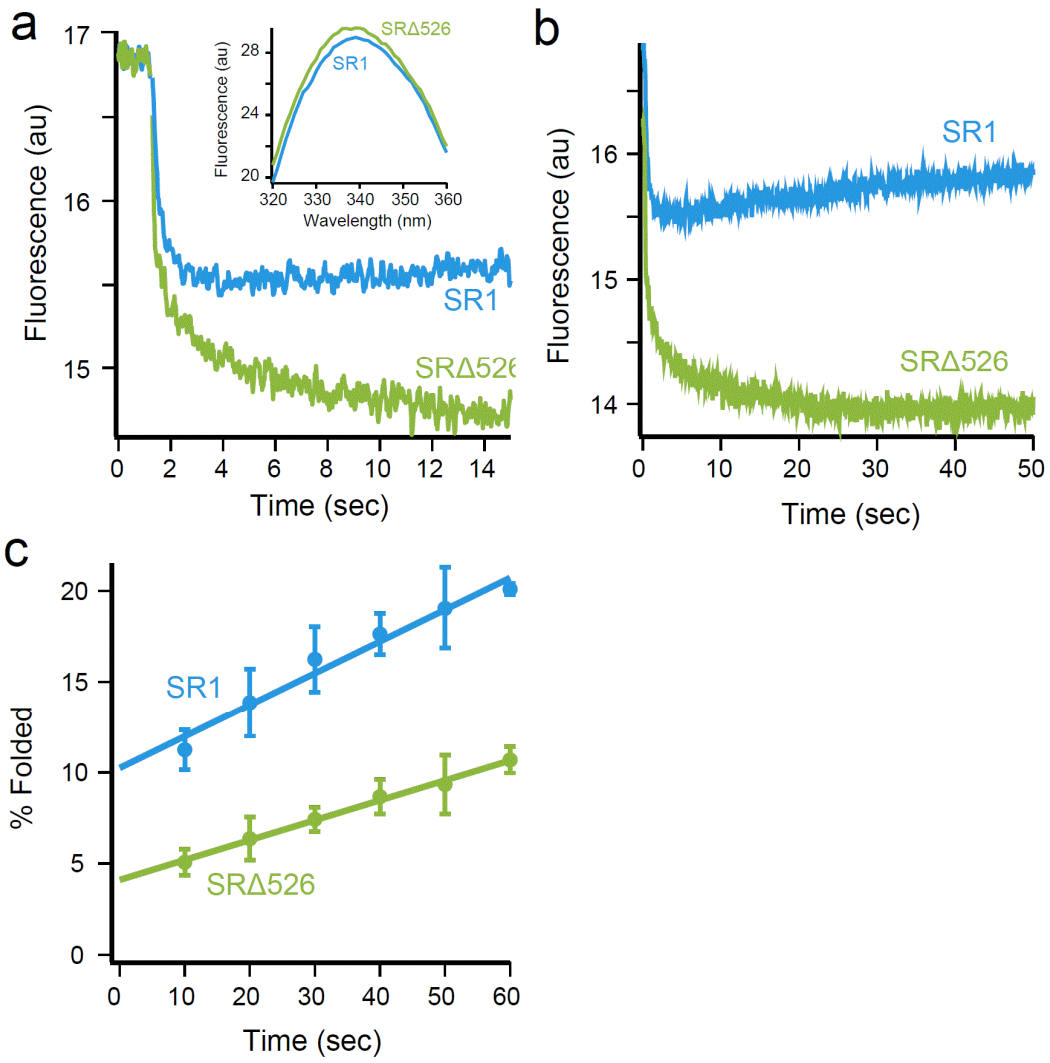


Figure 3.7: GroEL C-Termini Enhance the Fraction of RuBisCO that Folds Rapidly upon Encapsulation beneath GroES. (A) intra-cavity folding of RuBisCO at early times monitored by changes in tryptophan fluorescence following addition of GroES and ATP to complexes of wild-type RuBisCO bound to SR1 (blue) or SRΔ526 (green). Conditions are as in Fig. 3.1B, except with sampling time of 20 ms. The steady-state, tryptophan fluorescence emission spectra of wild-type RuBisCO bound to SR1 (blue) or SRΔ526 (green) are shown (inset). (B) same as in (A) but with the time scale expanded to show the full 50 s of folding monitored in this experiment. (C) RuBisCO folding at early times with SR1 and SRΔ526 monitored by regain of enzymatic activity. Chemically denatured, wild-type RuBisCO (100 nM) was bound to either 500 nM SR1 (blue) or SRΔ526 (green), to which excess GroES (1 μM) and ATP (2mM) were added. Folding was quenched by adding EDTA and simultaneously lowering the temperature to nonpermissive folding conditions. AU, arbitrary units. Error bars represent the standard deviation from n = 3 experiments.

Given the more extensive unfolding and compaction of the RuBisCO monomer seen with SR1 (Figure 3.4 and 3.5), the reduced perturbation in tryptophan fluorescence over the same time period is surprising. A simple explanation is suggested by the increase in tryptophan fluorescence that accompanies the committed folding step, which follows release of the less fluorescent, non-native monomer into the cavity (Figures 3.1, 3.7A and 3.7B) [163]. In SR1, the committed step is completed much sooner than in SRΔ526 (Figure 3.7A and 3.7B). If a subpopulation of the RuBisCO monomers achieves the committed monomer conformation very rapidly, then formation of small amounts of this more fluorescent, committed conformation would partially compensate for the drop in fluorescence of the more slowly progressing subpopulation, thereby accounting for the smaller amplitude of the tryptophan fluorescence drop within SR1. The larger apparent drop in tryptophan fluorescence within SRΔ526 would then be the result of less extensive unfolding and a correspondingly smaller subpopulation of a rapidly folding monomer. This explanation predicts that folding within SRΔ526 should result in a substantially reduced fraction of RuBisCO monomers that commit rapidly upon addition of ATP and GroES. Rapid commitment is detected as a dead-time burst in native enzyme activity in single turnover folding experiments [166]. As shown in Figure 3.7C, we observe a greater than 2-fold drop in the early fraction of rapidly committed RuBisCO monomers in comparison with SR1. The fact that this prediction is met suggests strongly that the C-terminal tails help stimulate RuBisCO folding through structural disruption of kinetically trapped RuBisCO monomers.

Discussion

The results we present here strongly support an active role for GroEL in assisted protein folding (Figure 3.8). We have shown that the unstructured C-terminal tails of GroEL bind and pull a partially structured RuBisCO folding intermediate toward the bottom of the GroEL cavity. In the absence of the C-termini, several distinct folding transitions of the RuBisCO monomer slow inside the GroEL-GroES cavity, suggesting that the conformational search executed by the RuBisCO folding intermediate is more efficient in the presence of the C-terminal tails. In addition, the C-termini participate in protein unfolding, both during the initial capture of a folding intermediate on an open GroEL ring, as well as during the forced expansion of the intermediate upon ATP binding. When unfolding is diminished by removal of the C-termini, the fraction of the RuBisCO monomers that populate the most rapid folding pathways is substantially reduced.

Our observations are consistent with a key role for substrate protein unfolding in GroEL-stimulated folding but inconsistent with an exclusive role for substrate protein confinement. Removal of the C-terminal tails in the Δ526 GroEL variant not only slows the rate of RuBisCO folding but also simultaneously increases the lifetime of the GroEL-GroES complex. If the stimulation of folding occurs mainly through substrate protein confinement within the GroEL-GroES cavity, then the longer lifetime of the Δ526-GroES complex presents a paradox; even taking into account the modest reduction in intra-cavity folding observed with SRΔ526 (Figures 3.1 and 3.6), the considerably

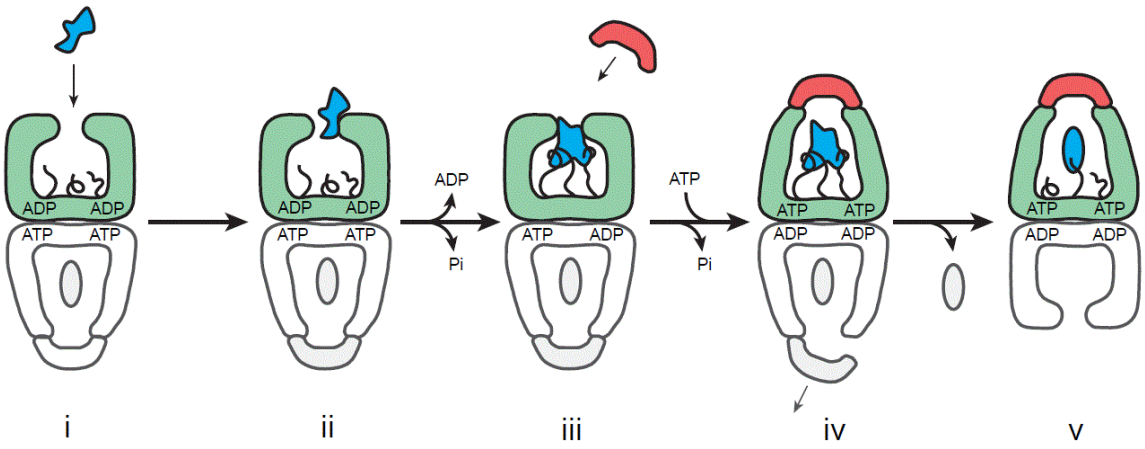


Figure 3.8: Model for the Role of the GroEL C-Termini in Substrate Protein Unfolding. A schematic is shown for the steps involved in substrate protein loading and the initiation of folding by GroEL. Step i, a non-native substrate protein (irregular blue shape) enters the GroEL reaction cycle on the open trans ring (green) of the ATP bullet complex [159]. Step ii, substrate protein binding accelerates both the release of ADP from the trans ring and ATP hydrolysis in the opposite, cis ring (gray) [156,158,160,340]. Step iii, binding of the non-native substrate protein by the C-terminal tails (black), helps pull the substrate protein into the GroEL cavity and, in combination with additional binding by multiple apical domains, results in substrate protein unfolding. Step iv, ongoing association of the C-termini with the substrate protein during ATP-driven encapsulation by GroES both retards pre-mature protein release [347] and provides an anchor point for forced expansion of the substrate protein as the apical domains shift to accommodate GroES binding. Assembly of the new folding cavity on the trans ring is directly coupled to the disassembly of the folding cavity on the opposite ring, potentially through a transient, symmetric intermediate [340,350-352]. Step v, a subsequent allosteric shift of the GroEL-GroES complex results in full ejection of the substrate protein in the enclosed GroEL-GroES cavity and the initiation of folding [347]. The C-terminal tails may continue to interact with the folding intermediate, influencing the spectrum of states populated during folding.

longer Δ526-GroES cavity lifetime (Figure 3.2) should result in better RuBisCO folding compared with full-length GroEL, not worse. This points to some other property of the GroEL machine that must contribute to the stimulation of RuBisCO folding. We previously demonstrated that a RuBisCO monomer is subjected to two phases of multiple axis unfolding prior to the initiation of productive folding inside the GroEL-GroES cis cavity as follows: 1) binding-driven unfolding upon capture of a folding intermediate by the GroEL ring, and 2) forced unfolding upon ATP binding to a RuBisCO-occupied GroEL ring [159,166,168]. The fraction of the RuBisCO population that populates efficient folding pathways is proportional to the magnitude of this unfolding [166]. Furthermore, the rapidly cycling GroEL-GroES system can achieve assisted folding rates that are substantially faster than what confined folding in the GroEL-GroES cavity can achieve alone [159]. These observations are consistent with predictions of the iterative annealing model of stimulated protein folding by GroEL [187,341].

Importantly, the behavior of the Δ526 GroEL variant also satisfies predictions of the iterative annealing model. If GroEL acts, at least in part, as an iterative annealing machine, disrupting inhibitory and kinetically trapped states through partial unfolding, then any modification to the machine that reduces the extent of unfolding and slows the rate of turnover should result in a reduced rate of stimulated folding [341]. As we have shown with observations of the Δ526 GroEL variant, these predictions are met. Not only does removal of the C-terminal tails reduce both binding-induced and forced unfolding, resulting in a substantial decrease in the fraction of RuBisCO that rapidly commits to the

native state, but removal of the tails also slows the rate at which the GroEL-GroES system cycles, reducing the frequency with which a given population of RuBisCO folding intermediates can be subjected to unfolding.

Although our observations with Δ526 GroEL are consistent with key elements of the iterative annealing model, they do not exclude a role for protein confinement. Indeed, we previously demonstrated that encapsulation of the RuBisCO monomer beneath GroES is associated with compaction of the folding intermediate, a conformational restriction that is, in principle, consistent with confinement-based models [159,168]. Notably, the magnitude and kinetics of this compaction event are compromised by removal of the GroEL C-termini (Figure 3.5). At the same time, intra-cavity folding of RuBisCO is measurably slower in the absence of the C-termini (Figures 3.1 and 3.6). It is possible that the reduced rate of RuBisCO folding inside the SRΔ526-GroES cavity is simply the result of reduced initial unfolding, so that the rate of all subsequent steps are impacted by the starting conformation of the protein at the moment of encapsulation. Indeed, it is remarkable that each of the kinetically distinct steps identified with the intramolecular FRET assay all slow to a similar extent in the absence of the C-termini (Figure 3.6 and Table 2). However, both cavity volume and character have also been suggested to impact intra-cavity folding, and both properties should be altered by deletion of the C-terminal tails [199,203,204,345]. Yet, the shift in the RuBisCO binding position upon removal of the C-terminal tails is inconsistent with a volume effect alone. If the C-termini primarily act via spatial constriction of the cavity, their removal would be expected to either have no effect on the RuBisCO binding

position or would allow it to occupy more of the cavity volume and thus bind more deeply into the cavity. However, we observe the opposite effect, i.e. removal of the C-terminal tails results in a more elevated average binding position (Figure 3.3). This observation, in combination with our earlier cryo-EM observations [347], suggests a direct binding interaction with the RuBisCO folding intermediate, whereby the C-termini help pull and stretch the monomer toward the bottom of the GroEL cavity.

In confinement-based mechanisms, the spectrum of partially folded intermediates in the GroEL-GroES cavity should differ from the ensemble of states populated in free solution. Precisely how interactions between the C-terminal tails and a protein folding intermediate could impact this distribution is not understood. Transient interactions between the weakly hydrophobic interior of the GroEL cavity and a folding intermediate have been proposed to assist productive folding through an annealing process that prevents or destabilizes kinetically trapped states [190,338]. Consistent with a role for the C-termini in such a process, we observe ongoing, although weakened, interactions between the RuBisCO folding intermediate and the C-terminal tails during early stages of intra-cavity folding [347]. Additionally, the amphipathic character of the C-terminal tails appears to be a key property of these unstructured segments as follows: modifications that make them either too polar or too hydrophobic inhibit folding [203,250]. Notably, RuBisCO displays a striking ability to fold inside the GroEL-GroES cavity under conditions where the monomer in free solution folds slowly or not at all [163,168,199]. More recently, a double mutant of the maltose-binding protein has been shown to display similar behavior [205]. For stringent substrate proteins like RuBisCO,

it is tempting to speculate that a combination of unfolding and confinement might provide the most efficient method of stimulating productive folding. In a combined mechanism, unfolding and disruption of kinetically trapped, misfolded states would increase the chance of opening a productive folding pathway. Subsequent confinement of the partially unfolded intermediate within the GroEL-GroES cavity could then provide a maximally conducive environment, in which the probability of populating inhibitory conformational states is minimized, at least for the short duration of the GroEL-GroES complex.

Given the clear impact of the GroEL C-termini on both protein encapsulation and folding, it is striking that these segments are not essential in vivo [246,247]. At the same time, the C-terminal tails are well conserved in the majority of chaperonins [354]. These observations, combined with the near universal essentiality of chaperonins across phyla, suggest that the efficiency of protein folding by chaperonins is a tight evolutionary constraint, a conclusion supported by recent studies on the linkage between both fitness and protein evolutionary rates and chaperonin activity [355-359]. In this view, even a modest loss of folding capacity, like that caused by the absence of the C-terminal tails, would result in a steep reduction in overall fitness. Indeed, one of the first studies to examine the role of the GroEL C-termini observed just such an effect [246]. In competition experiments between otherwise identical E. coli strains expressing either a full-length or a C-terminally truncated GroEL, cells forced to rely on a truncated GroEL rapidly lost out. Modifications of the C-termini may also be linked to shifts in substrate specificity or activity of different chaperonin subtypes. Many microbial species

maintain multiple, distinct chaperonin variants in the same cytoplasm, with the actinobacteria of particular note [213]. In most cases, the essential housekeeping chaperonin (Cpn60.2) has a C-terminal tail that retains the sequence character and Gly-Gly-Met motifs common to eubacterial chaperonins like GroEL. Strikingly, a secondary chaperonin (Cpn60.1), thought to be a specialized variant important for biofilm formation and pathogenesis, possesses a modified tail in which the sequence character has been dramatically altered, and the typically conserved Gly-Gly-Met motifs have been replaced with sequences enriched in His [213]. Thus, although the results we report here shed additional light on the role of the GroEL C-termini, our understanding of the role of these important and characteristic chaperonin domains remains incomplete.

CHAPTER IV

STRUCTURAL BASIS OF SUBSTRATE SELECTIVITY OF THE E. COLI PROLIDASE*

Prolidases, also known as Xaa-Pro dipeptidases, are metalloproteases that catalyze the hydrolysis of dipeptides containing a C-terminal proline residue. These enzymes are conserved in prokaryotes and eukaryotes, including not only single-celled organisms, such as yeast, but also humans and higher plants [360-367]. In higher organisms, prolidase serves a critical role in the recycling of collagen, as the penultimate products of collagen catabolism include the dipeptides Ala-Pro and Gly-Pro [368-370]. In humans, specific mutant alleles of prolidase have been linked to a wide array of physiological problems, which are known collectively as prolidase deficiency [369-372]. Despite the importance of human prolidase and the disease states associated with various mutations of the gene, knockout and knockdown studies in several eukaryotic model organisms have yet to reveal an essential role for prolidase [373-376]. Therefore, further studies are required for insight into the role of prolidase in collagen metabolism and human health.

In contrast to the human enzyme, there are no observable phenotypes for *Escherichia coli* prolidase mutants [377]. While a physiological role for prolidase in

*Reproduced with permission from Weaver J, Watts T, Li P, Rye HS (2014) PLoS One 9: e111531. Copyright 2014 Weaver, et al. For the original publication: JW conceived and designed experiments, performed experiments, analyzed data, and contributed to the writing of the manuscript.

bacteria remains to be established, the enzyme is known to possess protective activity against toxic organophosphates [365,378-380]. The *E. coli* enzyme may also play a role similar to that of human prolidase – the breakdown of dipeptides stemming from protein catabolism – or an additional, regulatory role [381]. In support of this theory, *Mycoplasma* species possess Xaa-Pro peptidases [214,382,383]. These bacteria, which evolved to retain only those cellular functions essential to their parasitic lifestyle, import most amino acids and lipids from the host cell [220]. The fact that *Mycoplasma* retain an enzyme for cleaving Xaa-Pro bonds suggests that prolyl peptide catabolism plays a broad and generally important physiological role.

Prolidases share a number of conserved sequence and structural features. These enzymes possess an N-terminal domain and a C-terminal catalytic domain, and form dimers through contact between both domains in a head-to-tail arrangement [364]. The catalytic site features a binuclear metal cluster in the center of a pita-bread fold that is a canonical feature of this family of enzymes [381,384]. While the identity and configuration of the coordinating ligands are conserved, the types of metals found in the active site vary widely, though manganese, cobalt and zinc appear to be the most common metals used [385]. Such metal variability has been observed in other pita-bread fold proteins [386,387]. Interestingly, the human prolidase can utilize magnesium, though to significantly lower extent than manganese – a feature not commonly seen in other prolidases [361,363,385,388]. Crystal structures of various prolidases, particularly those with bound substrates or inhibitors, have provided important structural insights

into how these enzymes bind substrate peptides and metals, though few members of this enzyme family have been thoroughly examined biochemically.

Members of the pita-bread fold family of proteins, which also includes other metalloproteases, share a number of sequence-specific features that permit robust structure/function prediction, despite the varying substrate specificities of different enzymes [384]. The first prolidase structure solved was from the archaea, *Pyrococcus furiosus* [364,389], which confirmed that prolidases possess many of the structural features common to the pita-bread fold superfamily. However, four large regions of primary structure, ranging from 13-25 amino acids in length, are found in the human prolidase that do not appear in the *P. furiosus* sequence [364]. Some of these regions are also absent from related pita-bread fold members, including methionine aminopeptidases, which cleave N-terminal methionine residues, as well as proline aminopeptidases, which cleave N-terminal residues that are followed by proline, from both bacterial and human sources [384,390].

Interestingly, the peptide regions absent in *P. furiosus* are present in prolidases from Gram negative bacteria, including *E. coli* and *Alteromonas sp.*[364], and include eleven residues highly conserved between humans and these two bacteria. *E. coli* PepQ, the only prolidase found in this organism [377], was previously characterized for activity against dipeptides, organophosphates and other small molecules [365], though the lack of an atomic structure for PepQ has prevented a detailed comparison to other prolidases. Examination of the *Alteromonas* prolidase structure, however, reveals an arginine residue reaching into the active site from one of the additional peptide segments. This

residue appears to be involved in positioning a structured water molecule and other active site residues and metals, and has been postulated to interact with the C-terminus of the substrate dipeptide [379], an interaction similar to that seen in a shifted location for proline aminopeptidase [391]. Because proline aminopeptidases cleave tripeptides, the positioning of this residue may have evolved to specify substrate length in pita-bread fold proteins.

Here we report the structure of *E. coli* PepQ, showing it to have the predicted pita-bread fold. We examine its ability to utilize various active site metals, including magnesium. Furthermore, we compare its sequence and structural similarity to proline aminopeptidase and other prolidases, showing that the position of the conserved arginine has, in fact, moved throughout evolution, likely to accommodate substrate peptide length. We further characterize the role of this arginine, demonstrating that it plays a critical role in substrate dipeptide binding.

Experimental Procedures

Cloning, Expression and Purification of PepQ and PepQ Mutants. The PepQ gene was PCR amplified from purified, chromosomal *E. coli* DNA, using primers adding a 5'-NdeI restriction site and a 3'-XhoI restriction site. The PCR product was sub-cloned into the pET21a vector (Novagen) and the sequence of this construct was verified by DNA sequencing. The R370E mutation of the PepQ gene was created via site-directed mutagenesis of the wild-type construct and was verified by DNA sequencing. Either 6 or 12 L of LB-Amp (100 mg/L) were inoculated 1:500 with overnight cultures

of BL21[DE3] cells transformed with either the wild-type or R370E PepQ plasmid. Upon reaching an A_{600} = 0.6-0.8, expression was induced with the addition of IPTG to a concentration of 400 μM. After four hours, the cells were centrifuged and the pellets were resuspended in cell disruption buffer (20 mM Tris, pH 8, 1 μM $MnCl_2$, 20% (w/w) sucrose, 4 mM DTT). Cells were lysed using a gas-driven cell-disruptor (Microfluidics Corporation; Newton, MA) and clarified by ultracentrifugation. The supernatant was loaded onto a fast-flow Q (GE Healthcare) anion exchange column. The column was washed with Buffer A (50 mM Tris, pH 7.4, 1 μM $MnCl_2$, 2 mM DTT) and washed with Buffer A containing 100 mM NaCl. A linear gradient was then developed from 100 mM to 500 mM NaCl. The fractions of the greatest PepQ purity were concentrated by precipitation with 70% (w/v) ammonium sulfate. The pellet was resuspended in a small volume of Buffer A containing 500 mM ammonium sulfate and loaded on a phenyl-sepharose hydrophobic interaction column (GE Healthcare). After washing with Buffer A containing 1 M ammonium sulfate, a linear gradient was developed from 1 M to 300 mM ammonium sulfate. Fractions of the greatest PepQ purity were again concentrated by precipitation with 70% (w/v) ammonium sulfate. The pellet was resuspended with a small volume of Buffer A and dialyzed against Buffer B (25 mM Tris, pH 7.4, 25 mM KCl, 1 μM $MnCl_2$, 2 mM DTT). Following addition of glycerol to 15% (v/v), the sample was aliquoted, snap frozen with liquid nitrogen and stored at -80˚C. Thawed samples showed no detectable loss of enzymatic activity.

Cloning, Expression and Purification of Alanine Dehydrogenase (AlaDH). The AlaDH gene was PCR amplified from purified, chromosomal *Bacillus subtilis str. 168*

DNA, using primers adding a 5'-NcoI restriction site (which required a mutation in the second codon, which was later reverted with site-directed mutagenesis) and a 3'-XhoI restriction site. The PCR product was sub-cloned into the pETDuet vector (Novagen) and the sequence of this construct was verified by DNA sequencing. Protein expression was conducted in 6 L of LB-Amp (100 mg/L) inoculated 1:500 with overnight cultures of BL21[DE3] cells transformed with the AlaDH plasmid. Upon reaching an A_{600} = 0.6-0.8, expression was induced with the addition of IPTG to a concentration of 400 μM. After four hours, the cells were centrifuged and the pellets were resuspended in cell disruption buffer (20 mM Tris, pH 8, 0.5 mM EDTA, 20% (w/w) sucrose, 4 mM DTT). Cells were lysed, clarified and loaded onto a fast-flow ion exchange column, as described above. The column was washed with Buffer C (50 mM Tris, pH 7.4, 0.5 mM EDTA, 2 mM DTT) containing 150 mM NaCl. A linear gradient was then developed from 150 mM to 500 mM NaCl. Fractions of the greatest AlaDH purity were concentrated by precipitation with 70% (w/v) ammonium sulfate. The pellet was then resuspended in a small volume of Buffer C containing 1 M ammonium sulfate. The sample was then loaded on a phenyl-sepharose hydrophobic interaction column (GE Healthcare). After washing with Buffer C containing 900 mM ammonium sulfate, a linear gradient was developed from 900 to 650 mM ammonium sulfate. Fractions of the greatest AlaDH purity were concentrated by precipitation with 70% (w/v) ammonium sulfate. The pellet was resuspended with a small volume of Buffer C and dialyzed against Buffer D (25 mM Tris, pH 7.4, 25 mM KCl, 0.5 mM EDTA, 2 mM DTT). Following addition of glycerol to 15% (v/v), the sample was aliquoted, snap frozen with

liquid nitrogen and stored at -80˚C. Thawed samples showed no detectable loss of enzymatic activity.

Crystallization and Refinement of PepQ. The PepQ sample was buffer-exchanged into 50 mM Tris, pH 7.4, 5 mM $MgCl_2$ and 5 mM DTT at a final concentration of 12 mg/ml. The protein was crystallized by hanging drop vapor diffusion method at 4°C using 20% PEG MME 5000 in 0.1 M Bis-Tris buffer at pH 6.5. The crystals were transferred stepwise to a cryobuffer containing 30% PEG 400, 20% PEG MME 5000, 0.1 M Bis-Tris at pH 6.5 and flash frozen in liquid nitrogen. The diffraction data were collected at beamline 7.1 at the Stanford Synchrotron Radiation Lightsource (SSRL) using a Quantum 315R CCD detector. The diffraction data were processed with the HKL2000 package [392]. The structure was determined by molecular replacement using Phaser in the Phenix package [393]. A homology model of pepQ generated using Swiss-Model based on the crystal structure of *Alteromonas macleodii* OpaA structure (PDB 3RVA) [379] was used as search model. The model was fine-tuned with Coot [394] and refined using the Phenix package [393]. Statistics of data collection and refinement are shown in Table 1.

Metal Usage. Metal usage of PepQ was directly monitored by the decrease in absorbance at 222 nm upon cleavage of the substrate peptide bond [365]. *E. coli* PepQ was diluted to 12.5 μM into 50 mM Tris, pH 7.4 and 10 mM EDTA. Following incubation at 25˚C for 30 min, this solution was then diluted 25-fold into 25 mM Tris, pH 7.4 containing either a divalent metal (1 mM), EDTA (5 mM) or no additional component. Samples were incubated at 25˚C for an additional 10 min. This sample was

then diluted 10-fold with 10 mM Tris, pH 8 and the substrate dipeptide AlaPro (TCI-America). The reaction was immediately assayed at 25˚C. The final concentration of PepQ was 50 nM and Ala-Pro was 0.25 mM in a final volume of 1 mL. All assays were conducted using a Perkin Elmer Lambda 35 spectrophotometer with a PCB 1500 water Peltier temperature control system.

Docking Simulations. Preparation of structure files and docking was done as described [395]. In brief, substrate and protein structure files were prepared using MGL Tools, in which polar hydrogens were added and flexible bonds were designated. Autodock Vina was then used to simulate the interaction of the small molecules with the active site of PepQ.

Enzyme Quaternary Structure. The stability of the dimeric structure of wild-type and R370E PepQ was determined using analytical gel filtration. PepQ (10 nM) in 50 mM Tris, pH 7.4, 50 mM KOAc, 10 mM $Mg(OAc)_2$ and 2 mM DTT was injected on a Superose 6 gel filtration column (GE), equilibrated in the same buffer, with a constant flow rate of 0.4 mL/min driven by an HPLC unit with a binary pump (Waters). The tryptophan fluorescence (excitation at 280 nm, emission at 340 nm) of the sample was measured using an in-line, post-column fluorescence detector (Waters).

Enzyme Stability. The thermodynamic stability of wild-type and R370E PepQ was determined by the red-shift in the tryptophan fluorescence peak as the protein unfolds with increasing concentrations of the chemical denaturant guanidinium-HCl. PepQ (50 nM) was incubated at room temperature for 60 minutes in solutions of 50 mM Tris, pH 7.4, 10 mM $Mg(OAc)_2$, 2 mM DTT and varying concentrations of guanidinium-

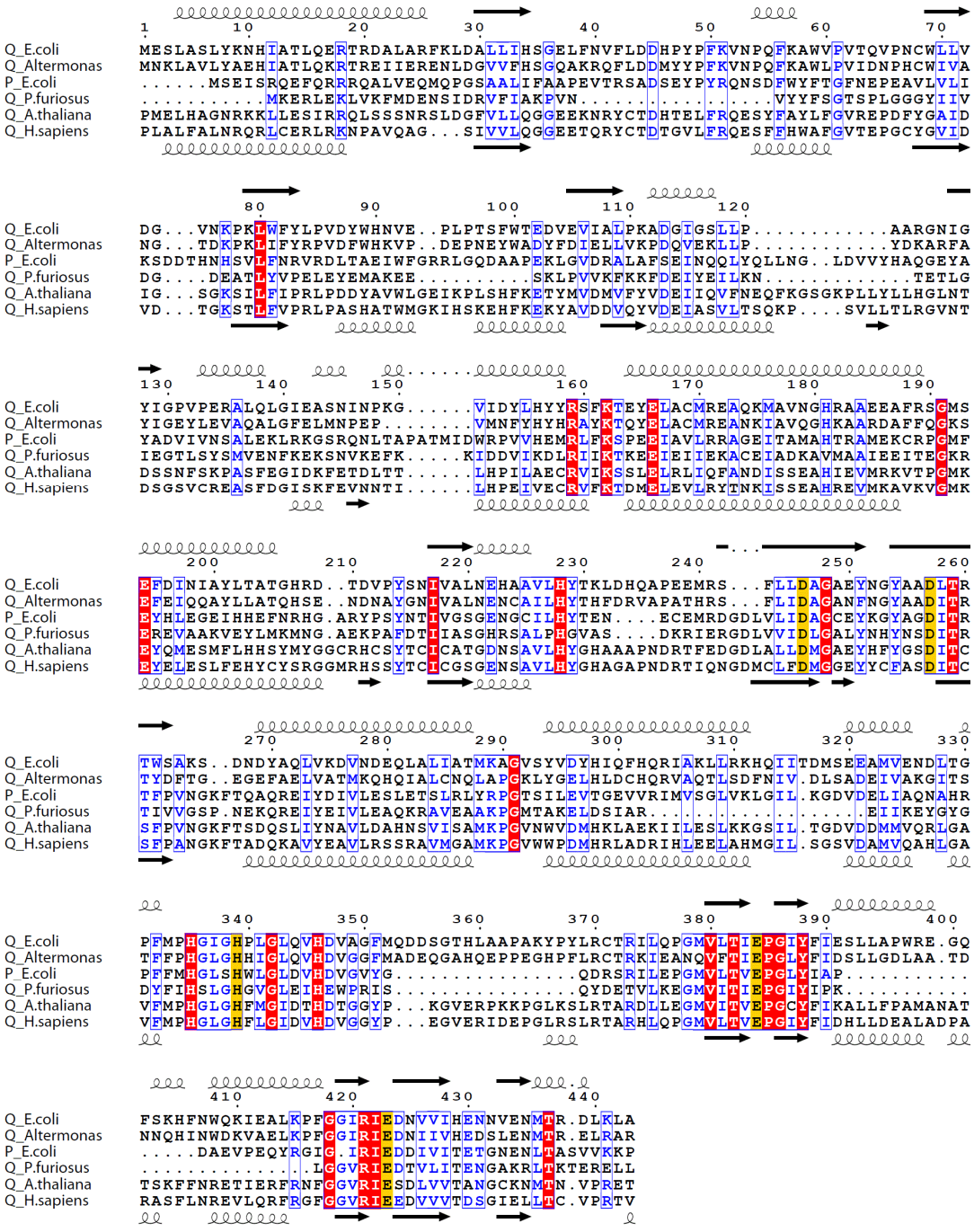
HCl. The tryptophan fluorescence was measured using a PTI fluorometer with excitation at 295 nm and emission from 315-375 nm. Solutions of buffer and guanidinium at each concentration, without protein, were also measured to account for changes in scattered light. The peak maximum and corresponding wavelength was determined using Microsoft Excel (MAX and VLOOKUP functions).

Enzyme Kinetics. The PepQ reaction rate was monitored by coupling the hydrolysis of the dipeptide AlaPro to the NAD-dependent oxidation of alanine [396]. These reactions were conducted in a 1 mL volume in 50 mM Tris, pH 8 and 20 mM $Mg(OAc)_2$ at 25˚C with varying concentrations of AlaPro-COOH (TCI America) or AlaPro-$CONH_2$ (Chem-Impex), supplemented with 1 μM AlaDH and 2 mM NAD^+ (Chem-Impex). The increase in absorbance at 340 nm was monitored as NADH was produced. All assays were conducted using a Perkin Elmer Lambda 35 spectrophotometer with a PCB 1500 water Peltier temperature control system.

Results

E. coli Prolidase Possesses an Expanded Sequence. To examine the extent of sequence conservation in the *E. coli* prolidase, PepQ, we collected primary structure information from organisms with sequenced genomes, including both higher plants and animals. Upon alignment (Figure 4.1), many regions of *E. coli* PepQ show sequence similarity (boxed) and identity (shaded) with the sequences of human and plant prolidase, illustrating the conservation of various elements of this protein family. Overall, *E. coli* PepQ shows high sequence identity (~30%) and similarity (~50%) with

Figure 4.1: Sequence Alignment of Prolidases. Sequence alignment of *E. coli* PepQ (accession number P21165) with eukaryotic and prokaryotic pita-bread fold enzymes was performed using CLUSTALW [237] and graphically organized with ESPript [238]. Completely conserved residues are highlighted in red and highly conserved residues or regions are boxed and shown in blue. Metal-chelating residues are highlighted with yellow. Numbering shown is for *E. coli* PepQ. Secondary structure assignments shown above the alignment are those from *E. coli* PepQ, while those shown below the alignment are from human PepD. The aligned proteins (with percent identity/similarity to *E. coli* PepQ, along with the number of aligned positions shown in parentheses; followed by the accession number of the sequence) are: *Alteromonas sp.* PepQ (50/67, 441), Q44238; *E. coli* PepP (31/46, 330), P15034; *Pyrococcus furiosus* PepQ (24/40, 337), P81535; *Arabidopsis thaliana* Xaa-Pro Dipeptidsae (34/51, 292), Q8L780; *Homo sapiens* PepD (29/45, 466), P12955. The degree of identity and similarity was determined by two-sequence alignment with BLAST [397].


Q_E.coli
Q_Altermonas
P_E.coli
Q_P.furiosus
Q_A.thaliana
Q_H.sapiens

the eukaryotic prolidases. Furthermore, the *E. coli* sequence shows good coverage of the human gene, with only one region of 10-15 residues missing (Figure 4.1, between *E. coli* residues 120-125). Although these additional regions may be shifted in our alignment, in a previous alignment [364], four regions of at least ten residues appeared in *E. coli* and human prolidase, but did not appear in *P. furiosus* prolidase (*E. coli* residues 35-53, 303-321, 360-372 and 391-415). In these regions, eleven residues (*E. coli* residues Gly36, Asp45, Phe50, Leu309, Ser319, Glu321, Leu369, Arg370, Glu391, Leu393 and Leu394) are conserved. Of these residues, all but two (Ser319 and Glu391) are also conserved among *E. coli*, humans and *Arabidopsis* (Figure 4.1). While shorter than ten residues, another additional region appears in all of the sequences, but not in *P. furiosus* PepQ – an N-terminal loop extension (94-101), though this region does not include any conserved residues.

To better understand the potential significance of sequence conservation between the *E. coli* and human prolidases, we solved the structure of the bacterial enzyme at 2.0 Å resolution (Figure 4.2A, Table 4.1). The protein is comprised of two sections – an N-terminal domain and a C-terminal catalytic domain. The catalytic domain features the predicted, canonical pita-bread fold common to this family of enzymes. At the center of the pita-bread fold is the active site, containing two metal ions chelated by five residues (metals shown in green). The asymmetric unit contains a single PepQ dimer, which is the native oligomeric structure of this protein [365], arranged head-to-tail with inter-dimer contacts made between both domains. With tertiary and quaternary features appearing as expected, we next focused our analysis on the regions of sequence not

Figure 4.2: PepQ Forms a Canonical Pita-Bread Fold with a Binuclear Active Site. (A) The PepQ dimer (PDB entry 4QR8) is shown with one monomer shown in yellow and one monomer colored by domain: N-terminal (residues 1-159, red) and catalytic (160-443, blue). The magnesium ions are colored green. Image was rendered in Pymol [398]. (B) The PepQ dimer with new regions of sequence (those not in *P. furiosus*) highlighted (residues 35-53, red; 303-321, blue; 360-372, pink; 391-415, yellow). (C) Electron density shows conserved active site residues coordinating two magnesium ions.

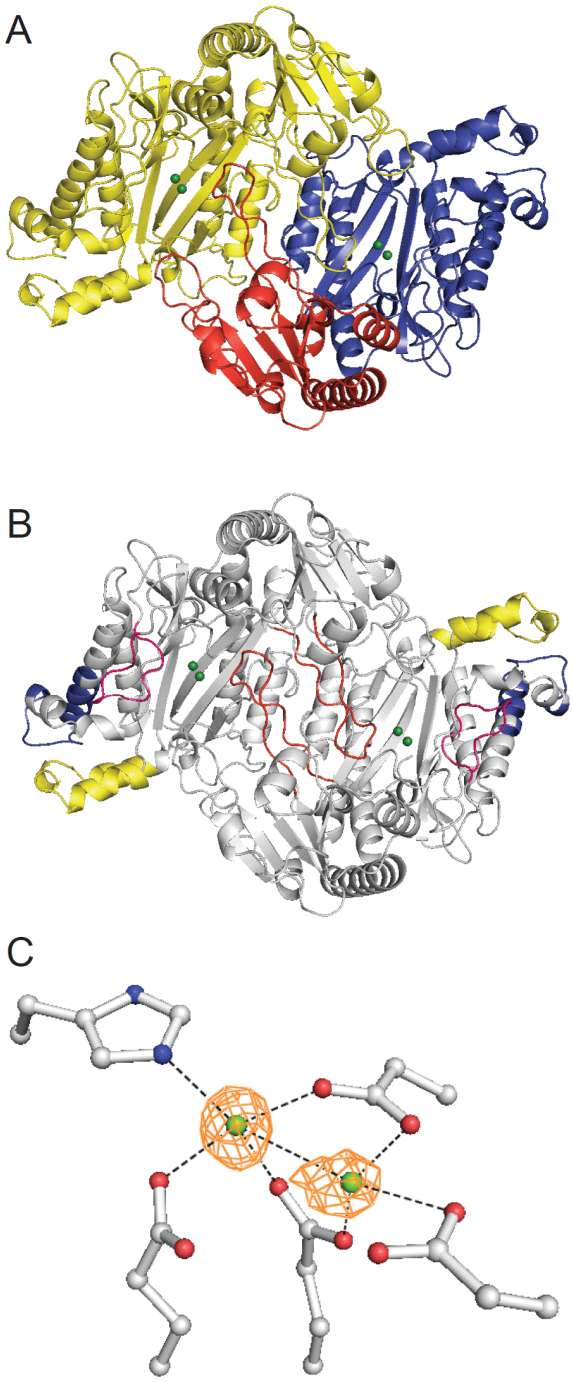

A
B
C

Table 4.1: Statistics of Crystallographic Analysis for PepQ.

PDB Entry	4QR8
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	72.57, 97.44, 126.94
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	2.00 (2.03 to 2.00)[1,2]
[3]R_{sym} or R_{merge}	11.6% (0.695)
$I / \sigma I$	18.0 (2.0)
Completeness (%)	97.0 (92.3)
Redundancy	4.5 (3.5)
Refinement	
Resolution (Å)	2.0
No. reflections	59597
[4]R_{work} / [5]R_{free}	17.39% / 21.1%
No. atoms	
Protein	7052
Ligand/ion	4
Water	1314
B-factors	
Protein	20.3
Ligand/ion	19.6
Water	31.2
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.80

[1]One crystal was used to collect each of the dataset.
[2]Values in parentheses are for highest-resolution shell.
[3]$R_{sym} = \sum_h \sum_i |I_{i,hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i |I_{i,hkl}|$, where $I_{hkl,i}$ is the intensity measured for a given reflection with Miller indices h, k, and l, and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection. [4]$R_{work} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively.
[5]R_{free} was calculated as R_{work} using a randomly selected subset (10%) of unique reflections not used for structure refinement.

found in *P. furiosus* and the residues in those regions that are conserved in other sequences.

These conserved regions consist of two helices and two loop structures, and three of these structural features are in the catalytic domain (Figure 4.2B). The N-terminal loop (highlighted in red) makes significant contact with the same loop from the other subunit. The loop in the catalytic domain (highlighted in pink) extends into the active site. The two helices in the catalytic domain (highlighted in blue and yellow) are on the outside edge of the domain, with both helices in contact with each other and one also in contact with the loop in the catalytic domain (pink). Given the location of these regions of sequence, it is not surprising that only two of the nine residues found in regions absent from *P. furiosus* (but conserved from *E. coli* into the eukaryia), are located near the active site of the enzyme (Asp45 and Arg370). The Arg370 equivalent residue in *Alteromonas* (also Arg370) has been predicted to play a role in organizing water in the active site and, possibly, interacting with the C-terminus of the substrate dipeptide [379]. Asp45, which reaches into the active site of one monomer from a loop region in the N-terminal domain of the other monomer, is seen in *E. coli* to be within interaction distance of Arg370, with the charged ends of the side chains approximately 3.5Å apart. The conservation of this interaction suggests co-evolution of these residues in support of additional known interactions in the active site.

The active site of *E. coli* PepQ also features canonical metal binding residues, Asp246, Asp257, His339, Glu384 and Glu423, chelating two metal ions. Because PepQ was crystallized in buffer containing magnesium, the density found in this region is most

likely derived from magnesium ions (Figure 4.2C). Additionally, the mF_0-DFc difference map shows greater density for one of the two metal ions (chelated by His339), consistent with reports from other pita-bread fold peptidases that this binding site has a higher affinity for metal ions [391,399]. The decreased occupancy at the second metal site is surprising, given that the magnesium concentration during crystallization was in the millimolar range. This observation suggests that the affinity for magnesium of either PepQ in general, or this site in particular, is not as high as seen for the preferred manganese ion in related proteins, reported to be in the low- or sub-micromolar range [381,391,400]. However, metal binding by prolidase does not necessarily convey enzymatic activity, leaving the functionality of magnesium-bound PepQ unresolved.

E. coli PepQ Can Utilize Mutliple Metals for Catalysis. Despite the shared pita-bread fold, prolidases, methionine aminopeptidases and proline amino peptidases from a range of taxa, display widely varying abilities to bind and utilize different metals for catalysis. The presence of magnesium ions in both metal binding sites of *E. coli* PepQ (Figure 4.2) suggests that this prolidase might be enzymatically active with this metal, though magnesium is not known to be the preferred metal of any pita-bread fold enzyme. We therefore examined the ability of PepQ to utilize various divalent cations – testing the dominant ions found in pita-bread fold proteases: manganese, zinc, cobalt, iron, nickel, copper, magnesium and calcium (Figure 4.3). As with many proteins in this family, manganese appears to be the optimal metal for PepQ activity, with cobalt a distant second. Nickel and copper are not generally employed by this family of proteases. Other metals, such as zinc and calcium, are known to require specific

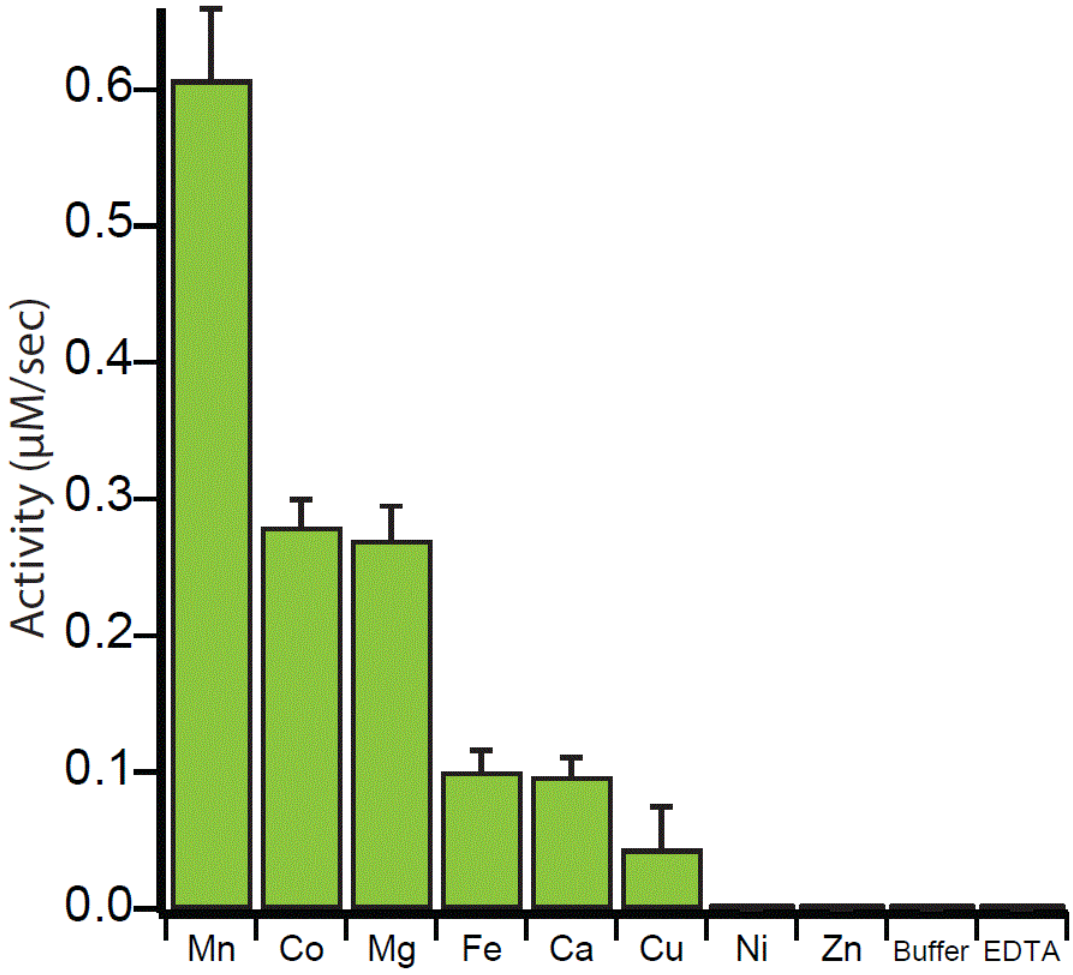


Figure 4.3: PepQ Utilizes Various Divalent Metals with Differing Efficiency. *E. coli* PepQ (50 nM) was assayed in the presence of various metals, in the absence of added metal (Buffer) or in the presence of EDTA. All metals used were in the form of metal-dichlorides. Error bars show the deviation of three independent samples.

coordination and spacing regimes that are not easily accessed in the active site of many pita-bread fold proteins [391], leading to little or no activity, consistent with our observations with PepQ (Figure 4.3). Magnesium, a metal that only rarely conveys activity in other pita-bread proteases, displayed significant levels of activity with PepQ, similar to cobalt. As expected for a metalloprotease, the addition of EDTA abolished the activity of PepQ.

To control for contaminating metal in the buffer, as well as for metal that was not removed from the active site prior to the experiment, PepQ was also tested in buffer in the absence of any residual metal (Figure 4.3). An absence of enzymatic activity indicates that the pre-incubation of PepQ with EDTA effectively stripped any remaining bound metal. Whether the metals that convey little or no activity do not bind, or bind, but are incapable of supporting catalysis, is unknown. Zinc, for example, has been shown to bind to the active site of some pita-bread fold peptidases and still not to convey activity [361,364,388,391]. It is possible that the inactivity of PepQ in the presence of zinc and nickel is the result of weak metal binding, which could, in principle, be examined by increasing in the concentrations of these metals in the PepQ assay. However, the concentrations of these metals are not thought to be higher *in vivo* than used here *in vitro* [401], implying that these metals are not likely used to support catalysis in the cell.

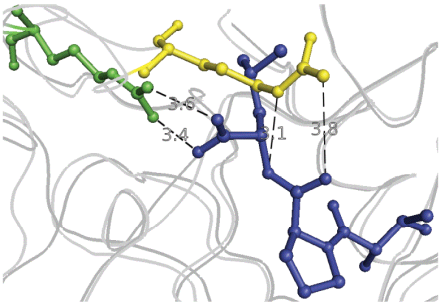
Ionic Interactions Favor the Substrate Peptide C-Terminus. While the metal-chelating residues of prolidase are well-described, the identity of these residues does not reliably predict metal usage. Likewise, *de novo* prediction of substrate specificity is

limited to dipeptides, as well as certain small molecules such as organophosphates that are hydrolyzed with far lower efficiency than peptide substrates. Different prolidases have varying affinities for dipeptides, though cleavage of collagen-catabolism products, such as GlyPro or AlaPro, seems to be conserved [362,365,388,402]. How dipeptide specificity is enforced by prolidase, as well as why these enzymes display a total lack of activity toward longer peptides, is not obvious, particularly given the high structural similarity between prolidase, which cannot cleave peptides longer than two amino acids, and proline aminopeptidase, which can cleave tripeptides (Xaa-Pro-Xaa) at the N-terminal side of proline. To further examine these differences, the structures of PepQ and PepP, the *E. coli* proline aminopeptidase, were aligned for comparison (Figure 4.4A). In the structure of PepP, which includes a bound tripeptide, Arg371 (of PepP) interacts with the C-terminus of the tripeptide. PepQ Arg370, which is projected further into the active site on one of the loop regions conserved in prolidases from higher organisms, is placed far enough into the active site that it would physically impede the binding of longer peptides, as seen in the overlap between this arginine and the PepP-bound substrate tripeptide. PepQ R370 is, however, in an appropriate position for the guanidinium group of the arginine to interact with the C-terminus of the proline residue.

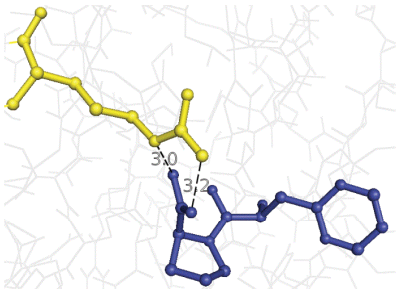
To further examine the potential role of R370 in dipeptide selection by PepQ, the structures of model dipeptides were docked into the active site of the PepQ structure [395]. Docking of dipeptides resulted in a configuration similar to that seen in PepP – the terminus of the dipeptide is in position to interact with Arg370 (Figure 4.4B). In order to experimentally test this interaction, a two-pronged approach was pursued. First,

Figure 4.4: Structural Alignment of Prolidases Reveals Conserved Arginine. (A) The PepQ catalytic domain (residues 160-443) was aligned with *E. coli* PepP, the proline aminopeptidase, with the bound substrate tripeptide ValProLeu (PDB entry 2BHA, residues 175-425; RMSD = 1.05 Å, 1020 atoms aligned). PepQ R370 is shown in yellow, PepP R371 is shown in green and the tripeptide is colored blue. The distances between PepP R371 and the C-terminal oxygens of the tripeptide measured at 3.4 and 3.6 Å. The distances between PepQ R370 and the prolyl-leucyl amide nitrogen and oxygen measured at 3.1 and 3.8 Å, respectively. (B) Docking simulations were performed between PepQ (yellow) and substrate dipeptides using AutoDock Vina [395]. Shown is the substrate PhePro (blue). The distances between R370 and the dipeptide C-terminal oxygens measured at 3.0 and 3.2 Å. (C) *E. coli* PepQ (yellow) and PepP (green) were aligned with *P. furiosus* prolidase (PDB entry 1PV9, residues 124-345, red). PepQ R370, PepP R371 and *P. furiosus* R295 are highlighted. ($RMSD_{EcoliQ\text{-}PfuriosusQ}$ = 0.92 Å, 816 atoms aligned; $RMSD_{EcoliP\text{-}PfuriosusQ}$ = 0.82 Å, 908 atoms aligned) (D) Structure alignment of catalytic domains of *E. coli* PepQ (yellow) and human PepD (PDB entry 2IW2, residues 187-470, purple; RMSD = 0.97 Å, 1179 atoms aligned). (E) R370 in PepQ (yellow) is sequentially and structurally conserved in humans (R398, purple). All structural alignments and distance measurements were performed with PyMOL [398].

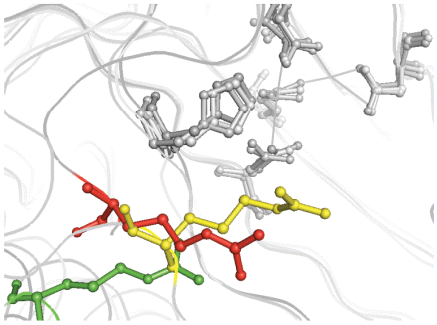
A



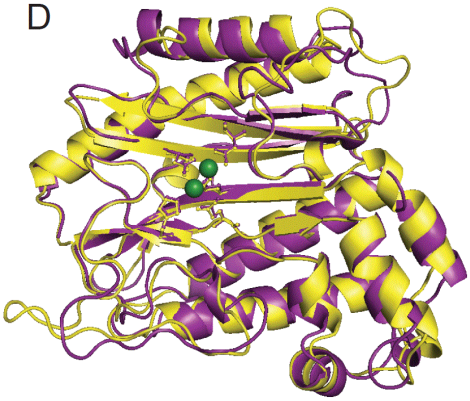
B



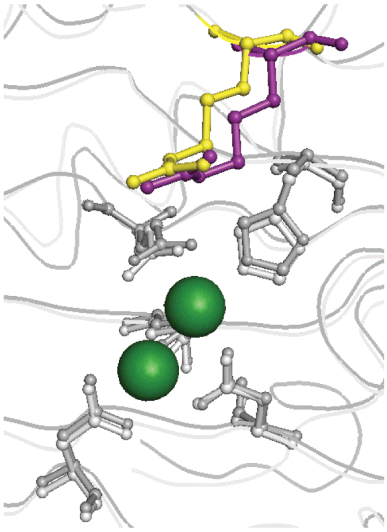
C



D



E



a charge-reversed mutant of PepQ (R370E) was made in which the predicted favorable interaction between the peptide carboxylate and R370 was replaced with an unfavorable interaction. The R370E mutant was expressed and purified following the protocols used for the wild-type enzyme, and eluted on gel filtration chromatography identically to wild-type PepQ (Figure 4.5A), suggesting that both the structure and dimer stability of the enzyme was not significantly compromised by the R370E mutation. To test this conclusion further, we examined the thermodynamic stability of the R370E mutant relative to wild type PepQ using guanidinium-induced unfolding at 25 °C. As shown in Figure 4.5B, the two proteins show essentially identical unfolding transitions, indicating that the thermodynamic stability of PepQ is not affected by the R370E mutation. Consequently, any changes in the activity of R370E relative to the wild-type enzyme are not likely due to secondary effects of the mutation on protein structure or stability.

As a second approach to examining the role of R370, we examined the activity of PepQ toward a substrate dipeptide featuring a terminal amide, rather than a carboxylic acid. With this modified substrate, the predicted interaction with Arg370 should remain favorable, as hydrogen bonding could still occur, though the favorable ionic interaction would be lost. Due to the partial positive charge of the amide nitrogen, a potentially favorable ionic interaction between the modified amide terminus of this substrate dipeptide and the glutamate of the mutant R370E remained a possibility. Kinetic analysis of both wild-type and R370E PepQ with both AlaPro-COOH and AlaPro-$CONH_2$ strongly supports the proposed model for the role of R370 (Table 4.2). R370E displayed a considerably higher K_m for the substrate AlaPro-COOH than the wild-type

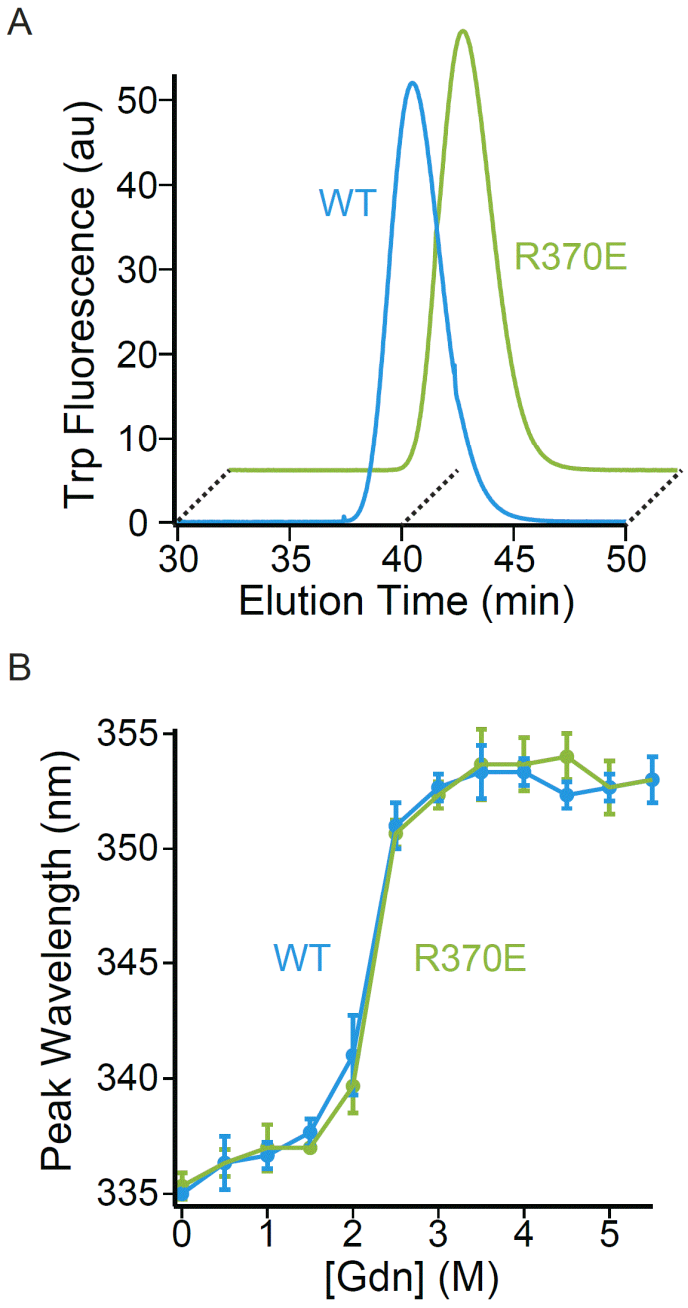


Figure 4.5: R370E Mutation Does Not Perturb PepQ Structure or Stability. (A) Wild-type (WT, blue) and R370E (green) PepQ (10 nM) were analyzed by size exclusion chromatography. (B) Wild-type (WT, blue) and R370E (green) PepQ (50 nM) were incubated with varying concentrations of guanidinium-HCl and the peak position of the tryptophan fluorescence emission spectrum of each was determined. Error bars indicate the deviation from three independent samples.

Table 4.2: Kinetic Parameters for the Hydrolysis of the Dipeptides Ala-Pro and Ala-Pro-NH_2 by Wild-Type and R370E PepQ.

	AlaPro-COOH			**AlaPro-$CONH_2$**		
	k_{cat}/Km ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/Km ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (mM)
WT	1.2×10^5	139.1±2.5	1.2±0.1	1.2×10^4	80.7±19.7	6.6±0.5
R370E	5.1×10^1	6.5±1.1	127±2	2.3×10^2	1.1±0.4	4.7±0.2

protein, while actually having a lower K_m for AlaPro-$CONH_2$, when compared to wild-type prolidase. The reduction in k_{cat} seen in R370E is likely due the role of this residue in the organization of water and other residues in the active site [379]. The changes in K_m, with information from both a charge-reversed protein and a charge-neutralized substrate, strongly suggest an interaction between the substrate carboxylate group and R370.

The placement of the loop arginine evolved for substrate selectivity. The location of the key R370 residue in PepQ, and similar arginine residues in other pita-bread fold enzymes, may have been an important factor in the evolution of prolidase. To examine this idea, the structures of *E. coli* PepP, *E. coli* PepQ and *P. furiosus* PepQ were aligned (Figure 4.4C). The loop region containing this arginine is absent from *P. furiosus* prolidase. While the archaeal prolidase retains an arginine in the same spatial location of the active site (Arg295), it appears to be in an intermediate position, relative to PepP and PepQ from *E. coli*. The active site residues of the three enzymes are nearly super-imposable, indicating that this change in position is not an artifact of the structure alignment. It thus seems reasonable that the addition of the expanded peptide regions containing the arginine in Gram-negative bacteria and eukaryotes could have resulted from evolutionary fine tuning of the enzyme for high specificity dipeptidase activity. While the loop residues are conserved in the sequence of human prolidase, we sought to verify the placement of this residue as a potential means of selecting for dipeptides. The structures of the catalytic domains of human and *E. coli* prolidases were aligned, showing high conservation in the secondary structure elements and an RMSD of less

than 1Å (Figure 4.4D, secondary structures shown in Figure 4.1). The critical arginine is observed in the active site of the human prolidase in a position nearly identical to the bacterial residue (Figure 4.4E). This suggests that after the initial evolution of the loop regions for the placement of this residue, no further optimization was necessary for the selection of dipeptides as the enzyme evolved further over the course of several billion years.

Discussion

The results presented here support a role for substrate length specificity in pita-bread fold enzymes through the positioning of an active site arginine. With a high-resolution structure of *E. coli* PepQ in hand, we were able to compare it to related enzymes, both bioinformatically and structurally. We found that the position of the active site arginine has changed during the evolution in this family of proteins, with a shift further into the active site leading to selection against peptide substrates greater than two residues in length. Not only does the placement of this arginine physically occlude longer peptides, as seen structurally, but also, kinetic analysis demonstrates the important role of the ionic interaction between this positively charged residue and the negatively charged C-terminus of the substrate dipeptide. We have also found that while this protein is maximally active with manganese, it can utilize other metals, including magnesium, an uncommon property for this family of metalloproteins.

Although they are very similar proteins, the members of this family of enzymes vary in a number of significant ways. Of particular note is the presence of several large

regions of additional residues in the prolidase sequences of Gram-negative bacteria, single-celled eukaryotes and higher plants and animals that are absent in the sequences of other bacterial prolidases, as well as proline aminopeptidase. When comparing the additional regions found in *E. coli* that are absent in *P. furiosus*, perhaps the most striking insert is the N-terminal loop. This loop not only makes significant contact with its counterpart on the adjacent subunit, but also contributes to the opening of the active site, relative to the loop-less *P. furiosus* structure. The role of the two helices inserted into the catalytic domain of *E. coli* is more difficult to surmise, given the distance from the active site and the other subunit. The structural rearrangements created through the insertion of the N-terminal loop may be stabilized by the presence of these helices, but this remains to be examined. However, both of these helices, as well as the N-terminal loop are found in the sequence and structure of *E. coli* PepP (Figure 4.1) [381,387,399]. This suggests that these changes may have occurred in an ancestor of this family before the divergence that led to separate substrate specificities of PepQ and PepP. The component found in neither PepP nor *P. furiosus* PepQ is the loop in the active site of the protein, which contains the conserved arginine.

Enzymes generally dictate specificity by utilizing binding pockets with specific interactions that favor some substrates and disfavor or occlude other substrates [403-406]. Pita-bread fold proteins are no exception – the occlusion of branched amino acids and selection against small amino acids in substrates has been observed previously in PepP [391] and charge interactions have been observed to dictate specificity in some prolidases [407]. Despite the high level of conservation among proteins with the pita-

bread fold, these enzymes are very specific for their substrates, at least in terms of peptide length. While the evolutionary benefit of selecting for dipeptides stems from the availability of byproducts of protein catabolism, like those derived from collagen, the movement of this residue also levies an advantage against certain small molecules. *E. coli* PepQ can not only hydrolyze at least thirteen different dipeptides, but also an assortment of organophosphates and other small molecules [365]. While these substrates vary considerably on the N-terminal side of the scissile bond, the C-terminal end of all previously tested substrates shared a negatively charged group, either a carboxylate or a nitro group [365]. Reactivity toward these substrates is likely dictated by the positioning of R370 in the active site of the enzyme. We have shown that the addition of a loop in the catalytic domain, near the active site, allowed for the substrate peptide length-determining residue to be repositioned, altering the specificity of the enzyme. Utilization of an arginine at the designated position in either PepP or PepQ for this selection likely stems from the ability of arginine to interact ionically with both oxygens in the C-terminus of the substrate peptide, as well as through hydrogen bonding. A lysine at this position is unlikely to interact with both oxygens due to spatial and angular limitations. Although the enzyme is still functional without this interaction, the activity is severely compromised, which is consistent with reports that *E. coli* PepP has minor activity against dipeptides [377]. Interestingly, the genomes of sequenced *Pyrococcus* species include PepQ sequences, but lack PepP annotations [408-414]. Given the intermediate positioning of the conserved arginine in *P. furiosus* PepQ, a dual functionality for cleaving di- and tripeptides may be predicted for that enzyme.

While many prolidases share various similarities, structural and biochemical data reveal that *E. coli* prolidase is more similar to the human enzyme than other enzymes. The catalytic domains of the *E. coli* and human prolidases align with an RMSD of less than 1.0 Å, and this bacterial enzyme utilizes magnesium to a similar extent as human prolidase, suggesting that the specific placement or conformational flexibility that influences metal coordination is shared between the enzymes of these two distantly related organisms. Variable metal usage has been postulated to serve as a regulatory role in aminopeptidases [386]. Other similarities between these two proteins may allow for the *E. coli* protein to provide insights into the functionality of the human protein. Not only is the placement of the critical arginine residue unchanged in the human prolidase, but many other residues are conserved between the two proteins, including some that are associated with disease alleles, for example, E412K and G448R in human prolidase deficiency [400].

Interestingly, despite its role in substrate selectivity, no mutation of the equivalent R370 residue has yet been associated with the onset of prolidase deficiency in humans [372,415-417]. It is possible that mutation of the same residue in the human enzyme results in a reduction of enzymatic activity too small to yield an observable phenotype. However, this seems unlikely, given that losing R370 in PepQ results in a decrease in enzymatic activity that is orders of magnitude more severe than caused by single residue mutants in the human enzyme with known phenotypes [400]. Notably, many disease associated mutations also (i) decrease the stability of the enzyme, (ii) have a reduced abundance *in vivo*, and (iii) perturb the dimer binding constant so that

formation of active enzyme requires protein concentrations that are much higher than needed for the wild-type protein [400]. It is possible that the impact of these mutations on folding and stability is, overall, more serious than the loss of activity seen with the arginine mutation alone, which has no effect on enzyme stability or folding. Alternately, loss of the active site arginine might have such severe developmental consequences that homozygous and many heterozygous genotypes are simply not viable. It also remains possible that the number of studied cases of prolidase deficiency is yet too small to have sampled every disease-associated allele. Although the critical active site arginine residue has yet to be associated with the physiological outcomes of reduced activity, observed defects in conserved regions in one enzyme generally predict similar defects in other, highly homologous enzymes. Utilizing *E. coli* PepQ may, therefore, be an effective strategy for studying prolidases in general and deficiencies in human prolidase specifically.

Many proteins evolve through the addition of loops or domains to gain solubility, new interactions or new activity [418]. Although it is not necessary for a prolidase to have the additional catalytic domain loop in order to place specificity-defining residues in the active site, as seen in the *P. furiosus* PepQ, the positioning of this residue in Gram-negative bacteria and higher organisms has remained constant during billions of years of evolution, indicating a preferred or optimal placement for activity. While *E. coli* PepQ may serve as a tool for studying prolidases in general, other prolidases may provide further insight into the role of these enzymes beyond collagen recycling in humans. One such protein of interest is the Xaa-Pro peptidase from the nearly

exclusively catabolic organism *Mycoplasma mobile*. Examination of this protein structurally and biochemically would reveal how this minimalist organism utilizes this enzyme, demonstrating the extent of its role in metabolism. Despite the continual advance of knowledge about prolidases – structurally, biochemically and genetically – much is left unknown about the role of these enzymes in metabolism and their connection to disease.

CHAPTER V

ALTERED FOLDING OF AN E. COLI SUBSTRATE PROTEIN BY GROEL

Most proteins must fold from linear polypeptides into three-dimensional structures that are suited to their intended function. While this process often occurs without incident, the folding of some proteins is intrinsically fraught with slowly converting or error-prone intermediates, and any protein can experience difficulties during cell stress [123,289,419]. If the rate of adopting a non-native conformation exceeds the rate of folding correctly, a protein will fail to acquire its native conformation and become kinetically-trapped or misfolded, or may interact with other non-native proteins to form aggregates [114,123,190,198,420]. To recover misfolded or aggregated proteins, all cells utilize molecular chaperones, proteins that aid other proteins in reaching or maintaining the native conformation. One chaperone system, the Hsp60s, alters the kinetics of folding, misfolding, and aggregation to promote native-state conformations and correct folding [127,239,240,290,339]. These chaperonins are essential to almost every organism across the kingdoms of life [133,213].

The Hsp60 in *Escherichia coli* is GroEL, which is a homotetradecamer composed of ~57kDa monomers that form two heptameric rings oriented with closed ends facing together and open ends facing the solvent [141,152]. Utilizing a large hydrophobic surface near the open end of one of these rings, GroEL binds to unfolded or misfolded substrate proteins [141,154,157,165,330,421,422]. In an ATP-binding-dependent manner, GroEL undergoes a large conformational change that expands the

apical domains, reducing the hydrophobic surface and causing the release of the substrate protein from the GroEL wall, while concomitantly exposing the binding sites of the cochaperone GroES, a heptamer totaling ~70kDa [131,154,157,163]. GroES binds a GroEL ring like a lid, encapsulating the substrate protein inside the newly-formed cavity, where folding can occur [154,157,206]. This closed GroEL-ES-ATP cavity, known as the *cis* ring, persists for 5-25 seconds, depending on conditions [157,159,163]. The ATP hydrolyzes during this period, but it is the binding of nucleotide to the opposite, or *trans*, ring of GroEL that causes an allosteric change in the protein, leading to the release of GroES, nucleotide, and the substrate protein from the other ring. This release occurs regardless of whether or not the substrate protein has folded correctly [173,178]. By the time the *trans* ring binds ATP, this ring has already bound a substrate protein that will subsequently be encapsulated, thus continuing the GroEL reaction cycle [157,163,173]. While much is known about the reaction cycle, how GroEL aids in the folding of substrate proteins has remained a contentious topic, even after decades of research, the popular consensus is that this molecular machine is believed to work by two primary mechanisms.

First, encapsulating a substrate protein inside the GroEL-ES cavity effectively removes it from the cytosol, preventing any interactions with other proteins. Here, GroEL serves as a passive- or Anfinsen-cage, where the chaperone exists solely to protect a folding monomer from intermolecular interactions and aggregation [192,193,249]. This model assumes that all proteins are capable of folding independently, as long as these folding proteins have no competing, exterior forces.

Under this assumption, GroEL exists only to permit this unperturbed folding to occur. Kinetically, the chaperonin would simply reduce the aggregation rate to zero.

In the second mechanism, interactions between the chaperone and the substrate protein may alter how the substrate protein folds. This active model is divided into two main sets of interactions: confinement inside the cavity, where the size and character of the cavity walls of GroEL and GroES influence potential conformations of the folding substrate protein, and iterative annealing, where the conformational change that occurs during each round of ATP binding induces unfolding in the bound substrate protein. This unfolding disrupts kinetically-trapped conformations and returns the substrate protein to a higher energy state, providing a new opportunity to fold [158,159,166-168,187,190]. The GroEL C-termini, which contribute to controlling cavity size, cavity character, the GroEL ATPase rate (and thus the cavity lifetime), and the unfolding action of the chaperone [203,204,209,244,250,347,423], are involved in both components of the active model. The active model stresses that interactions between the chaperone and the substrate protein change the availability of intermediate folding conformations, whether it is the diminishing of non-native forms or the promotion of on-pathway intermediates. While the active model highlights the enhancement of folding rates and the reduction of misfolding rates, it generally acknowledges that the aggregation rate of the encapsulated protein is also zero.

In vivo, GroEL interacts with hundreds of different proteins, including 50-100 proteins that require the chaperone to fold correctly [89,90]. Research on GroEL, however, has focused on a handful of substrate proteins that have been studied

extensively *in vitro*. While these various substrate proteins have different tendencies for folding, misfolding and aggregation, and are influenced by GroEL differently, nearly all share the feature of not being *E. coli* substrate proteins of GroEL [148,149,200,205,224,342]. The availability of chaperone inside the cell influences how proteins evolve [357,359], so examining the folding of proteins that evolved in the presence of the studied chaperone would seem critical. Recently, a true substrate protein of GroEL, *E. coli* 4-hydroxy-tetrahydrodipicolinate synthase (DapA), was examined, and its folding was shown to be altered by GroEL in a manner that mimics the folding of a non-*E. coli*, non-GroEL-utilizing homolog [88]. This finding demonstrates the importance of examining substrate proteins that are native to the chaperonin under study.

Here we introduce the *E. coli* prolidase, PepQ, as a new model folding substrate protein of GroEL and present evidence that its folding is actively enhanced by the chaperonin. PepQ requires the chaperone *in vivo* [89,90], and under conditions where this protein does not aggregate, folding is stimulated 15-20 fold by GroEL-ES. GroEL induces changes during the early stages of PepQ folding that occur on the timescale relevant to the cavity lifetime, and these alterations in folding are dependent on interactions with the GroEL C-termini, which influence PepQ unfolding prior to the start of folding. We find that the unfolding action of GroEL is critical for maximizing the stimulation of PepQ folding, demonstrating a clear role for the iterative annealing mechanism of GroEL function.

Experimental Procedures

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 [158,159,166,168,423]. The cysteine mutant of PepQ, A24C, was generated via site-directed mutagenesis and the sequence was verified by DNA sequencing. This mutant was expressed and purified following the protocol for wild-type PepQ.

Labeling of PepQ. A24C PepQ was labeled using either 5-iodoacetamidofluorescein (fluorescein, F), 5-(2-acetamidoethyl) aminonaphthalene 1-sulfonate (EDANS, ED), or tetramethylrhodamine-5-iodoacetamide dihydroiodide (tetramethylrhoadmine, TMR). The extent of labeling was determined by protein quantification by the Bradford assay (Bio-Rad) and dye quantification under denaturing conditions using known molar extinction coefficients [157,316]. Site-specific labeling was verified through denaturing anion-exchange chromatography and analysis of proteolytic fragments [316].

Enzymatic Refolding. 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 minutes prior to further use. CD spectra show a complete loss of secondary structure under these conditions (data not shown). 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. Chaperone-mediated folding

reactions using either wild-type or mutant tetradecameric GroEL began with a 50-fold dilution of denatured PepQ into TKM buffer containing chaperone. GroES and ATP were added to initiate the reaction cycle and the reaction was quenched with hexokinase and glucose [159,163,166,424]. 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 [163,199,203]. After quenching, all samples were incubated for 60 minutes at room temperature to allow for dimerization. The enzyme activity of all samples was measured through a coupled reaction described previously [423].

Fluorescence and Light Scattering. Light scattering and fluorescence measurements were conducted with a PTI fluorometer, with temperature regulation through a jacketed cuvette holder (Neslab). For both types of experiments, the assay was initiated with chemically-denatured PepQ being diluted 50-fold into temperature-equilibrated TKM buffer. Tryptophan fluorescence was monitored with excitation at 295 nm and emission read at 340 nm. The excitation and emission wavelengths were both 340 nm for light scattering experiments.

Stopped-flow Fluorescence. Stopped-flow experiments were conducted as described previously [166,168,316,424], using an SFM-400 rapid mixing unit (BioLogic) equipped with a custom-designed two-channel fluorescence detection system. Mixing was done using two syringes, one containing GroEL-PepQ complexes and one containing GroES and ATP.

Steady-state FRET. Steady-state fluorescence measurements were conducted with a PTI fluorometer, with temperature regulation through a jacketed cuvette holder (Neslab). FRET was calculated from donor-side fluorescence in the presence of unlabeled or acceptor-labeled molecules as previously described for other GroEL substrate proteins [157,316].

Sample Preparation for Fluorescence Correlation Spectroscopy. 24-TMR PepQ was diluted greater than 40-fold (to 5 μM) into 8 M urea, 25 mM glycine phosphate, pH 2 and incubated for 20 minutes at room temperature. For spontaneous folding reactions, this 24-TMR PepQ 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 μL to 50 μL of 1 μM GroEL in TKM buffer. For GroEL-mediated folding, 5 μM denatured 24-TMR PepQ was diluted to 100 nM in TKM buffer containing GroEL (200 nM final concentration). After a 10 minute incubation at room temperature, this solution was diluted into TKM buffer containing GroEL, GroES and an ATP-regeneration system [159]. Folding was initiated by the addition of ATP. The final concentration of ATP was 2 mM, GroEL was 1 μM, and GroES was 2 μM. Folding was quenched by the addition of 20 μL of the reaction mixture with an equal volume of hexokinase and glucose. In refolding assays at both 1 nM and 2 nM, PepQ dimerization was not observed, even after eight hours (by lack of activity, data not shown).

Fluorescence Correlation Spectroscopy. 10 μL of the quenched reaction mixture was dispensed on a BSA-blocked coverslip mounted on the objective of a custom-built

confocal microscope [425] and covered with a humidified chamber to prevent evaporation. Autocorrelation curves were collected for each sample for two minutes with a 500 μsec sampling time using a 50 μW laser at 561 nm. Autocorrelation curves were normalized in mean amplitude between 10^{-6} and 10^{-5} seconds 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 24-TMR in buffer), and as a native monomer (obtained by allowing a GroEL-mediated folding reaction with 1 nM PepQ to continue an hour) were also determined. Each autocorrelation curve was fit using a multi-component model [426,427] to account for populations of freely diffusing and GroEL-bound PepQ. Each autocorrelation curve was fit using two different sets of parameters. 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 and was normalized using the observed diffusion coefficients for free and bound PepQ. These methods yielded statistically consistent results.

Protease Protection. The protease sensitivity of non-native PepQ bound to a GroEL ring was conducted as described previously for the substrate protein RuBisCO [158,166]. Briefly, 24F-PepQ was bound to asymmetric GroEL-GroES ADP bullets. Chymotrypsin was added, and time points were taken, with the reaction stopped with PMSF. 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 a coupled reaction as reported previously [158,159,424,428,429]. The rate of spontaneous ATP hydrolysis under each condition without GroEL was also determined to control for effects on the coupling system.

PepQ Unfolding in Guanidinium. The unfolding of PepQ was monitored by the decrease in tryptophan fluorescence with increasing concentrations of guanidinium-HCl. PepQ (50 nM) was incubated at room temperature for 60 minutes in solutions of 50 mM Tris, pH 7.4, 10 mM Mg(OAc)2, 2 mM DTT and varying concentrations of guanidinium-HCl. The tryptophan fluorescence was measured using a PTI fluorometer with excitation at 295 nm and emission at 340 nm. Solutions of buffer and guanidinium alone were also measured to control for changes in scattered light.

Results

GroEL stimulates the folding of PepQ. PepQ is an enzyme that catalyzes the hydrolysis of dipeptides containing a C-terminal proline residue. It forms a homodimer, with each monomer (~50.2 kDa) containing N-terminal and C-terminal domains, the latter of which adopts a pita-bread fold and contains the active site [384,423]. PepQ was predicted by two independent proteomics studies to require GroEL-ES *in vivo* [89,90], but to determine whether PepQ requires GroEL to reach its native conformation, the extent of folding of PepQ was measured *in vitro* over time, both with and without the chaperonin system. Upon dilution from chemical denaturant, PepQ is capable of spontaneously folding at room temperature to a yield of 50-60% with a halftime of ~20

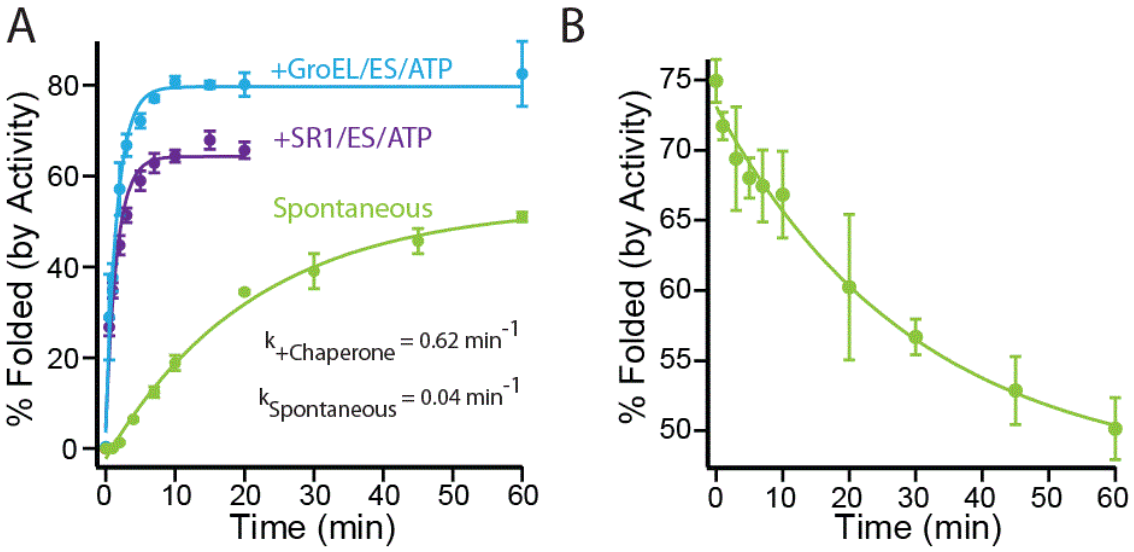


Figure 5.1: GroEL Enhances the Rate and Yield of PepQ Folding. (A) The refolding of PepQ was monitored by the recovery of enzymatic activity. Chemically denatured, wild-type PepQ was diluted either directly into buffer (100 nM, Spontaneous; green) or bound to wild-type GroEL (200 nM) and refolded in the presence of GroES (400 nM) and ATP (2 mM) (+GroEL/ES/ATP; blue). Alternatively, the same 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) (+SR1/ES/ATP; purple). The data were fit to a single-exponential rate law (solid lines), resulting in the observed rates of 0.62 $\pm$ 0.05 min^{-1} for GroEL-mediated folding, 0.62 $\pm$ 0.09 min^{-1} for SR1-mediated folding and 0.035 $\pm$ 0.005 min^{-1} for spontaneous folding. Error bars show the standard deviation from three independent folding experiments. (B) The rate that PepQ becomes refractory to GroEL activity was monitored by the recovery of enzymatic activity. Chemically denatured, wild-type PepQ was diluted directly into buffer (100 nM; green). At given time-points, samples of this solution were mixed with an equal volume of a solution containing wild-type GroEL (400 nM) and GroES (800 nM). ATP (2mM) was added immediately thereafter and the reactions were incubated at room temperature for 60 minutes before quenching. The data were fit to a single-exponential rate law (solid line), resulting in the observed rates of 0.033 $\pm$ 0.006 min^{-1}. n = 3 replicates.

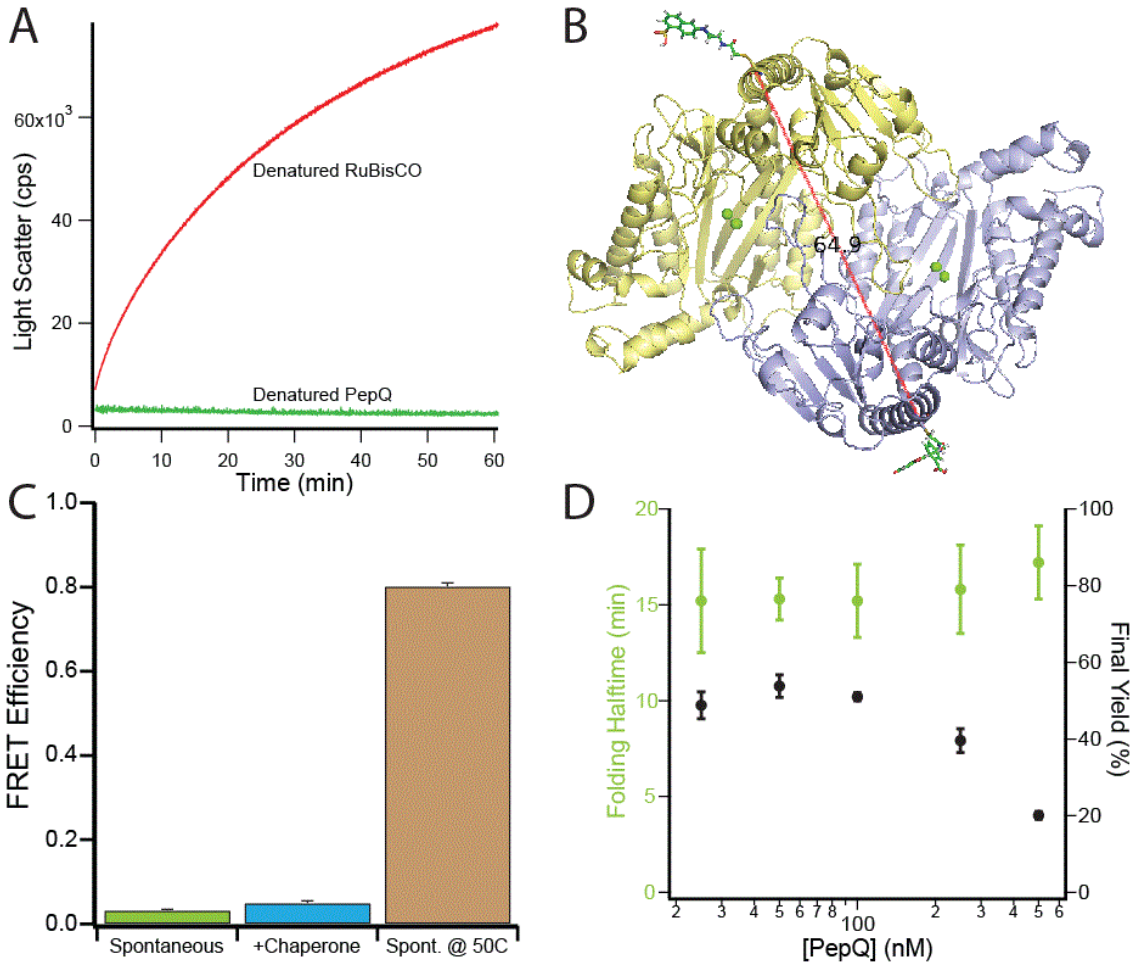
minutes (**Figure 5.1A**). The GroEL-mediated folding of PepQ under the same conditions achieved a yield of 80-90% and was completed with a halftime of ~1 minute (**Figure 5.1A**). Although PepQ does not require GroEL to fold, the chaperonin stimulates the folding rate by 15-20 fold, while also increasing the yield. To determine whether confinement inside the GroEL cavity is critical to the stimulation of PepQ folding, the single ring mutant of GroEL (SR1) was utilized. Upon encapsulation inside SR1, PepQ folds inside the SR1-ES cavity until exposure to conditions that disrupt this chaperone complex; the use of this non-cycling GroEL mutant removes the variables associated with the cycling of the chaperone machinery [206]. PepQ folding inside the SR1-ES cavity achieved a rate similar to wild-type GroEL, though to a lower yield (**Figure 5.1A**).

Although PepQ can fold to some extent spontaneously, the observed decrease in rate and yield suggests that some fraction of the folding PepQ population is acquiring a non-native state. This conformation may be at equilibrium with productive intermediates formed during PepQ folding; the intercoversion between these two populations could results in an observed decrease in the folding rate. To probe for a non-native conformation, spontaneously folding PepQ was subjected to the chaperonin system at various time-points during the course of spontaneous folding so that any protein that remains in an uncommitted state can bind to GroEL and be refolded by the chaperone machinery. If PepQ adopts a conformation that cannot be refolded by GroEL, a decrease in yield over time would be predicted. As shown in **Figure 5.1B,** PepQ becomes refractory to the chaperone at a similar rate as its spontaneous folding. In the

initial characterization of the substrate protein RuBisCO from *Rhodospirillum rubrum* [87], this same strategy showed a rapid loss of RuBisCO's capability to bind and be refolded by GroEL, which was shown to be the result of intermolecular aggregation. Knowing that aggregation could be the source of the reduced spontaneous folding of PepQ, we next sought to examine whether PepQ also suffers from a propensity to aggregate.

To determine whether PepQ folding is limited by aggregation, a number of approaches were employed. Protein aggregation requires the coalescence of smaller particles, such as non-native monomers, into larger particles. This transition can be observed by light scattering, where larger particles scatter more light than smaller particles. When PepQ is diluted from chemical denaturant, no increase in light scatter is observed (**Figure 5.2A**). As controls, native PepQ was diluted from buffer and showed no increase in light scatter, while unfolded RuBisCO was diluted from chemical denaturant, and showed a large increase in light scatter (**Figure 5.2A**). Although PepQ produced no aggregates observable by light scattering, we understood that this method is not very sensitive to small aggregates. Therefore, a more sensitive approach was employed that utilizes FRET to monitor protein aggregation. This method has been used effectively in the study of other GroEL-dependent substrate proteins [168]. Using fluorescently labeled variants of PepQ, with an introduced cysteine on the N-terminal helix labeled with either a donor or acceptor dye (position 24, with 5-iodoacetamidofluorescein (fluorescein, 24F) or 5-(2-acetamidoethyl) aminonaphthalene 1-sulfonate (EDANS, 24ED), respectively), FRET was utilized to examine aggregation

Figure 5.2: PepQ is Not Prone to Aggregation under Permissive Folding Conditions. (A) The aggregation of PepQ was monitored by light scattering. Chemically denatured, wild-type PepQ (100 nM; green) or chemically denatured, wild-type RuBisCO (100 nM; red) was diluted directly into buffer and the static light scattering was monitored at 340 nm and 90 degrees. The traces shown are the average of 3 replicates. (B) Structure of the PepQ dimer showing the position of the dyes. The PDB file (4QR8) was altered using PyMol to visualize the mutation of residue 24 to cysteine and attachment of small fluorescent dyes (EDANS on the yellow monomer, fluorescein on the blue monomer). The distance between the beta-carbons of the cysteines was measured with PyMol to be 64.9 Å. (C) Aggregation of PepQ was monitored by intermolecular FRET. PepQ was labeled at an introduced cysteine (24C) with either IAEDANS (donor, ED) or 5-IAF (acceptor, F). Chemically denatured, donor-labeled PepQ (50 nM) and either chemically denatured, unlabeled PepQ (50 nM) or chemically denatured, acceptor-labeled PepQ (50 nM) were mixed and allowed to fold spontaneously at room temperature (green), fold spontaneously at 50˚C (brown), or fold with GroEL (200 nM), GroES (400 nM) and ATP (2 mM) (blue). The FRET efficiency was calculated from the donor quenching using matched donor-only and donor-acceptor samples. Error bars show the deviation of n = 3 replicates. (D) The spontaneous folding of PepQ was monitored as a function of PepQ concentration. Chemically denatured, wild-type PepQ was diluted 50-fold into buffer to yield a folding reaction at the designated final concentration. Each rate (blue) was derived from a fit of the folding data to a single- exponential rate law. Each yield (black) is the yield achieved after folding for 60 minutes. Error bars show the standard deviation from three independent folding experiments.


A
60x10^3
40
20
0
Light Scatter (cps)
Denatured RuBisCO
Denatured PepQ
0
10
20
30
40
50
60
Time (min)
B
64.9
C
1.0
0.8
0.6
0.4
0.2
0.0
FRET Efficiency
Spontaneous
+Chaperone
Spont. @ 50C
D
20
15
10
5
0
Folding Halftime (min)
100
80
60
40
20
0
Final Yield (%)
2
3
4
5
6
7
8
100
2
3
4
5
6
[PepQ] (nM)

by measuring the proximity of the differently-labeled monomers (**Figure 5.2B**). In these experiments, both labeled proteins, or donor-labeled and unlabeled PepQ, were unfolded together (at a 1:1 ratio) in chemical denaturant and then diluted into buffer for a spontaneous folding reaction, a GroEL-mediated folding reaction, and a spontaneous folding reaction conducted under non-permissive conditions. Non-permissive conditions are achieved by raising the temperature to 50˚C, which leads to aggregation as seen by a lack of activity and increased light scattering (data not shown). The matched donor-fluorescence from each sample was used to calculate the FRET efficiency (**Figure 5.2C**). The lack of FRET in the GroEL-folded sample is due to two factors: 1) the beta-carbons of the labeled residue in the native dimer are ~65Å apart, well outside the Förster distance for this dye pair [430] and 2) only 50% of the dimers contain one of each differently-labeled monomer. The folding reaction conducted at 50˚C shows a large FRET signal, indicating the formation of substancial aggregates. The spontaneous reaction conducted at room temperature, however, shows a similar FRET efficiency as the GroEL-mediated sample, which strongly suggests that slow and inefficient folding PepQ folding is not caused by aggregation.

To further examine whether PepQ forms aggregates, the spontaneous folding of PepQ was measured over a range of protein concentrations. As aggregation is a multimolecular process, it is highly concentration-dependent; therefore, under conditions where aggregation is occurring, the rate and yield should be significantly decreased. Over the range of 25-500 nM PepQ, the halftime of folding is unchanged, though a decrease in the yield is observed at 250 and 500 nM (**Figure 5.2D**). The decrease in

yield indicates that aggregation is occurring at these higher concentrations; however, the standard folding conditions used for most of the work presented here (e.g. **Figure 5.1**) were 100 nM PepQ, which shows no difference in folding rate or yield from the folding reactions conducted with lower protein concentrations. In this experiment, the lower bound of protein concentration was limited by the able to conduct enzyme assays. Examining the folding of PepQ below 25 nM was precluded by the timescale of PepQ dimerization, a process that is necessary for activity. Although the reactions at 25 and 100 nM PepQ showed the same rates and yields, we sought to further expand the range of sampled concentrations.

In order to examine folding at even lower concentrations, fluorescence correlation spectroscopy (FCS) was utilized. Fluorescently-labeled PepQ (24TMR) was diluted 50-fold from chemical denaturant to 2 nM and allowed to fold spontaneously. At given time points, reaction samples were quenched by the addition of a large excess of GroEL. Using FCS, the apparent diffusion time of each sample was measured. As controls, samples of PepQ bound to GroEL and native PepQ were also measured to establish the diffusive behavior of the reference states. The normalized autocorrelation curves of these controls are shown in **Figure 5.3A**. Using the bounds established by the minimum and maximum diffusion times, the relative fraction bound to GroEL in each sample was determined (**Figure 5.3B**). The fraction bound to GroEL is proportional to the fraction that remains unfolded, and this was used to determine the folding halftime, which for spontaneous folding was ~53 minutes and for GroEL-mediated folding was ~3.6 minutes, which is an enhancement of ~15 fold. These rates match the folding rates

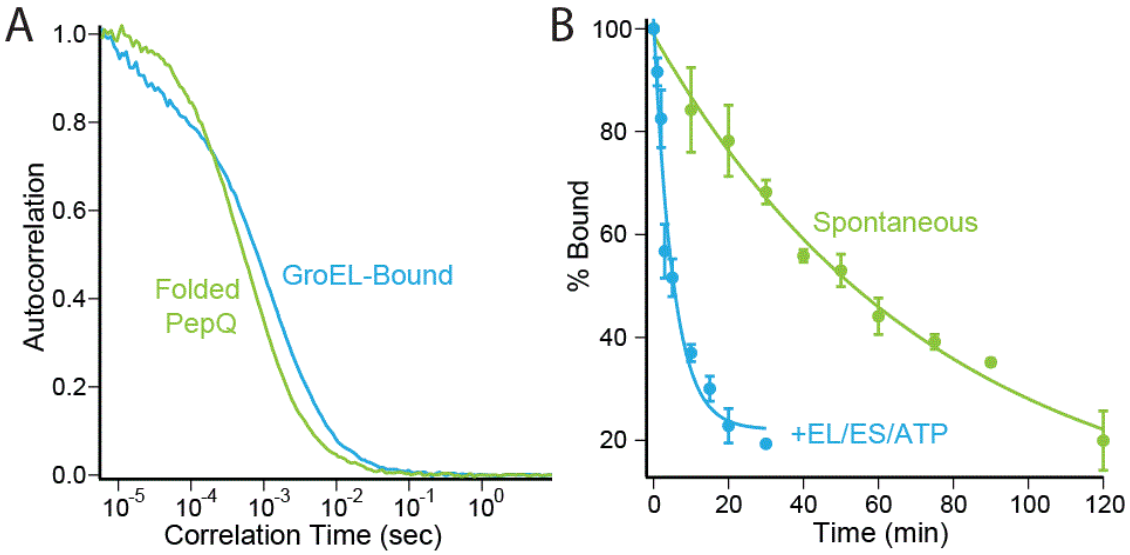


Figure 5.3: GroEL Stimulates the Folding of the PepQ Monomer. (A) Example FCS curves. Shown are the curves for folded, monomeric PepQ (green) and GroEL-bound PepQ (blue). (B) The refolding of PepQ was monitored by the shift in the apparent diffusion time. PepQ was labeled at an introduced cysteine (24C) with 5-TMRIA (TMR). Chemically denatured, TMR-labeled PepQ was diluted either directly into buffer (2 nM, Spontaneous; green) or bound to wild-type GroEL (1 uM) and refolded in the presence of GroES (2 uM) and ATP (2 mM) (+GroEL/ES/ATP; blue). The data were fit to a single-exponential rate law (solid lines), resulting in the observed rates of 0.19 ± 0.04 min^{-1} for GroEL-mediated folding and 0.013 ± 0.002 min^{-1} for spontaneous folding. n = 3 replicates.

of the same labeled PepQ molecule at 100 nM (data not shown). This observation further demonstrates that the process that limits the spontaneous folding of PepQ is not concentration-dependent. Slow folding of PepQ cannot be due to aggregation and must, therefore, be due to misfolding.

GroEL alters the folding trajectory of PepQ. As it is not required to block aggregation, GroEL must enhance PepQ folding through some other mechanism. Recently, it was shown that GroEL enhances the folding of only some specific sections of the *E. coli* protein DapA [88]. To investigate whether GroEL alters how PepQ folds, the tryptophan fluorescence of PepQ was observed during different folding reactions. Importantly, PepQ has multiple tryptophans, while both GroEL and GroES contain none. The tryptophan fluorescence of spontaneously folding PepQ shows a single, downward transition (**Figure 5.4A**). Tryptophan fluorescence often increases during protein folding in response to the burial of these residues inside the protein resulting in an increase in quantum yield upon shielding from solvent and burial in more non-polar environments. Unfolding PepQ with chemical denaturant supports this concept, as the tryptophan fluorescence decreases with increasing concentrations of denaturant (**Figure 5.5**). The most important difference between the folding and the unfolding experiments is the oligomeric state of PepQ at the beginning of measurement. The change in tryptophan fluorescence observed with denaturant could be dominated by alterations in the dimer structure. Alternatively, some tryptophans may remain quenched by nearby residues until a much later conformation, including the native dimer, is formed. It is also

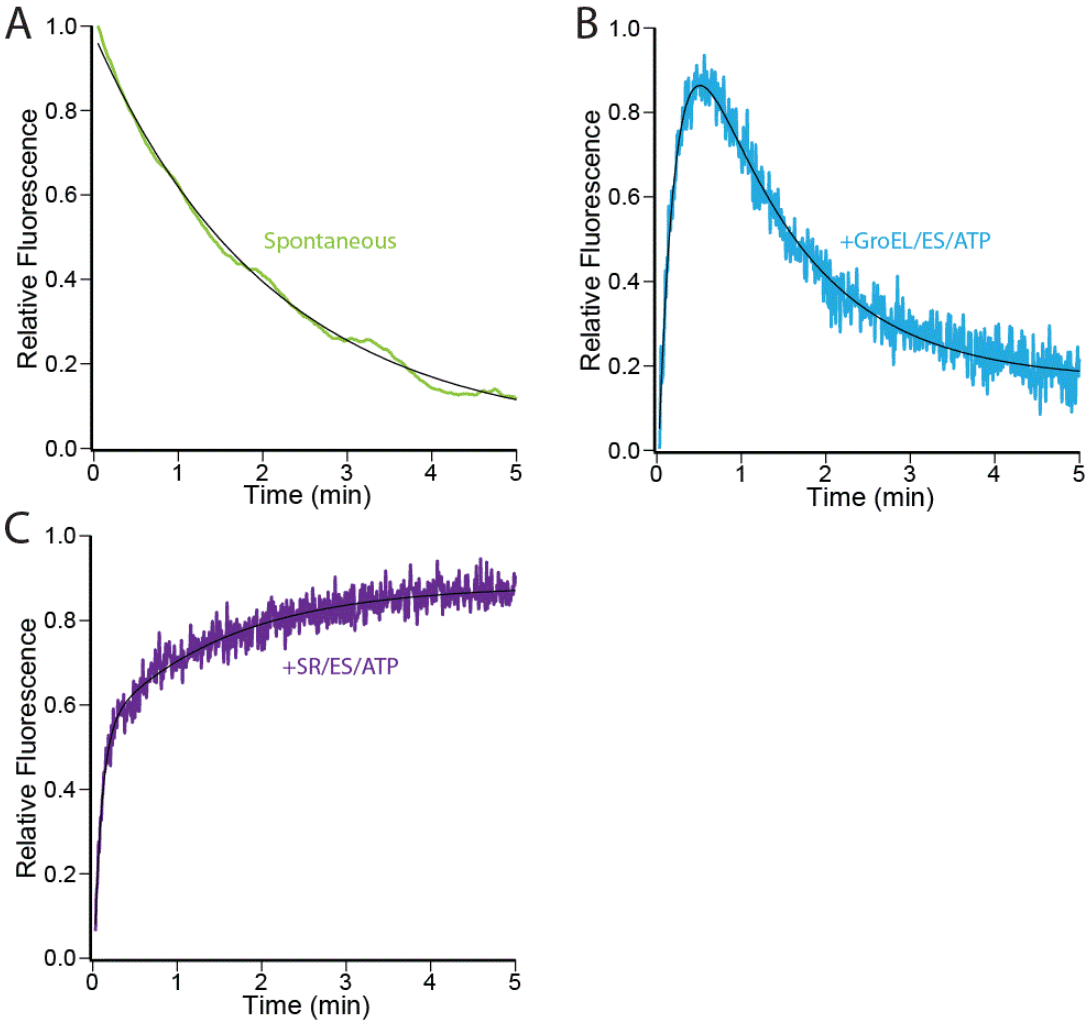


Figure 5.4: GroEL Alters the Folding Trajectory of PepQ. (A) The folding of PepQ was monitored by tryptophan fluorescence. Chemically denatured, wild-type PepQ (100 nM) was diluted directly into buffer. The fluorescence of tryptophan was observed (excitation at 295 nm, emission at 340 nm). The traces are the average of 10 replicates. (B) Chemically denatured, wild-type PepQ (100 nM) was bound to wild-type GroEL (200 nM) and refolded in the presence of GroES (400 nM) and ATP (2 mM). The fluorescence of tryptophan was observed (excitation at 295 nm, emission at 340 nm). The traces are the average of 10 replicates. (C) Chemically denatured, wild-type PepQ (100 nM) 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). The fluorescence of tryptophan was observed (excitation at 295 nm, emission at 340 nm). The traces are the average of 10 replicates.

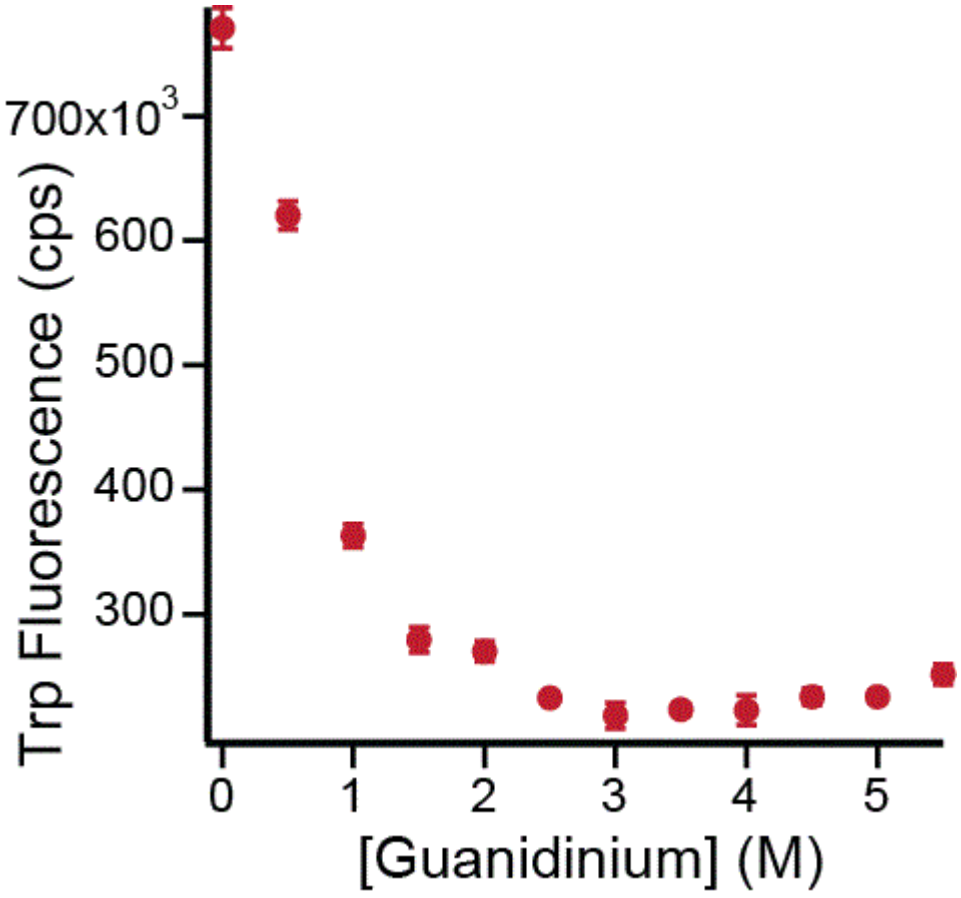


Figure 5.5: PepQ Unfolds in Response to the Chemical Denaturant Guanidinium. PepQ (50 nM) was incubated with varying concentrations of guanidinium-HCl and the tryptophan fluorescence at 340 nm each sample was determined. Error bars indicate the standard deviation from three independent samples.

possible that the changes in tryptophan fluorescence of misfolding PepQ offset observable changes occurring in correctly folding PepQ.

Unlike the fluorescence change during spontanoues PepQ folding, an increase in fluorescence is seen during the early phase of GroEL-mediated folding, which is followed by a subsequent decrease (**Figure 5.4B**). This increase in fluorescence must be reporting on some aspect of PepQ folding, as it is too slow to simply be the response to GroES binding or release of the bound substrate protein from the GroEL cavity wall [158,166,168]. Although folding in this system quickly becomes asynchronous, the transition between increasing and decreasing fluorescence occurs after roughly one cavity lifetime [157,163]. This suggests that the folding transition that is reported by the increase in fluorescence occurs inside the GroEL-ES cavity. To investigate this, the single-ring mutant was employed. Folding of PepQ inside SR1 only showed the increase in fluorescence, but not the subsequent decrease (**Figure 5.4C**), which suggests that the subsequent decrease in fluorescence is reporting on some aspect of folding that occurs after PepQ has been released from the GroEL cavity.

PepQ benefits from increased unfolding. We next sought to determine the interaction between GroEL and PepQ during folding that results in the a new folding transition reported by tryptophan fluorescence. Recently, the intrinsically disordered C-termini of GroEL were shown to alter the early folding behavior of the stringent substrate RuBisCO [424]. To investigate whether these C-terminal tails are involved in altering the folding of PepQ, C-terminal truncation mutants of wild-type GroEL and SR1 were used to examine chaperone-mediated folding. These mutants, known as Δ526 and

SRΔ526 respectively, lack the C-terminal 23 residues and have previously been shown to form stable tetradecamers, which are capable of efficiently encapsulating and folding substrate proteins [424].

Comparing the intracavity folding of PepQ inside full-length and truncated single-ring mutants shows that the largest component of the increase in tryptophan fluorescence is no longer observed in the absence of the GroEL C-termini (**Figure 5.6A**). If this transition reports on a process that is important for the stimulation of folding, then a decrease in the folding rate in the C-terminal truncation mutant would be predicted. Examining the folding with both the single-ring and tetradecameric truncation mutants shows decreased folding rates, with folding halftimes in SRΔ526 of ~2.1 minutes and in Δ526 of ~6.3 minutes (**Figure 5.6B**). Although folding with both truncated chaperones is several fold slower than in the full-length counterpart, the deficiency in Δ526 is significantly greater than in SRΔ526, indicating that the role of the C-terminal tails is more important in the cycling system. The key difference between tetradecameric GroEL and the single-ring mutant is that substrate proteins are iteratively bound and released with the wild-type chaperone, whereas the single-ring mutant only undergoes this process once.

When a substrate protein is bound by GroEL, the substrate protein undergoes a binding-driven conformational change where it adopts a more unfolded structure, and the extent of substrate protein unfolding is directly correlated to the subsequent rate of early, productive folding [166]. To examine whether the conformation of PepQ is

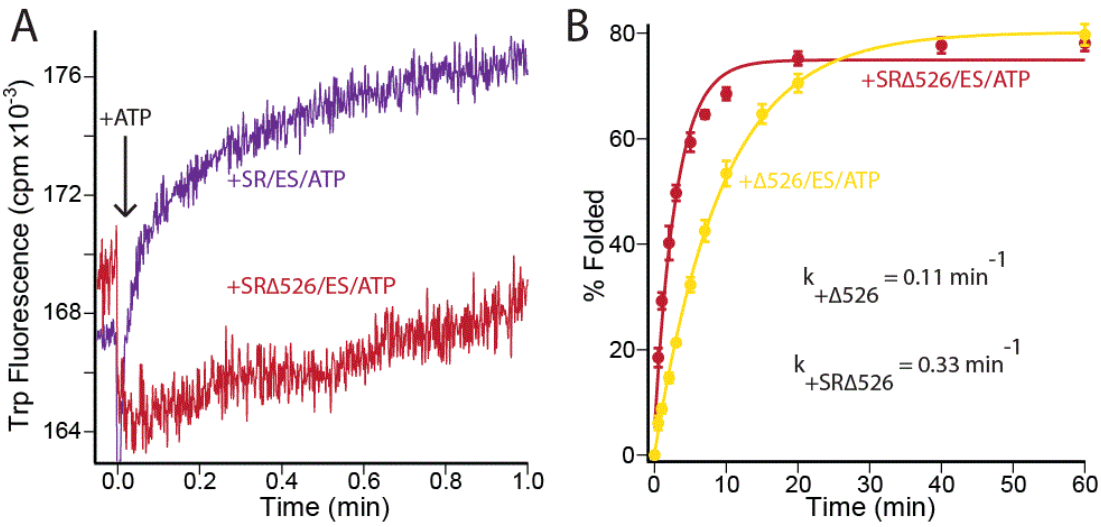


Figure 5.6: The Interactions with the GroEL C-Termini are Required for Maximal Enhancement**.** (A) Intra-cavity folding of PepQ at early times monitored by changes in tryptophan fluorescence following addition of GroES and ATP to complexes of wild-type PepQ bound to SR1 (blue) or SRΔ526 (*green*). Chemically denatured, wild-type PepQ (100 nM) was bound to SR1 (300 nM) and was rapidly mixed with an equal volume of GroES (600 nM) and ATP (2 mM) in a stopped-flow apparatus and monitored with a sampling time of 20 ms. Traces are the average of 20 replicates. (B) Chemically denatured PepQ was bound to the C-terminal truncation mutant of GroEL, Δ526 (200 nM, pink) or the single ring truncation mutant, SRΔ526 (300 nM, orange) and refolded in the presence of GroES (400 and 600 nM, respectively) and ATP (2 mM). The data were fit to a single-exponential rate law (solid lines), resulting in the observed rates of 0.106 ± 0.003 min^{-1} for Δ526-mediated folding and 0.332 ± 0.038 min^{-1} for SRΔ526-mediated folding. Error bars show the standard deviation from three independent folding experiments.

altered by the GroEL C-termini, the susceptibility of PepQ to protease degradation when bound to full-length or C-terminally-truncated GroEL-ES ADP bullets was determined. PepQ bound to the full-length GroEL ring was degraded by chymotrypsin ~2.5 fold faster than PepQ bound to the Δ526 ring (**Figure 5.7**). This result indicates that PepQ has a more unfolded conformation when the GroEL C-termini are present.

Although the substrate protein is more unfolded when bound to a wild-type GroEL ring than when bound to Δ526, a correlation between the unfolding and productive folding of PepQ has not been established. To examine whether PepQ benefits from being unfolded on GroEL, the chaperone-mediated folding rate of PepQ was examined under conditions where the cycling rate of GroEL is increased. Previous studies have established that the cycling rate of the chaperone is accelerated in the presence of substrate protein [157,176,244], so the addition of other proteins into the folding reaction could be one mechanism for increasing the cycling rate. To maintain a constant concentration of PepQ, and to reduce the additional competition for GroEL binding, native bovine serum albumin (BSA), rather than unfolded or non-native protein, was added to the reaction. Although BSA is a folded protein, its ability to bind lipids could potentially give it the necessary binding surfaces to transiently interact with the chaperone. As a measure of the cycling rate, the ATPase rate of the cycling chaperone was measured in the presence of varying concentrations of BSA, which showed that the addition of BSA does accelerate the cycling rate of GroEL by as much as 50% (**Figure 5.8a**). Under the same conditions, the folding rate of PepQ with the cycling chaperone is increased, while there is no significant effect on spontaneous folding (**Figure 5.8b**).

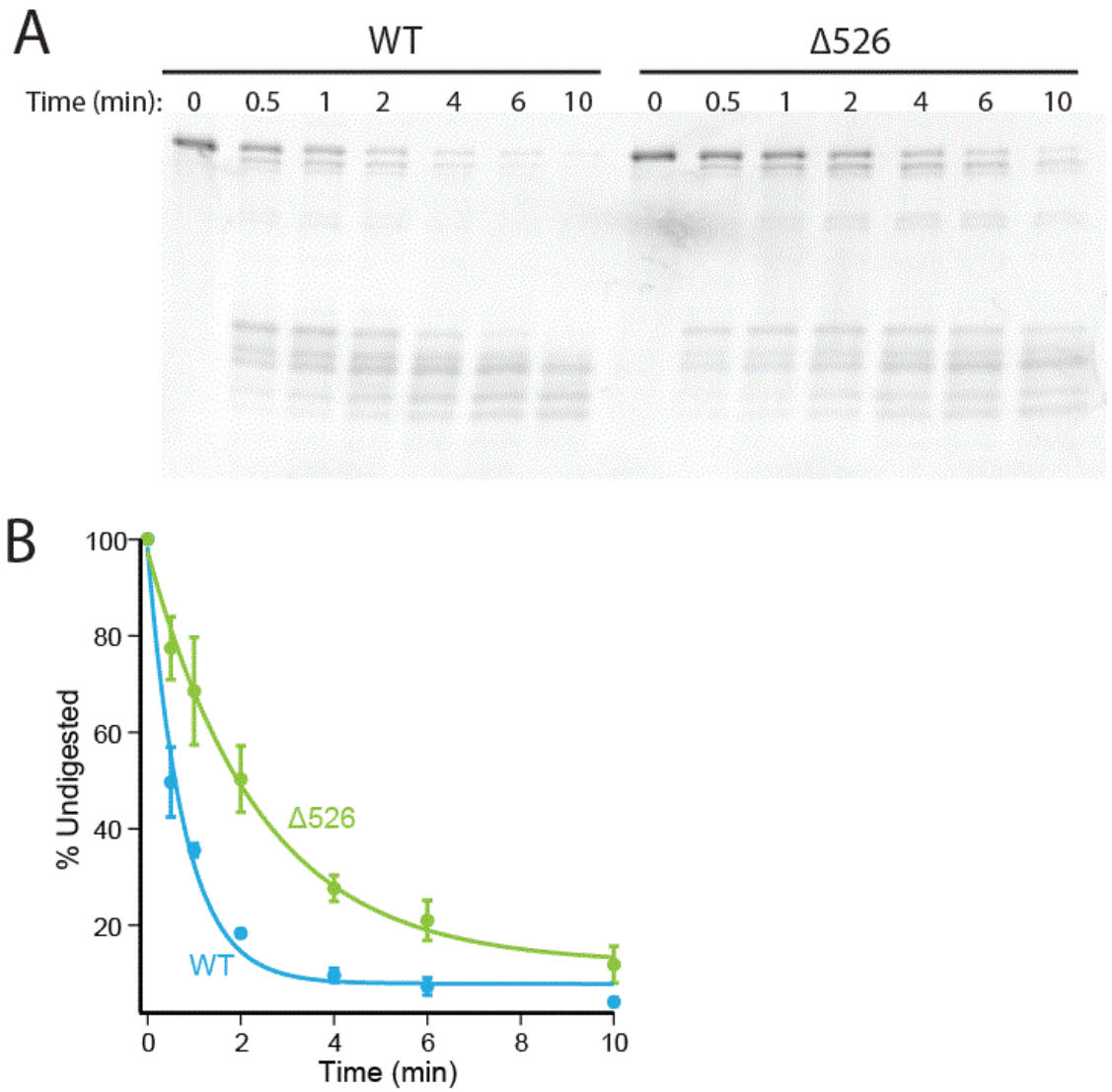


Figure 5.7: The GroEL C-Termini Modify the Conformation of Bound PepQ. (A) The extent of residual secondary structure of GroEL-bound PepQ was examined by protease susceptibility. Fluorescently labeled PepQ (24F; 100 nM) was chemically denatured and bound to the *trans* ring of either GroEL or Δ526 ADP bullets (120 nM) and then treated with chymotrypsin for the indicated times before quenching with PMSF (1mM). The samples were analyzed by SDS-PAGE and laser-excited fluorescence gel scanning. (B) The amount of full-length PepQ remaining at each time point in (A) was quantified and plotted as a function of time. The data were fit to a single-exponential rate law. The half-time for the digestion of PepQ bound to full-length ADP bullets was 0.53 $\pm$ 0.06 min (EL, *blue*) and 1.66 $\pm$ 0.17 min for C-terminal deletion ADP bullets (Δ526, *green*). Error bars represent the standard deviation of three independent experiments.

The enhanced folding seen with the wild-type chaperone over the truncation mutant, as well as the enhanced folding seen with faster cycling of the chaperone machinery – both of which are conditions with greater extent or rate of unfolding – suggests that GroEL aids PepQ folding through the disruption of kinetically-trapped intermediates.

Discussion

The results presented here strongly support an active role for GroEL in the folding of the *E. coli* protein PepQ. Under conditions where aggregation is not limiting, the rate of PepQ folding is enhanced 15-20 fold by the chaperone over spontaneous folding. Distinct changes in early folding transitions of PepQ are mediated by the chaperone, although these alterations and much of the stimulation of folding disappear in the absence of the GroEL C-termini. We show that these C-terminal tails alter the conformation of the bound substrate protein prior to encapsulation beneath GroES, leading to a more unfolded state of PepQ. The role of the unfolding action of the chaperone was examined further by modulating the rate of unfolding. Under conditions where the GroEL cycling rate is stimulated, the chaperone-mediated folding of PepQ is also stimulated, although spontaneous folding is not improved. Importantly, the massive improvement in the folding rate of PepQ, from ~20 minutes to ~1 minute, would prevent partially-folded conformations from residing in the cell for extended periods of time, which could divert other chaperones to bind this substrate protein or lead PepQ to be degraded.

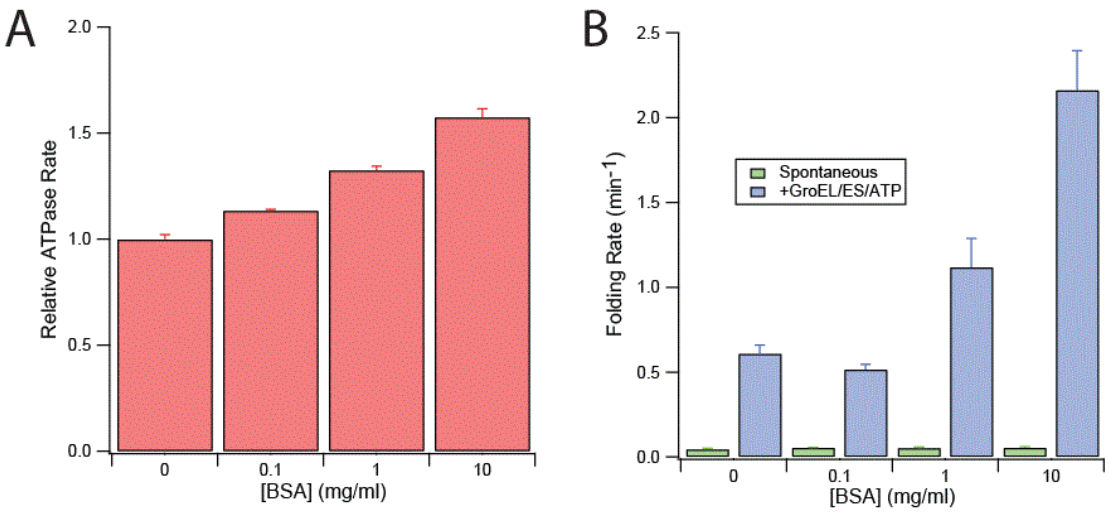


Figure 5.8: PepQ Folding Benefits from the Repetitive Unfolding Action of GroEL. (A) The GroEL ATPase rate was examined in the presence of different concentrations of BSA. The steady-state rate of ATP hydrolysis by GroEL (200 nM) in the presence of GroES (400 nM) and ATP (2 mM) was measured in the presence of varying concentrations of native BSA. Error bars show the standard deviation of three independent experiments. (B) The rate of spontaneous and GroEL-mediated PepQ folding was examined in the presence of different concentrations of BSA. As in Figure 5.1A, except with native BSA in the buffer. Error bars show the standard deviation of three independent experiments.

In the context of the common models for GroEL function (passive protection from aggregation, confinement inside the cavity, and iterative annealing), these observations clearly demonstrate an active role for GroEL in the folding of PepQ. An exclusive role for the chaperone as an Anfinsen cage can be eliminated, as PepQ does not depend upon the chaperone to protect folding monomers from aggregation (**Figure 5.2&5.3**). The stimulation of PepQ monomer folding at very low concentrations by GroEL indicates that the chaperone enhances folding by changing how the protein folds, suggesting a role for confinement or iterative annealing. The GroEL cavity is essential for the stimulation of folding, and inside this cavity changes occur in the folding transitions of PepQ (**Figure 5.4**). Interactions between GroEL and this substrate protein significantly enhance its folding, as seen through the differences between full-length and truncated chaperone (**Figure 5.6**). However, confinement inside the GroEL-ES cavity alone is not sufficient for the maximal enhancement of folding, as the stimulation of the folding rate is further improved by accelerating the rate of cavity turnover (**Figure 5.8**). This observation suggests that another mechanism contributes to PepQ folding and is consistent with the iterative annealing model of GroEL function, which predicts that any enhancement of the magnitude or rate of unfolding should also enhance the folding rate of the substrate protein [188]. Under this model, with each new cycle of the chaperone machinery, the residual structure of the bound substrate protein is disrupted, providing a new opportunity for the substrate protein to fold from a higher energy state.

A role for substrate protein unfolding is central to the interactions between the GroEL C-terminal tails and the bound substrate protein [424]. The Δ526 mutation

removes 161 amino acids and ~14.4 kDa of mass from the bottom of the GroEL cavity, which could greatly impact the environment of a protein folding inside the chaperonin cavity. Despite the influence of this alteration of cavity size and character, the changes in folding induced by this mutation fit into the paradigm of iterative annealing more than cavity confinement. Prior to encapsulation, the GroEL C-termini enhance the unfolding of the chaperone-bound PepQ (**Figure 5.7**); this observation predicts that the folding of PepQ would be slowed in the absence of these C-termini. Comparison of intracavity folding between full-length SR1 and truncated SRΔ526 shows that folding preceded by more extensive unfolding is roughly twice as fast in the case of PepQ. Adding the effects of cycling in the tetradecameric protein yields different predictions for the cavity confinement and iterative annealing models. In the confinement model, there should be no difference in the folding rates between cycling and non-cycling systems, as the mutation affects both cavities equally. However, a further decrease in the folding rate is observed in the truncated chaperone. The exacerbation of the difference between folding in cycling, full-length and truncated chaperonin further supports a role for the iterative annealing model, as the rate of unfolding is also decreased in the more slowly cycling Δ526 machine [424]. While the perturbation of the unfolding action of the chaperone machinery, either by mutation or by exogenously altering the cycling rate, has a direct influence on the folding of substrate proteins, the importance of this mechanism does not exclude all possible roles for substrate protein confinement.

Although only changes in the folding of PepQ are reported here, other studies with the substrate protein RuBisCO from *R. rubrum* have shown the impact of

chaperone-mediated unfolding and of the interactions with the GroEL C-termini on that protein [158,159,166,168,424]. While PepQ and RuBisCO are only two proteins among thousands that utilize the chaperonin machinery, common themes are observed in the mechanisms by which their folding is enhanced, despite the critical differences between the structures and the folding propensities of the two proteins. While it is not possible to compare the GroEL-mediated folding and the spontaneous folding of RuBisCO, both proteins undergo processes that are enhanced by the GroEL C-termini, including an unfolding action before encapsulation and changes in the earliest observed transitions in folding. Although neither the GroEL C-termini nor unfolding were analyzed therein, a recent study of the folding of the *E. coli* protein DapA showed clear changes in the early folding of this protein with GroEL relative to its spontaneous folding [88]. Unlike RuBisCO and DapA, PepQ does not have a TIM-barrel fold; rather, the catalytic domain is formed by a pita-bread fold [384,423]. The changes in substrate protein folding seen in these three proteins highlight a potentially universal mechanism for protein folding inside the GroEL cavity that is independent of the eventual protein structure or fold, where the removal of the kinetics barriers that may slow or block productive folding through forced unfolding is coupled with inhibition of off-pathway conformations through confinement inside the cavity.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The purpose of this study was to examine the role of the intrinsically-disordered C-terminal residues of the *E. coli* chaperonin GroEL in the various aspects of the folding reaction cycle. The GroEL-ES chaperonin system, which is highly conserved among bacteria, mitochondria and chloroplasts, aids substrate protein folding though several different mechanisms: protecting folding intermediates from intermolecular aggregation, restricting the allowable volume to limit or promote folding intermediates, and unfolding unproductive folding intermediates to restart the folding process. The first mechanism is described as the passive model, where GroEL serves as an Anfinsen Cage to allow intramolecular folding to occur without intermolecular interactions. Conversely, the second and third mechanisms dictate an active role for GroEL, where specific conformational changes in the chaperone or interactions between the chaperone and the substrate protein alter the configuration of available folding intermediates. While no single proposed mechanism for the function of GroEL is sufficient for the folding of every protein in the proteome, the necessity of each mechanism is unknown, and determining the contribution of each mechanism to the folding of any one substrate protein requires extensive *in vitro* experimentation. In order to better understand roles of both the passive and active functions of the chaperonin system, the folding of two different substrate proteins was examined in the context of these mechanisms, with a focus on the influence of the C-terminal tail of GroEL on each mechanism.

The substrate protein RuBisCO was previously shown to have an extremely high propensity for aggregation and to require the chaperonin system *in vivo*, both in plant chloroplasts and in the cytoplasm of photosynthetic bacteria [87,91,252]. The RuBisCO used in my studies, which came from the bacterium *Rhodospirillum rubrum*, does not fold independently *in vitro*, even under conditions where it does not aggregate [148,163,168]. GroEL-mediated folding proceeds under the same conditions with a halftime of roughly 3.5 minutes [163]. While RuBisCO does fold in the single-ring mutant of GroEL, SR1, at the same rate as in wild-type GroEL at room temperature, folding in the cycling system is accelerated at higher temperatures, while folding in non-cycling SR1 mutant is unchanged [159]. Because the intracavity interactions would change equally in the cycling and non-cycling systems, the increased folding rate must be correlated with the faster turnover of the cycling system at higher temperatures.

Faster turnover of the GroEL-ES complex shortens the duration of intracavity folding for substrate proteins and increases the rate at which substrate proteins will reinitiate the folding process. Therefore, the most dramatic difference between the cycling and non-cycling systems is the frequency of chaperone-mediated unfolding of the substrate protein. This implies that the iterative unfolding action of GroEL, rather than confinement inside the GroEL-ES cavity, may be the critical mechanism for the folding of RuBisCO, as the folding rate and cavity lifetime are inversely proportional. Simply stated, spending more time inside the GroEL cavity does not always directly translate into more folded protein. This observation supports the notion that the

unfolding mechanism of GroEL is critical, while the other models predict exactly the opposite result.

In the examination of the role of the GroEL C-terminal tail, the same inverse correlation was observed: increasing the cavity lifetime of GroEL corresponded to slower folding of RuBisCO. However, this mutation alters variables other than the cavity lifetime. Removal of the C-terminal tail also abolishes interactions between those GroEL residues and the substrate protein. These interactions are likely important for the function of the chaperonin, as the length and character of the C-terminus, which is comprised of 23 amino acids, have been maintained throughout evolution. The mildly hydrophobic and intrinsically disordered nature of this C-terminus, combined with the visualization of the interaction with substrate proteins, led to the hypothesis that these residues in the chaperone represent a binding region outside of the GroEL apical domains [424]. This binding site provides for additional interactions with substrate proteins outside of the GroEL apical domains [347,424]. One important component of this interaction is that it leads to the migration of the substrate protein deeper into the cavity, which increases the time required to diffuse away from the chaperone, thus improving the probability of encapsulating the substrate protein inside the GroEL-ES cavity. The same interaction that leads to this migration is also critical in unfolding the substrate protein during the binding process. This unfolding of GroEL-bound substrate proteins is correlated to productive early folding. Additionally, changes in the subsequent, slower folding processes are observed, indicating that the starting state of a protein dictates how a protein can fold. The concept that not all intermediate

conformations are available in every possible folding pathway of a protein has been suggested by computational models [431-433].

The effects of the GroEL C-terminal tail observed in the folding of RuBisCO are also seen in the folding of the substrate protein PepQ, which is a metalloprotease from *E. coli*. PepQ was chosen for examination as a new model substrate protein, because all other model substrate proteins at the time of choice were derived from other organisms or do not utilize the chaperone *in vivo*. Additionally, PepQ was predicted to require GroEL *in vivo* by two independent studies [89,90]. As no studies had been published on the folding of PepQ or a PepQ homolog, and the only research published about this protein focused on its enzymology [365], PepQ was a tabula rasa and an ideal target for introduction as a new substrate protein. Folding experiments bore out that PepQ does not require the chaperonin *in vitro*, though its folding rate is massively improved by GroEL. Importantly, unlike RuBisCO, PepQ shows no propensity for aggregation under the conditions used, suggesting that the enhancement of its folding is due to another mechanism of GroEL. Like RuBisCO, the binding-driven conformational changes in PepQ prior to folding are critical in optimizing the ability of the protein to fold, and the GroEL C-terminus has an important role in inducing these conformational changes. As likewise observed with RuBisCO, an early folding process occurs in GroEL-mediated folding but is not observed in spontaneous folding or folding mediated by the GroEL C-terminal truncation mutant.

For both RuBisCO and PepQ, three distinct effects arise from interactions with the GroEL C-terminal tail: the substrate protein adopts a more unfolded conformation

prior to folding, the processes by which the protein folds are altered, and folding is accelerated. These changes may appear disconnected, although it is very likely that the changes in starting conformation alter the availability of intermediates during the folding. Through this kinetic alteration of the folding landscape, the chaperone-mediated changes provide a more efficient pathway for productive folding. Although many of the predictions of the cavity confinement model are contested by experimental observations, it is impossible to rule out any role for confinement. For example, the explanation of the decrease in the folding rate of both RuBisCO and PepQ in the single-ring truncation mutant cannot exclude a combined effect of both unfolding and confinement. In fact, confinement inside the GroEL-ES cavity is essential for the chaperone-mediated folding of many proteins, but due to its role in folding hundreds of different proteins *in vivo* [89,90], GroEL could not have evolved to make specific, beneficial interactions inside the cavity with any of these proteins without potentially reducing the enhancement in folding other proteins [224]. With the broad specificity of the substrate binding region on the apical domain [157,165], this chaperone retains the ability to bind and subsequently unfold virtually all kinetically-trapped or partially-folded intermediates. Iterative annealing and unfolding, however, is also insufficient as an exclusive mechanism for GroE, as folding without GroES or folding in the presence of other chaperones (such as Hsp70s) does not lead to the same extent of stimulation of folding. Although such experiments have not carefully tested the variables of chaperone concentration and cycling rate on the protein folding rate, current evidence supports a requirement for coupled unfolding and confinement. Significantly, while this research

and other studies highlight the critical role for the active mechanisms of GroEL, many GroEL substrate proteins are highly aggregation-prone, and the cell is a crowded environment.

In the context of the cellular environment, proteins benefit from a number of different chaperones, although GroEL is unique in isolating substrate proteins in a closed cavity. Examining chaperonin function in the context of the passive model, the GroEL-ES cavity provides a setting for proteins to fold without the complication of intermolecular aggregation. Aggregation into non-native conformations may not be the only intermolecular interaction that is successfully avoided. Other chaperone systems, especially DnaK/DnaJ/GrpE and ClpB in *E. coli*, are critical in disassembling aggregates and unfolding proteins [120,124,197,434]; the latter is a mechanism shared by GroEL. In the case of DnaK, multiple chaperones could be bound to the same polypeptide simultaneously, which could inhibit folding both locally and globally. These chaperones are likely released only upon the threading of the substrate protein through the chaperone ClpB [124,434-436]. For slowly folding proteins, GroEL not only can unfold misfolded intermediates, thus resetting the energy landscape, but may also serve to block other chaperones from interacting with the folding substrate protein, which could potentially be recognized by exposed hydrophobic regions. In this scenario, GroEL is utilizing a well-accepted mechanism for other chaperones – unfolding – and is also coupling the existence of this newly unfolded conformation of the substrate protein to the creation of an environment in which it can fold without detrimental intermolecular interactions.

Here, a passive function for the chaperone is only successful after the utilization of an active mechanism.

This study on the folding of the substrate proteins RuBisCO and PepQ has provided observations in support of the active roles of GroEL. The unfolding action of the chaperone clearly benefits the subsequent intracavity folding of the substrate protein. The GroEL C-terminal tail, although important in many roles, is critical for modulating the unfolding action of the chaperonin, in terms of both magnitude and rate. Despite the observations herein that argue against GroEL simply serving as a passive Anfinsen-cage and against substrate protein confinement as the primary mechanistic role for the chaperone, evidence supports both of these models as well [189,191,198,437]. It is important to acknowledge that all three mechanisms contribute to the folding of the proteome, which contains hundreds of proteins. These proteins experience many different folding conditions over the lifetime of the organism and require different forms of assistance at different times. GroEL serves as an equal-opportunity chaperone.

The Future

There is much remaining to be studied concerning the role of the C-terminal tail of GroEL in stimulating the folding of substrate proteins. While *E. coli* has been the model organism for Hsp60-mediated folding, many organisms contain variants of GroEL with specific alterations to the C-terminus that are important for its function. Increasingly, organisms that contain multiple copies of the chaperonin are being scrutinized for the roles of each variant of GroEL. For example, alleles of GroEL

homologs containing significantly altered C-terminal sequences have been found in cyanobacteria and actinobacteria and shown as essential for viability, biofilm formation and pathogenesis [213,226,276]. Alternatively, in the mitochondria of *S. cerevisiae*, the C-terminus of the GroEL homolog contains several interspersed prolines, which are absent in the common model organisms elsewhere across the hierarchy of life [255-260,438-440]. Examining the folding of substrate proteins by these GroEL homologs or with *E. coli* GroEL containing an altered C-terminal sequence may not only provide insight as to the specific roles of these alterations on the folding of critical proteins in their host organisms, but may also reveal greater detail about the sophisticated allostery of the GroEL reaction cycle or the mechanisms by which GroEL folds proteins.

The ability to examine PepQ folding has been limited by an inability to label the protein with fluorescent dyes without greatly perturbing its folding, as well as by a cryptic site in PepQ which is unproductively labeled by maleimide-containing dyes. Attempts to identify this site by tryptic digest, peptide purification, and mass spectrometry have failed, presumably due to an incompatibility of the labeled peptide with the MALDI-TOF mass spectrometry method. Utilization of a different maleimide-containing dye may yield peptide fragments that can be analyzed by this method. Mutating this cryptic site does not solve the challenge that many labeled sites greatly perturb folding. Although many sites have been analyzed, more will need to be tested to identify PepQ mutants that can be labeled without this complication. These mutants will allow for the use of intramolecular FRET, which can be used not only to examine the

conformation of PepQ bound to GroEL, but also to monitor the folding of the two domains of PepQ.

This capability may offer insight into the structural constraints that make PepQ GroEL-dependent when the two other pita-bread fold-containing proteins in *E. coli*, PepP and MetAP, do not require the chaperone. These three proteins share very high structural and sequence homology in the catalytic domain, though they vary substantially in the N-terminal domain. MetAP has an N-terminal domain containing ~30 amino acids, while the N-terminal domain of PepQ is comprised of ~160 amino acids [255,384]. Individually expressing and purifying these domains may provide insight into the GroEL-dependence of PepQ folding. When coupled with intramolecular FRET, the study of these domains could reveal specific details about the intramolecular interactions that dominate PepQ folding, both spontaneously and when mediated by GroEL. Attempts to create these constructs have only yielded insoluble or degraded proteins, even in the presence of excess chaperone. Further testing of expression conditions and alterations of the constructs may allow for the folding of these domains to be analyzed.

Many other research routes have already been sampled, and some of these research areas show some promise. While the truncation of the GroEL C-terminus leads to changes in multiple aspects of chaperone activity, examining the impact of cavity size on the folding of substrate proteins through the utilization of GroES homologs that either increase or decrease the cavity size may prove to be a less complicated alternative for studying the role of substrate protein confinement. Preliminary work done with gp31,

the GroES homolog from bacteriophage T4 that is known to increase the cavity size relative to *E. coli* GroES, showed major alterations in intracavity folding of RuBisCO by FRET, though the overall folding rate appeared unchanged. And despite its ability to fold RuBisCO, attempts at studying the GroES homolog from *Sulfurospirillum deleyianum*, which by bioinformatics was predicted to produce a smaller cavity than *E. coli* GroES when bound to GroEL, was complicated by instability of the heptameric structure. Revisiting these GroES homologs may prove fruitful, as both showed behavior that was dissimilar to GroES. Additionally, the use of PepQ as a substrate protein does not suffer the same challenges of using RuBisCO, which would make some aspects of studying these GroES homologs less difficult.

While the study of individual substrate proteins can provide key biochemical and biophysical data to support the various models of chaperone function, the chaperone interacts with hundreds of proteins *in vivo*, and the effect of GroEL on any one protein cannot be immediately translated to effects on every protein. To understand the various aspects of GroEL across all of its natural substrate proteins, examination of the GroEL folding proteome would be ideal. Alterations to the system, such as introducing mutations that have known effects on substrate protein binding, the ATPase activity of the chaperone, or the allostery of the machinery, could provide insight into how each of these variables impacts the folding of each of the individual proteins that rely on GroEL. The initial proteomics study, which examined the folding proteome with and without the GroEL C-terminus, showed promise through the early stages and could potentially identify proteins that require either the unfolding action or the interactions of the C-

terminal tail for correct folding. A similar proteomics study could also be done with the GroES homologs to identify proteins that have specific confinement requirements for folding. While such experiments are inherently prone to failure, the potential reward of identifying such proteins would cement the various active mechanisms into the dogma of GroEL-mediated folding.

While GroEL has been the object of scientific study for several decades, insights into the functioning of the chaperone still provide clues as to the problems that arise during protein folding and serve as valuable tools in the study of other proteins. As a tool, GroEL is most commonly used to aid in the expression of proteins, but this chaperone is capable of more sophisticated use [441]. For example, recent studies utilized GroEL in determining the structure and mechanism of the active anthrax toxin [442]. The chaperonin serves as a great implement in other fields, but studying how the cell manages and repairs the off-pathway products of protein folding could lead to a new prospective for combating protein folding diseases, such as cystic fibrosis, amyotrophic lateral sclerosis, and Huntington's, Alzheimer's and Parkinson's diseases. The leap from studying chaperone-mediated protein folding and addressing these ailments may seem titanic, but no one knows what the next great breakthrough will be until it happens.

REFERENCES

1. Hartley H (1951) Origin of the Word 'Protein'. Nature 168: 244-244.

2. Mulder GJ (1838) Sur la composition des quelques substance animales. Bulletin des Sciences Physiques et Naturelles en Néerlande: 104.

3. Mulder GJ (1839) Ueber die Zusammensetzung einiger thierischen Substanzen. Journal fuer Praktische Chemie 16: 129-152.

4. Liebig J (1842) Animal Chemistry or Organic chemistry in its applications to physiology and pathology. London: Taylor and Walton.

5. McCoy RH, Meyer CE, Rose WC (1935) Feeding experiments with mixtures of highly purified amino acids - VIII. Isolation and identification of a new essential amino acid. Journal of Biological Chemistry 112: 283-302.

6. Vickery HB, Schmidt CLA (1931) The history of the discovery of the amino acids. Chemical Reviews 9: 169-318.

7. Fischer E (1902) Über die Hydrolyse der Proteinstoffe. Chemiker-Zeitung 26: 939-940.

8. Hofmeister F (1902) Über Bau und Gruppierung der Eiweisskörper. Ergebnisse der Physiologie 1: 759-802.

9. Fruton JS (1979) Early theories of protein structure. Ann N Y Acad Sci 325: xiv, 1-18.

10. Abderhalden E, Brockmann H (1930) Stufenweisen Abbau von Polypeptidderderivaten. Biochem Z 225: 386-408.

11. Block RJ (1945) The Amino Acid Composition of Food Proteins. In: Anson ML, John TE, editors. Advances in Protein Chemistry: Academic Press. pp. 119-134.

12. Fox SW (1945) Terminal Amino Acids in Peptides and Proteins. In: Anson ML, John TE, editors. Advances in Protein Chemistry: Academic Press. pp. 155-177.

13. Martin AJP, Synge RLM (1945) Analytical Chemisty of the Proteins. In: Anson ML, John TE, editors. Advances in Protein Chemistry: Academic Press. pp. 1-83.

14. Sanger F (1945) The free amino groups of insulin. Biochem J 39: 507-515.

15. Snell EE (1945) The Microbiological Assay of Amino Acids. In: Anson ML, John TE, editors. Advances in Protein Chemistry: Academic Press. pp. 85-118.

16. Gornall AG, Bardawill CJ, David MM (1949) Determination of serum proteins by means of the biuret reaction. J Biol Chem 177: 751-766.

17. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193: 265-275.

18. Sanger F (1952) The Arrangement of Amino Acids in Proteins. Advances in Protein Chemistry. pp. 1-67.

19. Edman P (1956) Mechanism of the Phenyl Isothiocyanate Degradation of Peptides. Nature 177: 667-668.

20. Sumner JB, Hand DB (1928) Krystallisierte Urease. Die Naturwissenschaften 16: 145-146.

21. Northrop JH (1929) Crystalline Pepsin. Science 69: 580.

22. Northrop JH (1930) Crystalline Pepsin. J Gen Physiol 13.

23. Svedberg T (1934) Molecular weight analysis in centrifugal fields. Science 79: 327-332.

24. Crowfoot D (1941) A review of some recent x-ray work on protein crystals. Chemical Reviews 28: 215-228.

25. Mirsky AE, Pauling L (1936) On the Structure of Native, Denatured, and Coagulated Proteins. Proc Natl Acad Sci U S A 22: 439-447.

26. Sanger F, Tuppy H (1951) The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. Biochem J 49: 463-481.

27. Sanger F, Tuppy H (1951) The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochem J 49: 481-490.

28. Anson ML (1945) Protein Denaturation and the Properties of Protein Groups. Advances in Protein Chemistry. pp. 361-386.

29. Fankuchen I (1945) X-Ray Diffraction and Protein Structure. In: Anson ML, John TE, editors. Advances in Protein Chemistry: Academic Press. pp. 387-405.

30. Mirsky AE, Anson ML (1930) Protein Coagulation and Its Reversal : Improved Methods for the Reversal of the Coagulation of Hemoglobin. J Gen Physiol 13: 477-481.

31. Anson ML, Mirsky AE (1931) The reversibility of protein coagulation. Journal of Physical Chemistry 35: 185-193.

32. Anson ML, Mirsky AE (1942) Communication to the editor - Reversibility of heat denaturation of protein. Journal of Physical Chemistry 46: 334-335.

33. Neurath H, Cooper GR, Erickson JO (1942) The denaturation of proteins and its apparent reversa1. Journal of Physical Chemistry 46: 203-211.

34. Neurath H, Greenstein JP, Putnam FW, Erickson JO (1944) The chemistry of protein denaturation. Chemical Reviews 34: 157-265.

35. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, et al. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181: 662-666.

36. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, et al. (1960) Structure of myoglobin: A three-dimensional fourier synthesis at 2A resolution. Nature 185: 422-427.

37. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, et al. (1960) Structure of Hæmoglobin: A three-dimensional fourier synthesis at 5.5A resolution, obtained by X-ray analysis. Nature 185: 416-422.

38. Dintzis HM (1961) Assembly of the peptide chains of hemoglobin. Proc Natl Acad Sci U S A 47: 247-261.

39. Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, et al. (1965) Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. Nature 206: 757-761.

40. Corey RB, Donohue J (1950) Interatomic Distances and Bond Angles in the Polypeptide Chain of Proteins. Journal of the American Chemical Society 72: 2899-2900.

41. Carpenter GB, Donohue J (1950) The Crystal Structure of N-Acetylglycine. Journal of the American Chemical Society 72: 2315-2328.

42. Pauling L, Corey RB (1951) The structure of synthetic polypeptides. Proc Natl Acad Sci U S A 37: 241-250.

43. Pauling L, Corey RB (1951) Atomic coordinates and structure factors for two helical configurations of polypeptide chains. Proc Natl Acad Sci U S A 37: 235-240.

44. Pauling L, Corey RB, Branson HR (1951) 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 37: 205-211.

45. Marsh RE, Corey RB, Pauling L (1955) An investigation of the structure of silk fibroin. Biochim Biophys Acta 16: 1-34.

46. Scheraga HA, Nemethy G, Steinberg IZ (1962) The contribution of hydrophobic bonds to the thermal stability of protein conformations. J Biol Chem 237: 2506-2508.

47. Scheraga HA (1961) Effect of Hydrophobic Bonding on Protein Reactions1. The Journal of Physical Chemistry 65: 1071-1072.

48. Klotz IM (1958) Protein hydration and behavior; many aspects of protein behavior can be interpreted in terms of frozen water of hydration. Science 128: 815-822.

49. Lumry R, Eyring H (1954) Conformation Changes of Proteins. Journal of Physical Chemistry 58: 110-120.

50. Moffitt W, Yang JT (1956) The Optical Rotatory Dispersion of Simple Polypeptides. I. Proceedings of the National Academy of Sciences 42: 596-603.

51. Doty P, Yang JT (1956) Polypeptides. Vii. Poly-γ-Benzyl-L-Glutamate: The Helix-Coil Transition in Solution1. Journal of the American Chemical Society 78: 498-500.

52. Yang JT, Doty P (1957) The Optical Rotatory Dispersion of Polypeptides and Proteins in Relation to Configuration1. Journal of the American Chemical Society 79: 761-775.

53. Cohen C (1955) Optical Rotation and Polypeptide Chain Configuration in Proteins. Nature 175: 129-130.

54. Sela M, Anfinsen CB, Harrington WF (1957) The correlation of ribonuclease activity with specific aspects of tertiary structure. Biochim Biophys Acta 26: 502-512.

55. Anfinsen CB, Haber E (1961) Studies on the reduction and re-formation of protein disulfide bonds. J Biol Chem 236: 1361-1363.

56. Anfinsen CB, Haber E, Sela M, White FH, Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47: 1309-1314.

57. Haber E, Anfinsen CB (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. J Biol Chem 236: 422-424.

58. Haber E, Anfinsen CB (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. J Biol Chem 237: 1839-1844.

59. Goldberger RF, Epstein CJ, Anfinsen CB (1963) Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. J Biol Chem 238: 628-635.

60. Potts JT, Jr., Young DM, Anfinsen CB (1963) Reconstitution of fully active RNase S by carboxypeptidase-degraded RNase S-peptide. J Biol Chem 238: 2593-2594.

61. Givol D, Goldberger RF, Anfinsen CB (1964) Oxidation and Disulfide Interchange in the Reactivation of Reduced Ribonuclease. J Biol Chem 239: PC3114-3116.

62. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223-230.

63. Levinthal C (1969) How to fold graciously. Mossbauer Spectroscopy in Biological Systems: 22-24.

64. Karplus M, Weaver DL (1976) Protein-folding dynamics. Nature 260: 404-406.

65. Kim PS, Baldwin RL (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51: 459-489.

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

67. Baldwin RL (1975) Intermediates in protein folding reactions and the mechanism of protein folding. Annu Rev Biochem 44: 453-475.

68. Epstein HF, Schechter AN, Chen RF, Anfinsen CB (1971) Folding of staphylococcal nuclease: Kinetic studies of two processes in acid renaturation. Journal of Molecular Biology 60: 499-508.

69. Epstein HF, Schechter AN, Cohen JS (1971) Folding of staphylococcal nuclease: magnetic resonance and fluorescence studies of individual residues. Proc Natl Acad Sci U S A 68: 2042-2046.

70. Kuwajima K (1977) A folding model of α-lactalbumin deduced from the three-state denaturation mechanism. Journal of Molecular Biology 114: 241-258.

71. Tsong TY, Karr T, Harrington WF (1979) Rapid helix--coil transitions in the S-2 region of myosin. Proc Natl Acad Sci U S A 76: 1109-1113.

72. Yutani K, Ogasahara K, Sugino Y (1980) pH dependence of stability of the wild-type tryptophan synthase α-subunit and two mutant proteins (Glu49 → Met or Gln). Journal of Molecular Biology 144: 455-465.

73. Zetina CR, Goldberg ME (1982) Kinetics of renaturation and self-assembly of intermediates on the pathway of folding of the β2-subunit of Escherichia coli tryptophan-synthetase. Journal of Molecular Biology 157: 133-148.

74. Schonbrunner N, Wey J, Engels J, Georg H, Kiefhaber T (1996) Native-like beta-structure in a trifluoroethanol-induced partially folded state of the all-beta-sheet protein tendamistat. J Mol Biol 260: 432-445.

75. Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem. pp. 83-229.

76. Jennings PA, Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science 262: 892-896.

77. Ballew RM, Sabelko J, Gruebele M (1996) Direct observation of fast protein folding: the initial collapse of apomyoglobin. Proceedings of the National Academy of Sciences 93: 5759-5764.

78. Matthews CR (1993) Pathways of protein folding. Annu Rev Biochem 62: 653-683.

79. Sali A, Shakhnovich E, Karplus M (1994) How does a protein fold? Nature 369: 248-251.

80. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 48: 545-600.

81. Cheung MS, Garcia AE, Onuchic JN (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc Natl Acad Sci U S A 99: 685-690.

82. Dobson CM (2003) Protein folding and misfolding. Nature 426: 884-890.

83. Bicout DJ, Szabo A (2000) Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. Protein Sci 9: 452-465.

84. Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. Science 338: 1042-1046.

85. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4: 10-19.

86. Schechter AN, Chen RF, Anfinsen CB (1970) Kinetics of folding of staphylococcal nuclease. Science 167: 886-887.

87. Goloubinoff P, Christeller JT, Gatenby AA, Lorimer GH (1989) Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. Nature 342: 884-889.

88. Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, et al. (2014) GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. Cell 157: 922-934.

89. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122: 209-220.

90. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO Journal 29: 1552-1564.

91. Barraclough R, Ellis RJ (1980) Protein synthesis in chloroplasts IX. Assembly of newly-synthesized large subunits into ribulose bishopshate carboxylase in isolated intact pea chloroplasts. BBA Section Nucleic Acids And Protein Synthesis 608: 19-31.

92. Jensen TJ, Loo MA, Pind S, Williams DB, Goldberg AL, et al. (1995) Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. Cell 83: 129-135.

93. Ward CL, Omura S, Kopito RR (1995) Degradation of CFTR by the ubiquitin-proteasome pathway. Cell 83: 121-127.

94. Farinha CM, Amaral MD (2005) Most F508del-CFTR is targeted to degradation at an early folding checkpoint and independently of calnexin. Mol Cell Biol 25: 5242-5252.

95. Welsh MJ, Ostedgaard LS (1998) Cystic fibrosis problem probed by proteolysis. Nature Structural Biology 5: 167-169.

96. Lukacs GL, Verkman AS (2012) CFTR: Folding, misfolding and correcting the ΔF508 conformational defect. Trends in Molecular Medicine 18: 81-91.

97. O'Sullivan BP, Freedman SD (2009) Cystic fibrosis. The Lancet 373: 1891-1904.

98. Davies JC, Ebdon AM, Orchard C (2014) Recent advances in the management of cystic fibrosis. Arch Dis Child 99: 1033-1036.

99. Jethva R, Bennett MJ, Vockley J (2008) Short-chain acyl-coenzyme A dehydrogenase deficiency. Mol Genet Metab 95: 195-200.

100. Waters PJ, Parniak MA, Akerman BR, Jones AO, Scriver CR (1999) Missense mutations in the phenylalanine hydroxylase gene (PAH) can cause accelerated proteolytic turnover of PAH enzyme: A mechanism underlying phenylketonuria. Journal of Inherited Metabolic Disease 22: 208-212.

101. Thomas PJ, Qu BH, Pedersen PL (1995) Defective protein folding as a basis of human disease. Trends in Biochemical Sciences 20: 456-459.

102. Georgiou G, Valax P, Ostermeier M, Horowitz PM (1994) Folding and aggregation of TEM beta-lactamase: analogies with the formation of inclusion bodies in Escherichia coli. Protein Sci 3: 1953-1960.

103. Chi EY, Krishnan S, Randolph TW, Carpenter JF (2003) Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. Pharmaceutical Research 20: 1325-1336.

104. Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, et al. (2001) Effect of Environmental Factors on the Kinetics of Insulin Fibril Formation: Elucidation of the Molecular Mechanism†. Biochemistry 40: 6036-6046.

105. Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Folding and Design 3: R9-R23.

106. Guo Z, Eisenberg D (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. Proc Natl Acad Sci U S A 103: 8042-8047.

107. Zhuravlev PI, Reddy G, Straub JE, Thirumalai D (2014) Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. J Mol Biol 426: 2653-2666.

108. Speed MA, Wang DI, King J (1996) Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. Nat Biotechnol 14: 1283-1287.

109. Kim D, Yu MH (1996) Folding pathway of human alpha 1-antitrypsin: characterization of an intermediate that is active but prone to aggregation. Biochem Biophys Res Commun 226: 378-384.

110. Fink AL (1995) Compact intermediate states in protein folding. Annu Rev Biophys Biomol Struct 24: 495-522.

111. Stefani M, Dobson CM (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J Mol Med (Berl) 81: 678-699.

112. Carrell RW, Lomas DA (1997) Conformational disease. Lancet 350: 134-138.

113. Koo EH, Lansbury P.T, Jr., Kelly JW (1999) Amyloid diseases: Abnormal protein aggregation in neurodegeneration. Proceedings of the National Academy of Sciences of the United States of America 96: 9989-9990.

114. Horwich A (2002) Protein aggregation in disease: A role for folding intermediates forming specific multimeric interactions. Journal of Clinical Investigation 110: 1221-1232.

115. Jucker M, Walker LC (2013) Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature 501: 45-51.

116. Takalo M, Salminen A, Soininen H, Hiltunen M, Haapasalo A (2013) Protein aggregation and degradation mechanisms in neurodegenerative diseases. Am J Neurodegener Dis 2: 1-14.

117. Mattson MP (2004) Pathways towards and away from Alzheimer's disease. Nature 430: 631-639.

118. Silveira JR, Raymond GJ, Hughson AG, Race RE, Sim VL, et al. (2005) The most infectious prion protein particles. Nature 437: 257-261.

119. Lesne S, Koh MT, Kotilinek L, Kayed R, Glabe CG, et al. (2006) A specific amyloid-beta protein assembly in the brain impairs memory. Nature 440: 352-357.

120. Walter S, Buchner J (2002) Molecular chaperones--cellular machines for protein folding. Angew Chem Int Ed Engl 41: 1098-1113.

121. Hendrick JP, Hartl FU (1993) Molecular chaperone functions of heat-shock proteins. Annu Rev Biochem 62: 349-384.

122. Ellis J (1987) Proteins as molecular chaperones. Nature 328: 378-379.

123. Powers ET, Powers DL, Gierasch LM (2012) FoldEco: A Model for Proteostasis in E. coli. Cell Reports 1: 265-276.

124. Zietkiewicz S, Krzewska J, Liberek K (2004) Successive and synergistic action of the Hsp70 and Hsp100 chaperones in protein disaggregation. J Biol Chem 279: 44376-44383.

125. Genevaux P, Keppel F, Schwager F, Langendijk-Genevaux PS, Hartl FU, et al. (2004) In vivo analysis of the overlapping functions of DnaK and trigger factor. EMBO Rep 5: 195-200.

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

127. Hendrick JP, Hartl FU (1995) The role of molecular chaperones in protein folding. FASEB J 9: 1559-1569.

128. Georgopoulos CP, Hendrix RW, Casjens SR, Kaiser AD (1973) Host participation in bacteriophage lambda head assembly. J Mol Biol 76: 45-60.

129. Sternberg N (1973) Properties of a mutant of Escherichia coli defective in bacteriophage lambda head formation (groE). II. The propagation of phage lambda. J Mol Biol 76: 25-44.

130. Georgopoulos CP, Hohn B (1978) Identification of a host protein necessary for bacteriophage morphogenesis (the groE gene product). Proc Natl Acad Sci U S A 75: 131-135.

131. Chandrasekhar GN, Tilly K, Woolford C, Hendrix R, Georgopoulos C (1986) Purification and properties of the groES morphogenetic protein of Escherichia coli. J Biol Chem 261: 12414-12419.

132. Tilly K, Murialdo H, Georgopoulos C (1981) Identification of a second Escherichia coli groE gene whose product is necessary for bacteriophage morphogenesis. Proceedings of the National Academy of Sciences of the United States of America 78: 1629-1633.

133. Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The Groes and Groel Heat-Shock Gene-Products of Escherichia-Coli Are Essential for Bacterial-Growth at All Temperatures. Journal of Bacteriology 171: 1379-1385.

134. Georgopoulos CP, Hendrix RW, Kaiser AD, Wood WB (1972) Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. Nat New Biol 239: 38-41.

135. Takano T, Kakefuda T (1972) Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. Nat New Biol 239: 34-37.

136. Coppo A, Manzi A, Pulitzer JF, Takahashi H (1973) Abortive bacteriophage T4 head assembly in mutants of Escherichia coli. J Mol Biol 76: 61-87.

137. Zweig M, Cummings DJ (1973) Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. J Mol Biol 80: 505-518.

138. Goldenberg D, King J (1982) Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. Proc Natl Acad Sci U S A 79: 3403-3407.

139. Brunschier R, Danner M, Seckler R (1993) Interactions of phage P22 tailspike protein with GroE molecular chaperones during refolding in vitro. Journal of Biological Chemistry 268: 2767-2772.

140. Gordon CL, Sather SK, Casjens S, King J (1994) Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. Journal of Biological Chemistry 269: 27941-27951.

141. Bochkareva ES, Lissin NM, Girshovich AS (1988) Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature 336: 254-257.

142. Kusukawa N, Yura T, Ueguchi C, Akiyama Y, Ito K (1989) Effects of mutations in heat-shock genes groES and groEL on protein export in Escherichia coli. EMBO J 8: 3517-3521.

143. Laminet AA, Pluckthun A (1989) The precursor of beta-lactamase: purification, properties and folding kinetics. EMBO J 8: 1469-1477.

144. Laminet AA, Ziegelhoffer T, Georgopoulos C, Plückthun A (1990) The Escherichia coli heat shock proteins GroEL and GroES modulate the folding of the β-lactamase precursor. EMBO Journal 9: 2315-2319.

145. Lecker S, Lill R, Ziegelhoffer T, Georgopoulos C, Bassford PJ, Jr., et al. (1989) Three pure chaperone proteins of Escherichia coli--SecB, trigger factor and

GroEL--form soluble complexes with precursor proteins in vitro. EMBO J 8: 2703-2709.

146. Hunt JF, Weaver AJ, Landry SJ, Gierasch L, Deisenhofer J (1996) The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. Nature 379: 37-45.

147. Touchette NA, Perry KM, Matthews CR (1986) Folding of dihydrofolate reductase from Escherichia coli. Biochemistry 25: 5445-5452.

148. Goloubinoff P, Gatenby AA, Lorimer GH (1989) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. Nature 337: 44-47.

149. Mendoza JA, Rogers E, Lorimer GH, Horowitz PM (1991) Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. Journal of Biological Chemistry 266: 13044-13049.

150. Mendoza JA, Lorimer GH, Horowitz PM (1991) Intermediates in the chaperonin-assisted refolding of rhodanese are trapped at low temperature and show a small stoichiometry. J Biol Chem 266: 16973-16976.

151. Gatenby AA (1991) Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. Biochemistry 30: 9716-9723.

152. Braig K, Otwinowskl Z, Hegde R, Boisvert DC, Joachimiak A, et al. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. Nature 371: 578-586.

153. Boisvert DC, Wang J, Otwinowski Z, Horwich AL, Sigler PB (1996) The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. Nat Struct Biol 3: 170-177.

154. Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 388: 741-750.

155. Martin J, Mayhew M, Langer T, Hartl FU (1993) The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. Nature 366: 228-233.

156. Kad NM, Ranson NA, Cliff MJ, Clarke AR (1998) Asymmetry, commitment and inhibition in the GroE ATPase cycle impose alternating functions on the two GroEL rings. Journal of Molecular Biology 278: 267-278.

157. Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, et al. (1999) GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. Cell 97: 325-338.

158. Madan D, Lin Z, Rye HS (2008) Triggering protein folding within the GroEL-GroES complex. Journal of Biological Chemistry 283: 32003-32013.

159. Lin Z, Puchalla J, Shoup D, Rye HS (2013) Repetitive protein unfolding by the trans ring of the GroEL-GroES chaperonin complex stimulates folding. Journal of Biological Chemistry 288: 30944-30955.

160. Grason JP, Gresham JS, Lorimer GH (2008) Setting the chaperonin timer: a two-stroke, two-speed, protein machine. Proc Natl Acad Sci U S A 105: 17339-17344.

161. Chen S, Roseman AM, Hunter AS, Wood SP, Burston SG, et al. (1994) Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. Nature 371: 261-264.

162. Roseman AM, Chen S, White H, Braig K, Saibil HR (1996) The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL. Cell 87: 241-251.

163. Rye HS, Burston SG, Fenton WA, Beechem JM, Xu Z, et al. (1997) Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. Nature 388: 792-798.

164. Todd MJ, Viitanen PV, Lorimer GH (1994) Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. Science 265: 659-666.

165. Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371: 614-619.

166. Lin Z, Madan D, Rye HS (2008) GroEL stimulates protein folding through forced unfolding. Nature Structural and Molecular Biology 15: 303-311.

167. Shtilerman M, Lorimer GH, Englander SW (1999) Chaperonin function: Folding by forced unfolding. Science 284: 822-825.

168. Lin Z, Rye HS (2004) Expansion and compression of a protein folding intermediate by GroEL. Molecular Cell 16: 23-34.

169. Zahn R, Spitzfaden C, Ottiger M, Wüthrich K, Plückthun A (1994) Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. Nature 368: 261-265.

170. Staniforth RA, Burston SG, Atkinson T, Clarke AR (1994) Affinity of chaperonin-60 for a protein substrate and its modulation by nucleotides and chaperonin-10. Biochem J 300 (Pt 3): 651-658.

171. Cliff MJ, Limpkin C, Cameron A, Burston SG, Clarke AR (2006) Elucidation of steps in the capture of a protein substrate for efficient encapsulation by GroE. Journal of Biological Chemistry 281: 21266-21275.

172. Ueno T, Taguchi H, Tadakuma H, Yoshida M, Funatsu T (2004) GroEL mediates protein folding with a two successive timer mechanism. Molecular Cell 14: 423-434.

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

174. Jackson GS, Staniforth RA, Halsall DJ, Atkinson T, Holbrook JJ, et al. (1993) Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: implications for the mechanism of assisted protein folding. Biochemistry 32: 2554-2563.

175. Weissman JS, Kim PS (1991) Reexamination of the folding of BPTI: predominance of native intermediates. Science 253: 1386-1393.

176. Martin J, Langer T, Boteva R, Schramel A, Horwich AL, et al. (1991) Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. Nature 352: 36-42.

177. Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, et al. (1995) Mechanism of GroEL action: Productive release of polypeptide from a sequestered position under groes. Cell 83: 577-587.

178. Burston SG, Weissman JS, Farr GW, Fenton WA, Horwich AL (1996) Release of both native and non-native proteins from a cis-only GRoEL ternary complex. Nature 383: 96-99.

179. Baumketner A, Jewett A, Shea JE (2003) Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. J Mol Biol 332: 701-713.

180. Hayer-Hartl M, Minton AP (2006) A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. Biochemistry 45: 13356-13360.

181. Agard DA (1993) To fold or not to fold. Science 260: 1903-1904.

182. Saibil HR, Zheng D, Roseman AM, Hunter AS, Watson GMF, et al. (1993) ATP induces large quaternary rearrangements in a cage-like chaperonin structure. Current Biology 3: 265-273.

183. Ellis RJ (1996) Revisiting the Anfinsen cage. Fold Des 1: R9-15.

184. Chan HS, Dill KA (1996) A simple model of chaperonin-mediated protein folding. Proteins: Structure, Function and Genetics 24: 345-351.

185. Zahn R, Perrett S, Fersht AR (1996) Conformational states bound by the molecular chaperones GroEL and SecB: A hidden unfolding (annealing) activity. Journal of Molecular Biology 261: 43-61.

186. Walter S, Lorimer GH, Schmid FX (1996) A thermodynamic coupling mechanism for GroEL-mediated unfolding. Proceedings of the National Academy of Sciences of the United States of America 93: 9425-9430.

187. Todd MJ, Lorimer GH, Thirumalai D (1996) 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 93: 4030-4035.

188. Thirumalai D, Lorimer GH (2001) Chaperonin-mediated protein folding. Annual Review of Biophysics and Biomolecular Structure. pp. 245-269.

189. Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 16: 574-581.

190. Lin Z, Rye H (2006) GroEL-mediated protein folding: Making the impossible, possible. Critical Reviews in Biochemistry and Molecular Biology 41: 211-239.

191. Grantcharova V, Alm EJ, Baker D, Horwich AL (2001) Mechanisms of protein folding. Current Opinion in Structural Biology 11: 70-82.

192. Apetri AC, Horwich AL (2008) Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proceedings of the National Academy of Sciences of the United States of America 105: 17351-17355.

193. Horwich AL, Apetri AC, Fenton WA (2009) The GroEL/GroES cis cavity as a passive anti-aggregation device. Febs Letters 583: 2654-2662.

194. Horwich AL (2014) Molecular chaperones in cellular protein folding: the birth of a field. Cell 157: 285-288.

195. Ellis RJ, Minton AP (2006) Protein aggregation in crowded environments. Biological Chemistry 387.

196. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of in vivo substrates of the chaperonin GroEL. Nature 402: 147-154.

197. Mayer MP, Bukau B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci 62: 670-684.

198. Horwich AL, Fenton WA (2009) Chaperonin-mediated protein folding: Using a central cavity to kinetically assist polypeptide chain folding. Quarterly Reviews of Biophysics 42: 83-116.

199. Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, et al. (2001) Dual function of protein confinement in chaperonin-assisted protein folding. Cell 107: 223-233.

200. Horst R, Fenton WA, Englander SW, Wuthrich K, Horwich AL (2007) Folding trajectories of human dihydrofolate reductase inside the GroEL GroES chaperonin cavity and free in solution. Proc Natl Acad Sci U S A 104: 20788-20792.

201. Tyagi NK, Fenton WA, Deniz AA, Horwich AL (2011) 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 Lett 585: 1969-1972.

202. Buchner J, Schmidt M, Fuchs M, Jaenicke R, Rudolph R, et al. (1991) GroE facilitates refolding of citrate synthase by suppressing aggregation. Biochemistry 30: 1586-1591.

203. Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, et al. (2006) Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein. Cell 125: 903-914.

204. Tang YC, Chang HC, Chakraborty K, Hartl FU, Hayer-Hartl M (2008) Essential role of the chaperonin folding compartment in vivo. EMBO Journal 27: 1458-1468.

205. Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, et al. (2010) Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding. Cell 142: 112-122.

206. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL (1996) Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell 84: 481-490.

207. Takagi F, Koga N, Takada S (2003) How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. Proceedings of the National Academy of Sciences of the United States of America 100: 11367-11372.

208. Jewett AI, Baumketner A, Shea JE (2004) Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. Proc Natl Acad Sci U S A 101: 13192-13197.

209. Farr GW, Fenton WA, Horwich AL (2007) 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 104: 5342-5347.

210. Daniel RM, Peterson ME, Danson MJ, Price NC, Kelly SM, et al. (2010) The molecular basis of the effect of temperature on enzyme activity. Biochem J 425: 353-360.

211. Farr GW, Fenton WA, Chaudhuri TK, Clare DK, Saibil HR, et al. (2003) Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes. EMBO J 22: 3220-3230.

212. Zheng X, Rosenberg LE, Kalousek F, Fenton WA (1993) GroEL, GroES, and ATP-dependent folding and spontaneous assembly of ornithine transcarbamylase. J Biol Chem 268: 7489-7493.

213. Lund PA (2009) Multiple chaperonins in bacteria--why so many? FEMS Microbiol Rev 33: 785-800.

214. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. (1995) The Minimal Gene Complement of Mycoplasma genitalium. Science 270: 397-404.

215. Shepard MC, Lunceford CD, Ford DK, Purcell RH, Taylor-Robinson D, et al. (1974) Ureaplasma urealyticum gen. nov., sp. nov.: Proposed Nomenclature for the Human T (T-Strain) Mycoplasmas. International Journal of Systematic Bacteriology 24: 160-171.

216. Kenny GE, Cartwright FD (1977) Effect of urea concentration on growth of Ureaplasma urealyticum (T-strain mycoplasma). J Bacteriol 132: 144-150.

217. Dunbar J, Yennawar HP, Banerjee S, Luo J, Farber GK (1997) The effect of denaturants on protein structure. Protein Sci 6: 1727-1733.

218. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293: 487-489.

219. Bennion BJ, Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. Proc Natl Acad Sci U S A 100: 5142-5147.

220. Pollack JD, Williams MV, McElhaney RN (1997) The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. Crit Rev Microbiol 23: 269-354.

221. Sherman MY, Qian SB (2013) Less is more: improving proteostasis by translation slow down. Trends Biochem Sci 38: 585-591.

222. Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM (2010) Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. J Mol Biol 396: 1310-1318.

223. Zeilstra-Ryalls J, Fayet O, Baird L, Georgopoulos C (1993) Sequence analysis and phenotypic characterization of groEL mutations that block lambda and T4 bacteriophage growth. J Bacteriol 175: 1134-1143.

224. Wang JD, Herman C, Tipton KA, Gross CA, Weissman JS (2002) Directed evolution of substrate-optimized GroEL/S chaperonins. Cell 111: 1027-1039.

225. Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Jr., et al. (2005) GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell 123: 861-873.

226. Huq S, Sueoka K, Narumi S, Arisaka F, Nakamoto H (2010) Comparative biochemical characterization of two GroEL homologs from the cyanobacterium Synechococcus elongatus PCC 7942. Biosci Biotechnol Biochem 74: 2273-2280.

227. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, Saibil HR (2006) An expanded protein folding cage in the GroEL-gp31 complex. J Mol Biol 358: 905-911.

228. van der Vies SM, Gatenby AA, Georgopoulos C (1994) Bacteriophage T4 encodes a co-chaperonin that can substitute for Escherichia coli GroES in protein folding. Nature 368: 654-656.

229. Kiljunen S, Hakala K, Pinta E, Huttunen S, Pluta P, et al. (2005) Yersiniophage phiR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. Microbiology 151: 4093-4102.

230. Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, et al. (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. Proc Natl Acad Sci U S A 110: 12798-12803.

231. Jang HB, Fagutao FF, Nho SW, Park SB, Cha IS, et al. (2013) Phylogenomic network and comparative genomics reveal a diverged member of the PhiKZ-related group, marine vibrio phage PhiJM-2012. J Virol 87: 12866-12878.

232. Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, et al. (2012) Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse PhiKZ-related phages. J Virol 86: 1844-1852.

233. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, et al. (2005) Genome comparison of Pseudomonas aeruginosa large phages. J Mol Biol 354: 536-545.

234. Kurochkina LP, Semenyuk PI, Orlov VN, Robben J, Sykilinda NN, et al. (2012) Expression and functional characterization of the first bacteriophage-encoded chaperonin. J Virol 86: 10103-10111.

235. Semenyuk PI, Orlov VN, Kurochkina LP (2015) Effect of Chaperonin Encoded by Gene 146 on Thermal Aggregation of Lytic Proteins of Bacteriophage EL Pseudomonas aeruginosa. Biochemistry (Mosc) 80: 172-179.

236. Tsuji A, Kaneko Y, Takahashi K, Ogawa M, Goto S (1982) The effects of temperature and pH on the growth of eight enteric and nine glucose non-fermenting species of gram-negative rods. Microbiol Immunol 26: 15-24.

237. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

238. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31: 3320-3323.

239. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475: 324-332.

240. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Ulrich Hartl F (2013) Molecular chaperone functions in protein folding and proteostasis. Annual Review of Biochemistry. pp. 323-355.

241. Li J, Wang Y, Zhang CY, Zhang WY, Jiang DM, et al. (2010) Myxococcus xanthus viability depends on groEL supplied by either of two genes, but the paralogs have different functions during heat shock, predation, and development. J Bacteriol 192: 1875-1881.

242. Wang Y, Zhang WY, Zhang Z, Li J, Li ZF, et al. (2013) Mechanisms involved in the functional divergence of duplicated GroEL chaperonins in Myxococcus xanthus DK1622. PLoS Genet 9: e1003306.

243. Wang Y, Li X, Zhang W, Zhou X, Li YZ (2014) The groEL2 gene, but not groEL1, is required for biosynthesis of the secondary metabolite myxovirescin in Myxococcus xanthus DK1622. Microbiology 160: 488-495.

244. Suzuki M, Ueno T, Iizuka R, Miura T, Zako T, et al. (2008) Effect of the C-terminal truncation on the functional cycle of chaperonin GroEL: Implication that the C-terminal region facilitates the transition from the folding-arrested to the folding-competent state. Journal of Biological Chemistry 283: 23931-23939.

245. Koike-Takeshita A, Arakawa T, Taguchi H, Shimamura T (2014) Crystal structure of a symmetric football-shaped GroEL:GroES2-ATP14 complex determined at 3.8A reveals rearrangement between two GroEL rings. J Mol Biol 426: 3634-3641.

246. McLennan NF, Girshovich AS, Lissin NM, Charters Y, Masters M (1993) The strongly conserved carboxyl-terminus glycine-methionine motif of the Escherichia coli GroEL chaperonin is dispensable. Molecular Microbiology 7: 49-58.

247. Burnett BP, Horwich AL, Low KB (1994) A Carboxy-Terminal Deletion Impairs the Assembly of Groel and Confers a Pleiotropic Phenotype in Escherichia-Coli K-12. Journal of Bacteriology 176: 6980-6985.

248. McLennan NF, McAteer S, Masters M (1994) The tail of a chaperonin: the C-terminal region of Esctierichia coli GroEL protein. Molecular Microbiology 14: 309-321.

249. Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. Chem Rev 106: 1917-1930.

250. Machida K, Kono-Okada A, Hongo K, Mizobata T, Kawata Y (2008) Hydrophilic residues 526KNDAAD531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. Journal of Biological Chemistry 283: 6886-6896.

251. Erbse A, Dougan DA, Bukau B (2003) A folding machine for many but a master of none. Nat Struct Biol 10: 84-86.

252. Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, et al. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333: 330-334.

253. Jindal S, Dudani AK, Singh B, Harley CB, Gupta RS (1989) Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. Mol Cell Biol 9: 2279-2283.

254. Levy-Rimler G, Bell RE, Ben-Tal N, Azem A (2002) Type I chaperonins: not all are created equal. FEBS Letters 529: 1-5.

255. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of Escherichia coli K-12. Science 277: 1453-1462.

256. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

257. Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26: 167-180.

258. Clary DO, Wolstenholme DR (1984) The Drosophila mitochondrial genome. Oxf Surv Eukaryot Genes 1: 1-35.

259. Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris suum. Genetics 130: 471-498.

260. Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet 15: 57-61.

261. Taylor SW, Fahy E, Zhang B, Glenn GM, Warnock DE, et al. (2003) Characterization of the human heart mitochondrial proteome. Nat Biotechnol 21: 281-286.

262. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res 33: D154-159.

263. Han W, Christen P (2004) cis-Effect of DnaJ on DnaK in ternary complexes with chimeric DnaK/DnaJ-binding peptides. FEBS Lett 563: 146-150.

264. Cyr DM, Langer T, Douglas MG (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends in Biochemical Sciences 19: 176-181.

265. Borges JC, Seraphim TV, Mokry DZ, Almeida FC, Cyr DM, et al. (2012) Identification of regions involved in substrate binding and dimer stabilization within the central domains of yeast Hsp40 Sis1. PLoS One 7: e50927.

266. Luke MM, Sutton A, Arndt KT (1991) Characterization of SIS1, a Saccharomyces cerevisiae homologue of bacterial dnaJ proteins. J Cell Biol 114: 623-638.

267. Philippsen P, Kleine K, Pohlmann R, Dusterhoft A, Hamberg K, et al. (1997) The nucleotide sequence of Saccharomyces cerevisiae chromosome XIV and its evolutionary implications. Nature 387: 93-98.

268. Cheetham ME, Caplan AJ (1998) Structure, function and evolution of DnaJ: Conservation and adaptation of chaperone function. Cell Stress and Chaperones 3: 28-36.

269. Mayer MP, Brehmer D, Gässler CS, Bukau B (2001) Hsp70 chaperone machines. Advances in Protein Chemistry. pp. 1-44.

270. Fan CY, Lee S, Cyr DM (2003) Mechanisms for regulation of Hsp70 function by Hsp40. Cell Stress and Chaperones 8: 309-316.

271. Caplan AJ, Douglas MG (1991) Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein. J Cell Biol 114: 609-621.

272. Bardwell JC, Tilly K, Craig E, King J, Zylicz M, et al. (1986) The nucleotide sequence of the Escherichia coli K12 dnaJ+ gene. A gene that encodes a heat shock protein. J Biol Chem 261: 1782-1785.

273. Lund PA (2001) Microbial molecular chaperones. Adv Microb Physiol 44: 93-140.

274. Karlin S, Brocchieri L (2000) Heat shock protein 60 sequence comparisons: duplications, lateral transfer, and mitochondrial evolution. Proc Natl Acad Sci U S A 97: 11348-11353.

275. Goyal K, Qamra R, Mande SC (2006) Multiple gene duplication and rapid evolution in the groEL gene: functional implications. J Mol Evol 63: 781-787.

276. Colaco CA, MacDougall A (2014) Mycobacterial chaperonins: the tail wags the dog. FEMS Microbiol Lett 350: 20-24.

277. Gould PS, Burgar HR, Lund PA (2007) Homologous cpn60 genes in Rhizobium leguminosarum are not functionally equivalent. Cell Stress & Chaperones 12: 123.

278. Fan M, Rao T, Zacco E, Ahmed MT, Shukla A, et al. (2012) The unusual mycobacterial chaperonins: evidence for in vivo oligomerization and specialization of function. Mol Microbiol 85: 934-944.

279. Wilson ML (2008) Reducing the global burden of mycobacterial infections: one more piece of the puzzle. Am J Clin Pathol 130: 849-852.

280. Kong TH, Coates AR, Butcher PD, Hickman CJ, Shinnick TM (1993) Mycobacterium tuberculosis expresses two chaperonin-60 homologs. Proc Natl Acad Sci U S A 90: 2608-2612.

281. Hu Y, Henderson B, Lund PA, Tormay P, Ahmed MT, et al. (2008) A Mycobacterium tuberculosis mutant lacking the groEL homologue cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. Infect Immun 76: 1535-1546.

282. Sielaff B, Lee KS, Tsai FT (2011) Structural and functional conservation of Mycobacterium tuberculosis GroEL paralogs suggests that GroEL1 Is a chaperonin. J Mol Biol 405: 831-839.

283. Qamra R, Mande SC (2004) Crystal structure of the 65-kilodalton heat shock protein, chaperonin 60.2, of Mycobacterium tuberculosis. J Bacteriol 186: 8105-8113.

284. Qamra R, Srinivas V, Mande SC (2004) Mycobacterium tuberculosis GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. J Mol Biol 342: 605-617.

285. Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, et al. (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. Molecular Microbiology 50: 463-473.

286. Slayden RA, Barry CE, 3rd (2002) The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in Mycobacterium tuberculosis. Tuberculosis (Edinb) 82: 149-160.

287. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, et al. (2006) Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci U S A 103: 15200-15205.

288. Otani M, Tabata J, Ueki T, Sano K, Inouye S (2001) Heat-shock-induced proteins from Myxococcus xanthus. J Bacteriol 183: 6282-6287.

289. Lindquist S (1986) The heat-shock response. Annu Rev Biochem 55: 1151-1191.

290. Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381: 571-579.

291. Dworkin M (1996) Recent advances in the social and developmental biology of the myxobacteria. Microbiol Rev 60: 70-102.

292. Burnside K, Rajagopal L (2012) Regulation of prokaryotic gene expression by eukaryotic-like enzymes. Curr Opin Microbiol 15: 125-131.

293. Konovalova A, Petters T, Sogaard-Andersen L (2010) Extracellular biology of Myxococcus xanthus. FEMS Microbiol Rev 34: 89-106.

294. D'Costa VM, McGrann KM, Hughes DW, Wright GD (2006) Sampling the antibiotic resistome. Science 311: 374-377.

295. Walsh CT (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science 303: 1805-1810.

296. Hibbing ME, Fuqua C, Parsek MR, Peterson SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8: 15-25.

297. Xiao Y, Wei X, Ebright R, Wall D (2011) Antibiotic production by myxobacteria plays a role in predation. J Bacteriol 193: 4626-4633.

298. Musgrove JE, Johnson RA, Ellis RJ (1987) Dissociation of the ribulosebisphosphate-carboxylase large-subunit binding protein into dissimilar subunits. European Journal of Biochemistry 163: 529-534.

299. Martel R, Cloney LP, Pelcher LE, Hemmingsen SM (1990) Unique composition of plastid chaperonin-60: α and β polypeptide-encoding genes are highly divergent. Gene 94: 181-182,IN183-IN184,183-187.

300. Nishio K, Hirohashi T, Nakai M (1999) Chloroplast chaperonins: Evidence for heterogeneous assembly of α and β Cpn60 polypeptides into a chaperonin oligomer. Biochemical and Biophysical Research Communications 266: 584-587.

301. Wastl J, Fraunholz M, Zauner S, Douglas S, Maier UG (1999) Ancient gene duplication and differential gene flow in plastid lineages: The groEL/cpn60 example. Journal of Molecular Evolution 48: 112-117.

302. Dickson R, Weiss C, Howard RJ, Alldrick SP, Ellis RJ, et al. (2000) Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. Journal of Biological Chemistry 275: 11829-11835.

303. Viitanen PV, Schmidt M, Buchner J, Suzuki T, Vierling E, et al. (1995) Functional-Characterization of the Higher-Plant Chloroplast Chaperonins. Journal of Biological Chemistry 270: 18158-18164.

304. Hill JE, Hemmingsen SM (2001) Arabidopsis thaliana type I and II chaperonins. Cell Stress & Chaperones 6: 190-200.

305. Peltier JB, Cai Y, Sun Q, Zabrouskov V, Giacomelli L, et al. (2006) The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts. Mol Cell Proteomics 5: 114-133.

306. Weiss C, Bonshtien A, Farchi-Pisanty O, Vitlin A, Azem A (2009) Cpn20: siamese twins of the chaperonin world. Plant Mol Biol 69: 227-238.

307. Peng L, Fukao Y, Myouga F, Motohashi R, Shinozaki K, et al. (2011) A chaperonin subunit with unique structures is essential for folding of a specific substrate. PLoS Biol 9: e1001040.

308. Mayhew M, Da Silva ACR, Martin J, Erdjument-Bromage H, Tempst P, et al. (1996) Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature 379: 420-426.

309. Burston SG, Ranson NA, Clarke AR (1995) The origins and consequences of asymmetry in the chaperonin reaction cycle. Journal of Molecular Biology 249: 138-152.

310. Yifrach O, Horovitz A (1995) Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. Biochemistry 34: 5303-5308.

311. Clare DK, Bakkes PJ, Van Heerikhuizen H, Van Der Vies SM, Saibil HR (2009) Chaperonin complex with a newly folded protein encapsulated in the folding chamber. Nature 457: 107-110.

312. Elad N, Farr GW, Clare DK, Orlova EV, Horwich AL, et al. (2007) Topologies of a Substrate Protein Bound to the Chaperonin GroEL. Molecular Cell 26: 415-426.

313. Falke S, Tama F, Brooks Iii CL, Gogol EP, Fisher MT (2005) The 13 Å structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. Journal of Molecular Biology 348: 219-230.

314. Kanno R, Koike-Takeshita A, Yokoyama K, Taguchi H, Mitsuoka K (2009) Cryo-EM Structure of the Native GroEL-GroES Complex from Thermus thermophilus Encapsulating Substrate Inside the Cavity. Structure 17: 287-293.

315. Clare DK, Vasishtan D, Stagg S, Quispe J, Farr GW, et al. (2012) ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. Cell 149: 113-123.

316. Rye HS (2001) Application of fluorescence resonance energy transfer to the GroEL-GroES chaperonin reaction. Methods 24: 278-288.

317. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. Journal of Structural Biology 128: 82-97.

318. Yang C, Jiang W, Chen DH, Adiga U, Ng EG, et al. (2009) Estimating contrast transfer function and associated parameters by constrained non-linear optimization. J Microsc 233: 391-403.

319. Chen DH, Song JL, Chuang DT, Chiu W, Ludtke SJ (2006) An Expanded Conformation of Single-Ring GroEL-GroES Complex Encapsulates an 86 kDa Substrate. Structure 14: 1711-1722.

320. Schröder GF, Brunger AT, Levitt M (2007) Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution. Structure 15: 1630-1641.

321. Scheres SHW, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. Nature Methods 9: 853-854.

322. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25: 1605-1612.

323. Chen DH, Luke K, Zhang J, Chiu W, Wittung-Stafshede P (2008) Location and Flexibility of the Unique C-Terminal Tail of Aquifex aeolicus Co-Chaperonin Protein 10 as Derived by Cryo-Electron Microscopy and Biophysical Techniques. Journal of Molecular Biology 381: 707-717.

324. Penczek PA, Yang C, Frank J, Spahn CMT (2006) Estimation of variance in single-particle reconstruction using the bootstrap technique. Journal of Structural Biology 154: 168-183.

325. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, et al. (2012) Outcome of the first electron microscopy validation task force meeting. Structure 20: 205-214.

326. Cong Y, Schröder GF, Meyer AS, Jakana J, Ma B, et al. (2012) Symmetry-free cryo-EM structures of the chaperonin TRiC along its ATPase-driven conformational cycle. EMBO Journal 31: 720-730.

327. Baker ML, Ju T, Chiu W (2007) Identification of Secondary Structure Elements in Intermediate-Resolution Density Maps. Structure 15: 7-19.

328. Ranson NA, Clare DK, Farr GW, Houldershaw D, Horwich AL, et al. (2006) Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. Nature Structural and Molecular Biology 13: 147-152.

329. Van Der Vies SM, Viitanen PV, Gatenby AA, Lorimer GH, Jaenicke R (1992) Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. Biochemistry® 31: 3635-3644.

330. Farr GW, Furtak K, Rowland MB, Ranson NA, Saibil HR, et al. (2000) Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. Cell 100: 561-573.

331. Horovitz A, Fridmann Y, Kafri G, Yifrach O (2001) Review: Allostery in chaperonins. Journal of Structural Biology 135: 104-114.

332. Sparrer H, Buchner J (1997) How GroES regulates binding of nonnative protein to GroEL. Journal of Biological Chemistry 272: 14080-14086.

333. Frank GA, Goomanovsky M, Davidi A, Ziv G, Horovitz A, et al. (2010) Out-of-equilibrium conformational cycling of GroEL under saturating ATP

concentrations. Proceedings of the National Academy of Sciences of the United States of America 107: 6270-6274.

334. Yifrach O, Horovitz A (2000) Coupling between protein folding and allostery in the GroE chaperonin system. Proceedings of the National Academy of Sciences of the United States of America 97: 1521-1524.

335. Badcoe IG, Smith CJ, Wood S, Halsall DJ, Holbrook JJ, et al. (1991) Binding of a chaperonin to the folding intermediates of lactate dehydrogenase. Biochemistry 30: 9195-9200.

336. Brockwell DJ, Radford SE (2007) Intermediates: ubiquitous species on folding energy landscapes? Current Opinion in Structural Biology 17: 30-37.

337. Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE (2009) Biological and chemical approaches to diseases of proteostasis deficiency. Annual Review of Biochemistry. pp. 959-991.

338. Jewett AI, Shea JE (2010) Reconciling theories of chaperonin accelerated folding with experimental evidence. Cellular and Molecular Life Sciences 67: 255-276.

339. Ranson NA, Dunster NJ, Burston SG, Clarke AR (1995) Chaperonins can catalyse the reversal of early aggregation steps when a protein misfolds. Journal of Molecular Biology 250: 581-586.

340. Ye X, Lorimer GH (2013) Substrate protein switches GroE chaperonins from asymmetric to symmetric cycling by catalyzing nucleotide exchange. Proceedings of the National Academy of Sciences of the United States of America 110: E4289-E4297.

341. Betancourt MR, Thirumalai D (1999) Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. Journal of Molecular Biology 287: 627-644.

342. Chen J, Walter S, Horwich AL, Smith DL (2001) Folding of malate dehydrogenase inside the GroEL-GroES cavity. Nature Structural Biology 8: 721-728.

343. Coyle JE, Texter FL, Ashcroft AE, Masselos D, Robinson CV, et al. (1999) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. Nature Structural Biology 6: 683-690.

344. Goldberg MS, Zhang J, Sondek S, Matthews CR, Fox RO, et al. (1997) Native-like structure of a protein-folding intermediate bound to the chaperonin GroEL.

Proceedings of the National Academy of Sciences of the United States of America 94: 1080-1085.

345. Sharma S, Chakraborty K, Müller BK, Astola N, Tang YC, et al. (2008) Monitoring Protein Conformation along the Pathway of Chaperonin-Assisted Folding. Cell 133: 142-153.

346. Motojima F, Motojima-Miyazaki Y, Yoshida M (2012) 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 109: 15740-15745.

347. Chen DH, Madan D, Weaver J, Lin Z, Schroder GF, et al. (2013) Visualizing GroEL/ES in the act of encapsulating a folding protein. Cell 153: 1354-1365.

348. James DR, Siemiarczuk A, Ware WR (1992) Stroboscopic optical boxcar technique for the determination of fluorescence lifetimes. Review of Scientific Instruments 63: 1710-1716.

349. Gray TE, Fersht AR (1991) Cooperativity in ATP hydrolysis by GroEL is increased by GroES. FEBS Letters 292: 254-258.

350. Sameshima T, Iizuka R, Ueno T, Funatsu T (2010) Denatured proteins facilitate the formation of the football-shaped GroEL-(GroES)2 complex. Biochemical Journal 427: 247-254.

351. Sameshima T, Iizuka R, Ueno T, Wada J, Aoki M, et al. (2010) Single-molecule study on the decay process of the football-shaped GroEL-GroES complex using zero-mode waveguides. Journal of Biological Chemistry 285: 23159-23164.

352. Yang D, Ye X, Lorimer GH (2013) Symmetric GroEL: GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. Proceedings of the National Academy of Sciences of the United States of America 110: E4298-E4305.

353. Murai N, Makino Y, Yoshida M (1996) GroEL locked in a closed conformation by an interdomain cross-link can bind ATP and polypeptide but cannot process further reaction steps. Journal of Biological Chemistry 271: 28229-28234.

354. Brocchieri L, Karlin S (2000) Conservation among HSP60 sequences in relation to structure, function, and evolution. Protein Science 9: 476-486.

355. Çetinbaş M, Shakhnovich EI (2013) Catalysis of Protein Folding by Chaperones Accelerates Evolutionary Dynamics in Adapting Cell Populations. PLoS Computational Biology 9.

356. Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E (2002) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. Nature 417: 398.

357. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-673.

358. Williams TA, Fares MA (2010) The effect of chaperonin buffering on protein evolution. Genome Biology and Evolution 2: 609-619.

359. Wyganowski KT, Kaltenbach M, Tokuriki N (2013) GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. Journal of Molecular Biology 425: 3403-3414.

360. Browne P, O'Cuinn G (1983) The purification and characterization of a proline dipeptidase from guinea pig brain. Journal of Biological Chemistry 258: 6147-6154.

361. Fernandez-Espla MD, Martin-Hernandez MC, Fox PF (1997) Purification and characterization of a prolidase from Lactobacillus casei subsp. casei IFPL 731. Appl Environ Microbiol 63: 314-316.

362. Jalving R, Bron P, Kester HC, Visser J, Schaap PJ (2002) Cloning of a prolidase gene from Aspergillus nidulans and characterisation of its product. Mol Genet Genomics 267: 218-222.

363. Lupi A, Della Torre S, Campari E, Tenni R, Cetta G, et al. (2006) Human recombinant prolidase from eukaryotic and prokaryotic sources. Expression, purification, characterization and long-term stability studies. FEBS J 273: 5466-5478.

364. Maher MJ, Ghosh M, Grunden AM, Menon AL, Adams MW, et al. (2004) Structure of the prolidase from Pyrococcus furiosus. Biochemistry 43: 2771-2783.

365. Park MS, Hill CM, Li Y, Hardy RK, Khanna H, et al. (2004) Catalytic properties of the PepQ prolidase from Escherichia coli. Arch Biochem Biophys 429: 224-230.

366. Sjostrom H, Noren O, Josefsson L (1974) Purification and specificity of pig intestinal prolidase. Biochimica et Biophysica Acta 327: 457-470.

367. Arabidopsis Genome I (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.

368. Surazynski A, Miltyk W, Palka J, Phang JM (2008) Prolidase-dependent regulation of collagen biosynthesis. Amino Acids 35: 731-738.

369. Kitchener RL, Grunden AM (2012) Prolidase function in proline metabolism and its medical and biotechnological applications. J Appl Microbiol 113: 233-247.

370. Myara I, Charpentier C, Lemonnier A (1984) Prolidase and prolidase deficiency. Life Sci 34: 1985-1998.

371. Endo F, Tanoue A, Hata A, Kitano A, Matsuda I (1989) Deduced amino acid sequence of human prolidase and molecular analyses of prolidase deficiency. J Inherit Metab Dis 12: 351-354.

372. Lupi A, Tenni R, Rossi A, Cetta G, Forlino A (2008) Human prolidase and prolidase deficiency: an overview on the characterization of the enzyme involved in proline recycling and on the effects of its mutations. Amino Acids 35: 739-752.

373. Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, et al. (2004) Identification of 315 genes essential for early zebrafish development. Proc Natl Acad Sci U S A 101: 12792-12797.

374. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, et al. (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421: 231-237.

375. Silva JM, Marran K, Parker JS, Silva J, Golding M, et al. (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. Science 319: 617-620.

376. White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, et al. (2013) Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell 154: 452-464.

377. Miller CG, Schwartz G (1978) Peptidase-deficient mutants of Escherichia coli. J Bacteriol 135: 603-611.

378. Vyas NK, Nickitenko A, Rastogi VK, Shah SS, Quiocho FA (2010) Structural insights into the dual activities of the nerve agent degrading organophosphate anhydrolase/prolidase. Biochemistry 49: 547-559.

379. Stepankova A, Duskova J, Skalova T, Hasek J, Koval T, et al. (2013) Organophosphorus acid anhydrolase from Alteromonas macleodii: structural study and functional relationship to prolidases. Acta Crystallogr Sect F Struct Biol Cryst Commun 69: 346-354.

380. Cheng TC, Rastogi VK, Defrank JJ, Sawiris GP (1998) G-type nerve agent decontamination by Alteromonas prolidase. Annals of the New York Academy of Sciences 864: 253-258.

381. Lowther WT, Matthews BW (2002) Metalloaminopeptidases: common functional themes in disparate structural surroundings. Chem Rev 102: 4581-4608.

382. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, et al. (1996) Complete Sequence Analysis of the Genome of the Bacterium Mycoplasma Pneumoniae. Nucleic Acids Research 24: 4420-4449.

383. Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, et al. (2004) The complete genome and proteome of Mycoplasma mobile. Genome Res 14: 1447-1461.

384. Bazan JF, Weaver LH, Roderick SL, Huber R, Matthews BW (1994) 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 91: 2473-2477.

385. Alberto ME, Leopoldini M, Russo N (2011) Can human prolidase enzyme use different metals for full catalytic activity? Inorg Chem 50: 3394-3403.

386. Wilcox DE (1996) Binuclear Metallohydrolases. Chem Rev 96: 2435-2458.

387. Wilce MC, Bond CS, Dixon NE, Freeman HC, Guss JM, et al. (1998) Structure and mechanism of a proline-specific aminopeptidase from Escherichia coli. Proc Natl Acad Sci U S A 95: 3472-3477.

388. Wang SH, Zhi QW, Sun MJ (2005) Purification and characterization of recombinant human liver prolidase expressed in Saccharomyces cerevisiae. Archives of Toxicology 79: 253-259.

389. Willingham K, Maher MJ, Grunden AM, Ghosh M, Adams MWW, et al. (2001) Crystallization and characterization of the prolidase from Pyrococcus furiosus. Acta Crystallographica Section D: Biological Crystallography 57: 428-430.

390. Lowther WT, Matthews BW (2000) Structure and function of the methionine aminopeptidases. Biochim Biophys Acta 1477: 157-167.

391. Graham SC, Bond CS, Freeman HC, Guss JM (2005) Structural and functional implications of metal ion selection in aminopeptidase P, a metalloprotease with a dinuclear metal center. Biochemistry 44: 13820-13836.

392. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 276: 307-326.

393. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.

394. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486-501.

395. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455-461.

396. Ito Y, Watanabe Y, Hirano K, Sugiura M, Sawaki S, et al. (1984) A fluorometric method for dipeptidase activity measurement in urine, using L-alanyl-L-alanine as substrate. J Biochem 96: 1-8.

397. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

398. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

399. Zhang LB, Crossley MJ, Dixon NE, Ellis PJ, Fisher ML, et al. (1998) Spectroscopic identification of a dinuclear metal centre in manganese(II)-activated aminopeptidase P from Escherichia coli: implications for human prolidase. Journal of Biological Inorganic Chemistry 3: 470-483.

400. Besio R, Gioia R, Cossu F, Monzani E, Nicolis S, et al. (2013) Kinetic and structural evidences on human prolidase pathological mutants suggest strategies for enzyme functional rescue. PLoS One 8: e58792.

401. Outten CE, O'Halloran TV (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. Science 292: 2488-2492.

402. Ghosh M, Grunden AM, Dunn DM, Weiss R, Adams MW (1998) Characterization of native and recombinant forms of an unusual cobalt-dependent proline dipeptidase (prolidase) from the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol 180: 4781-4789.

403. Fersht A (1999) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. New York: W. H. Freeman. 631 p.

404. Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem 70: 209-246.

405. Shao Z, Arnold FH (1996) Engineering new functions and altering existing functions. Curr Opin Struct Biol 6: 513-518.

406. Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. Protein Sci 4: 337-360.

407. Hu K, Tanaka T (2009) S1 site residues of Lactococcus lactis prolidase affect substrate specificity and allosteric behaviour. Biochim Biophys Acta 1794: 1715-1724.

408. Cohen GN, Barbe V, Flament D, Galperin M, Heilig R, et al. (2003) An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi. Mol Microbiol 47: 1495-1512.

409. Jun X, Lupeng L, Minjuan X, Oger P, Fengping W, et al. (2011) Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon Pyrococcus yayanosii CH1. J Bacteriol 193: 4297-4298.

410. Jung JH, Lee JH, Holden JF, Seo DH, Shin H, et al. (2012) Complete genome sequence of the hyperthermophilic archaeon Pyrococcus sp. strain ST04, isolated from a deep-sea hydrothermal sulfide chimney on the Juan de Fuca Ridge. J Bacteriol 194: 4434-4435.

411. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, et al. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3. DNA Res 5: 55-76.

412. Lecompte O, Ripp R, Puzos-Barbe V, Duprat S, Heilig R, et al. (2001) Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. Genome Res 11: 981-993.

413. Lee HS, Bae SS, Kim MS, Kwon KK, Kang SG, et al. (2011) Complete genome sequence of hyperthermophilic Pyrococcus sp. strain NA2, isolated from a deep-sea hydrothermal vent area. J Bacteriol 193: 3666-3667.

414. Maeder DL, Weiss RB, Dunn DM, Cherry JL, Gonzalez JM, et al. (1999) Divergence of the hyperthermophilic archaea Pyrococcus furiosus and P. horikoshii inferred from complete genomic sequences. Genetics 152: 1299-1305.

415. Forlino A, Lupi A, Vaghi P, Cornaglia AI, Calligaro A, et al. (2002) Mutation analysis of five new patients affected by prolidase deficiency: The lack of enzyme activity causes necrosis-like cell death in cultured fibroblasts. Human Genetics 111: 314-322.

416. Ledoux P, Scriver CR, Hechtman P (1996) Expression and molecular analysis of mutations in prolidase deficiency. Am J Hum Genet 59: 1035-1039.

417. Ledoux P, Scriver C, Hechtman P (1994) Four novel PEPD alleles causing prolidase deficiency. Am J Hum Genet 54: 1014-1021.

418. Tawfik DS (2006) Biochemistry. Loop grafting and the origins of enzyme species. Science 311: 475-476.

419. Craig EA (1985) The heat shock response. CRC Crit Rev Biochem 18: 239-280.

420. Mogk A, Tomoyasu T, Goloubinoff P, Rudiger S, Roder D, et al. (1999) Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. EMBO J 18: 6934-6949.

421. Braig K, Simon M, Furuya F, Hainfeld JF, Horwich AL (1993) A polypeptide bound by the chaperonin groEL is localized within a central cavity. Proceedings of the National Academy of Sciences of the United States of America 90: 3978-3982.

422. Lin Z, Schwarz FP, Eisenstein E (1995) The hydrophobic nature of GroEL-substrate binding. Journal of Biological Chemistry 270: 1011-1014.

423. Weaver J, Watts T, Li P, Rye HS (2014) Structural basis of substrate selectivity of E. coli prolidase. PLoS One 9: e111531.

424. Weaver J, Rye HS (2014) The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. J Biol Chem 289: 23219-23232.

425. Puchalla J, Krantz K, Austin R, Rye H (2008) Burst analysis spectroscopy: a versatile single-particle approach for studying distributions of protein aggregates and fluorescent assemblies. Proc Natl Acad Sci U S A 105: 14400-14405.

426. Sengupta P, Garai K, Balaji J, Periasamy N, Maiti S (2003) Measuring Size Distribution in Highly Heterogeneous Systems with Fluorescence Correlation Spectroscopy. Biophysical Journal 84: 1977-1984.

427. Pal N, Dev Verma S, Singh MK, Sen S (2011) Fluorescence correlation spectroscopy: an efficient tool for measuring size, size-distribution and polydispersity of microemulsion droplets in solution. Anal Chem 83: 7736-7744.

428. Poso D, Clarke AR, Burston SG (2004) A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. J Mol Biol 338: 969-977.

429. Kreuzer KN, Jongeneel CV (1983) Escherichia coli phage T4 topoisomerase. Methods Enzymol 100: 144-160.

430. Wu PG, Brand L (1994) Resonance Energy Transfer: Methods and Applications. Analytical Biochemistry 218: 1-13.

431. Vendruscolo M, Paci E, Karplus M, Dobson CM (2003) Structures and relative free energies of partially folded states of proteins. Proc Natl Acad Sci U S A 100: 14817-14821.

432. Bollen YJ, Sanchez IE, van Mierlo CP (2004) Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. Biochemistry 43: 10475-10489.

433. Voelz VA, Jager M, Yao S, Chen Y, Zhu L, et al. (2012) Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. J Am Chem Soc 134: 12565-12577.

434. Doyle SM, Hoskins JR, Wickner S (2007) Collaboration between the ClpB AAA+ remodeling protein and the DnaK chaperone system. Proc Natl Acad Sci U S A 104: 11138-11144.

435. Rosenzweig R, Moradi S, Zarrine-Afsar A, Glover JR, Kay LE (2013) Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. Science 339: 1080-1083.

436. Doyle SM, Shastry S, Kravats AN, Shih YH, Miot M, et al. (2015) Interplay between E. coli DnaK, ClpB and GrpE during Protein Disaggregation. J Mol Biol 427: 312-327.

437. Ellis RJ, Hartl FU (1996) Protein folding in the cell: competing models of chaperonin function. FASEB J 10: 20-26.

438. Foury F, Roganti T, Lecrenier N, Purnelle B (1998) The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae. FEBS Letters 440: 325-331.

439. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 390: 249-256.

440. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res 6: 283-290.

441. Naik S, Haque I, Degner N, Kornilayev B, Bomhoff G, et al. (2010) Identifying protein stabilizing ligands using GroEL. Biopolymers 93: 237-251.

442. Katayama H, Janowiak BE, Brzozowski M, Juryck J, Falke S, et al. (2008) GroEL as a molecular scaffold for structural analysis of the anthrax toxin pore. Nat Struct Mol Biol 15: 754-760.