Dissertation zur Erlangung des Doktorgrades der Fakultät für

Chemie und Pharmazie der Ludwig-Maximilians-Universität

München

Catalysis of protein folding by the chaperonin GroEL/ES

Amit Kumar Singh

aus

New Rajnagar, M.P., India

2022

Acknowledgement

I would like to express the deepest gratitude to my supervisor, Prof. Dr. Franz Ulrich Hartl, who provided me an opportunity to work on this project and constantly supported my work. I learned a lot during the years of my PhD and grew on both personal and professional level.

I am very thankful to Dr. Manajit Hayer-Hartl, Dr. David Balchin and Dr. Andreas Bracher for their valuable advice and support. Great thanks to Dr. David who taught me the H/DX-MS technique and other members of the chaperonin office whose valuable input and fruitful discussions were so helpful to complete this project. I am thankful to all members in the department who worked with me during my PhD: Javaid, Goran, Andy, Rahmi, Liang, Chandhuru, Lucas, Itika, Prasad, Gopal. I would like to acknowledge the technical staff in my department: Albert, Romy, Nadine, Emmanuel, Darija, Silvia for their help and support.

I also want to express gratitude to members of the examination board for critical evaluation of my PhD thesis.

And finally, my deepest thanks to my friends and family who always encouraged me and helped me at every step of my life.

Contents

List of abbreviations

ADP	Adenosine 5'-diphosphate
ATP	Adenosine 5'-triphosphate
Bis-ANS	4,4'-Dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt
CDTA	*trans* -1,2-Diaminocyclohexane-N, N, N', N'-tetra acetic acid
Cys	Cysteine
D	Diffusion coefficient
DapA	Dihydrodipicolinate synthase
dcFCCS	Dual color fluorescence cross-correlation spectroscopy
DM-MBP	Double mutant maltose binding protein
DNA	Deoxyribonucleic acid
DOL	Degree of labeling
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FCS	Fluorescence correlation spectroscopy
FRET	Forster resonance energy transfer
GuHCl	Guanidine hydrochloride
HDX	Hydrogen/deuterium exchange
HEPES	2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid
Hsp	Heat shock protein
IPTG	Isopropyl β-D-thiogalactopyranoside
LB	Lysogeny broth

LC-IMMS	Liquid chromatography-ion mobility mass spectrometry
M-532	Alexa532
M-647	Alexa647
MetF	5,10 - methylenetetrahydrofolate reductase
NADH	Nicotinamide adenine dinucleotide
PCR	Polymerase chain reaction
PET	Photoinduced electron transfer
PMSF	Phenylmethylsulfonyl fluoride
RuBisCO	Ribulose 1,5-biphosphate carboxylase/oxygenase
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
smFRET	Single molecule FRET
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TF	Trigger factor
Tris-Cl	2-Amino-2-hydroxymethyl-propane-1,3-diol; hydrochloride
Trp	Tryptophan

1 Summary

The chaperonins (Hsp60s) are essential components of the cellular machinery in all three domains of life. They are double-ring complexes with ATPase activity that function by enclosing a single molecule of non-native protein into a nano-cage for folding to occur unimpaired by aggregation. The GroEL/ES chaperonin system has been studied intensively for more than two decades, but it remains unclear and controversial whether its cavity functions solely as a passive cage, preventing off-pathway aggregation, or actively enhances folding kinetics of some proteins by smoothing their folding energy landscape. Previous studies using heterologous protein substrates showed that folding requires multiple ATP-dependent encapsulation cycles by GroEL/ES with only a few percent of protein folded per cycle. These observations suggested that the GroEL/ES complex is an inefficient ATP-consuming machinery. Here, we analyzed the spontaneous and chaperonin assisted folding of the *E. coli* enzyme 5,10-methylenetetrahydrofolate reductase (MetF), an obligate GroEL/ES substrate, in order to elucidate the mechanism of GroEL/ES action. We found that MetF, a homotetramer of 33-kDa subunits with $(\beta/\alpha)_8$ TIM-barrel fold, populates a kinetically trapped folding intermediate(s) (MetF-I) during spontaneous folding that fails to convert to the native state, even in the absence of aggregation. However, GroEL/ES recognizes MetF-I and folds it rapidly, with more than 50% of MetF reaching the native (FAD binding) state in a single round of encapsulation in the GroEL/ES cage. Moreover, analysis by hydrogen/deuterium exchange (H/DX) and mass spectrometry (MS) at peptide resolution showed that the MetF monomer folds in the GroEL/ES nano-cage reaching a near-native like state that binds the cofactor flavin adenine dinucleotide. Rapid folding required the net negative charges of the wall of the GroEL/ES cavity, as shown using the GroEL mutant EL(KKK2) in which the net negative charge is removed. These findings uncover a prominent capacity of the GroEL/ES system to catalyse, by global encapsulation, the folding of an endogenous protein that would have coevolved with the chaperonin system.

2 Introduction

2.1 Overview of protein structure

Proteins are involved in almost every biological process. They are synthesized on ribosomes as linear chains of amino acids in a specific sequence determined by the genetic information. Proteins are linear polymers of amino acids, with each amino acid residue covalently joined through a peptide bond between the α-carboxyl group of one amino acid and the α-amino group of the next. In the late 1930s, Linus Pauling and Robert Corey concluded that the peptide C-N bonds, because of their partial double-bond character that keeps the entire six atom peptide group in a rigid planar configuration, cannot rotate freely. Rotation about the N-C$_\alpha$ and C$_\alpha$-C bonds define two torsion angles, denoted ф and ψ, respectively (Fig. 1). The values of ф (phi) and ψ (psi) for each residue in the polypeptide chain suffice to completely specify the three-dimensional structure of the peptide backbone. The conformational space accessible to a polypeptide chain is described by the Ramachandran plot, a plot of the allowed values of the torsional angles- ф (phi) and ψ (psi) of the main chain of the polypeptide. The map applies to all amino acid residues except glycine. Glycine is the only amino acid residue often found in a conformation outside these regions. Four levels of protein structure are known. The primary structure is the sequence of amino acid residues. The local conformation of the polypeptide chain is referred to as the secondary structure which is mainly stabilized by hydrogen bonding between the amide and carbonyl groups of the main chain. The most common secondary structures are the *α* helix, β conformation and β turns. Tertiary structure is the complete three-dimensional structure of a polypeptide chain (fibrous and globular). When a protein has two or more polypeptide chains, their structural patterns in space is referred to as quaternary structure. A structural motif, also called a fold, is a distinct structural pattern containing two or more elements of secondary structure and the connection(s) between them. The protein folds into a particular three-dimensional shape, determined by its amino acid sequence and stabilized primarily by noncovalent interactions. All polypeptide chains are synthesized by the ribosome in a vectorial manner from the N-terminus towards the C-terminus but exactly when and how the newly synthesized polypeptide chain folds into its native three-dimensional structure remains one of the fundamental questions in biochemistry.
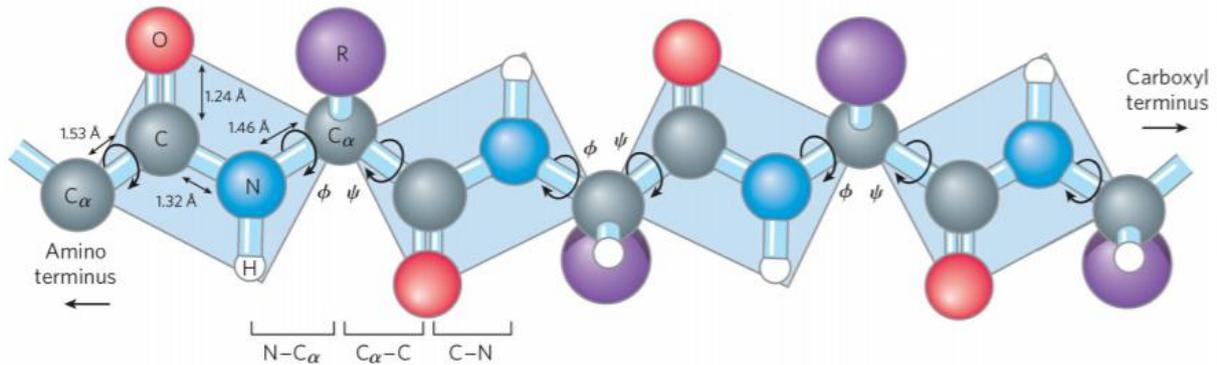
Figure 1: A short polypeptide showing the dihedral angles.

The N—C$_\alpha$ and C$_\alpha$—C bonds can rotate, defined by the dihedral angles ϕ and ψ, respectively. Because of its partial double-bond character, the peptide C—N bonds cannot rotate freely (figure adapted from Nelson, D.L. *et al.*, Lehninger Principles of Biochemistry, 2008).

2.2 How do proteins fold?

Protein folding is one of the most complex problems in protein science which includes the prediction of the three-dimensional structure of a protein from its amino acid sequence and the process by which the unfolded polypeptide chain attains its native conformational state? A major breakthrough in protein science was provided by Anfinsen's pioneering experiments, which showed that ribonuclease-A folds reversibly *in vitro* on removal of denaturant. This finding demonstrated that the amino acid sequence of a protein determines its three-dimensional native structure and that the native state of a protein is a thermodynamically stable state. Thus, the native conformation has a lower free energy than the unfolded states, making folding energetically favorable [1]. Proteins are synthesized on ribosomes as long linear polymers of amino acids and must fold into a unique three-dimensional structure to execute their cellular functions. Understanding how this process is achieved has fascinated researchers for decades. Protein folding could not be a random, trial and error process. The sheer number of possible conformations around every single bond in its backbone prohibits folding to occur by sampling all of them. This would take an astronomical length of time and this problem, as first pointed out by Cyrus Levinthal in 1968 (also called Levinthal's paradox). Accordingly, Levinthal suggested that proteins must fold by specific folding pathways. Hence, protein folding must occur by some directed process, encoded by the primary structure.

Introduction

Several plausible models were proposed to explain how the polypeptide chain may arrive at its native conformation without having to sample a large number of conformers:

(A) The diffusion collision model (framework model) suggests that secondary structural elements form first and undergo collisions with other secondary structure elements to form the native structure [2-4] (Fig. 2). (B) The nucleation condensation model posits that a general compaction of the protein chain occurs first, leading to the nucleation of the native secondary and tertiary structure [3, 4] (Fig. 2). (C) The nucleation propagation model states that local interactions generate a small fraction of secondary structure, which functions as a nucleus for the propagation of further native structure [3] (Fig. 2). (D) According to the hydrophobic collapse model, protein folding is initiated by the formation of a collapsed intermediate or molten globule state in which hydrophobic amino acid side chains are clustered in a protein's interior, away from water. Subsequently, the fully folded native state develops by searching within this compact state [2, 3, 5] (Fig. 2).
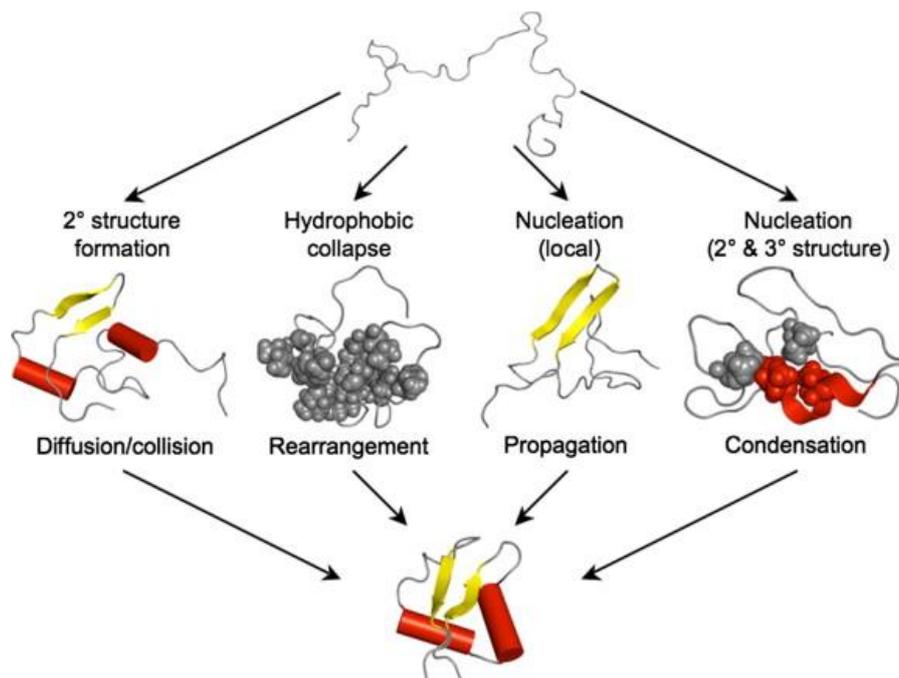
Figure 2: Classical models of protein folding.

The framework model suggests that local elements of secondary structure form first. These then dock into the correct tertiary structure. In the hydrophobic collapse model, a protein collapses rapidly around its hydrophobic side-chains, forming a molten globule state, and the native state develops from this conformationally restricted intermediate state. The nucleation propagation model implies that local interactions develop a small portion of secondary structure, which acts as a nucleus for the generation of native structure. The nucleation-condensation model postulates that protein folding is

4

Introduction

initiated by the presence of a weakly-structured nucleus that acts as a template from which the native structure develops through fast condensation of further structure around it. Figure adapted from Ref. [3].

(E) Folding funnel model – Because a protein chain can adopt several possible conformations, protein folding reactions are highly complex and heterogeneous, relying on the cooperation of many weak, noncovalent interactions [6, 7]. Among these, the hydrophobic effect is the major driving force in chain collapse and the protein's interior core is densely packed with hydrophobic amino acid side chains, minimizing the conformational search during folding [6, 7]. After decades of research, it is now clear that there is not a single, sequential folding route, as was implicit in some early models; instead, the process of protein folding can be described as a kind of free energy funnel [6, 8-10] (Fig. 3). The unfolded proteins are structurally heterogeneous, representing the starting point of folding, depicted by a high degree of conformational entropy and relatively high free energy [2, 9] (Fig. 3). As folding proceeds along several downhill routes, the funnel reduces the number of states present (decreases entropy), reaching the native conformation that represents the minimum of free energy. Native proteins are thermodynamically stable under physiological conditions. Free-energy funnels can have a rugged free-energy surface. These small depressions in the energy landscape represent kinetically stable intermediates and misfolded states that can slow the folding and often require molecules to cross substantial kinetic barriers [2, 9]. This problem is less pronounced for small proteins which often fold rapidly without populating stable intermediates [11]. Folding intermediates may have a high degree of configurational flexibility, increasing the search time required for productive folding. The propensity of proteins to populate such kinetically trapped intermediates increases with larger, more complex α/β and α+β domain topologies that are stabilized by many long-range interactions. Such proteins are mostly chaperone dependent for folding [12]. Misfolded states are characterized by the presence of non-native interactions that typically expose hydrophobic residues and regions of flexible polypeptide backbone to the solvent, the features that promote aggregation in a concentration-dependent manner [7, 9, 13]. Although aggregates are primarily amorphous structures mostly driven by hydrophobic interactions, a subset of proteins can aggregate into amyloid fibrils. Amyloids are characterized by cross β-structure, where β-strands run perpendicular to the fibril axis. These structures are thermodynamically highly stable state and are associated with many diseases, especially neurodegenerative disorders [9, 10, 13].

5
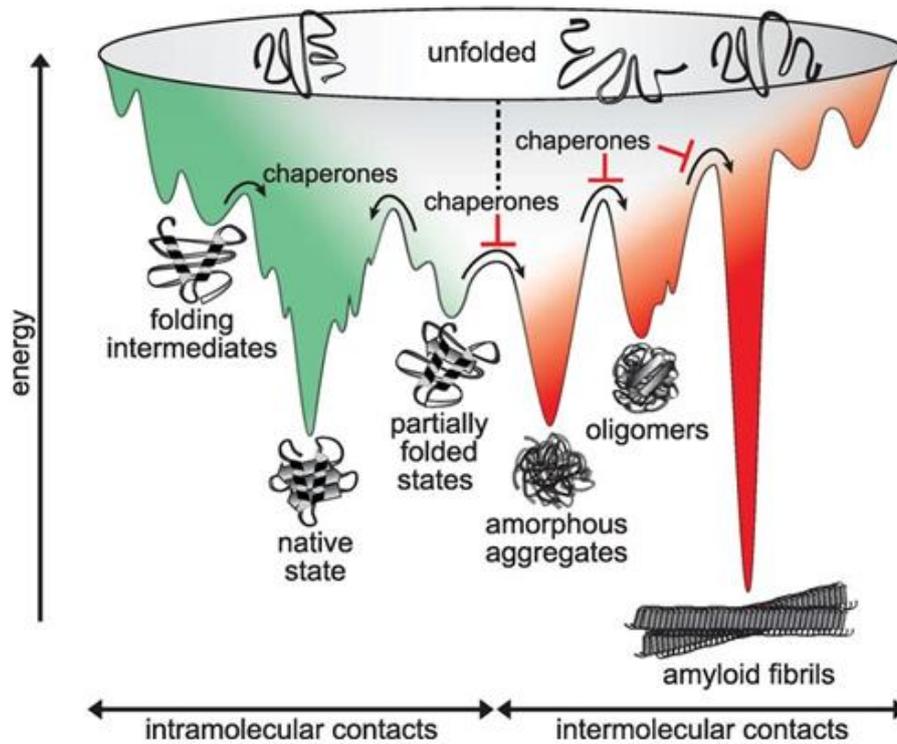
Figure 3: A schematic energy landscape diagram of protein folding.

Unfolded polypeptide chains at the top of the funnel explore multiple conformations while moving toward the thermodynamically favorable native state (green), some of which involve kinetically trapped on or off pathway intermediates with low energy (folding intermediates and partially folded states). Molecular chaperones assist in protein folding by reducing free-energy barriers and preventing aberrant intermolecular contacts (red), which can lead to various forms of aggregates (amorphous, oligomeric, fibrillar). Amyloid fibrils represent thermodynamically the most stable state. Figure adapted from Ref. [9].

2.3 Hydrogen exchange (HX) mass spectrometry

Hydrogen/deuterium exchange (HDX) coupled to mass spectrometry (MS) is a technique for the study of protein conformational dynamics [14]. This technique is made possible because backbone amide hydrogens in the protein can exchange with deuterium atoms when the protein is incubated in a D_2O solution. The rate of this exchange is dependent on four factors- pH, temperature, hydrogen bonding and solvent accessibility. Because deuterium has a mass of 2.0141 Da and hydrogen a mass of 1.0078 Da, the subsequent increase in protein mass over time can be measured with a mass spectrometer. The location of the deuterium uptake is determined by monitoring deuterium incorporation in digested peptide fragments [15]. Every amino acid except proline has a backbone amide hydrogen in the polypeptide sequence [16](Fig. 4A). Regions that are highly dynamic and solvent-exposed will exchange quickly while those parts of the protein that are less dynamic and/or involved in hydrogen bonding or buried within the core of the protein (such as β-sheets or α-helices) will exchange more slowly [15, 17](Fig. 4B).

(A)

Gln—Asp—His—Pro—Lys—Leu

(B)

○ hydrogen
● deuterium

after 1 minute
Dynamic and exposed amide hydrogens exchange rapidly

after 2 hours
Protected amide hydrogens exchange slower
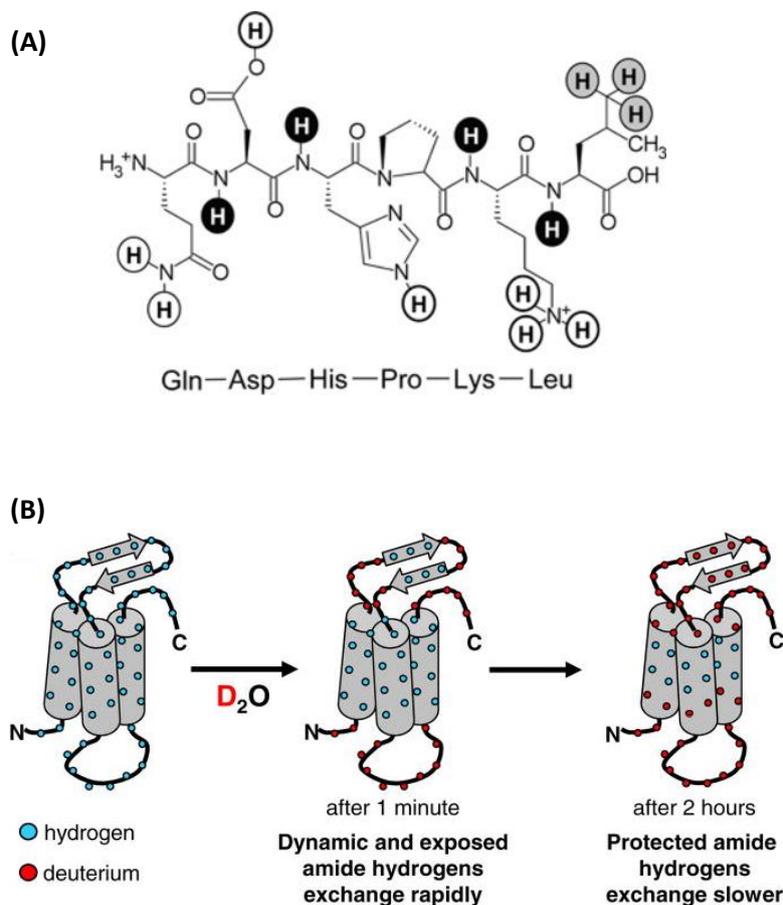
Introduction

Figure 4: Overview of hydrogen exchange in proteins.

(A) Different types of hydrogens in proteins. The six-residue peptide Gln-Asp-His-Pro-Lys-Leu represents backbone amide hydrogens (black), nonexchangeable hydrogens bonded to carbon (gray), and exchangeable hydrogens which cannot be measured with HX-MS (white). Figure adapted from Ref. [16]. (B) Once the protein is placed in a D_2O solution, solvent-exposed and dynamic regions (like the loops connecting the α-helices) will exchange rapidly while regions that are protected or buried will exchange more slowly. Figure adapted from Ref. [17].

There are two types of labeling methods: continuous and pulse labeling. Continuous labeling experiments are usually performed on the native state of proteins and provide information of protein structural dynamics under steady-state conditions. A typical scheme for a continuous labeling experiment measured with mass spectrometry is shown in Figure 5. Protein samples are incubated at the desired temperature, pH and in all H_2O buffer. Protein solutions are then diluted 10 to 20-fold into the identical buffer containing D_2O for labelling. The exchange reaction is allowed to proceed for various time intervals and is quenched by lowering the pH and temperature (pH 2.5 and 0°C). The labelled protein can then be analysed with or without digestion with pepsin (pH 2.5) by injection into a liquid chromatography system coupled to a mass spectrometer. After analysis of the mass spectra, the location of deuteration can be determined over time and plotted, either for the intact protein, or for each of the peptic fragments [14, 17].
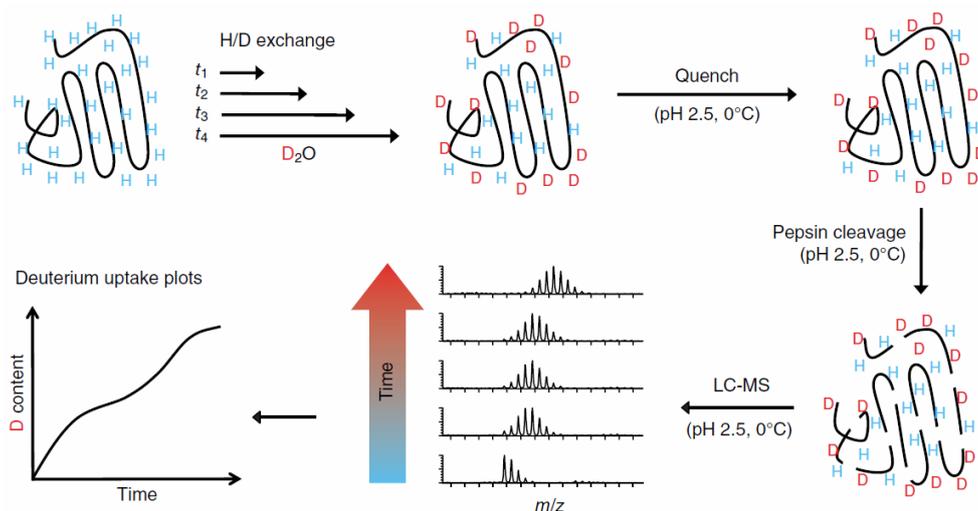
Figure 5: A general workflow of continuous labeling HDX-MS experiment.

For labeling, equilibrated protein solutions are diluted (typically 10-fold or more) into a D_2O-containing buffer. The exchange reaction proceeds for various time and is quenched by reducing the pH to 2.5 and the temperature to 0°C. Global HDX (protein level) can then be measured either directly by liquid

chromatography (LC) and mass spectrometry (MS) analysis of the intact protein, or the local HDX (peptide level) can be measured by pepsin cleavage and subsequent LC-MS analysis of the digested peptides. The mass spectra are analyzed and the deuterium uptake over time plotted, either for the intact protein or for each peptide (figure adapted from David D. Weis, Hydrogen Exchange Mass Spectrometry of Proteins).

In pulse labeling experiments, the protein is forced to undergo a conformational change by addition of a perturbant. Perturbants are often a chaotropic agent (urea, guanidine hydrochloride), although changes in pH, temperature or binding to substrates can also be used. The protein sample is then exposed to deuterium for a very brief period of time, typically 10 sec [14]. Pulse-labeling experiments have been used to study protein folding mechanisms as well as to identify intermediate states in a folding reaction [18]. If the refolding rate (k_{-1}) of the unfolded protein is slower than the deuterium labeling rate, the unfolded regions of a protein can easily deuterate and show a higher mass than the folded species. The resulting deuterium uptakes then reveal the population of folded and unfolded species [17](Fig. 6).
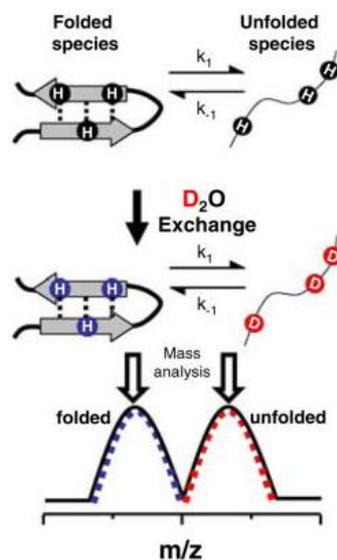
Figure 6: Using HDX-MS to distinguish different populations of protein in solution.

Two populations in the mass spectra are showing folded (blue distribution) and unfolded states (red distribution). These two distributions occur when the rate of interconversion of the folded and unfolded state is slower than the amide exchange rate. Figure adapted from Ref. [17].

2.4 Single molecule spectroscopy in protein folding

Single molecule techniques have a great advantage for probing protein structure and dynamics, especially in the analysis of structurally heterogeneous populations. Besides providing ensemble-averaged values, energetic and kinetic information as well as information on hidden intermediates can be determined in equilibrium. Traditional biochemical experiments involve billions of molecules that gives the average value with a high signal to noise ratio, but lack information due to ensemble averaging [19]. Signal averaging can be suppressed by analysis of individual molecules one at a time. Thus, the single molecule methods are very useful in the study of isolated proteins in real time, particularly revealing the conformation and dynamics of chaperones and their impact on the substrates during the folding process. Single molecule experiments could be performed by diluting the protein sample to a concentration of about 10-100 pM. In this concentration range, aggregation can be excluded completely [20]. Two types of single-molecule techniques are currently applied to study aggregation and protein folding at the single molecule level: force spectroscopy - atomic force microscopy (AFM) or optical tweezers and fluorescence methods [19, 21].

Here, we focus on single molecule fluorescence spectroscopy; most notably, single molecule Forster resonance energy transfer (smFRET), fluorescence correlation and cross-correlation spectroscopy (FCS and FCCS) to analyze the aggregation as well as protein folding. These methods require labelling with fluorescent dyes at specific positions in the protein or chaperone. FRET is based on the radiation-less energy transfer between a donor and an acceptor dye when they are in close proximity (2nm to 10nm) (Fig. 7). FRET efficiency varies as the sixth power of the distance between the donor and acceptor dyes. Single-molecule FRET is a very sensitive method to probe intramolecular distance distributions and conformational dynamics of single proteins. Also, this method has been used to investigate unfolded proteins and intrinsically disordered proteins [21]. smFRET could be performed either in solution, where single molecules freely diffuse through the confocal volume, or the proteins of interest are immobilized on a glass surface and analyzed by TIRF microscopy [19].

Fluorescence correlation spectroscopy (FCS) measurements are based on the analysis of time-dependent intensity fluctuations in the fluorescence intensity signal primarily by fluorescent molecules diffusing through a femtoliter detection volume of a confocal microscope (Fig. 7). Correlation analysis of the observed fluctuating fluorescence signal from excited molecules

Introduction

within the focal volume provide the information about diffusion coefficients, local concentration, chemical rate constants, structural dynamics and several other molecular processes [22, 23]. In fluorescence cross-correlation spectroscopy (dual-color FCCS), an extension of FCS that measures the interaction of two differently labeled binding partners, the fluorescence signals of two molecules are cross-correlated with each other from two spectral channels (Fig. 7). A cross-correlation curve is only formed if the two species interact and co-diffuse through the FCCS detection volume. FCCS can be used to calculate rate constants of molecular interactions and mobility of the bound complex as well as the concentrations of single and double-labeled molecules simultaneously [24] .
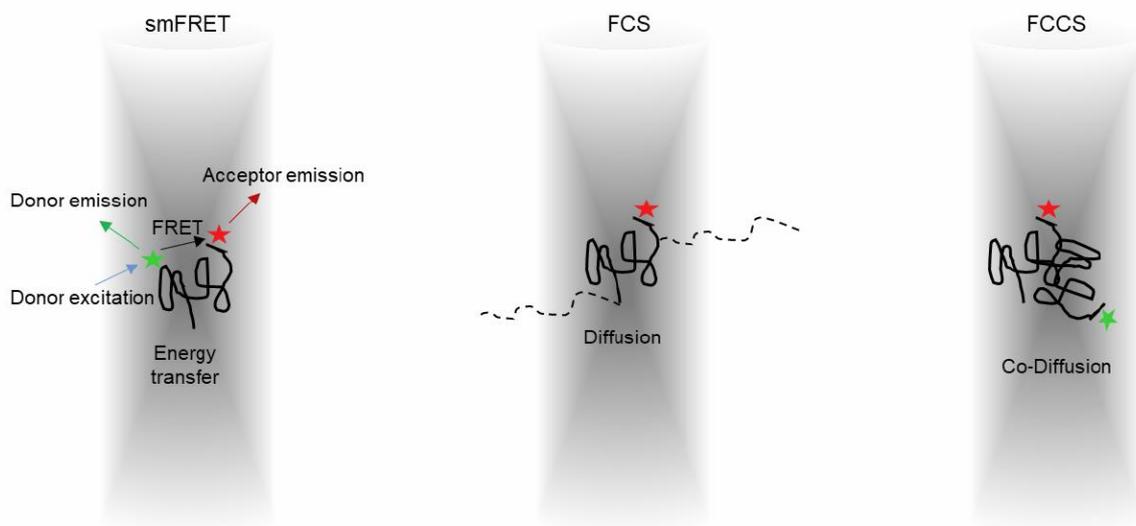
Figure 7: Single molecule fluorescence methods.

Microscopy methods utilized for observation of single-molecule experiments. In an smFRET experiment, protein molecules are labeled with a donor and an acceptor dye at two different positions. Upon excitation of the donor, energy is transferred to the acceptor. FCS measurements are executed on fluorescent molecules diffusing into or out of the detection volume of a confocal microscope, giving rise to intensity fluctuations over time. For FCCS, two different dyes are employed to label the interacting molecules to elucidate the molecular interactions. If they are bound to each other and co-diffuse through the focal volume, they show a positive cross-correlation read out (reproduced from Rahmi's thesis).

2.5 Molecular chaperones for protein folding

Anfinsen's pioneering work in the 1950s provided the first evidence that *in vitro* a small protein can fold spontaneously without the help of additional factors [1]. However, many proteins do not fold spontaneously when they are synthesized in the cell. Research over the last two decades has firmly established the indispensable role of molecular chaperones in the cellular environment. These proteins are required to allow newly synthesized proteins to fold efficiently and at a biologically relevant time scale [25, 26]. To what extent and how chaperones alter the energy landscapes of folding remains to be established. A molecular chaperone can be defined as any protein that interacts with, stabilizes or assists another protein to reach its native, biologically active conformation, without being part of its final structure [27, 28]. Most small proteins (< 100 amino acids) fold rapidly at a millisecond to second time scale *in vitro* [11, 29]. Larger, multidomain proteins and proteins with complex fold topologies tend to fold slowly populating long-lived, partially folded intermediate structures which are prone to misfolding and aggregation [30-32]. The folding of such proteins *in vivo* becomes considerably more challenging as the interior of the cell is characterized by very high macromolecular concentrations (~200 to 300 mg/mL of total protein in *E. coli*). The resultant excluded volume effects (macromolecular crowding) strongly enhance the propensity of folding intermediates and misfolded states to aggregate [33]. Consequently, the fundamental function of molecular chaperones is to prevent the aberrant intermolecular interaction that may lead to aggregation [7, 9, 28] (Fig. 8).
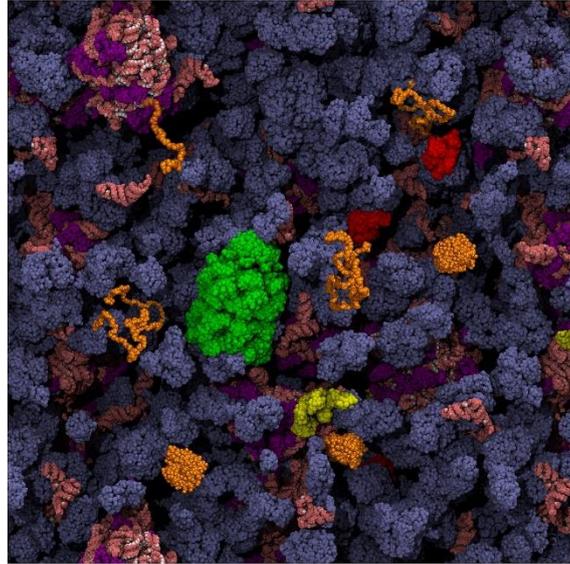
Figure 8: Schematic representation of the crowded interior of an *E. coli* cell in which folding of newly synthesized protein proceeds.

Features that influence the folding of a protein of interest (in orange) are: the resulting excluded volume effect due to macromolecular crowding, the presence of various chaperones that interact with nascent chains and misfolded proteins (GroEL in green, DnaK in red, and trigger factor in yellow) and the probability of co-translational folding on the ribosome (ribosomal proteins are purple; all RNA is salmon). Figure adapted from Ref. [34].

2.6 Chaperone families in the cytosol

Several evolutionary conserved families of molecular chaperones exist in all organisms and in all cellular compartments. They prevent non-native (off-pathway) interactions between not-yet folded proteins that lead to aggregation and direct their substrates along productive folding, transport or degradation pathways. Their members are also referred to as stress proteins or heat shock proteins (Hsps), because their expression is upregulated in stress conditions (e.g., heat shock or oxidative stress), which structurally destabilize a subset of cellular proteins. The major chaperone families are classified according to their molecular weight: Hsp40s, Hsp60s, Hsp70s, Hsp90s, Hsp100s, and the small Hsps. Chaperone networks in the cytosol are highly conserved throughout evolution. Protein folding can start co-translationally when the nascent peptide is still bound to the ribosome and is completed post-translationally upon chain release from the ribosome. In all domains of life (bacteria, archaea and eukarya), ribosome-associated chaperone systems [Trigger factor (TF) in bacteria; ribosome-associated complex (RAC) and nascent chain–associated complex (NAC) in eukarya] interact first with the nascent polypeptide, followed by ATP-dependent chaperones acting downstream of ribosome-binding chaperones. A fraction of cytosolic proteins that are unable to fold with the ribosome associated chaperones (≥30% of the proteome) require further assistance by the Hsp70 system (DnaK/DnaJ in bacteria; Hsp70/Hsp40 in eukarya) or require transfer to the chaperonin (GroEL/ES in bacteria; TRiC in eukarya) or the Hsp90 system (HtpG in bacteria) [9] (Fig. 9). In bacteria and eukaryotic cells, Hsp70s do not bind directly to the ribosome and cooperate in co-or post-translational folding, providing connections to the downstream chaperones, chaperonins and Hsp90s [35]. In eukaryotic cells, the co-chaperone Hop (Hsp70/90 organizing protein) facilitates physical interaction and substrate transfer by acting as an adaptor between Hsp70 and Hsp90, although such co-chaperones have not been reported for *E. coli* Hsp90 (HtpG) [36] (Fig. 9B). Eukaryotic Hsc70 forms a stable complex with the chaperonin TRiC/CCT and may transfer the unfolded substrate from Hsc70 to the chaperonin TRiC/CCT [37]. In *E. coli,* the DnaK/DnaJ system maintains a subset of proteins in a folding competent state and transfers them to the chaperonin GroEL/ES for subsequent folding to the native state [12, 38, 39] (Fig. 9A). Besides the Hsp70s, the ~90 kDa oligomeric prefoldin (Pfd) complex, consisting of two α and four β subunits, may bind to nascent chains and transfer them to the chaperonins for folding in the archaeal and eukaryotic cytosol [40]

Introduction

(Fig. 9B). The ATP-independent small heat shock proteins (sHsps) exhibit holdase activity, that is they bind and sequester early unfolding intermediates of substrates and facilitate their refolding by ATP-dependent Hsp70 and Hsp100 chaperones [41]. In addition to these general chaperone systems, specific assembly chaperones are also present that are required for assembly of folded protein subunits [42].
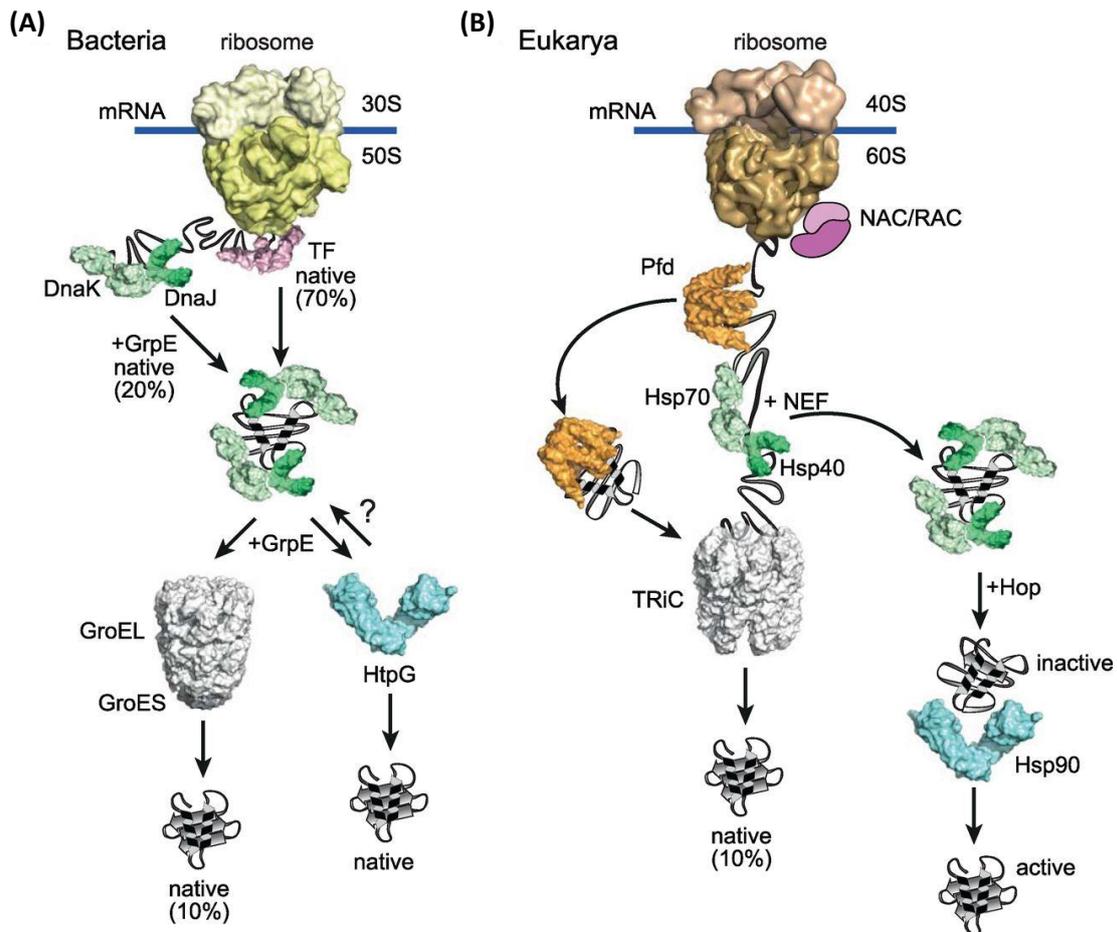
Figure 9: Chaperone networks involved in *de novo* protein folding.

For ~70% of bacterial (A) and eukaryotic (B) proteins, the ribosome and associated chaperones [trigger factor (TF) in bacteria; nascent chain–associated complex (NAC) and ribosome associated complex (RAC) in eukaryotes] assist in co-translational protein folding. Hsp70s (DnaK in bacteria) collaborate with Hsp40s (DnaJ in bacteria) and nucleotide exchange factors (NEFs; GrpE in bacteria), mediate co- and posttranslational folding (~20% of total proteome). About ~10% of the proteome is transferred to the chaperonins (GroEL/ES in bacteria and TRiC in eukaryotes) for folding. In eukaryotes, prefoldin (Pfd) recognizes specific nascent chains and transfers them directly to TRiC. This alternative pathway mainly presents in archaea, as only some archaeal species contain the Hsp70 system. Hsp90 is another important chaperone system that receives its clients via the Hsp70 system and co-chaperone Hop for completion of folding and conformational regulation in eukaryotes. The bacterial Hsp90, HtpG, is thought to participate in protein folding without known co-chaperones. Figure adapted from Ref. [9].

Introduction

The dependence of a large fraction of proteins on molecular chaperones for their initial folding is well established. A major function of the chaperone family includes aggregation prevention, refolding of misfolded states, and actively dissociating certain protein aggregates. Clearance of terminally misfolded proteins is mainly performed by the ubiquitin proteasome system (UPS) and plays an important role in maintaining proteome integrity (Fig. 10). Larger protein aggregates resisting disassembly can be removed by autophagy and lysosomal degradation. Dysfunction of autophagy leads to inclusion body formation and neurodegeneration [7, 9, 43].
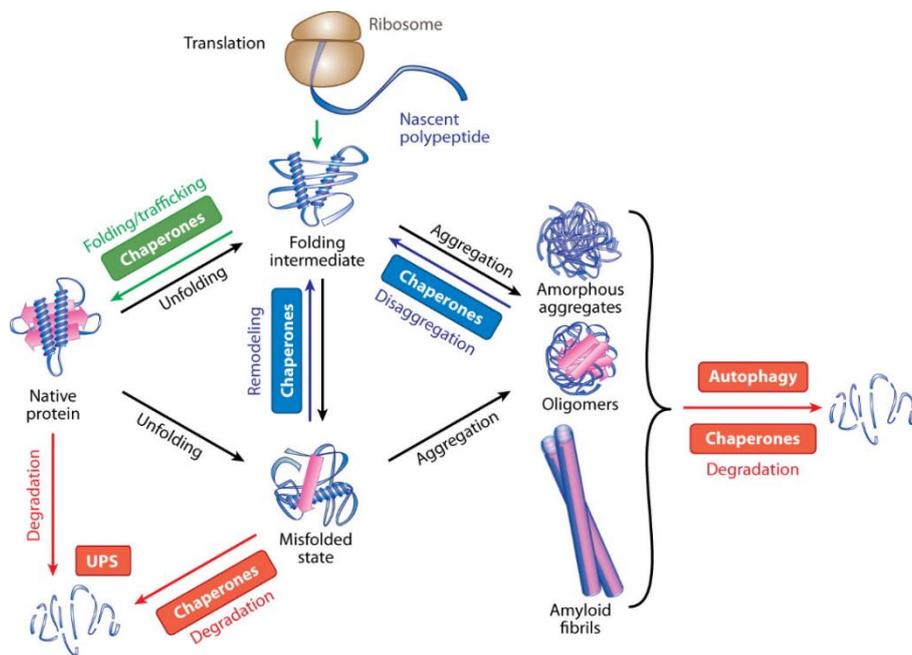
Figure 10: The proteostasis network.

The proteostasis network includes chaperone for the assisted folding of newly translated proteins, the remodeling of misfolded proteins, and the dissociation of certain aggregates. Chaperone cooperates with degradation machinery (UPS, autophagy and lysosomal degradation) in the removal of terminally misfolded proteins. Figure adapted from Ref. [7].

Cells acquire an extensive network of cellular components to maintain protein homeostasis (proteostasis). Operationally, this network could be classified into three branches: protein biogenesis, conformational maintenance and protein degradation (Fig. 11). Failure of proteostasis is associated with aging and neurodegenerative disease. Proteostasis is maintained by ~1400 different proteins involved in protein biogenesis (~400), conformational

maintenance (~300), and degradation (~700), with several proteins being part of more than one pathway in mammalian cells [7, 43] (Fig. 11).
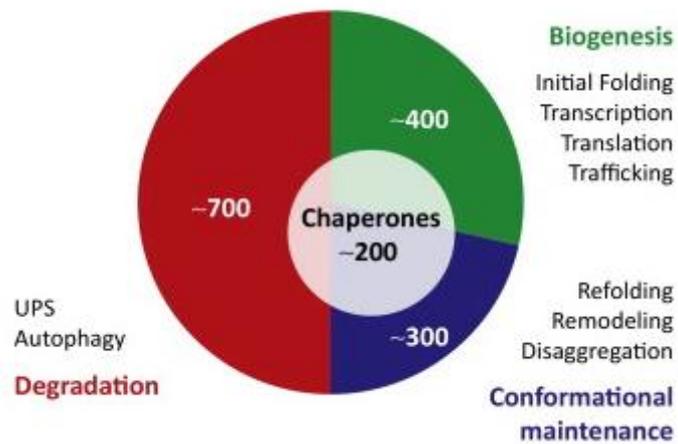
Figure 11: Chaperones play crucial roles in all three domains of the proteostasis network.

The approximate numbers of proteostasis network components in the human proteome are indicated. Figure adapted from Ref. [43].

2.7 Chaperone action at the ribosome

During protein synthesis on the ribosome, newly synthesized proteins are sensitive to misfolding and aggregation or degradation by proteases, because they emerge from the ribosome in an unfolded state. In bacteria, nascent chains interact first with the ribosome-associated chaperone trigger factor (TF) which facilitates their folding at the ribosome. TF is only present in bacteria and chloroplasts but is absent from the cytosol of archaea and eukaryotes. Although deletion of TF does not influence the viability of *E. coli* cells, combined loss of both TF and the Hsp70 chaperone DnaK leds to the build-up of protein aggregates, which causes cell death at growth temperatures above 30°C [35, 44]. *E. coli* TF (48 kDa) folds into a unique dragon shape, consisting of an N-terminal ribosome-binding (RB) domain, a peptidyl-prolyl-*cis/trans* isomerase (PPIase) domain and a C-terminal portion [44, 45] (Fig. 12 A, B, C). TF has been shown to enhance the folding yield of multidomain proteins by reducing the co-translational folding speed relative to translation [46]. TF has multiple hydrophobic contact sites throughout its interior which may bind to a nascent peptide in an unfolded state at the ribosome exit tunnel and protect it from aggregation by shielding its exposed hydrophobic regions (Fig. 12D). TF can bind small proteins up to a length of 130 amino acids [44]. A similar mechanism of folding has been observed for the ATP-independent periplasmic chaperone Spy, which allows the folding of model proteins while they remain associated with the chaperone [47]. In eukaryotes, two conserved ribosome-associated chaperone systems, RAC and NAC may have a similar role to TF in co-translational folding [9, 48]. These ribosomes associated factors are best characterized in *S. cerevisiae*. In yeast, the ribosome-associated complex (RAC) is a stable heterodimeric complex of the Hsp40 chaperone, Zuo (MPP11 in humans) and the Hsp70, Ssz (Hsp70L1 in humans) and functions together with a ribosome-bound Hsp70, Ssb [45, 49]. The RAC complex and Ssb form a functional chaperone triad. RAC acts as a co-chaperone and stimulates the ATPase activity of Ssb, thereby increasing the affinity of Ssb for unfolded polypeptides, eventually assisting *de novo* folding of the nascent polypeptide chain [45, 48-50]. NAC is a heterodimeric protein complex composed of α (αNAC) and β (βNAC) subunits, whereas in archaea NAC is formed by a homo-dimeric α-subunits [45, 48, 49]. Eukaryotic NAC interacts with nascent polypeptides as they emerge from ribosomes and prevents them from misfolding and aggregations; however, ribosome binding is solely

mediated by the β subunit [48, 50]. Compelling evidence supports a chaperone function of NAC at the ribosome but its precise role in folding is not fully understood.
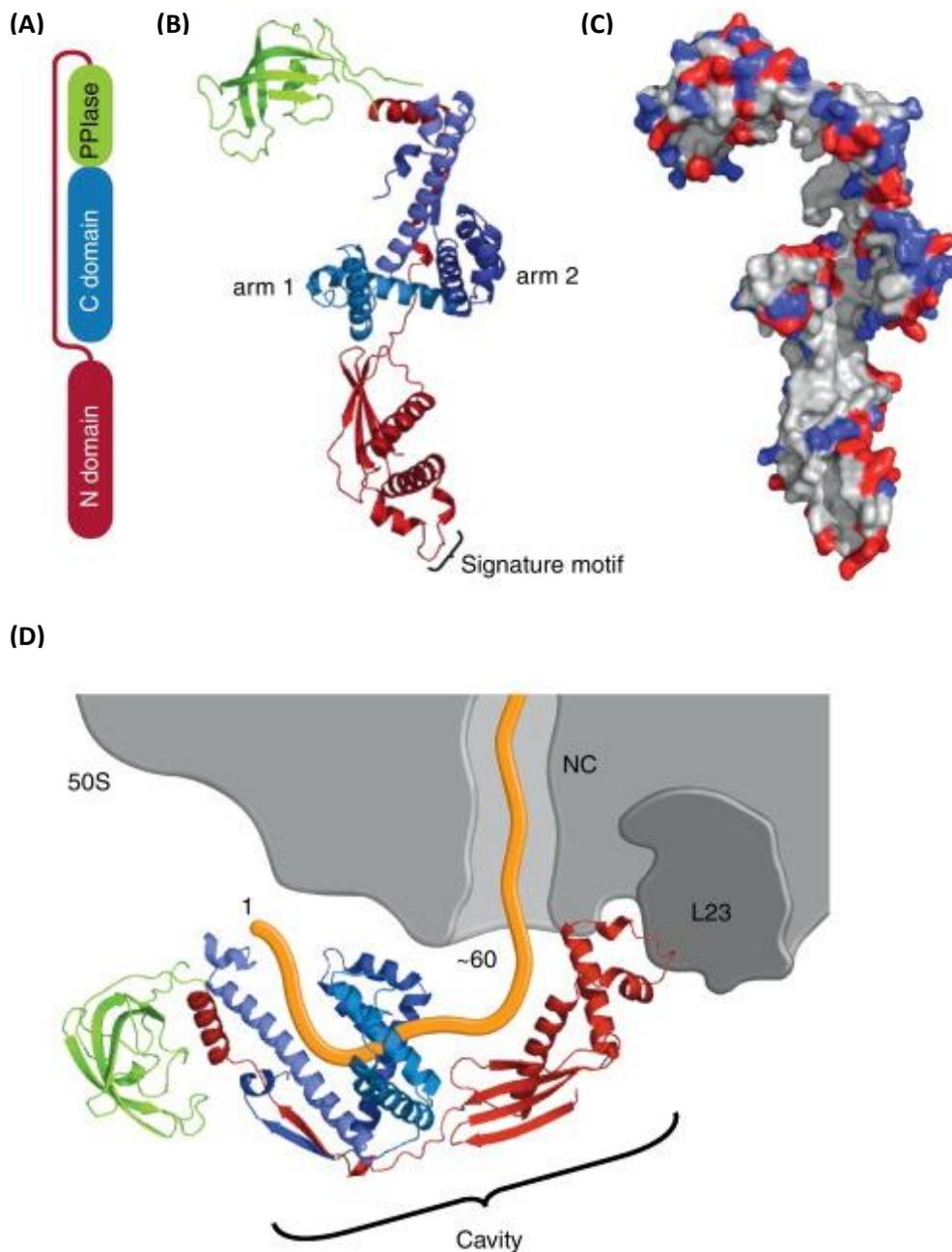
Figure 12. Structure of Trigger factor (TF) from *E. coli*.

(A) Three domains of TF: an N domain, a peptidylprolyl *cis/trans* isomerase (PPIase) domain and a C domain. (B) The crystal structure of TF (PDB 1W26). The N domain (red) contains a conserved signature motif (GFRxGxxP) in a loop region for the binding of ribosomal protein L23 and is connected to the PPIase domain (green) via an extended linker. The C domain (blue) is located in the middle of the molecule, together with the N domain forming a surface for nascent polypeptide chain binding. (C) Surface charge distribution of TF. Positively and negatively charged residues are shown in blue and red

respectively. (D) Structural model of TF bound to the large subunit of ribosome (grey). The main contact between TF and the ribosome involves the conserved signature motif in the N domain (red) of TF and the ribosomal protein L23 (dark grey) close to the exit tunnel. *In vivo*, TF prefers to bind to emerging nascent chains (orange) only once they have exposed at least ~ 60 amino acids outside the ribosomal tunnel. Figure adapted from Ref. [45].

2.8 The Hsp70 chaperone system

The 70-kDa heat shock proteins (Hsp70s) are ubiquitous chaperones that participate in a range of cellular activities, including the *de novo* folding of nascent chains, the translocation of proteins into organelles, the disassembly of protein complexes, regulation of protein activity and transfer of polypeptides to the more specialized downstream chaperones, Hsp90s and chaperonins. Moreover, Hsp70 chaperones also take part in stress-related activities to prevent aggregation and promote refolding and disaggregation. They also cooperate with cellular degradation machineries to clear protein aggregates [51].

Most of the work elucidating the allosteric mechanism of Hsp70s is based on a detailed characterization of the *E. coli* Hsp70, DnaK. Hsp70s consist of an N-terminal ~45-kDa nucleotide-binding domain (NBD) and a C-terminal ~30-kDa substrate-binding domain (SBD), connected by a conserved hydrophobic, flexible linker. The SBD is composed of a 15kDa substrate binding subdomain (SBDβ), a 10kDa α-helical lid subdomain (SBDα) and a disordered C- terminal region. SBDβ has a strong ability to bind 5-7 amino acid, segments that are enriched in hydrophobic residues and flanked by positively charged amino acids, generally exposed by proteins in non-native or misfolded states [7, 52] (Fig. 13A). The substrate peptide binds to DnaK in an extended conformation in the β sandwich domain mediated by van der Waals contacts from its side chains and hydrogen bonds [53]. The NBD has an actin-like fold and is composed of four subdomains (IA, IB, IIA and IIB), which are arranged into two lobes. Nucleotides bind at the deep crevice between subdomains IB and IIB and are coordinated by all four subdomains [7, 54] (Fig. 13A).

Hsp70 family members do not function alone but generally collaborate with co-chaperones, Hsp40s (J domain proteins; DnaJ in *E. coli*) and nucleotide exchange factors (NEFs; GrpE in *E. coli*) in an ATP dependent reaction cycle (Fig. 13B). Hsp70s exist in open and closed conformational states, depending on the nucleotide state (Fig. 13A). In the ATP-bound state of Hsp70s, the α-helical lid detaches from the SBDβ, and both SBD subdomains become tightly associated with the NBD. As a result, the SBD adopts an open conformation with high on/off

Introduction

rates and low affinity for the substrates. Hsp40s recruit non-native substrates to the open, ATP bound state of Hsp70 and strongly stimulate the hydrolysis of the bound ATP to ADP. In the ADP-bound state of DnaK, the NBD and SBD behave independently of each other, similar to the individual domains joined by a flexible linker. In this state, the SBD adopts a closed conformation with the α-helical lid locked onto the peptide bound to the SBDβ, resulting in a low on/off rate for substrates. Hsp70s exhibit high substrate-binding affinity in the ADP-bound state (closed state). Subsequent NEF (GrpE in bacteria; Bag, HspBP1 or Hsp110 in eukaryotes) binding to the NBD promotes ADP-ATP exchange and substrate release, with the resulting folding intermediate either progressing directly to the native state or being transferred to the downstream chaperone systems, chaperonins or Hsp90s, for folding [7, 9, 55] (Fig. 13B).

A basic function of the Hsp70 chaperone system is to prevent off-pathway aggregation by shielding exposed hydrophobic stretches in non-native proteins and facilitating (re)folding via a kinetic partitioning mechanism [56]. Notably, a recent study has shown that the Hsp70 system can also accelerate the folding of multi-domain proteins by preventing inter-domain misfolding [32].

(A)

Closed state:

Nucleotide binding domain

IB

IA

Linker

Substrate binding domain

IIB

ADP

IIA

Bound peptide (NRLLLTG)

α-helical lid

Open state:

α-helical lid

ATP

β-sandwich domain

(B)

Native

ADP

Other chaperone systems

Hsp70

Open

High k_{on}/k_{off} for substrate

ATP

Nonnative protein

NEF +ATP

Hsp40

ADP

Closed

Low k_{on}/k_{off} for substrate

Open

ATP

+P$_i$
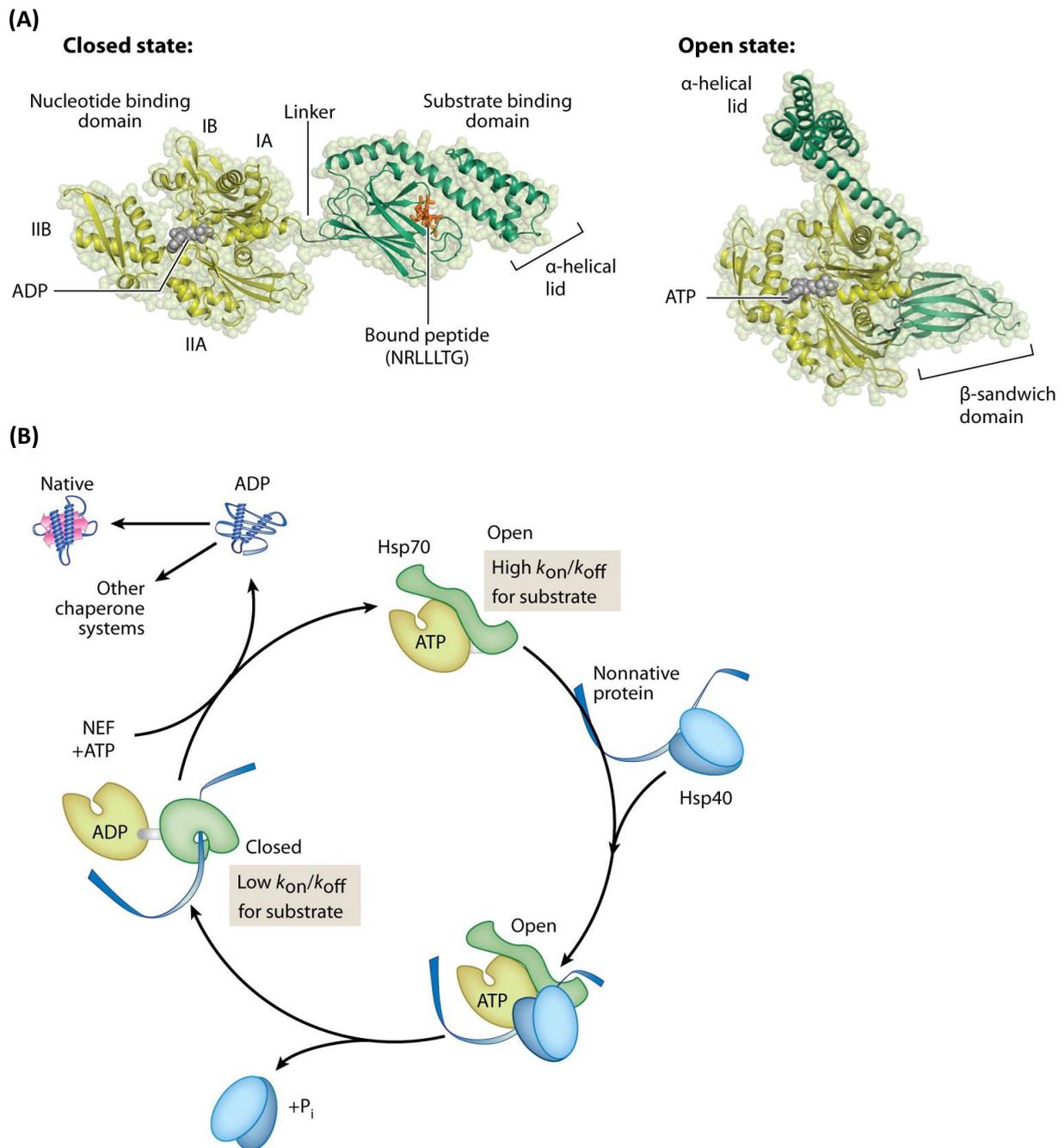
Figure 13: Structure and reaction cycle of Hsp70.

(A) Domain organization of Hsp70. Hsp70 consists of two domains, the nucleotide-binding domain (NBD) and the substrate-binding domain (SBD), connected by a conserved flexible linker, shown in yellow and green cartoons, respectively. (B) Reaction cycle of the Hsp70 chaperone system. Hsp40 delivers non-native protein to ATP-bound Hsp70 (open state) and stimulates the Hsp70 ATPase activity, generating the closed ADP-bound state. Nucleotide exchange factor (NEF) facilitates ADP/ATP exchange and promotes the release of substrates from Hsp70. Figure adapted from Ref. [7].

2.9 The chaperonins

Chaperonins are essential for protein folding in all three kingdoms of life. They are large ~1 MDa, ATP-driven cylindrical complexes consisting of two rings of ~ 60 kDa subunits stacked back-to-back. Their basic function is to provide an isolated chamber for the folding of single protein molecules to occur, unimpaired by aggregation. Two families of chaperonins, group I and group II, are being distinguished [57, 58] (Fig. 14).

2.9.1 Group I chaperonins

Group I chaperonins are found in bacteria (GroEL) and endosymbiotically related organelles, mitochondria (Hsp60) and chloroplasts (Cpn60). They consist of double-ring assemblies with 7-fold symmetry and employ a detachable "lid" protein (GroES in bacteria, Hsp10 in mitochondria and Cpn10/Cpn20 in chloroplasts) that encapsulates a non-native substrate protein in the sequestered chaperonin cavity where productive folding occurs. The GroEL/ES chaperonin system of *E. coli* has been investigated most extensively. The chaperonins provide a nano-cage for single protein molecule to fold in isolation. However, whether GroEL/ES is a catalyst of protein folding, actively accelerating folding speed, remains controversial [59].

2.9.2 Group II chaperonins

Group II chaperonins are found in the cytosol of archaea (thermosomes) and eukaryotic cells (TRiC/CCT). Thermosomes are composed of two stacked octameric or nonameric rings containing one to three different subunits, while the eukaryotic cytosolic chaperonin, TRiC has eight different subunits per ring. Group II chaperonins do not require a GroES-like cofactor but instead carry an extra helix (helical protrusion) located on their apical domains that closes the internal cavity in nucleotide dependent manner [58-61].
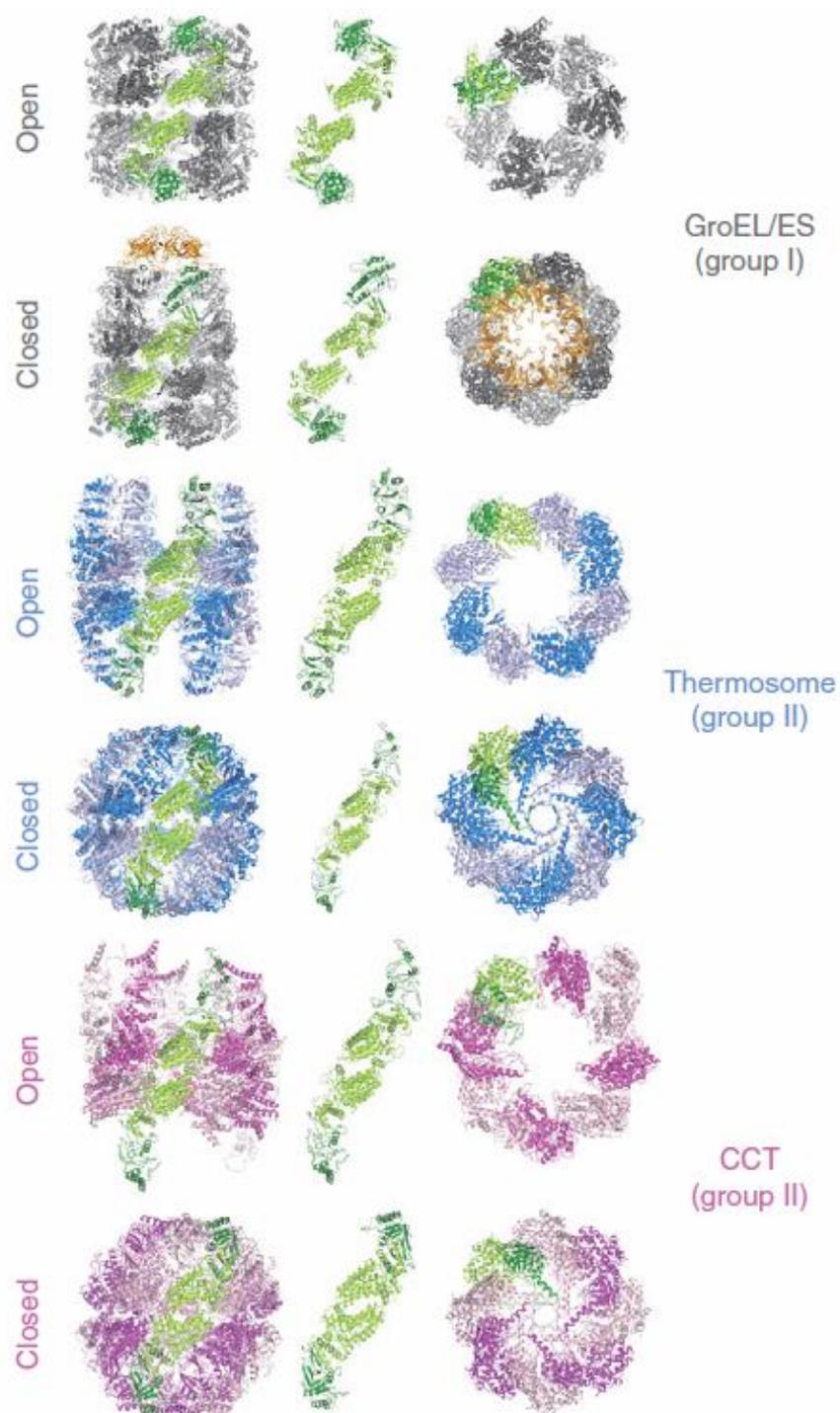
Figure 14: Structural comparison between group I and group II chaperonins.

Chaperonins are arranged as double-ring complexes, stacked back to back. The alternative open and closed conformations are shown for each chaperonin group: GroEL (PDB: 3E76), GroEL/ES (PDB: 1AON), open (PDB: 3KFK) and closed (PDB: 1A6D) conformations of thermosome, open (PDB: 2XSM) and closed (PDB: 3IYG) conformations of CCT/TriC are shown as side view in the first column. In the middle column equatorial and intermediate domains are shown in lime green whereas the apical

domains in dark green color. The third column depicts the top view of open and closed conformation. Figure adapted from Ref. [58].

2.10 Structure of GroEL/ES complexes

The structures of the GroEL oligomer and its complexes with nucleotide and GroES have been determined by X-ray crystallography and cryo-EM [62-64]. The first high-resolution structure of GroEL was determined at 2.8 Å resolution [65]. GroEL of *E. coli* is a tetradecamer of 57 kDa subunits that are assembled as two back-to-back heptameric rings, each containing a large, solvent-filled cavity (Fig. 15B). Each subunit of GroEL consists of three distinct domains: an equatorial ATP-binding domain (residues 6–133, 409–523), an intermediate hinge domain (residues 134–190, 377–408), and an apical domain (residues 191–376) that binds non-native substrate proteins and GroES (Fig. 15A). The C-terminal 23 residues (ending with four GGM repeats; disordered in the crystal structure) (525-548) extend from the equatorial domains into the central cavity, providing a barrier to prevent free passage between the GroEL ring cavities [59] (Fig. 15A). However, the role of these flexible C-tails is still not well characterized. While the C-terminal tails of the GroEL subunit are shown to be dispensable for basic GroEL function, interestingly, they are highly conserved in almost all GroEL homologues, acting as a 'floor' inside the GroEL ring, interacting with the substrate protein and affecting ATPase activity upon deletion [66, 67]. The equatorial domains provide all the intra- and inter-ring contacts between subunits and contain the sites of nucleotide binding and inter-ring communication (Fig. 15A, D). The apical domains form the entrance to the ring cavity and are connected to the equatorial domains via the intermediate domains, which transmit conformational changes triggered by nucleotide binding and hydrolysis [59, 64] (Fig. 15A). Helices H (residues 233–243) and I (residues 255–267) of the apical domains expose hydrophobic residues for the binding of a non-native polypeptide with exposed hydrophobic surfaces. GroES is a homo-heptameric ring of ~10 kDa subunits that binds to the ATP bound GroEL, forming the nano-cage in which substrate protein is encapsulated for folding (Fig. 15E). The crystal structure of *E. coli*, GroES has been solved at 2.8 Å resolution [68]. Each subunit of GroES is folded into a single domain, containing nine β-strands with a mobile loop region (Glu16 to Ala32), which is highly disordered in the free GroES heptamer but becomes more ordered upon interaction with helices H and I of the GroEL apical domains [59, 62] (Fig. 15D).
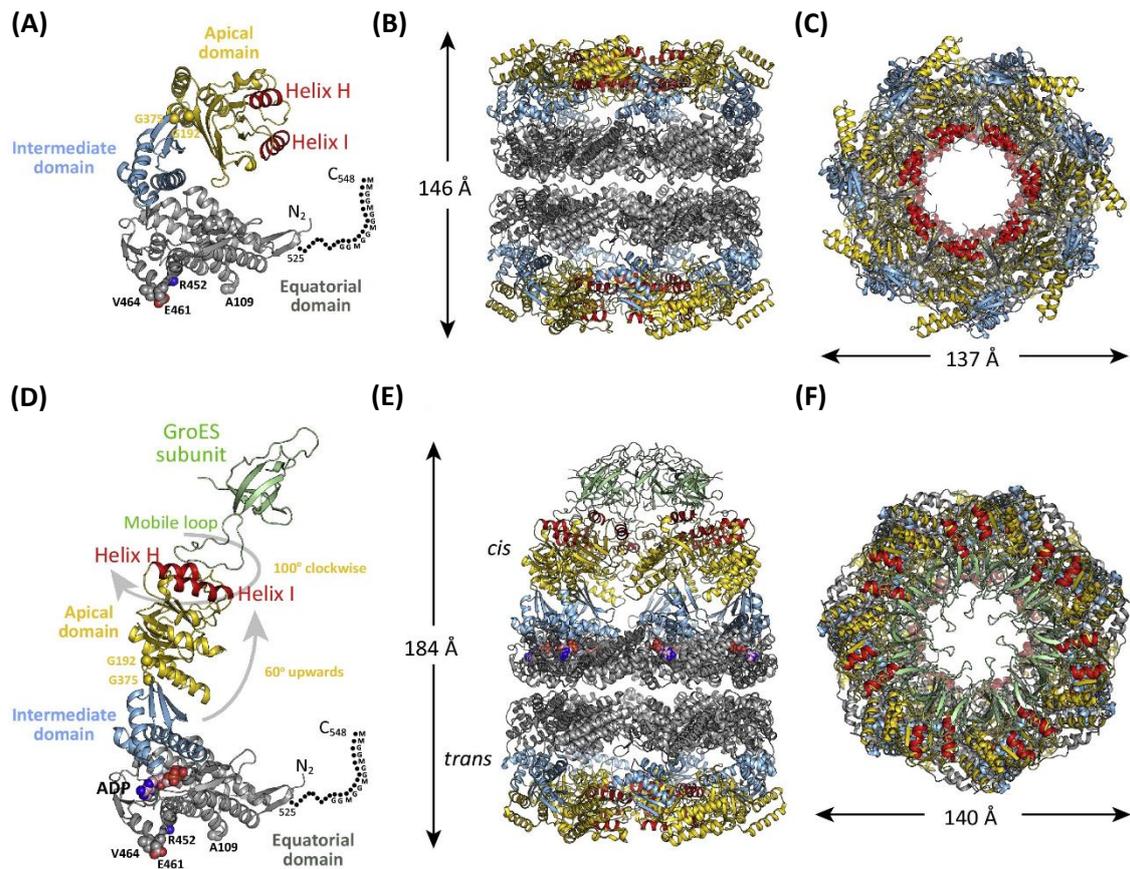
Figure 15: Crystal structures of *E. coli* chaperonin, GroEL and the GroEL/ES complex.

(A) Structure of the individual subunit of GroEL (PDB:1SS8). Three domains of the GroEL subunit: apical (gold), intermediate (blue) and equatorial (gray) domains are shown. Helices H and I of the apical domains bind substrate protein and GroES are shown in red. Hinge residues (Gly192 and Gly375) between intermediate and apical domains and residues forming the inter-ring contacts (Ala109, Arg452, Glu461, and Val464) are shown in space-filling representation. The 23 residues at the C terminus (ending with Gly-Gly-Met repeats; disordered in the crystal structure) are represented as dots. (B) Side view of the two homo-heptameric rings of GroEL (PDB:1SS8). (C) Top view of the GroEL cylinder. (D–F) Equivalent views of GroEL/ES/ADP complex (PDB:1PF9). The bound ADP is shown as space-filling models. The gray arrows indicate the reorientation of the apical domain upon binding of ATP and GroES. In the *cis*-ring, helices H and I of the apical domains interact with the mobile loops of GroES. The opposite *trans*-ring might be the same conformation as in apo-GroEL. Figure adapted from Ref. [59].

2.11 Endogenous substrates of GroEL

GroEL/ES is indispensable for *E. coli* viability under all cellular conditions [69], explaining the presence of essential obligate substrate proteins that depend on GroEL/ES for folding. Hundreds of GroEL substrates have been identified by a combination of biochemical analysis and quantitative proteomics [12, 70]. About ~250 of the ~2400 cytosolic proteins interact with GroEL in *E. coli,* corresponding to ~10% of total *E. coli* cytosolic proteins, including 67 essential proteins. These substrates were classified into three classes depending on their dependence on GroEL for folding (Fig. 16): Class I substrates fold spontaneously. They have a low propensity to aggregate during spontaneous folding and are not dependent on the chaperonin for refolding. Class II substrates are partially GroEL dependent and do not fold spontaneously. However, the DnaK system (DnaK, DnaJ, GrpE) mediates their refolding efficiently at 37°C. Class III substrates are fully dependent on GroEL/ES for folding due to rapid aggregation upon dilution from denaturant. The DnaK system alone could not mediate their refolding, however, DnaK was able to bind and stabilize these substrates in a non-aggregated state and transfer them to GroEL/ES for subsequent folding [12, 38, 39, 71, 72]. Around ~84 obligate substrate proteins are identified, based on their enrichment on GroEL. They typically have complex α/β or $\alpha + \beta$ domain topologies. Such proteins, including the $(\beta\alpha)_8$ TIM-barrels, are stabilized by many long-range contacts in their native states and have a strong tendency to populate kinetically trapped intermediates during folding [12, 73]. Most GroEL substrates are ~ 35-60 kDa in size, which is compatible with the volume of the GroEL cavity [12]. Some larger proteins are also found which may utilize GroEL for folding without encapsulation [74].
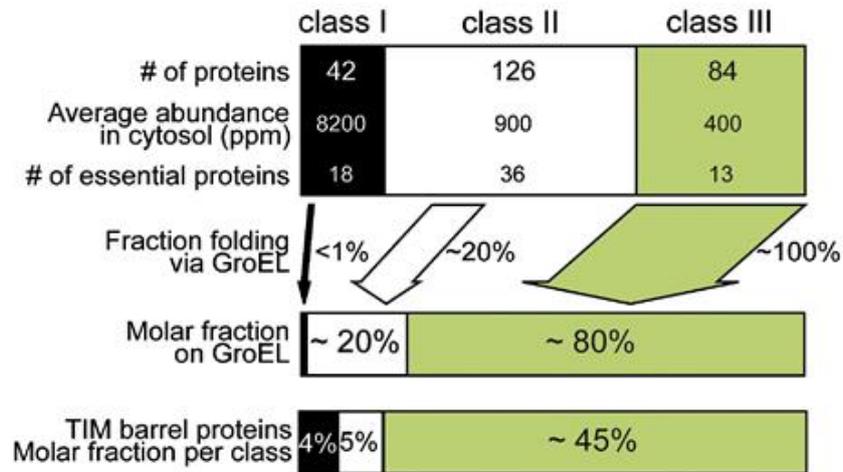
Figure 16: Classification of GroEL substrates in *E. coli.*

GroEL substrates can be divided into three classes based on their dependence on GroEL for folding. Class I substrates can fold spontaneously. Class II substrates are partially dependent on GroEL; however, the DnaK system can mediate their refolding efficiently. Class III substrates are obligate GroEL substrates and can only fold in the presence of GroEL/ES. Notably, 84 proteins are assigned to Class III substrates and occupy ~80% of the available GroEL capacity in the cell. Figure adapted from Ref. [12].

2.12 Reaction cycle of GroEL/ES

GroEL receives its substrates from TF and the DnaK system (DnaK/DnaJ/GrpE). The apical domains of the GroEL subunits form the entrance to the cavity and expose hydrophobic amino acid residues for the binding of molten globule-like folding intermediates [71, 72](Fig. 17).

In the nucleotide-free state, the GroEL subunits are in equilibrium between the T-state (low affinity for ATP) and the R-state (high affinity for ATP). ATP binding with positive cooperativity within the GroEL ring stabilizes the R-state which facilitates GroES binding (*cis*-ring). The opposite *trans*-ring has low binding affinity for nucleotide (T-state), resulting from negative cooperativity between rings. As a result, two heptameric rings of GroEL function sequentially as a two-stroke machine and the asymmetric GroEL/ES complex actively participates in folding reactions [59, 63, 75]. GroES binds at one or both ends of the GroEL cylinder in the presence of adenine nucleotides. Binding of 7 ATP to an equatorial domain of GroEL subunits and GroES to the substrate bound ring causes the displacement of non-native protein into a central cavity forming the *cis*-ring. Encapsulated proteins (up to ~60 kDa in size) are now free to fold unimpaired by aggregation inside a hydrophilic chamber during the time required to hydrolyze 7 ATP in the GroEL *cis*-ring (~5-10 sec at 25°C) [20]. After completion of ATP hydrolysis, ATP binding to the GroEL *trans*-ring then induces an allosteric signal that causes dissociation of ADP and GroES. Folded protein is released, while misfolded or not yet folded molecules may be recaptured for another folding cycle [20, 59, 71, 72] (Fig. 17).

However, symmetric GroEL-GroES$_2$ complexes with GroES bound to both GroEL rings have also been observed in the absence and presence of substrate protein but their functional significance in the reaction cycle is still a matter of ongoing research [59, 75-77].

Both rings of GroEL separate and exchange between complexes. Ring separation is a result of inter-ring negative allostery upon ATP binding. Transient ring separation prevents the formation of functionally impaired GroEL-GroES$_2$ complexes [78].

The hydrophobic surfaces of the apical domains mediating initial substrate binding are buried in the GroEL/ES complex and the hydrophilic inner surface of the cavity provides permissive conditions for folding. The wall of the GroEL/ES cavity has a net negative charge of 42 (189 negatively and 147 positively charged amino acids). Several negative charges (residues E252, D253, E255, D359, D361, and E363) cluster in two circular layers. Most of them are highly conserved among GroEL homologs, although they are not involved in substrate or GroES

binding. To investigate whether these charged residues are important in protein folding, a GroEL mutant, KKK2 has been employed in which three negatively charged residues (D359, D361 and E363) are mutated to lysines, resulting in a cavity net charge of 0. The KKK2 mutant impairs the ability of GroEL/ES to fold several substrate proteins [79, 80].

The volume of the GroEL/ES cavity (175 000 Å^3) has more than double the size of a GroEL ring in the absence of GroES and thus can accommodate substrate proteins up to ~60 kDa [59, 62]. Proteins that exceed the size limit of the GroEL cage may utilize the DnaK system for folding [12, 46] or undergo multiple rounds of binding and release from GroEL without GroES encapsulation [74, 81].
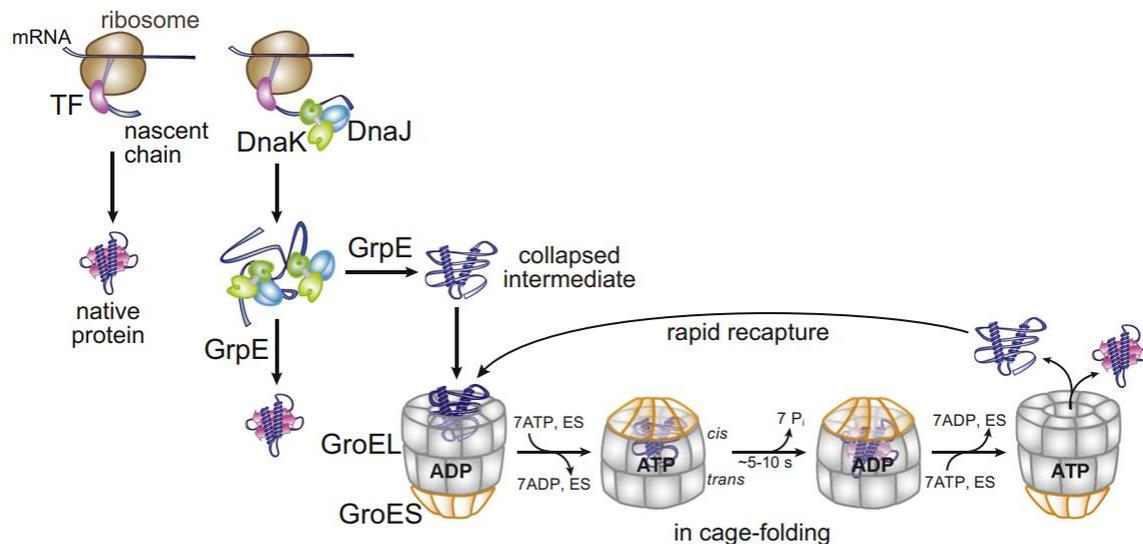
Figure 17: Pathways of chaperone-assisted protein folding in the cytosol of *E. coli.*

The ribosome-binding chaperone Trigger factor (TF) interacts first with the nascent polypeptide chain. Large multidomain proteins and proteins with complex α/β or α+β domain topologies interact subsequently with the Hsp70 chaperone system (DnaK and its co-chaperone DnaJ). The ATP-driven conformational cycle of DnaK is coordinated by DnaJ and nucleotide exchange factor (NEFs; GrpE in bacteria), which facilitate folding of a subset of proteins. The DnaK system stabilizes the obligate substrates of the GroEL/ES in a nonaggregated, but kinetically trapped state and transfers them to the GroEL/ES for folding. Substrate protein as a collapsed intermediate is captured by the *trans*-ring of the GroEL/ES/ADP complex. Upon ATP binding to the substrate-bound ring causes a conformational change in the apical domains, allowing substrate encapsulation into a GroES-capped *cis*-cavity. At the same time ADP and GroES dissociate from the opposite *trans*-ring. The encapsulated substrate is now free to fold within the hydrophilic cage for the time needed to hydrolyze the seven ATP molecules bound to the GroEL *cis*-ring. ATP binding followed by GroES binding to the GroEL *trans*-ring then triggers release of ADP, GroES and folded substrate protein while incompletely folded protein is rapidly recaptured by GroEL to proceed a new cycle until the protein reaches its native state.

2.13 MetF-methylenetetrahydrofolate reductase, a class III GroEL substrate

E. coli MetF catalyzes the reduction of 5,10-methylenetetrahydrofolate (CH_2-H_4folate) to 5-methyltetrahydrofolate (CH_3-H_4folate) using NADH. In this reaction, NADH transfers reducing equivalents to a FAD cofactor, which in turn transfers them to CH_2-H_4folate. MetF synthesizes CH_3-H_4folate that is utilized by methionine synthase to convert homocysteine to methionine [82]. The native MetF ($\beta_8\alpha_8$ barrel) consists of four identical 33-kDa subunits with noncovalently bound FAD at the C-termini of the β-strands. All these four subunits are arranged in a planar rosette with 2-fold symmetry [83] (Fig. 24A). Two subunits of MetF associate extensively with a strong interface to form dimers which further interact via a weak interface to form the homo-tetramer. Gel filtration was used to demonstrate that this enzyme can easily dissociate to dimers upon dilution and bound FAD is released on extensive unfolding of the holo-dimer to monomers [84] (Figure 18). MetF is a class III obligate substrate of GroEL which cannot refold spontaneously and requires GroEL/ES/ATP to fold. Each subunit folds inside the GroEL/ES cavity and assembles to tetramers to reach the final native state [12, 79].
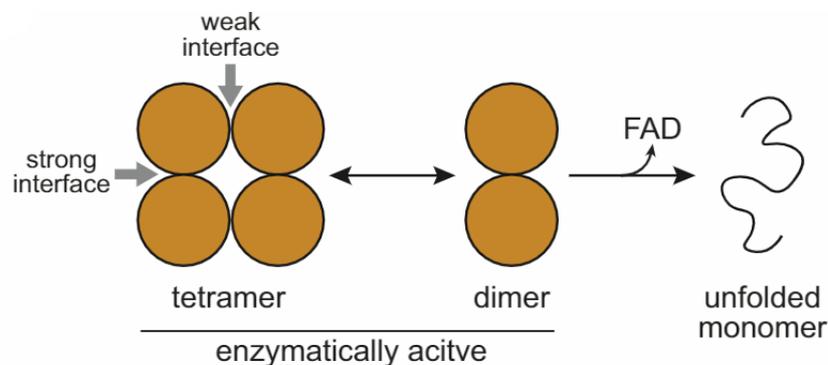
Figure 18: Schematic diagram of the *E. coli* methylenetetrahydrofolate reductase tetramer.

Schematic diagram summarizing the steps of dissociation and FAD release on unfolding of MetF. The homo-tetramer exists in equilibrium with dimer at low concentrations and both forms are present in the FAD-bound, enzymatically active state. The tetramer has 2-fold symmetry with two types of interfaces. FAD dissociates only on unfolding of the dimer to monomer.

2.14 Models of GroEL/ES assisted protein folding

The refolding of GroEL substrate proteins can be analyzed in bulk solution (spontaneous folding) and in the presence of GroEL/ES (assisted folding). It has been reported previously that the GroEL/ES chaperonin system can accelerate folding for a subset of substrate proteins. Does the GroEL cavity wall actively modify folding reaction, or does it simply act passively in preventing aggregation? How the chaperonin system assists in protein folding is still a matter of debate. Three models have been proposed to explain how the GroEL/ES reaction cycle promotes protein folding.

2.14.1 Passive cage model

The passive cage (also called Anfinsen cage) model suggests that the GroEL/GroES cavity functions mainly as a passive anti-aggregation device. Protein substrates fold inside the cavity with the same kinetics as in free solution at infinite dilution, that is, in the absence of (irreversible or reversible) aggregation. GroEL substrates fold at a biologically relevant timescale as long as aggregation is prevented. The model also postulates that GroEL/ES seemingly accelerates folding by prevents reversible aggregation phenomena that may reduce folding speed in free solution [85-87]. However, only rarely is aggregation a reversible process. Typically, aggregation is irreversible and thus only reduces the yield but does not affect the rate of spontaneous folding. Moreover, recent experiments using fluorescence correlation spectroscopy (FCS) to formally test the passive cage model, revealed that GroEL/ES can accelerate the folding of various substrate proteins ~10 to 100-fold above their spontaneous folding rate at very low concentrations of substrate proteins, where aggregation is completely excluded [20, 39, 71, 88]. All these cases provide compelling evidence that the slower rate of spontaneous folding was not due to transient aggregation, the GroEL/ES can modify the folding energy landscape for these substrates [56].

2.14.2 Active cage model

The active cage model posits that, besides preventing aggregation, the physical environment of the GroEL/ES nanocage alters the folding energy landscape, resulting in an accelerated folding kinetics for certain substrate proteins, thereby adjusting the folding speed relative to the rate of protein synthesis. This is attributed to an effect of steric confinement in the

Introduction

GroEL/ES cavity that restricts the conformational freedom of dynamic folding intermediates which slow or block the folding of the protein in free solution [20, 39, 71]. Recent studies have shown that GroEL/ES accelerates the folding of various proteins above their spontaneous folding rate. These include MBP mutant [20], *E. coli* prolidase PepQ [88], *Rhodospirillum rubrum* Rubisco [89], bacterial proteins with topological knots [90] and other *E. coli* $(\beta\alpha)_8$TIM-barrel proteins [39, 71]. In all cases, slow spontaneous folding was not due to reversible aggregation, indicating fact the GroEL/ES cage environment accelerates the folding kinetics of these substrates by smoothing their folding energy landscape [56]. The precise mechanism by which GroEL/ES catalyses protein folding is not clear but folding speed may depend on the following features: (a) Forced unfolding of substrate proteins upon binding to GroEL and additional stretching upon ATP-induced apical domain movements in GroEL cause transient further unfolding, thereby reversing misfolding [72, 91, 92]; (b) the limited volume of the GroEL/ES cavity relative to the size of the substrates, allowing for conformational confinement during folding (Fig. 19A); (c) the net negative charge distribution of the cavity wall, enforcing protein compaction due to charge repulsion effects (Fig. 19B); (d) and the flexible C-terminal tails that extend from the equatorial domains of each subunit into the cage, perhaps facilitating conformational rearrangement steps (Fig.19A). The contribution of these individual factors may be substrate specific [79]. These considerations are supported by multiple lines of evidence. Photo-induced electron transfer/fluorescence correlation spectroscopy and H/DX-MS experiments have shown the GroEL/ES cage minimizes the chain mobility of mutant MBP upon encapsulation, thereby facilitating the conversion of dynamic folding intermediate to the native state [20, 93]. Double-mutant maltose binding protein (DM-MBP) folds slowly due to the presence of kinetically trapped intermediate. Introducing two disulfide bonds that mimic the steric confinement of the chaperonin cage reduced the entropic barrier of the trapped intermediate, thereby accelerating native state formation [80]. Kinetic analyses showed that GroEL/ES catalyses the folding of the obligate *in vivo* substrate DapA, a TIM-barrel protein [71], by lowering the entropic component of the energy barrier, consistent with an effect of steric confinement exerted by the chaperonin cavity. Evidence was also presented that the negative charge clusters of the GroEL/ES *cis*-cavity are crucial for accelerated folding of certain substrate proteins [39, 79]. Molecular dynamics modelling suggested an ordering effect of the negative charges on water structure that enhances the hydrophobic effect [94]. However, experimental data of cavity-confined water is still missing

33

Introduction

[95]. The C-terminal tails of GroEL also contribute to accelerated folding by an unclear mechanism [67, 79, 88, 96]. In summary, compelling evidence indicates that the chaperonin cage acts as a powerful folding catalyst for a set of proteins that otherwise fail to reach their native state at a biologically relevant timescale.
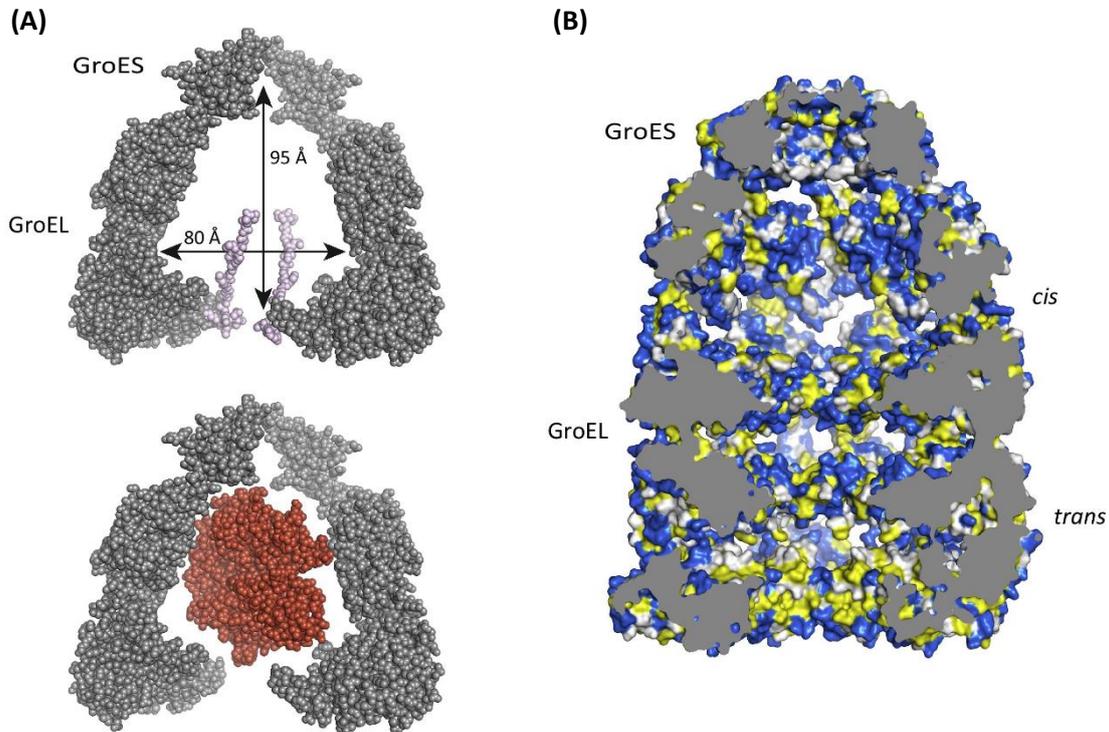
Figure 19: Structural features of the asymmetric complex of GroEL/ES.

(A) Upper panel: space-filling representation of the GroEL/ES/ADP complex. The C-terminal 23 residues are shown in pale pink. The approximate maximum dimensions of the central cavity are highlighted. Lower panel: the native structure of substrate protein S-adenosylmethionine synthase (MetK) is shown in space-filling model and red colored inside the GroEL/ES cavity. A molten globule state of MetK would have an ~30% expanded volume. (B) A vertical cross-section of the asymmetric GroEL/ES/ADP complex (PDB: 1PF9), showing the charged and hydrophobic residues in blue and yellow, respectively. Figure adapted from ref. [59].

2.14.3 Iterative annealing model

Finally, the iterative annealing model implies that the function of GroEL/ES is to actively unfold misfolded proteins that are kinetically stabilized by non-native contacts through ATP-driven cycles of binding and release, while folding may occur either inside or outside the chaperonin cage. In this model, accelerated folding requires unfolding of misfolded proteins that partition between productive and unproductive folding pathways. In this model, the transient

Introduction

encapsulation of the substrate does not promote folding, in fact the GroEL cavity retains the unfolded state of the protein, thus providing the protein a new opportunity for folding in the bulk solution [97, 98]. Does folding to the native state occur within the GroEL cavity? Interestingly, an accelerated folding mechanism has been seen not only under conditions of active GroEL/ES cycling, but also inside the single ring mutant of GroEL (SREL). SREL allows only one round of ATP hydrolysis upon GroES binding and forms a stable SREL/ES complex with substrate encapsulated. SREL/ES accelerates the folding of certain substrates with full yield upon a single round of encapsulation, suggesting that in this case the substrate undergoes only one round of forced unfolding, which plays a minor role (if at all) in accelerated folding mechanism [20, 71]. Moreover, recent studies indicate that substrate folding must occur inside the GroEL/ES cage in order to be accelerated [39], consistent with previous findings that substrate proteins spend ~80% of the duration of the GroEL/ES cycling reaction in the encapsulated state and less than 5% in free bulk solution [20].

2.15 Aim of this study

GroEL/ES of *E. coli* is one of the most extensively studied chaperonin systems in terms of structure and function. Numerous essential proteins depend on GroEL/ES for efficient folding, but how the chaperonin promotes their folding remains controversial. Specifically, the mechanisms underlying accelerated folding by GroEL/ES remain unclear. Three different models were proposed to explain how the GroEL/ES system provides kinetic assistance to the folding process. The passive or Anfinsen-cage model [66], the active cage model [56], and the iterative annealing model [97]. In order to test whether GroEL acts solely as a passive aggregation prevention device or actively speeds up the folding of certain proteins, we have re-examined the mechanism of GroEL/ES-assisted protein folding using the obligate chaperonin substrate protein, MetF of *E. coli*. We reason that the chaperonin system has co-evolved with its endogenous protein substrates. Consequently, a biologically relevant understanding the GroEL/ES mechanism is best achieved by studying how the chaperonin interacts with its endogenous substrates. The aim of this study is to test the existing models of the GroEL/ES mechanism with a specific focus on understanding how the properties of the central GroEL cavity impact the folding process of encapsulated protein.

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals

Acetic acid	Merck
Acetonitrile	VWR
Acrylamide/Bis Solution, 37.5:1	SERVA Electrophoresis GmbH
Adenosine triphosphate, disodium salt (ATP)	Sigma-Aldrich
Agarose	Lonza
Ammonium sulfate	Merck
Ammonium persulfate (APS)	Sigma-Aldrich
Ampicillin sodium salt	Carl Roth
Alexa-647 maleimide	Invitrogen
Alexa-532 maleimide	Invitrogen
Bacto agar	BD Biosciences
Bacto trypton	Gibco, Thermo Scientific
Bacto yeast extract	BD
Beryllium sulfate tetrahydrate ($BeSO_4$ $4H_2O$)	Sigma
Bromophenol blue sodium salt	Sigma-Aldrich
Calcium chloride ($CaCl_2$)	Sigma
Complete EDTA-free protease inhibitor	Roche
Coomassie Brilliant Blue R 250	SERVA
Deuterium oxide (D_2O)	Silantes

Materials and Methods

trans-1,2-Diaminocyclohexane-*N, N, N', N'*- tetraacetic acid monohydrate (CDTA)	Sigma-Aldrich
4,4'-Dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt (Bis-ANS)	Sigma-Aldrich
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich
Dithiothreitol (DTT)	Thermo Scientific
Dodecylsulfate.Na-salt (SDS pellets)	SERVA
Ethanol	Honeywell
Ethylenediaminetetraacetic acid, disodium salt, dihydrate (EDTA)	Merck
Flavin adenine dinucleotide disodium salt hydrate (FAD)	Sigma-Aldrich
Formic acid	Merck
Glycerol	Carl Roth
8M Guanidine HCl solution (GuHCl)	Thermo Scientific
HEPES	Carl Roth
Hydrochloric acid 37%	VWR Chemicals
Imidazole	Carl Roth
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Carl Roth
4x Laemmli Sample Buffer	BIO-RAD
Magnesium acetate tetrahydrate	Merck
Magnesium chloride hexahydrate	Sigma
Menadione	Sigma-Aldrich
2-Mercaptoethanol	Sigma-Aldrich
Methanol	Carl Roth
Morpholinoethane sulfonic acid (MES)	SERVA

38

Materials and Methods

β-Nicotinamide adenine dinucleotide, reduced disodium salt	Sigma-Aldrich
2-Phenylbenzothiazole (PBT)	Sigma-Aldrich
Phenylmethylsulfonyl fluoride (PMSF)	Sigma
Potassium dihydrogen phosphate (KH_2PO_4)	Carl Roth
Potassium chloride	Carl Roth
Potassium hydroxide	Carl Roth
Sodium acetate (NaOAc)	VWR
Sodium chloride	Carl Roth
Sodium fluoride (NaF)	Sigma
Sodium hydroxide	VWR Chemicals BDH
SYBR Safe DNA Gel Stain	Invitrogen
N, N, N', N'-Tetramethyl ethylenediamine (TEMED)	Sigma-Aldrich
Thioglycolic acid	Sigma
Tris-base	Sigma-Aldrich
Tris-(2-carboxyethyl) phosphine, hydrochloride (TCEP)	Thermo Fisher
Tween-20	Sigma-Aldrich
Urea	Sigma-Aldrich

3.1.2 Proteins and enzymes

Benzonase	Novagen
Lysozyme	Sigma
Pfu DNA polymerase	Promega
DpnI	NEW ENGLAND BioLabs
Bovine serum albumin (BSA)	Sigma

Materials and Methods

3.1.3 Instruments

Supplier	Material
Beckman Coulter (Pasadena, USA)	Benchtop centrifuge GS-6
	Ultracentrifuge rotor type 45 Ti
Biometra (Göttingen, Germany)	PCR thermocycler
Eppendorf (Hamburg, Germany)	Benchtop centrifuges 5427 R and 5424
	Research plus pipette (2.5 µL, 10 µL, 20 µL, 100 µL, 200 µL, 1 mL)
	Thermomixer comfort
	Biospectrometer
Amersham	Typhoon Biomolecular Imager
GE Healthcare (München, Germany)	Äkta Explorer, Äkta Purifier, chromatography columns (S30Q, MonoQ, Sephacryl S200, Heparin)
Waters Synapt	HDMS ESI-QTOF mass spectrometer
HORIBA Jobin Yvon	FluoroLog Spectrofluorometer
Ibidi (Martinsried, Germany)	µ-slide 8 well chambered microscope coverslip
BIO-RAD	PowerPac HV Power Supply
Epson Perfection V750 Pro	Photo Scanner
Jasco (Gross-Umstadt, Germany)	V-560 Spectrophotometer
Mettler-Toledo (Greifensee, Switzerland)	Balances AG285 and PB602
Milipore (Bedford, USA)	Amicon centrifuge filter units, steritop filter Units
Misonix (Farmingdale, USA)	Sonicator 3000
New Brunswick Scientific	Innova 44 incubator
Olympus (Tokio, Japan)	IX71 microscope body, microscope objective (60X W, NA 1.2)
PicoQuant (Berlin, Germany)	MicroTime 200 time resolved, confocal fluorescence microscope
Carl Roth (Karlsruhe, Germany)	ZelluTrans dialysis membrane
Scientific Industries	Vortex Mixer Genie 2

| PEQ Lab, Biotechnologie, GmbH | NanoDrop 1000 |
| inoLab pH7310 | pH meter |

3.1.4 Buffers and media

Lysogeny broth medium (Bertani, 1951)

10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl. The pH was found to be ~7.0 and not adjusted with NaOH to avoid pH variation during cultivation

Coomassie gel staining solution

40% ethanol, 8% acetic acid, 0.16% (w/v) Serva Coomassie Blue R-250

Coomassie gel destaining solution

12% methanol, 7% acetic acid, 81% H_2O

5X SDS-PAGE running buffer

15 g/L Tris, 72 g/L Glycine, 5 g/L SDS

Sample buffer SDS-PAGE

4x Laemmli sample buffer (BIO-RAD)

3.1.5 Strains and plasmids

The *E. coli* strains DH5α and BL21 (DE3) Gold (Stratagene) were used for cloning and protein expression, respectively. The *E. coli* genes *groEL* and *groES* were cloned into pET11a using BamH1 and NdeI restriction sites. EL(KKK2) (D359K, D361K, E363K) [79] variant of GroEL were constructed by site directed mutagenesis. The gene *metF* from *E. coli* was cloned into the vector pBAD18 (arabinose pBAD promoter). The MetF mutant, MetF (S35C) was generated by site-directed mutagenesis of the wild-type *metF* gene. The authenticity of each construct was confirmed by DNA sequencing.

Materials and Methods

3.2 Protein biochemistry methods

3.2.1 Purification of GroEL and mutants

Buffer A [30 mM Tris-Cl pH 7.5, 30 mM NaCl, 1 mM EDTA]

Buffer B [30 mM Tris-Cl pH 7.5, 500 mM NaCl, 1 mM EDTA]

Buffer C [30 mM Tris-Cl pH 7.5, 30 mM NaCl, 1.2M $(NH_4)_2SO_4$)]

Buffer D [30 mM Tris-Cl pH 7.5, 30 mM NaCl]

Buffer E [30 mM Tris-Cl pH 7.5, 500 mM NaCl]

Buffer F [30 mM Tris-Cl pH 7.5, 100 mM NaCl, 10% glycerol]

GroEL and variants were purified from *E. coli* BL21 gold strain as described previously [99] with some modifications. Cells were grown at 37°C to an OD_{600nm} of ~ 0.8. Protein expression was induced upon addition of 1 mM IPTG (Roth). The cells were furthermore grown for 4 h at 37°C, harvested by centrifugation (Beckman J6-MI, 3200 x g, 45 min, 4°C) and subsequently frozen in liquid nitrogen as a suspension in buffer A. Thawed cells were lysed by incubation for 1 h at 4°C in the presence of complete protease inhibitor (Roche), 1 mg/mL of Lysozyme (Sigma) and Benzonase (Novagen) and subsequent sonication with a tip sonicator (Misonix Sonicator 3000, power output 5 in pulse mode, 30 sec on, 40 sec off, total process time is 15 min), while the suspension was cooled on ice to prevent protein precipitation. After removal of cell debris and membranes by ultracentrifugation (Beckman L-90K, Ti45 rotor, 185,677 x g for 1 h at 4 °C), the supernatant was fractionated by a weak anion exchange column (DEAE Sepharose FF, GE Healthcare) equilibrated in buffer A and eluted with a gradient of NaCl from 30 to 500 mM. Fractions enriched for GroEL were pooled, and the protein was precipitated with 1.2 M ammonium sulfate. The precipitate was dissolved in buffer C and applied to a hydrophobic interaction column (Phenyl Sepharose FF, GE Healthcare) equilibrated in the same buffer and eluted with a gradient of decreasing ammonium sulfate from 1.2M to 0M. Fractions enriched for GroEL were pooled, dialyzed against buffer D and applied to a strong anion exchange column (Mono Q, GE Healthcare) equilibrated in buffer D and eluted with a gradient of NaCl 30 to 500 mM. The peak fractions containing GroEL were combined and buffer exchanged by HiPrep Desalting 26/10 column (GE Healthcare), equilibrated in buffer F. After every chromatography step, the purity of proteins was checked by SDS-PAGE. Finally, all purified

Materials and Methods

GroEL were checked for ATPase activity in presence and absence of GroES [100], MetF aggregation prevention and refolding [79].

3.2.2 Purification of GroES

Buffer A [50 mM Tris-Cl pH7.5, 20 mM NaCl, 1 mM EDTA, 1 mM DTT]

Buffer B [50 mM Tris-Cl pH 7.5, 1 mM EDTA, 1 mM DTT]

Buffer C [50 mM Tris-Cl pH 7.5, 1M NaCl, 1 mM EDTA, 1 mM DTT]

Buffer D [20 mM Imidazole pH 5.8]

Buffer E [20 mM Imidazole pH 5.8, 1 M NaCl]

Buffer F [30 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM EDTA]

Buffer G [50mM NaOAc pH 4.6]

Buffer H [30 mM Tris-Cl pH 7.5, 100 mM NaCl, 10% glycerol]

GroES was purified from *E. coli* BL21 gold strain. Cells were grown at 37°C to an OD_{600nm} of 0.8. Protein expression was induced upon addition of 1 mM IPTG (Roth). The cells were furthermore grown for 4 h at 37°C, harvested by centrifugation (Beckman J6-MI, 3200 x g, 45 min, 4°C) and subsequently frozen in liquid nitrogen as a suspension in buffer A. Thawed cells were lysed by incubation for 1 h at 4°C in the presence of complete protease inhibitor (Roche), 1 mg/mL of Lysozyme (Sigma) and Benzonase (Novagen) and subsequent sonication with a tip sonicator (Misonix Sonicator 3000, power output 5 in pulse mode, 30 sec on, 40 sec off, total process time is 15 min), while the suspension was cooled on ice to prevent protein precipitation. After removal of cell debris and membranes by ultracentrifugation (Beckman L-90K, Ti45 rotor, 185,677 x g, 45 min, 4°C), the supernatant was fractionated by chromatography on Source 30Q (GE Healthcare), equilibrated in buffer B with a gradient to buffer C. Fractions containing GroES were pooled and adjusted to buffer D, and next applied to a MonoQ HR 16/10 column (GE Healthcare), equilibrated in buffer D with a gradient to buffer E. GroES containing fractions were again pooled and subjected to Superdex 200 HiPrep 16/60 (GE Healthcare) gel filtration chromatography in buffer F. Fractions enriched for GroES were pooled, dialyzed against buffer G and applied to a cation exchange column (SP, GE Healthcare), pre-equilibrated with buffer G and eluted with a linear gradient of NaCl 0 M to 0.5 M. The peak fractions containing GroES were pooled and buffer exchanged by HiPrep Desalting 26/10 column (GE Healthcare), pre-equilibrated with buffer H. Fractions containing

Materials and Methods

pure GroES were concentrated at 4°C using Vivaspin (MWCO 10 kDa; GE Healthcare), followed by flash-freezing and stored at -80°C. After every chromatography step, the purity of proteins was controlled by SDS-PAGE. Finally, all GroES purifications were controlled for ATPase activity originating from impurities, efficient inhibition of GroEL ATPase activity [100] and MetF refolding [79].

3.2.3 Expression and purification of Hsp70 system

E. coli DnaK, DnaJ and GrpE were expressed and purified as previously described [12, 32, 79].

3.2.4 Purification of Apyrase

Apyrase was isolated from red skin potato (Solanum tuberosum) variety Desiree and purified as previously described [101, 102] with the following modifications. After ammonium sulfate $((NH_4)_2SO_4)$ precipitation, the pellet was dissolved in buffer [40 mM MES (pH 6.2), 1 mM $CaCl_2$, 1 mM thioglycolic acid, 1.6 M $(NH_4)_2SO_4$] and applied to a hydrophobic interaction chromatography column (Phenyl Sepharose 26/10; GE Healthcare) equilibrated in the same buffer and eluted with a $(NH_4)_2SO_4$ gradient from 1.6 to 0 M. Fractions were analyzed for apyrase activity using the Malachite Green colorimetric assay for protein phosphatase activity [103], and the most active fractions were pooled and subjected to $(NH_4)_2SO_4$ precipitation. The pellet was dissolved in buffer [40 mM MES (pH 6.2), 1 mM $MgCl_2$, 1 mM thioglycolic acid] and applied to a size-exclusion chromatography column (Superdex 75, GE Healthcare) equilibrated in the same buffer. Fractions were analyzed by SDS-PAGE and apyrase activity assay. The most active fractions were pooled and concentrated using Vivaspin centrifugal concentrator (MWCO 30 kDa, Satorius AG), followed by flash freezing and storage at −80 °C. The purified apyrase had an activity of 2.5 U μl^{-1}.

3.2.5 Purification of MetF and cysteine variants

Buffer A [30 mM Tris-Cl pH 7.2, 30 mM NaCl, 2 mM DTT, 1 mM EDTA]

Buffer B [30 mM Tris-Cl pH 7.2, 30 mM NaCl, 2 mM DTT, 10% (w/v) $(NH_4)_2SO_4$]

Buffer C [30 mM Tris-Cl pH 7.2, 30 mM NaCl, 2 mM DTT]

Buffer D [30 mM Tris-Cl pH 7.2, 30 mM NaCl, 10 mM DTT, 10% glycerol]

MetF and MetF(S35C) were expressed from an inducible plasmid (pBAD18) in *E. coli* BL21(DE3) cells grown in Luria -Bertani broth at 30 °C. After cell disruption by lysozyme (1 mg/ml) (Sigma)

Materials and Methods

treatment and sonication, the crude lysate was clarified by ultracentrifugation (185,677 x g for 1 h at 4°C). The supernatant was applied to a weak anion exchange column (DEAE Sepharose FF, GE Healthcare) equilibrated in buffer A and eluted with a gradient of NaCl from 50 to 500 mM. The fractions containing MetF were identified by absorbance at 447 nm, as MetF has an intense yellow color due to bound FAD. Fractions enriched for MetF were pooled, and the protein was precipitated with 10% (w/v) ammonium sulfate. The precipitate was dissolved in buffer B and applied to a hydrophobic interaction column (Phenyl Sepharose FF, GE Healthcare) equilibrated in the same buffer and eluted with a gradient of decreasing ammonium sulfate from 757 to 0 mM. Fractions enriched for MetF were pooled, dialyzed against buffer C and applied to a strong anion exchange column (Mono Q, GE Healthcare) equilibrated in buffer C and eluted with a gradient of NaCl 50 to 500 mM. The peak fractions containing MetF, as determined by SDS -PAGE, were combined and buffer exchanged by size -exclusion chromatography (HiLoad 26/60 Superdex 200, GE Healthcare) into buffer D. MetF containing fractions were concentrated to ~400 µM using Vivaspin centrifugal concentrator (MWCO 30 kDa, Satorius AG), followed by flash freezing and storage at -80°C. MetF concentration was determined using the molar absorbance coefficient of 14300 M^{-1} cm^{-1} of FAD at 447 nm [82].

3.2.6 Fluorophore labeling

For FCS and FCCS experiments, MetF(S35C) was labeled with either Alexa532-maleimide (M-532) or Alexa647-maleimide (M-647) (Invitrogen). Before labeling, cysteine residues of MetF(S35C) were reduced by incubation with 5 mM TCEP, which was subsequently removed using a Nap5 column (GE Healthcare). Labeling was performed with a 2-fold molar excess of dye over MetF in 30 mM Tris-Cl pH 7.2, 30 mM NaCl, 0.5 mM FAD at 4°C for ~ 10 h. Unreacted dye was removed by size -exclusion chromatography (Superdex 200; 3.2/300 column, GE Healthcare Life Sciences) pre -equilibrated with buffer (30 mM Tris-Cl pH 7.2, 30 mM NaCl, 10 mM DTT) using the Ettan LC system. Wild -type MetF has three cyteines that are buried and remained unlabeled. All labelling was performed with folded protein in native conditions. The degree of labeling (DOL) was measured spectrophotometrically (Nanodrop; PeQ Lab, Biotechnologie GmbH) using the following extinction coefficients (MetF, ε_{447} =14,300$M^{-1}cm^{-1}$, Alexa532, ε_{max} = 78,000 $M^{-1}cm^{-1}$, cf_{280} = 0.093; Alexa647, ε_{max}= 265,000 $M^{-1}cm^{-1}$, cf_{280} = 0.023) and the equation,

$$DOL = \frac{A_{dye} \times \varepsilon_{dye}}{(-(A_{dye} \times cf_{280}) + A_{280}) \times \varepsilon_{protein}}$$

DOL was >80%. The absence of free dye in the sample was confirmed by FCS.

As a control for FCCS experiments, double -labeled firefly luciferase (FLuc -DL) was prepared by incubation of the double cysteine mutant FLuc(D19C/S504C) with 1.5 -fold excess of an equimolar mixture of Alexa532 and Alexa647 -maleimide. Before labeling, cysteine residues of FLuc(D19C/S504C) were reduced by incubation with 2 mM TCEP, which was subsequently removed using a Nap5 column (GE Healthcare). Labeling was performed in buffer [25 mM HEPES-KOH (pH 7.5), 100 mM KCl, 10 mM Mg (OAc)$_2$, 5 mM ATP and 100 µM phenylbenzothiazole (PBT)] for 2 h on ice. After labeling, free dye was removed using a Nap5 column pre-equilibrated with buffer [25 mM Tris-acetate (pH 7.8), 1 mM EDTA, 0.2 M ammonium sulfate, 15% glycerol, 30% ethylene glycol, and 2 mM TCEP]. Labeling efficiencies for each dye were typically ~70%.

3.2.7 MetF aggregation

MetF was denatured in 6M GuHCl/10 mM DTT for 1 h at 25°C and diluted 200 -fold into buffer [20 mM Tris-Cl pH 7.2, 200 mM KCl, 5 mM Mg (OAc)$_2$, 50 µM FAD, 10 mM DTT] to the final concentrations indicated in figure. MetF (0.5 µM) aggregation was also measured in the presence of GroEL (0.5 µM and 1 µM) or EL -KKK2 (0.5 µM and 1 µM). Turbidity was recorded immediately upon dilution of denatured MetF at 320 nm using a UV-visible spectrophotometer. Background absorbance of the buffer alone was subtracted.

3.2.8 MetF folding, assembly and enzymatic assay

MetF was denatured in 6M GuHCl/10mM DTT for 1 h at 25°C. Spontaneous folding was initiated upon 200-fold dilution into buffer [20 mM Tris-Cl pH 7.2, 200 mM KCl, 5 mM Mg (OAc)$_2$, 100 µM FAD]. For assisted folding, denatured MetF was diluted 200 -fold into buffer containing 2 µM GroEL (or 2 µM EL-KKK2) and folding was initiated upon addition of 4 µM GroES and 5 mM ATP. Folding was stopped by addition of 50 mM CDTA. To measure the function of the *E. coli* Hsp70 system in MetF folding, denatured MetF was diluted into buffer containing DnaK (2 µM), DnaJ (1 µM) and 5 mM ATP, and the reaction initiated by addition of GrpE (2 µM). Enzyme activities were measured as previously described [82]. GroEL/ES-assisted

folding of MetF was also performed at 25°C and 37°C in the presence of excess GroEL (D87K) mutant (GroEL-trap) [89, 104]. Denatured MetF (100 μM) was diluted 200 -fold into buffer containing GroEL (1 μM), followed by the addition of GroES (3 μM) and folding initiated by the addition of 5 mM ATP with 10 μM GroEL -trap (10 -fold access over GroEL) to capture not yet folded MetF released from GroEL. Reactions were incubated for 1 h at 25°C to allow for complete assembly prior to enzyme assay.

3.2.9 Analysis of MetF encapsulation by GroEL/ES

Encapsulation of MetF inside the GroEL/ES cavity was analyzed by proteinase K (PK) protection assay. Alexa532-labeled MetF(S35C) (M-532) was denatured in 6M GuHCl/10 mM DTT for 1 h at 25°C. For encapsulation, GroEL -bound MetF was prepared by dilution of denatured M-532 (final concentration 150 nM) into buffer [20 mM Tris-Cl pH 7.2, 200 mM KCl, 5 mM Mg (OAc)$_2$, 20 μM FAD] containing 0.6 μM GroEL/ 2.4 μM GroES and BeF$_x$ (1 mM BeSO$_4$/ 6 mM NaF), followed by addition of 0.2 mM ATP. For the control reactions with GroEL-bound M-532, ATP/BeF$_x$ was omitted from the buffer. Any aggregated protein was removed by centrifugation at 20,000 x g for 20 min. Treatment with PK (5 μg/ml; Roth) was for 10 min at 25°C, followed by inhibition of PK with 2 mM phenylmethylsulfonyl fluoride (PMSF) and incubation on ice for 10 min. Reactions were analyzed by 4-20% SDS-PAGE (Bio-Rad) by Coomassie staining and fluorescence imaging using a Typhoon 5 imager (GE Life Sciences).

3.3 Biophysical methods

3.3.1 Tryptophan fluorescence

Intrinsic Trp fluorescence spectra of MetF in the presence and absence of GroEL (GroEL does not contain Trp) in buffer [20 mM Tris-Cl pH 7.2, 200 mM KCl, 5 mM Mg (OAc)$_2$, 100 μM FAD, 10 mM DTT] were recorded using a FluoroLog 3 Spectrofluorometer (Horiba) with excitation at 295 nm (5 nm slit width) and emission from 310 nm to 450 nm (10 nm slit width). Each spectrum represents the average of three consecutive scans. A peltier thermostat was used to maintain the temperature at 25°C. Background fluorescence of chemically identical samples lacking MetF was subtracted.

Materials and Methods

3.3.2 Bis-ANS fluorescence

The concentration of Bis-ANS (Sigma) was estimated using an extinction coefficient of 16790 M^{-1} cm^{-1} at 385 nm in water. MetF was denatured in 6M GuHCl/10mM DTT and diluted 200-fold into a folding buffer [20 mM Tris-Cl (pH 7.2), 200 mM KCl, 5 mM Mg $(OAc)_2$, 1 µM FAD] at a final concentration of 100 nM. Various forms of MetF (100 nM) were incubated with a 20-fold molar excess of Bis-ANS over the protein for 5 min at 25°C. The experiments were performed using a FluoroLog 3 Spectro-fluorometer (Horiba), with the excitation wavelength set to 390 nm (2 nm slit width). The emission spectra were recorded from 400-650 nm (2 nm slit width). Each spectrum represents the average of three consecutive scans. A peltier-thermostat was used to maintain the temperature at 25°C during the measurement. Each spectrum was corrected for background fluorescence caused by Bis-ANS and FAD or GuHCl in reactions lacking the MetF substrate.

3.3.3 Microscale thermophoresis (MST)

Binding of FAD with encapsulated MetF inside the chaperonin cage was calculated using Monolith NT. 115 Microscale Thermophoresis (NanoTemper Technologies GmbH, Munich, Germany). MetF(S35C) was labeled with Alexa Fluor 647 Maleimide (Invitrogen) fluorescent dye. MetF (M-647) was denatured in 6 M GuHCl/10 mM DTT for 1h at 25°C. Denatured M-647 was diluted 150-fold at a final concentration of 50 nM in folding buffer [20 mM Tris-Cl (pH 7.2), 200 mM KCl, 5 mM Mg $(OAc)_2$, 1 mM $BeSO_4$, 6 mM NaF] containing 0.2 µM GroEL, 0.8 µM GroES and 0.2 mM ATP. FAD with varying concentrations (from 0.3052 nM to 10000 nM) was incubated with MetF (M-647) encapsulated inside the GroEL: $GroES_2$ cage at 25°C for 15 min. The sample was loaded into Nano Temper standard treated capillaries. Measurements were performed at 25°C using 25% RED LED power and 60% MST power. All experiments were repeated three times for each measurement. Data analyses were carried out using Nano Temper MO Affinity Analysis software.

3.3.4 Fluorescence correlation and dual-color cross-correlation spectroscopy

FCS and dcFCCS measurements using pulsed interleaved excitation [105] were performed on a Microtime 200 inverse time-resolved fluorescence microscope (PicoQuant) as previously described [20]. Experiments were performed at a constant temperature of 20°C unless

Materials and Methods

otherwise stated. Pulsed diode lasers at 530 nm (LDH-PFA-530) and 640 nm (LDH-PC-640B) were used for excitation of Alexa532 and Alexa 647, respectively. For FCS measurements, the laser power was set to 60 µW measured before the major dichroic filter. For FCCS measurements, each laser was set to 40 µW. The lasers were pulsed with a rate of 26.7 mHz. Measurements were performed using a water immersion objective (60 × 1.2 NA, Olympus) in the sample cuvette (Ibidi). The emitted fluorescence was separated from excitation light by a dichroic mirror (Z532/635RPC), guided through a pinhole (75 µm) and in case of cross-correlation split according to wavelength by a beam splitter (600 DCXR) onto photon avalanche diodes (SPADs) (PDM series, MPD). The emission light was filtered by emission bandpass filters (HQ 690/70 and HQ 580/70, Chromas) in front of the respective detector. Detection was performed using time-correlated single-photon counting, making it possible to correlate any given photon with the excitation source. For FCS measurements, after-pulsing artifacts were removed by fluorescence lifetime filters (Symphotime, PicoQuant) [106]. Correlation plots were fitted with a triplet diffusion equation using Symphotime 32 software (PicoQuant).

$$G(\tau) = \left[1 - T + T \times e^{\left(\frac{\tau}{\tau_T} \right)} \right]$$

$$\times \left[\rho \times \left(1 + \frac{\tau}{\tau_D} \right)^{-1} \times \left(1 + \frac{\tau}{\tau_D \times \kappa^2} \right)^{-1/2} \right]$$

The mean diffusion time τ_D of particles through the focal spot is described by the structural parameter $\kappa = z_0/w_0$, where z_0 and w_0 denote the axial and radial dimension of the confocal volume, respectively. The amplitude of the correlation function is denoted by ρ. The first term is used to compensate for fast dynamics arising from dye photophysics such as triplet blinking with the amplitude T on the timescale τ_T [20].

The kinetics of assisted and spontaneous folding by FCS were measured using M-647. M-647 was denatured in 6 M GuHCl/10 mM DTT for 1 h at 25°C. Spontaneous folding was initiated by 200-fold dilution of denatured M-647 into refolding buffer [20 mM Tris–HCl (pH 7.2), 200 mM KCl, 5 mM Mg (OAc)$_2$, 1 µM FAD, 10 mM DTT, 0.05% Tween 20] at a final concentration of 100 pM. For assisted folding, denatured M-647 (100 pM) was diluted into the same buffer

49

Materials and Methods

containing 1 µM GroEL and 2 µM GroES and the folding reaction was initiated by the addition of 1 mM ATP. The assisted folding reaction was stopped at different time points with excess apyrase (25 U). After stopping the reaction, not-yet folded M-647 binds to GroEL, whereas already folded, native MetF does not. Spontaneous folding was stopped by addition of 1 µM GroEL. The shift in diffusion time allows us to determine the fraction of folded protein. For rescue experiments, denatured M-647 was diluted into buffer [20 mM Tris–Cl (pH 7.2), 200 mM KCl, 5 mM Mg (OAc)$_2$, 1 µM FAD, 10 mM DTT, 0.05% Tween 20]. After 1 h of incubation, 1 µM GroEL was added, and the diffusion was measured. GroES (2 µM) and 1 mM ATP were then added to initiate folding to be analysed as above. All diffusion measurements were performed for 30 min and diffusion coefficients were calculated using the following equation:

$$D = \frac{\left(V_{\text{eff}} \times \pi^{-3/2} \times \kappa^{-1} \right)^{2/3}}{4 \times \tau_D}$$

The data were fit with a one-triplet one-diffusion equation using Symphotime (PicoQuant) and the mean diffusion time of particles through the focal spot (τ_D) was extracted. The confocal volume (V_{eff}) was calibrated daily with ATTO655 dye.

FCCS measurements were performed to measure intermolecular association during spontaneous folding of MetF. M-532 and M-647 were denatured in 6 M GuHCl and 10 mM DTT for 1 h. The proteins were then diluted to different final concentrations in buffer [20 mM Tris–Cl (pH 7.2), 200 mM KCl, 5 mM Mg (OAc)$_2$, 1 µM FAD, 10 mM DTT, 0.05% Tween 20] maintaining a 1:1 molar ratio of labelled proteins. Note that higher concentrations of FAD interfere with the green laser. Where indicated, reactions were supplemented with different concentrations of denatured, unlabeled MetF (M-UL). The residual concentration of GuHCl was 30 mM in the reaction. Spontaneous folding was performed 30 min at 25°C before FCCS measurement. FCCS was recorded for 30 min at 25°C. Reactions were maintained at 25°C using a temperature-controlled chamber (Ibidi Heating System). FLuc-DL (0.55 nM) was used as a positive control.

Materials and Methods

3.3.5 Hydrogen/deuterium exchange (H/DX)

Native MetF was prepared at a final concentration of 5 µM in folding buffer supplemented with 50 µM FAD. GroEL:MetF complexes were prepared by 200-fold dilution of GuHCl-denatured MetF to a final concentration of 0.5 µM, in folding buffer containing 1 µM GroEL and 50 µM FAD. Complexes were then concentrated 10-fold using a Vivaspin centrifugal concentrator (MWCO 50-kDa, Satorius AG) to a final concentration of 5 µM MetF/10 µM GroEL, and centrifuged at 20,000g for 10 min to remove aggregated material. GroEL:ES$_2$ encapsulated MetF was prepared by adding 30 mM NaF, 5 mM BeSO$_4$, 1 mM ATP, and 40 µM GroES to GroEL:MetF complexes prepared as described above. For conditions in the absence of FAD, residual FAD was removed from GuHCl-denatured MetF by buffer exchange into folding buffer using a Biospin 6 column (Biorad), and FAD was omitted from reaction buffers. Deuterium exchange was initiated by 10-fold dilution into deuterated folding buffer, followed by incubation at 25°C for different times (10, 30, 100, 300, or 1000 sec). For reactions prepared in the presence of FAD, the FAD was also included in the deuteration buffer. Similarly, for analyses of GroEL: GroES$_2$ encapsulated MetF, ATP, NaF, and BeSO$_4$ were included in the deuteration buffer. Exchange reactions were quenched by addition of an equal volume of ice-cold quench buffer [100 mM potassium phosphate (pH 2.1), 10 mM TCEP, 6 M urea], to a final pH of ~2.5. Three independent replicates were performed for each condition, each using a new preparation of the GroEL: MetF complex. Quenched samples were injected into a Waters ACQUITY UPLC M-class with H/DX maintained at 0°C, via a 50-µl sample loop. Proteins were digested using an Enzymate BEH-pepsin column (Waters) at a flow rate of 100 µl min^{-1} and temperature of 20°C. Peptides were trapped and desalted for 3 min at 100 µl min^{-1} on an Acquity UPLC BEH C18 VanGuard pre-column (Waters) before elution onto an ACQUITY UPLC peptide BEH C18 column (Waters) Peptides were eluted over 7 min at 40 µl min^{-1} with an 8%–40% acetonitrile gradient in 0.1% formic acid at pH 2.5. Mass analysis was performed on a Waters Synapt G2-Si in HD-MSe mode. Source and ion guide settings were adjusted to minimize gas-phase back exchange during ion mobility separation as described previously [107]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [108] partner repository with the dataset identifier PXD016666. Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2020.02.031.

4 Results

4.1 Rapid and efficient refolding of MetF by GroEL/ES

Multiple model substrates have been used to demonstrate an enhancement of folding speed by steric confinement in the GroEL/ES cavity [20, 89, 109], although catalysis of folding by chaperonin has been observed for a relatively limited set of endogenous obligate proteins [71, 88]. We therefore reasoned that further experiments with endogenous substrate proteins are required to understand the detailed mechanism underlying the function of GroEL/ES as a folding catalyst. One important obstacle in conducting such experiments is the very high aggregation propensity of most of the class III obligate substrates during attempted spontaneous folding. In order to compare spontaneous and assisted folding by GroEL/ES, permissive conditions for spontaneous folding need to be established, *i.e.,* conditions under which aggregation is minimized and folding to the native state is energetically favorable. Here, we identified the homo-tetrameric class III enzyme, 5,10-methylenetetrahydrofolate reductase (MetF, ~33-kDa) as a suitable obligate substrate for which rapid and efficient refolding by GroEL/ES can be followed by enzymatic activity at temperatures ≥ 25°C; however, MetF does not fold spontaneously under standard folding conditions and forms kinetically trapped intermediate states. Similar to numerous other class III obligate substrates, MetF is a $\beta_8\alpha_8$ TIM-barrel protein [83] with bound FAD at the C-termini of the β-strands (Fig. 24a). MetF is therefore structurally representative of many natural obligate substrates of the chaperonin system.

Results
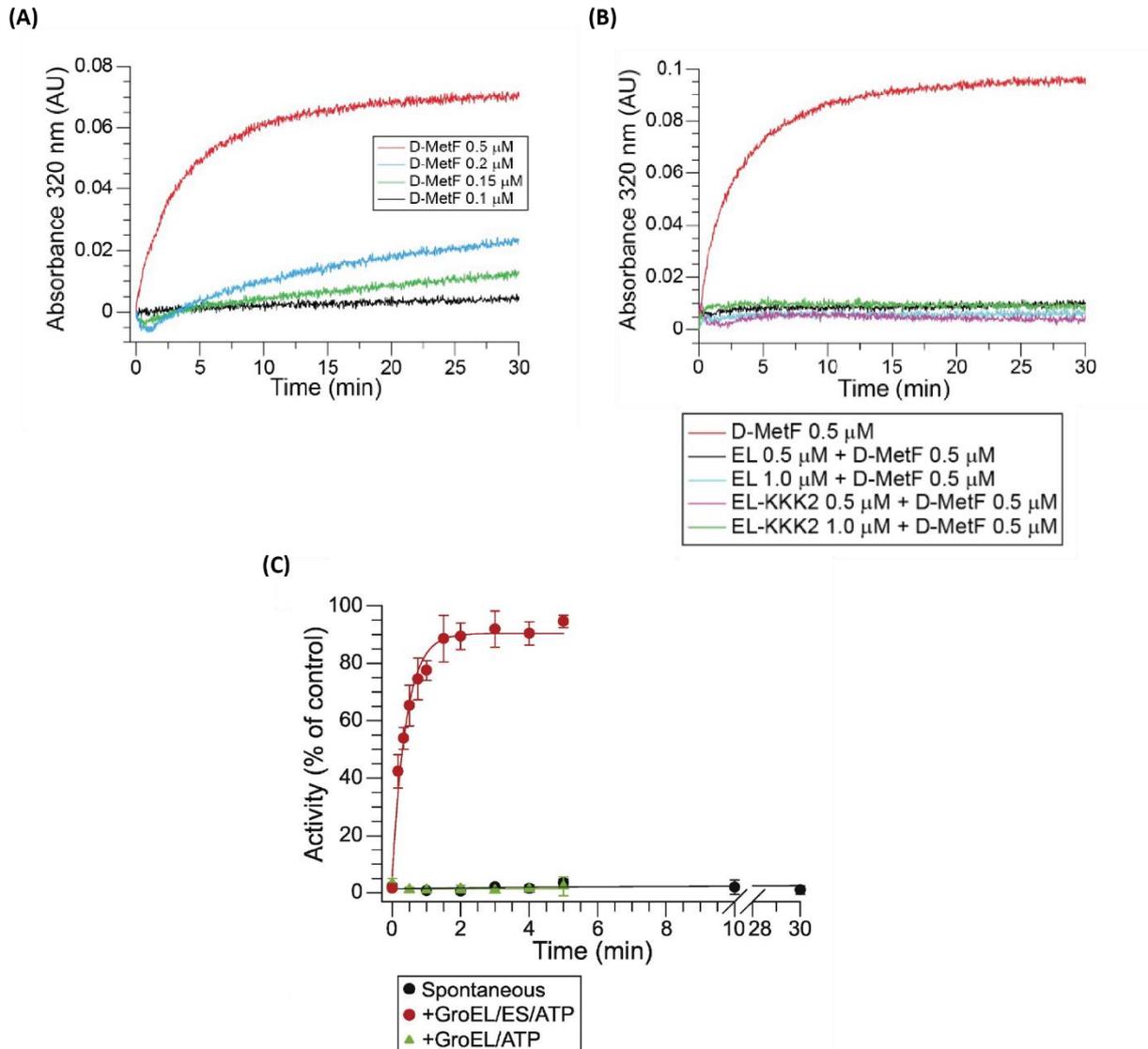
Figure 20: Fast and efficient folding of MetF by GroEL/ES/ATP.

(A) Aggregation of MetF was monitored by turbidity measurements at 320 nm using a UV-visible spectrophotometer. Denatured MetF (unfolded in 6M GuHCl/10 mM DTT) was diluted 200-fold into a refolding buffer to the final concentrations indicated and turbidity followed over time at 25°C. Background absorbance prior to addition of denatured MetF was subtracted. Representative data from 3 independent experiments are shown. **(B)** Aggregation prevention of MetF by GroEL and GroEL mutant, EL-KKK2. MetF (0.5 μM) aggregation was measured by turbidity assay at 320 nm as in (B) in the absence and presence of GroEL (0.5 μM and 1 μM) or EL -KKK2 (0.5 μM and 1 μM). Turbidity was recorded immediately upon dilution of denatured MetF. Background absorbance of the buffer alone was subtracted. Representative data from 3 independent experiments are shown. **(C)** Spontaneous and GroEL/ES-assisted folding of MetF. Spontaneous folding was monitored upon dilution of denatured MetF to 0.5 μM subunit in refolding buffer containing ATP at 25°C. For GroEL/ES-assisted folding, denatured MetF was diluted into refolding buffer containing 2 μM GroEL and 4 μM GroES and folding initiated by addition of 5 mM ATP. Folding reactions were stopped after different time points with 50 mM CDTA. Reactions were incubated for 1 h to allow for assembly, followed by enzyme assay.

Results

The activity of native MetF is set to 100%. Averages ± SD from three independent experiments are shown.

Unlike monomeric substrates, production of MetF enzymatic activity requires the folding of individual subunits and subsequent tetrameric assembly (Fig. 18). Purified *E. coli* MetF was denatured in 6 M GuHCl/10 mM DTT and diluted 200-fold into refolding buffer [20 mM Tris-HCl, pH 7.2, 200 mM KCl, 5 mM Mg (OAc)$_2$, 100 μM FAD] with or without GroEL/ES. Consistent with previous reports [12, 79], MetF aggregated upon dilution from denaturant, as measured by turbidity assay (Fig. 20A). As expected, aggregation was highly concentration dependent. No significant turbidity was measured at a final concentration of MetF or 0.1 μM. GroEL efficiently prevented the aggregation of denatured MetF (Fig. 20B). However, even at low concentration MetF did not fold spontaneously and instead populated kinetically trapped folding intermediate(s) (MetF-I) that were aggregation prone (Fig. 20C). In contrast, GroEL/ES-assisted folding of MetF was highly efficient (>90% yield) and occurred at a fast rate ($t_{1/2}$ ~15 sec), indicating high refolding yields after only ~3 chaperonin cycles (Fig. 20C). Folding was dependent on the presence of GroES, indicating that it required encapsulation in the GroEL/ES cage. We decided to use a combination of biophysical approaches to further characterize the formation of MetF-I formed by MetF during spontaneous folding.

4.2 MetF is efficiently refolded by GroEL/ES but not the Hsp70 system

Like other Class III obligate substrates of GroEL/ES, MetF does not fold in *E. coli* cells upon depletion of the chaperonin and instead forms insoluble aggregates [12]. However, *in vitro* refolding experiments showed that MetF aggregation was completely prevented when either the Hsp70 system (DnaK/DnaJ/GrpE) or GroEL/ES or both were present in the reaction mixtures (Fig. 21). In order to investigate whether the Hsp70 system alone could refold denatured MetF to its native state or only prevented its aggregation, we measured MetF enzymatic activity. MetF was unable to regain enzymatic activity in the presence of the Hsp70 system (DnaK, DnaJ, GrpE with ATP), indicating that it stabilized MetF in a non-aggregated, yet non-native state (Fig. 21). When GroEL/ES/ATP was added 5 min after diluting denatured MetF into buffer containing the Hsp70 components, refolding to enzymatically active MetF was ~75% efficient, compared to ~95% efficiency when GroEL/ES was present upon dilution from denaturant (Fig. 21). The folding yield decreased further when GroEL/ES was added 15 min after dilution of MetF (Fig. 21). This gradual loss of refolding competence indicated that MetF aggregated at 0.5 µM concentration during binding and release cycles from Hsp70, as confirmed by turbidity measurements (Fig. 20A, B). Thus, the Hsp70 system maintains MetF in a refoldable state and allows its transfer to GroEL/ES, but the ability of Hsp70 to stabilize non-native MetF is limited. These results are consistent with previous observations using other substrate proteins, which had shown that the apparent refolding rate in the presence of the Hsp70 system is significantly slower than folding with GroEL/ES [71, 72, 79]. Multiple DnaK/DnaJ molecules were found to stabilize non-native substrates in an expanded conformation [32, 72, 110-112]. However, in the case of MetF, release from DnaK/DnaJ does not result in refolding. Thus, the ability to promote MetF folding is unique to GroEL/ES, suggesting that the chemical and physical environment of the GroEL/ES cavity is critical in avoiding the formation of misfolded or kinetically trapped folding intermediates of MetF.
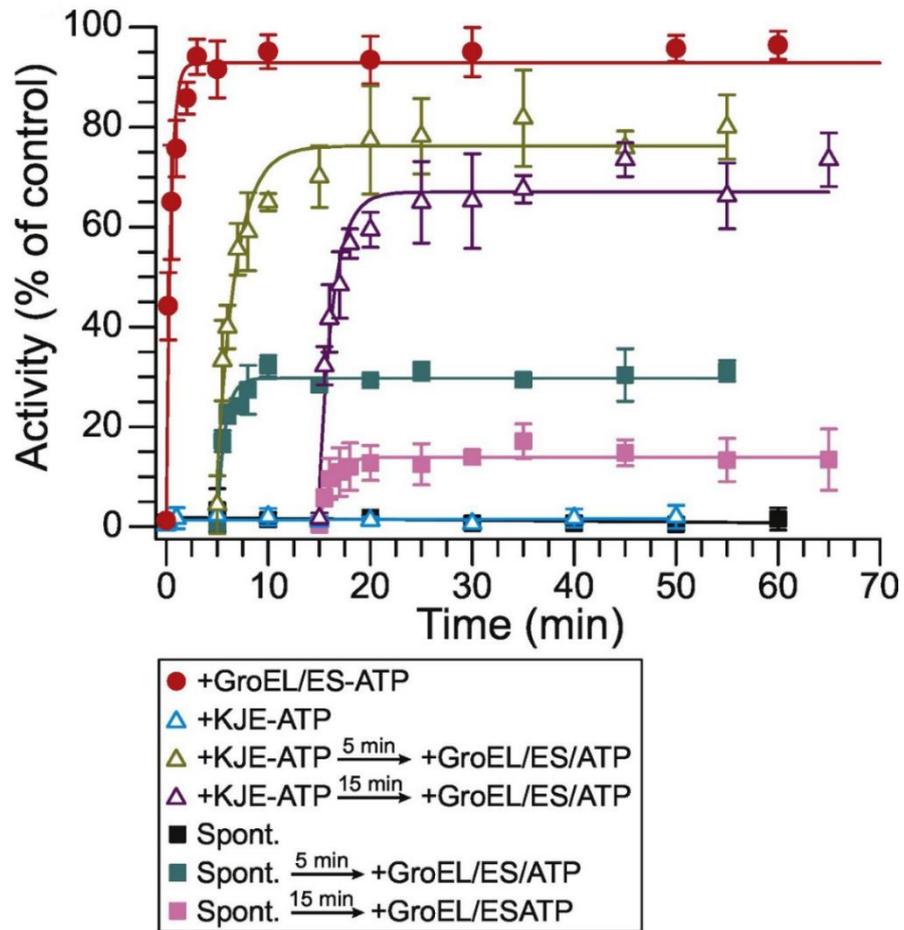
Figure 21: The *E. coli* Hsp70 system (DnaK, DnaJ, and GrpE) stabilizes non-native MetF in a refolding competent state but does not mediate folding by cycles of binding and release. GuHCl-denatured MetF was diluted into refolding buffer containing the Hsp70 system (2 µM DnaK, 1 µM DnaJ and 2 µM GrpE) and ATP. In other reactions, GroEL/ES was added 5 or 15 min after dilution of denatured MetF into buffer containing DnaK/DnaJ/GrpE/ATP or into buffer alone (spont.). Enzyme assays were performed as in figure 20C. Averages ± SD from three independent experiments are shown.

4.3 A single round of MetF encapsulation by GroEL/ES allows efficient folding

Different models have been proposed for how GroEL/ES promotes folding. The iterative annealing model suggests that cycles of substrate protein binding and release from GroEL accelerate folding by repeatedly unfolding kinetically trapped intermediates, thereby affording them the opportunity for productive folding, which may occur either inside or outside the GroEL/ES cage [97]. A critical question is therefore: Does substrate protein fold inside the expanded cavity of GroEL/ES or in solution after ejection? To address this question, we measured the fractions of MetF that folds in a single round of encapsulation by GroEL/ES. We first diluted denatured MetF (unfolded in 6 M GuHCl/10 mM DTT) into refolding buffer containing GroEL/ES, generating the GroEL: MetF complex. Before initiating folding (and GroES binding to GroEL) with ATP, we added a 10-fold excess of the GroEL-trap mutant, GroEL (D87K) over wild type GroEL (Fig. 22 & 23). GroEL (D87K) is an ATPase-deficient mutant of GroEL that binds non-native substrate proteins but does not release it [89, 104]. As a result, >90% of MetF that is released from GroEL in a non-native state per encapsulation cycle will be captured by GroEL-trap and prevented from folding. Hence, the yield of refolding will reflect the fraction of MetF that folds within the chaperonin cage in a single round of encapsulation. We found that >50% of MetF folded in a single round encapsulation cycle at 25°C and ~35% at 37°C (Fig. 22 & 23), indicating that the environment of the GroEL/ES cage is highly conducive in committing MetF to a productive folding trajectory and demonstrating that the majority of molecules need only one to three cycles of encapsulation inside the chaperonin cage to reach the native state. Note that the ATPase is accelerated at 37°C, shortening the time (~2.2 sec) a substrate protein spends in the encapsulated state per cycle from ~7 sec at 20°C to ~2.2 sec [20]. These results argue against iterative annealing as the major principle underlying MetF folding by chaperonin and support the view that encapsulation inside the GroEL/ES cage is functionally critical in promoting MetF folding.
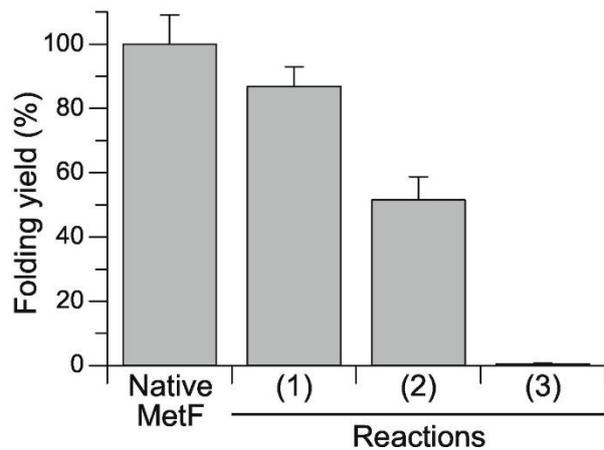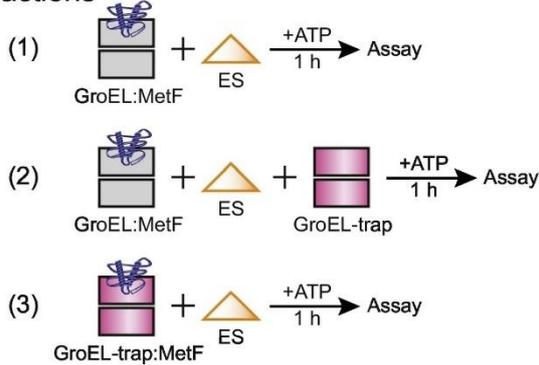
Figure 22: MetF folding yield in a single round of encapsulation by GroEL/ES. Experimental conditions are shown schematically on the left. MetF folding by GroEL/ES was measured either in the absence (reaction 1) or in the presence of GroEL-trap (reaction 2) at 25°C. In reaction 1, GuHCl-denatured MetF (0.5 µM) was first diluted 200-fold into a refolding buffer containing GroEL (1 µM)/GroES (3 µM), followed by addition of ATP (5 mM) to initiate the folding reaction. In reaction 2, ATP was added simultaneously with GroEL-Trap (10 µM). In reaction 3, GuHCl-denatured MetF was diluted into a refolding buffer containing GroEL-trap (1 µM) to form the GroEL-trap: MetF complex, followed by addition of GroES (3 µM) and ATP (5 mM). Enzyme assays were performed as in figure 20C. The activity of native MetF is set to 100%. Averages ± SD from three independent experiments are shown.
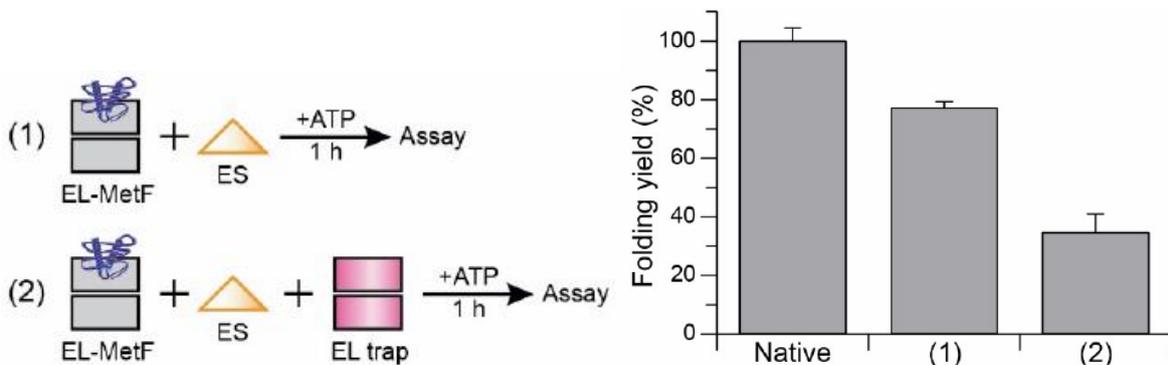
Figure 23: MetF folding yield in a single cycle of encapsulation by GroEL/ES at the physiological temperature of 37°C. Experimental conditions are shown schematically on top. GroEL/ES-assisted folding of MetF was performed either in the absence (reaction 1) or in the presence of GroEL-trap (reaction 2). Assays were performed as in figure 20C. The activity of native MetF is set to 100%. Averages ± SD from 3 independent experiments are shown.

Results

4.4 A highly dynamic folding intermediate is populated during spontaneous refolding of MetF

In order to understand how GroEL/ES promotes protein folding, it is important to determine why obligate chaperonin substrates fail to fold spontaneously. Previous studies attempted to distinguish the properties of GroEL substrates from other non-substrate proteins but no specific sequence signatures or structural properties could be identified [113]. Based on physicochemical properties of the amino acids, computational tools predict that obligate chaperonin substrates have high aggregation propensities [114]. In order to obtain more insight into the features that render MetF GroEL/ES dependent, we analyzed the properties of the metastable intermediates populated along the pathway of spontaneous MetF-folding. We probed the conformation of the MetF folding intermediate(s) (MetF-I) that accumulates during attempted spontaneous folding by performing Bis-ANS binding and intrinsic tryptophan (Trp) fluorescence measurements [99, 115]. The fluorescent dye Bis-ANS probes the exposure of hydrophobic residues on a protein, resulting in increased fluorescence [116]. Collapsed, molten globule-like folding intermediates devoid of ordered tertiary structure are typically Bis-ANS positive, allowing the dye to enter the dynamic hydrophobic core [99, 115]. We found that MetF-I exposed hydrophobic regions accessible to Bis-ANS, as reflected in a blue shift in fluorescence and a substantial increase in fluorescence intensity at 495 nm, whereas Bis-ANS fluorescence was hardly detectable for the native and guanidine-denatured forms of MetF (Fig. 24A). We concluded that MetF populates molten globule-like intermediate states during attempted refolding. Next, we examined the intrinsic tryptophan (Trp) fluorescence of MetF-I as a measure of tertiary structure. Each subunit of MetF contains three Trp residues at positions 42, 107, and 234 (Fig. 24B). Native MetF (100 nM) with bound FAD shows a Trp fluorescence emission maximum (λ_{max}) at 345 nm (Fig. 24C). This λ_{max} for the native protein is consistent with the Trp residues being partially solvent exposed with relatively low fluorescence intensity, presumably due to quenching interactions. In the presence of 6M GuHCl, the emission spectrum of MetF undergoes a redshift from 345 nm to 356 nm, accompanied by a substantial increase in fluorescence intensity (Fig. 24C), indicating dequenching and solvent exposure of the Trp residues upon complete unfolding. Interestingly, when denatured MetF was diluted into a refolding buffer, MetF-I exhibited a λ_{max} of 345 nm

Results

and a high fluorescence intensity compared to native state was detected immediately upon dilution (Fig. 24C).

These measurements were performed at a final concentration of MetF of 100 nM, where the protein remains ~80-85% competent for refolding by GroEL/ES for at least 5 min (Fig. 24D), suggesting that aggregates have not yet formed during this time. These spectroscopic properties suggest that the Trp residues of MetF-I are collapsed into a more hydrophobic environment, consistent with the conformational properties associated with flexible compact intermediate or molten globule states [99, 115, 117]. Taking advantage that GroEL lacks Trp residues [115, 118], we also measured the Trp fluorescence of GroEL-bound MetF, the emission maximum (λ_{max}) and relative fluorescence intensity of GroEL-bound MetF was almost identical to that of MetF-I, consistent with GroEL stabilizing non-native MetF protein in a conformation without ordered tertiary structure. This finding agrees with the previous conformational characterization of other GroEL-bound proteins [71, 119].

Results

(A)

(B)

(C)

(D)
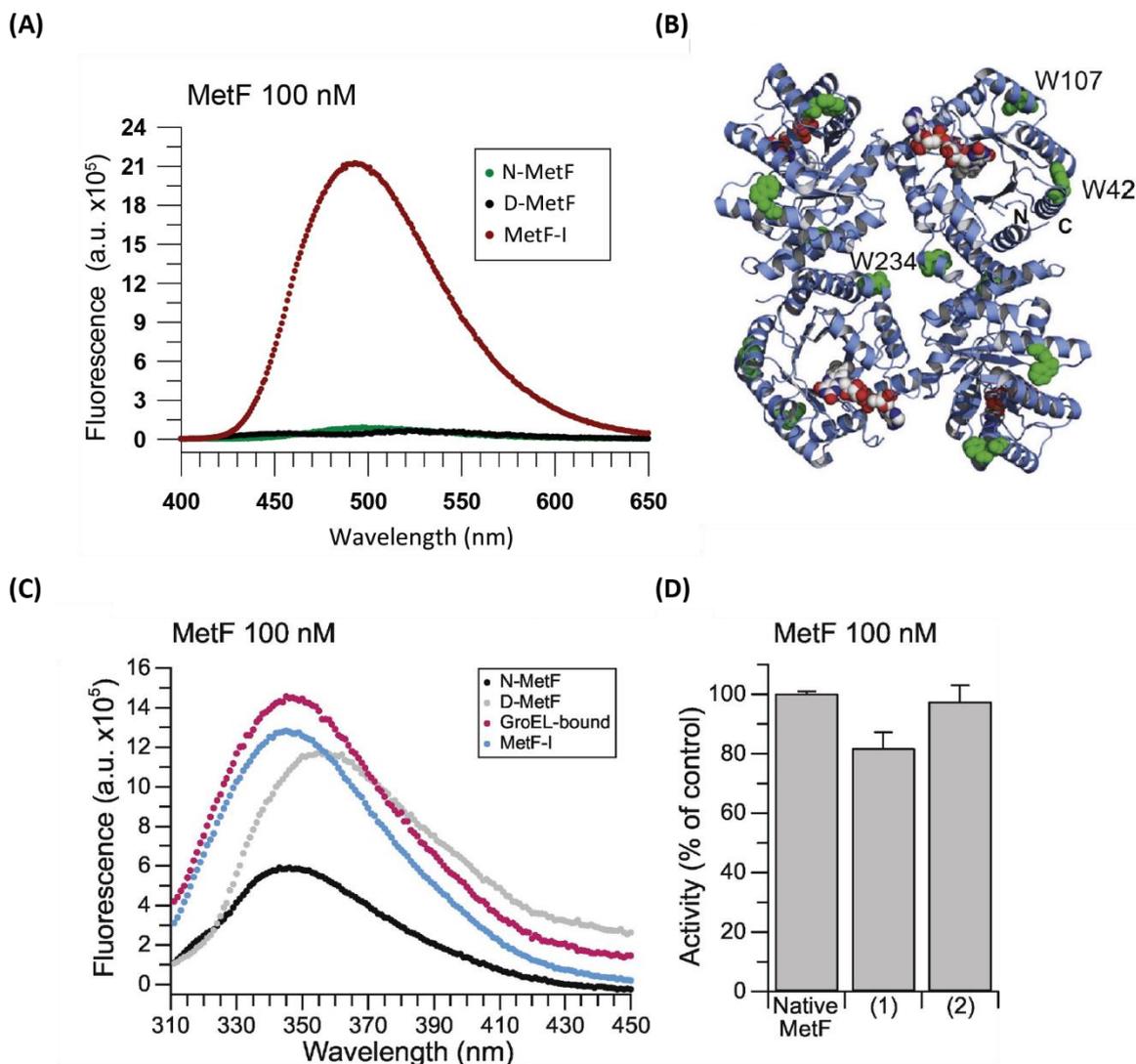
Figure 24: MetF forms a kinetically trapped intermediate. (A) Bis-ANS fluorescence spectra upon binding to different conformational states of MetF. Bis-ANS fluorescence spectra in presence of the denatured state of MetF (in 6 M GuHCl/10mM DTT, black), the native state (blue) and the intermediate state (red), formed upon 200-fold dilution from denaturant, were recorded with an excitation wavelength of 390 nm. Each spectrum represents the average of three consecutive scans. Each spectrum was corrected for background fluorescence caused by Bis-ANS and FAD or GuHCl in reactions. Exemplary curves are shown. (B) Crystal structure of homotetrameric MetF (PDB: 1B5T) with FAD-bound in each subunit in ribbon representation. Trp residues W42, W107, and W234 are indicated in green in space-filling representation. (C) Intrinsic Trp fluorescence spectra of native MetF (N-MetF), denatured MetF (D-MetF), MetF-I and GroEL-bound MetF. The final concentration of MetF was 100 nM. MetF-I was analysed immediately after dilution of denatured MetF from 6M GuHCl into refolding buffer. Background fluorescence of chemically identical reactions lacking MetF protein was subtracted. Representative measurements of three independent repeats are shown. (D) MetF-I remains competent for GroEL/ES assisted refolding. Denatured MetF was diluted into buffer A to a final concentration of 100 nM. GroEL (0.4 µM)/GroES (0.8 µM) was added after 5 min and folding initiated by addition of ATP (2 mM) (reaction 1). As a control, GuHCl-denatured MetF was diluted directly into

61

GroEL/GroES containing buffer and assisted folding initiated upon addition of ATP (reaction 2). The activity of native MetF is set to 100%. Averages ± SD from three independent experiments are shown.

4.5 Evidence for native-like conformation of encapsulated substrate

Refolding assays indicated that MetF folds very efficiently with assistance by the GroEL/ES system, requiring 2-3 chaperonin cycles to reach > 90% folding yield (Fig. 20C). In order to validate whether MetF folds to completion forming its substrate-binding active site while inside the chaperonin cage, we performed microscale thermophoresis (MST) measurements to detect FAD binding to encapsulated apo-MetF. Stable encapsulation of MetF was achieved by forming the symmetric GroEL: $GroES_2$ complex in the presence of BeFx/ATP. We used BeF_X in the presence of ATP to mimic the ATP-bound state of GroEL, thereby generating MetF encapsulated in a long-lived symmetric GroEL: $GroES_2$ complex that maintains ADP.BeFx bound state [78, 120, 121]. Previously this symmetrical complex GroEL: $GroES_2$-$(ADP.BeF_X)_{14}$ has been shown to be highly stable for a longer period of time and used for the crystallization with and without enclosed substrate protein [76, 78, 120].

As a negative control, we probed the binding of FAD to GroEL-bound MetF. MST is a powerful technique for measuring binding affinities for biomolecular interactions in solution [122]. It is based on the thermophoresis effect, the phenomenon of molecules migrating along a temperature gradient in a solution. We monitored the thermophoretic behavior of the Alexa 647 labeled MetF (M-647), which will change upon the binding of a ligand (FAD) in titration experiments, allowing the extraction of binding affinity information as a K_d value. MetF (M-647) was encapsulated inside the GroEL: $GroES_2$ cage, and changes in thermophoresis were measured over a range of FAD ligand concentrations (from ~0.3 nM to 10000 nM). We obtained a K_d value for FAD binding of 351.2 nM (Fig. 25). This data suggests that the active site of MetF folds correctly inside the chaperonin cage that binds FAD tightly.

Results
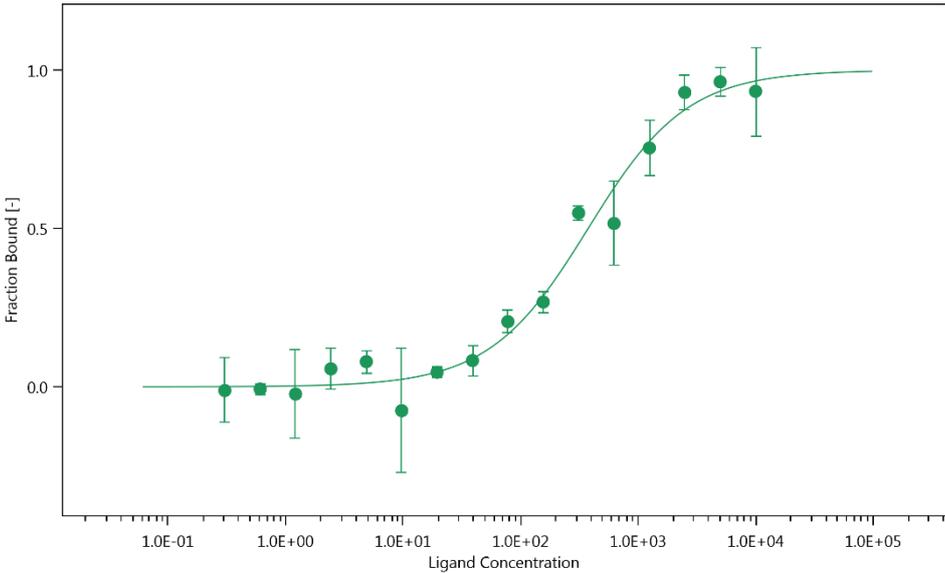
Figure 25: Binding curve of FAD ligand (nM) titrated against encapsulated MetF inside the chaperonin cage. Data shown are representative of three independent microscale thermophoresis experiments in the presence of FAD. Errors bars, s.d. (n=3).

Results

4.6 MetF does not aggregate during folding at single molecule level

Upon dilution from denaturant to a low concentration of 100 nM, MetF remained competent for refolding by GroEL/ES for several minutes (Fig. 24D), suggesting that aggregation was not the cause why spontaneous refolding failed. In order to rule out aggregation completely, we measured MetF folding under single molecule conditions at a final concentration of 100 pM [20, 123]. Absence of aggregation was established using dual-color fluorescence cross-correlation (dc-FCCS) experiments [20, 71]. We generated a MetF mutant with solvent exposed cysteine for fluorescent labeling by replacing serine 35 with cysteine (MetF-S35C) (Fig. 26A). MetF-S35C was labeled either with the fluorescent dye Alexa647 (M-647) or Alexa532 (M-532). Labeled MetF-S35C remained enzymatically active (Fig. 27). Equimolar mixtures of the differentially labeled proteins were denatured in 6M GuHCl and diluted into refolding buffer to analyze dc-FCCS. No cross-correlation signal was observed at concentrations of 100 pM and 1 nM, (Fig. 26B). Thus, at these low concentrations MetF remains monomeric. Higher concentrations of denatured MetF were analyzed by mixing labeled and unlabeled (M-UL) MetF molecules. A weak cross-correlation signal was seen at 100 nM of MetF concentration and a strong cross-correlation signal at 150 nM (Fig. 26B). As a positive control, we used a double-cysteine mutant of firefly luciferase protein, FLuc (D19C/S504C) which was labeled simultaneously with Alexa647 and Alexa532 (FLuc-DL) dye. A very strong cross-correlation signal was seen with 0.55 nM FLuc-DL. To estimate the sensitivity of this method in detecting aggregation, we analyzed a mixture of 50 pM FLuc-DL with 1 nM single-labeled MetF (0.5 nM MetF-647 and 0.5 nM MetF-532). We observed a clear cross-correlation signal (Fig. 26B), indicating that this method is very sensitive, allowing the detection of aggregate formation by 5%–10% of molecules. To measure spontaneous and chaperonin-assisted refolding under the single molecule conditions, we took advantage of an assay established previously that utilizes the change in diffusion coefficient (D) upon binding of non-native protein to GroEL (~800 kDa) [20, 71]. Non-native fluorescent-labeled MetF, i.e. protein that did not fold spontaneously, will bind to GroEL, resulting in slower diffusion (D=~34 $\mu m^2 s^{-1}$) while native MetF does not bind and diffuses rapidly (D=~64 $\mu m^2 s^{-1}$), as determined by fluorescence correlation spectroscopy (FCS) [20, 71, 80, 124](Fig. 26C). Attempted spontaneous refolding was performed at 100 pM MetF and reactions were stopped by adding 1 μM GroEL at different time points. Spontaneous refolding was not detectable, as GroEL

Results

recognized the non-native MetF (M-647) and shifted it to a slower diffusion time (Fig. 26D). Similarly, we performed assisted refolding of denatured M-647 by dilution at 100 pM into the buffer containing 1 µM GroEL, 2 µM GroES and 1 mM ATP. Folding reactions were stopped at different time points with apyrase to rapidly hydrolyze ATP (25 U). Assisted folding resulted in a time-dependent gradual shift of the slow-diffusing complex with GroEL to fast free diffusing native M-647 (Fig. 26D). Folded MetF subunits did not bind back to GroEL and remained monomeric, as assembly did not occur at 100 pM. Assisted refolding was GroES and ATP dependent and occurred with very fast kinetics (Fig. 26D), consistent with the results of the enzymatic assays (Fig. 20C).

Thus, upon dilution from denaturant, MetF (M-647) populates conformationally dynamic intermediate(s) (MetF-I), which remain monomeric at low concentration. MetF-I maintained competence for folding by GroEL/ES-ATP for at least 1 h after dilution from denaturant (Fig. 26E), based on diffusion time (ms) measurements by FCS (Fig. 26F).
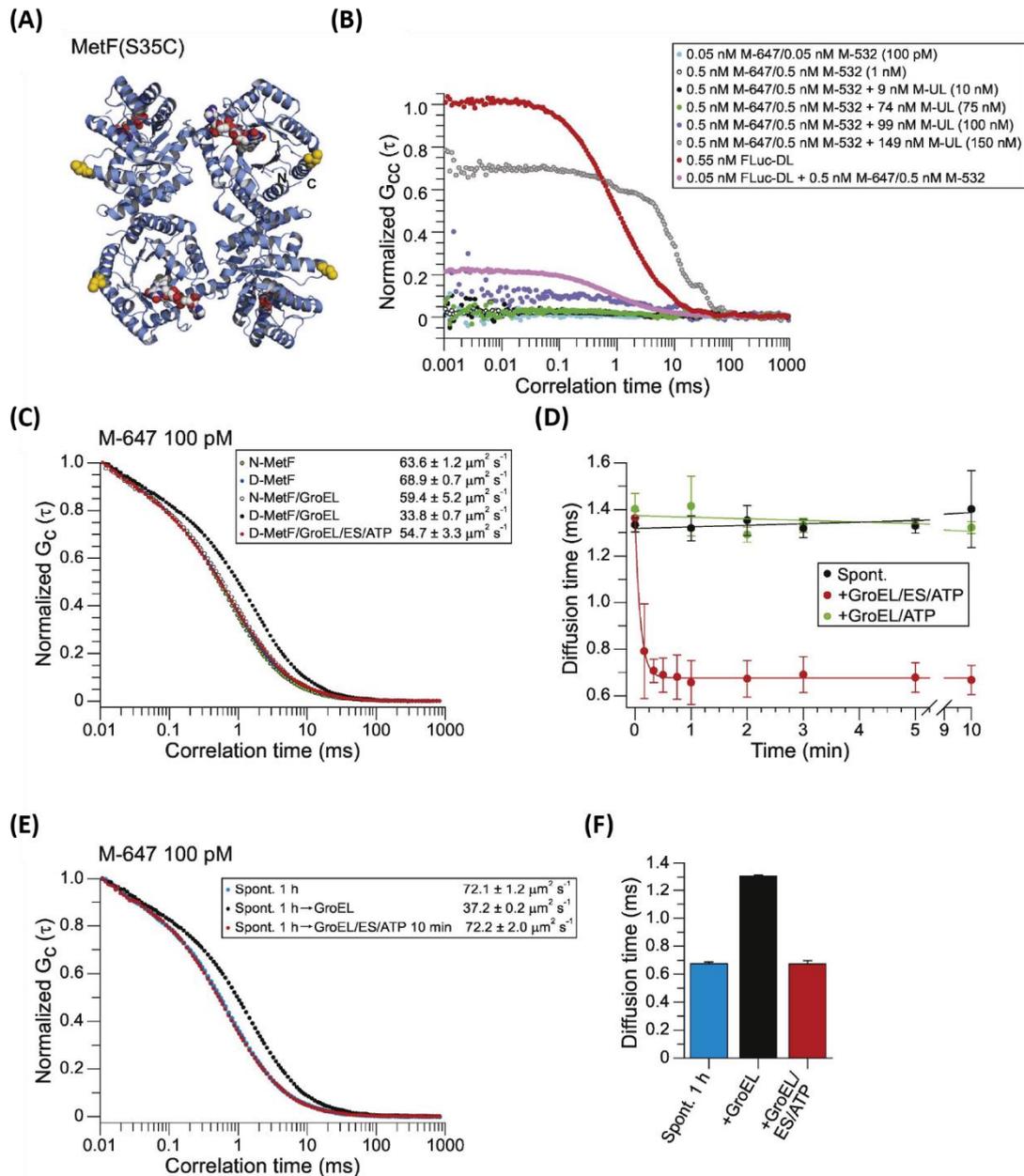
Figure 26: Analysis of spontaneous and GroEL/ES-assisted folding of MetF at single molecule level. (a) Crystal structure in ribbon representation of homotetrameric MetF (PDB: 1B5T) with FAD-bound at the C-termini of the β-strands in each subunit. The structure lacks the 20 N-terminal and two C-terminal residues. The position of the cysteine at position 35 used for Alexa dye labeling is indicated in yellow (space-filling representation). (b) MetF does not aggregate during spontaneous folding below 100 nM at 20°C. A 1:1 mixture of MetF(S35C) labeled with either Alexa647 (M-647) or Alexa532 (M-532) was denatured in 6M GuHCl/10mM DTT and diluted into refolding buffer to a final concentration of 100 pM to 1 nM, and intermolecular association was recorded using dc-FCCS. As a positive control, FLuc-DL (0.55 nM) labeled with both dyes was used to mimic the presence of dimeric species (red). Alternatively, MetF aggregation was analysed by gradually increasing the total protein concentration to 75, 100, or 150 nM by addition of unlabeled GuHCl-denatured MetF (M-UL). The dcFCCS signal G_{cc} (τ) of 50 pM FLuc-DL was also measured in the presence of equimolar mixture of single-labeled MetF (0.5 nM M-647 and 0.5 nM M-532) to show the sensitivity of the FCCS method in detecting aggregates.

Results

Samples were incubated at 25°C for 30 min before recording dc-FCCS for 30 min. Representative measurements of three independent repeats are shown. (c) Representative autocorrelation curves of Alexa 647 labeled native MetF (N-MetF) and denatured MetF (D-MetF) at 100 pM either in the presence or absence of 1 µM GroEL in refolding buffer as indicated. Additionally, GuHCl-denatured M-647 was diluted to 100 pM in a buffer containing 1 µM GroEL, 2 µM GroES and 1 mM ATP. FCS was measured after the end point of assisted folding (10 min) at 20°C. The diffusion coefficients (averages ± SD) are indicated. Representative measurements of three independent repeats are shown. (d) Spontaneous and assisted folding of M-647 in the absence of aggregation. Folding kinetics showing spontaneous and assisted folding of M-647 as measured by the average diffusion time obtained from the FCS curves, plotted against refolding time. M-647 was denatured in 6M GuHCl/10 mM DTT for 1 hr at 25°C. Spontaneous folding (Spont.) was initiated by 200-fold dilution of denatured M-647 into refolding buffer alone at a final concentration of 100 pM. Spontaneous folding was stopped at different time points by the addition of 1 µM GroEL. For assisted folding, denatured M-647 was diluted into the same buffer containing chaperonin (1 µM GroEL and 2 µM GroES) and folding was initiated by the addition of 1 mM ATP. Assisted folding was stopped by addition of excess apyrase (25 U). As a control, assisted folding was performed with GroEL and ATP but in the absence of GroES. The difference in diffusion time (ms) between GroEL-bound M-647 and free native-M-647 monomers in solution was analysed by FCS, resulting in rates of subunit folding. Error bars indicate SD values from three independent experiments. (e) GroEL can bind and rescue kinetically trapped folding intermediate, MetF-I, in the presence of GroES, and ATP after attempted spontaneous folding. GuHCl-denatured M-647 was diluted to 100 pM into buffer alone for spontaneous folding. After 1 h of incubation at 20°C, either 1 µM GroEL or 1 µM GroEL/2 µM GroES/1 mM ATP was added for 10 min. FCS was subsequently analysed for 30 min. Diffusion coefficients are indicated as averages ± SD from three independent experiments. Representative measurements are shown. (f) Diffusion times in ms of the data in (e). Error bars represent ± SD from three independent experiments.
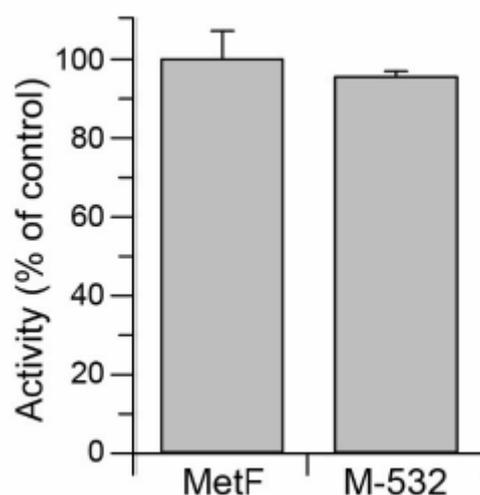
Figure 27: Homo-tetrameric MetF(S35C) mutant labeled with Alexa532 (M-532) is enzymatically active. Enzymatic activities were measured for 200 nM protein and are given in percent of wild-type control. Averages ± SD from 3 independent experiments are shown.

Results

4.7 Characterization of MetF inside the chaperonin GroEL/ES cage by H/DX-MS

As we have seen in enzymatic assays, the MetF subunit folds efficiently in a few ATP-driven GroEL/ES reaction cycles (Fig. 22,23). To provide more detailed information on the structural properties of MetF encapsulated in the GroEL/ES cage, we analyzed stably encapsulated MetF by hydrogen/deuterium-exchange measurements coupled to mass spectrometry (H/DX-MS) at peptide resolution. The degree of deuterium uptake correlates with structural flexibility, as backbone amide hydrogens are shielded from exchange when involved in stable secondary structure (α-helices and β-sheets), parts of the protein buried in the core of a folded protein or at a protein-protein interfaces [14, 17, 125, 126].

Stable encapsulation inside the GroEL: GroES$_2$ cavity was confirmed by demonstrating MetF protection against digestion by proteinase K (PK), whereas the GroEL-bound MetF complex was highly sensitive to PK digestion (Fig. 30).

Native MetF (with bound FAD), GroEL-bound MetF and MetF inside the GroEL: GroES$_2$ cage were diluted 10-fold into D$_2$O buffer for increasing periods of time (10-1000 sec) to allow for exchange of exposed hydrogens to deuterium [126]. At different time points, the exchange reactions were quenched by lowering the pH and temperature (pH ~2.5 and 0°C) to minimize the exchange of exposed backbone hydrogens to deuterium, followed by pepsin digestion and analysis of deuterium uptake into individual peptides by liquid chromatography-ion mobility spectrometry-mass spectrometry (LC-IMS-MS) (Fig. 28A). Note that we could not analyze MetF-I because it aggregates at the higher concentrations required for H/DX-MS.

We analyzed 109 unique and overlapping peptides of MetF in the native (92.6% sequence coverage), 47 peptides (89.2% sequence coverage) in the GroEL-bound, 49 peptides (87.8% sequence coverage) in the encapsulated state with bound FAD and 37 peptides (77% sequence coverage) in the encapsulated state without FAD (Fig. 31). All the peptides analyzed displayed unimodal exchange kinetics (i.e., a single binomial isotope distribution; Figure 32 and PRIDE identifier PXD016666), consistent with a single population of protein molecules [125, 127]. Unimodal mass spectra indicate that the MetF protein did not undergo large folding-unfolding transitions during deuterium labeling (Fig. 32). Native MetF (with bound FAD) exhibited overall low deuterium incorporation indicative of stable secondary structure in the protein (Fig. 28B). In contrast, the GroEL-bound MetF was globally destabilized compared to native MetF (Fig.

68

Results

28C). Similar properties were previously reported for other GroEL-bound substrates [71, 119], however in some cases residual structure was detected in the bound state [128, 129].

Notably, encapsulated MetF inside the GroEL: GroES$_2$ cavity in the presence of FAD showed almost the same low level of deuterium uptake as native MetF, except for the regions that are present at the strong dimer interface (peptides 239–255, 275–283, and 284–296) (Fig. 28D). All these peptides showed high and rapid deuterium incorporation relative to the native tetramer. Note that FAD molecules can diffuse into the GroEL: GroES$_2$ cage through windows in the GroEL structure. To provide additional evidence that the active site of MetF was correctly formed inside the chaperonin cavity, we also analyzed the encapsulated MetF inside the GroEL: GroES$_2$ in the absence of FAD (Fig. 29). Without FAD, several peptides were considerably destabilized, including peptides 38-48 and 150-166, which are involved in FAD binding sites (Fig. 29B, C). These peptides exhibited increased levels of deuterium incorporation without FAD, but in the presence of FAD showed exchange properties almost similar to the FAD-bound native MetF tetramer. In contrast, peptides 239–255 and 284–296 at the strong dimer interface were FAD insensitive (Fig. 29D, E). In conclusion, upon folding inside the GroEL:ES$_2$ cage, individual subunits of MetF reach a native like conformation with a fully formed active site that is competent in binding FAD. These folded MetF monomers are released from GroEL/ES into the solution for assembly of homo-tetramers.

Results
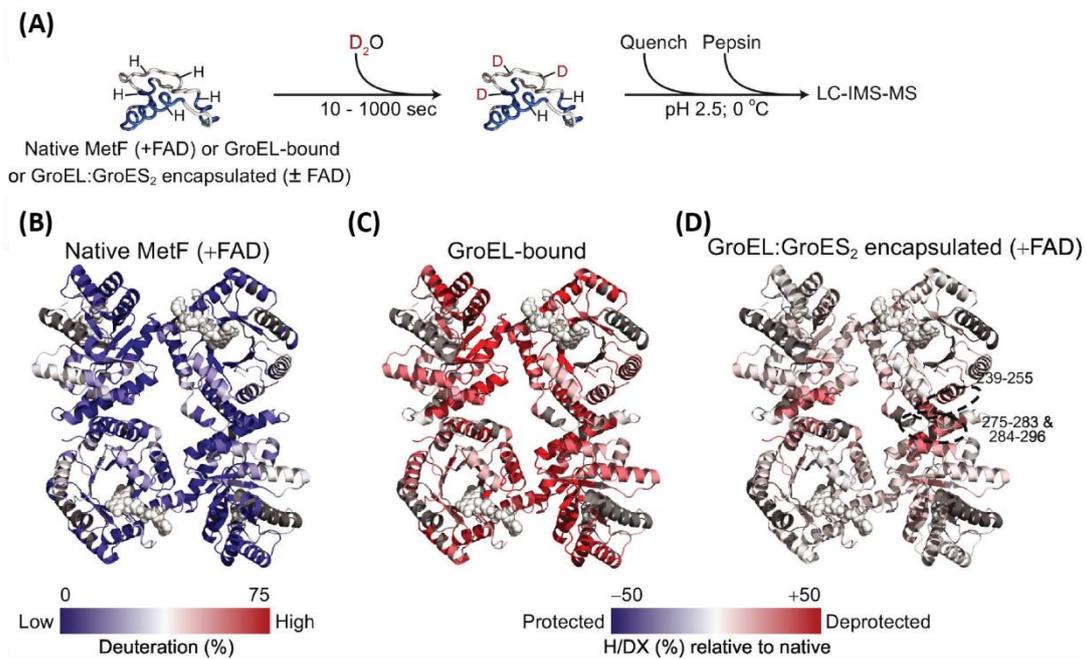
Figure 28: Conformational dynamics of GroEL-bound and GroEL: GroES$_2$ encapsulated MetF. (a) Schematic representation of the work-flow of the equilibrium H/DX-MS experiment. (b) Peptide-level deuterium exchange of native MetF (+FAD) after exposure to deuterium for 100 s. Relative fractional deuterium exchange for each peptide is mapped onto the ribbon structure of MetF (PDB: 1B5T) as a gradient from blue (0%) to red (75%). (c–d) Difference in deuterium exchange between native MetF and GroEL-bound (c) or GroEL:ES$_2$ encapsulated MetF (d), following exposure to deuterium for 100 sec. Deuteration differences are scaled from blue (−50%) to red (+50%). Red colored regions are deprotected when MetF is bound to GroEL or encapsulated in the symmetric GroEL:ES$_2$ complex. Peptides 239–255, 275–283, and 284–296 at the strong subunit interface are indicated. H/DX data are the average of three independent repeats.

Results
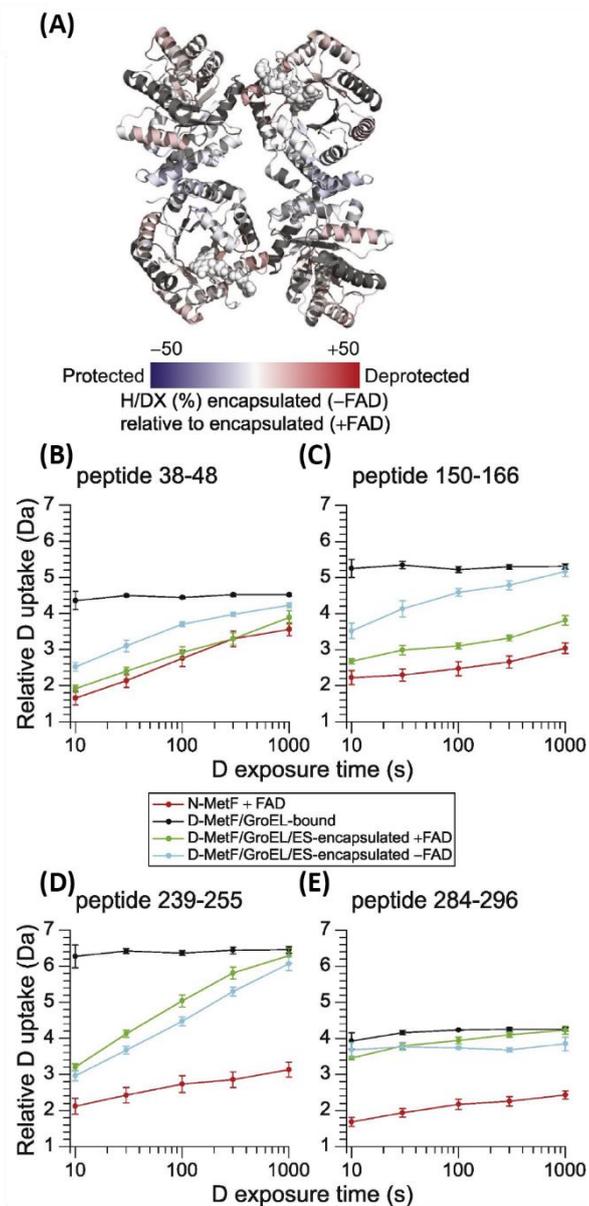
Figure 29: MetF is able to bind FAD inside the chaperonin GroEL: GroES$_2$ cage.
(a) Difference in deuterium exchange between GroEL: ES$_2$ encapsulated MetF prepared in the presence or absence of FAD, following exposure to deuterium for 100 sec. Deuteration differences are scaled from blue (−50%) to red (+50%). Red colored regions are deprotected in the absence of FAD. (b–e) Deuterium uptake plots for MetF peptides 38–48 (b), 150–156 (c), 239–255 (d), and 284–296 (e). Relative deuterium uptake is shown as a function of deuterium exposure time for MetF in the native state (red), bound to GroEL (black), encapsulated by GroEL:ES$_2$ in the presence of FAD (green), and encapsulated by GroEL:ES$_2$ in the absence of FAD (cyan). Averages ± SD from three independent experiments are shown.
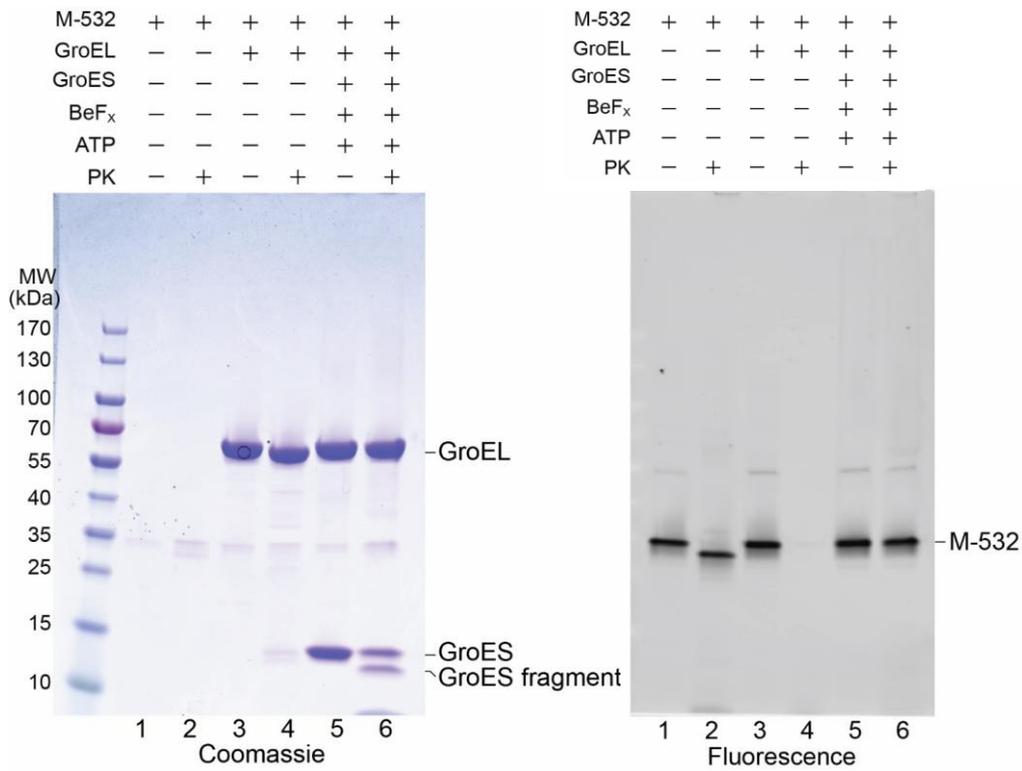
Figure 30: Stable encapsulation of MetF inside the chaperonin GroEL: ES$_2$ cage. Denatured M-532 (150 nM) was bound to GroEL (0.6 μM) upon dilution from denaturant. When indicated, GroES (2.4 μM) was added in the presence of BeFx and ATP to generate stable GroEL: ES$_2$ complexes with encapsulated MetF (see Methods for details). Proteinase K (PK) treatment was performed for 10 min at 25°C, followed by addition of PMSF to stop PK action. Native M-532 was analyzed as control. Reactions were separated by SDS-PAGE followed by Coomassie staining to detect GroEL and GroES (left), and by fluorescence imaging to detect M-532 (right). Note that GroEL-bound GroES is PK-resistant, whereas free GroES is fragmented with PK. Native M-532 is largely PK resistant, as expected for the removal of 20 amino acids from the N-terminus (lanes 1, 2). GroEL-bound M-532 is PK sensitive (lanes 3, 4) while the encapsulated protein is protected (lanes 5, 6). Representative data from 3 independent experiments are shown.

Results

(A) Native MetF +FAD

XSFFHASQRDALNQSLAEVQGQINVSFEFFPPRTSEMEQTLWNSIDRLSSLKPKFVSVTYGANSGERDRTHSIIKGIKDRTGLEAAPHLTCIDATPDELR
5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95   100

TIARDYWNNGIRHIVALRGDLPPGSGKPEMYASDLVTLLKEVADFDISVAAYPEVHPEAKSAQADLLNLKRKVDAGANRAITQFFFDVESYLRFRDRCVS
105   110   115   120   125   130   135   140   145   150   155   160   165   170   175   180   185   190   195   200

AGIDVEIIPGILPVSNFKQAKKFADMTNVRIPAWMAQMFDGLDDDAETRKLVGANIAMDMVKILSREGVKDFHFYTLNRAEMSYAICHTLGVRPGL
205   210   215   220   225   230   235   240   245   250   255   260   265   270   275   280   285   290   295

Total: 109 peptides, 92.6% coverage, 3.88 redundancy

(B) GroEL-bound +FAD

XSFFHASQRDALNQSLAEVQGQINVSFEFFPPRTSEMEQTLWNSIDRLSSLKPKFVSVTYGANSGERDRTHSIIKGIKDRTGLEAAPHLTCIDATPDELR
5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95   100

TIARDYWNNGIRHIVALRGDLPPGSGKPEMYASDLVTLLKEVADFDISVAAYPEVHPEAKSAQADLLNLKRKVDAGANRAITQFFFDVESYLRFRDRCVS
105   110   115   120   125   130   135   140   145   150   155   160   165   170   175   180   185   190   195   200

AGIDVEIIPGILPVSNFKQAKKFADMTNVRIPAWMAQMFDGLDDDAETRKLVGANIAMDMVKILSREGVKDFHFYTLNRAEMSYAICHTLGVRPGL
205   210   215   220   225   230   235   240   245   250   255   260   265   270   275   280   285   290   295

Total: 47 peptides, 89.2% coverage, 1.69 redundancy

(C) GroEL/ES ATP/BeF$_X$ -encapsulated +FAD

XSFFHASQRDALNQSLAEVQGQINVSFEFFPPRTSEMEQTLWNSIDRLSSLKPKFVSVTYGANSGERDRTHSIIKGIKDRTGLEAAPHLTCIDATPDELR
5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95   100

TIARDYWNNGIRHIVALRGDLPPGSGKPEMYASDLVTLLKEVADFDISVAAYPEVHPEAKSAQADLLNLKRKVDAGANRAITQFFFDVESYLRFRDRCVS
105   110   115   120   125   130   135   140   145   150   155   160   165   170   175   180   185   190   195   200

AGIDVEIIPGILPVSNFKQAKKFADMTNVRIPAWMAQMFDGLDDDAETRKLVGANIAMDMVKILSREGVKDFHFYTLNRAEMSYAICHTLGVRPGL
205   210   215   220   225   230   235   240   245   250   255   260   265   270   275   280   285   290   295

Total: 49 peptides, 87.8% coverage, 1.79 redundancy

(D) GroEL/ES ATP/BeFₓ -encapsulated -FAD

X S F F H A S Q R D A L N Q S L A E V Q G Q I N V S F E F F P P R T S E M E Q T L W N S I D R L S S L K P K F V S V T Y G A N S G E R D R T H S I I K G I K D R T G L E A A P H L T C I D A T P D E L R
5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95   10(

T I A R D Y W N N G I R H I V A L R G D L P P G S G K P E M Y A S D L V T L L K E V A D F D I S V A A Y P E V H P E A K S A Q A D L L N L K R K V D A G A N R A I T Q F F F D V E S Y L R F R D R C V S
105   110   115   120   125   130   135   140   145   150   155   160   165   170   175   180   185   190   195   200

A G I D V E I I P G I L P V S N F K Q A K K F A D M T N V R I P A W M A Q M F D G L D D D A E T R K L V G A N I A M D M V K I L S R E G V K D F H F Y T L N R A E M S Y A I C H T L G V R P G L
205   210   215   220   225   230   235   240   245   250   255   260   265   270   275   280   285   290   295

Total: 37 peptides, 77.0% coverage, 1.53 redundancy

Figure 31: H/DX-MS peptide coverage maps.

(A–D) Peptide coverage maps were generated in DynamX for native MetF (A), GroEL-bound MetF (B), GroEL:ES₂ encapsulated MetF with FAD (C) and GroEL:ES₂ encapsulated MetF without FAD (D). Peptides are from 3 independent experiments. The redundancy score is a measure of the average number of peptides that cover each residue in the sequence.
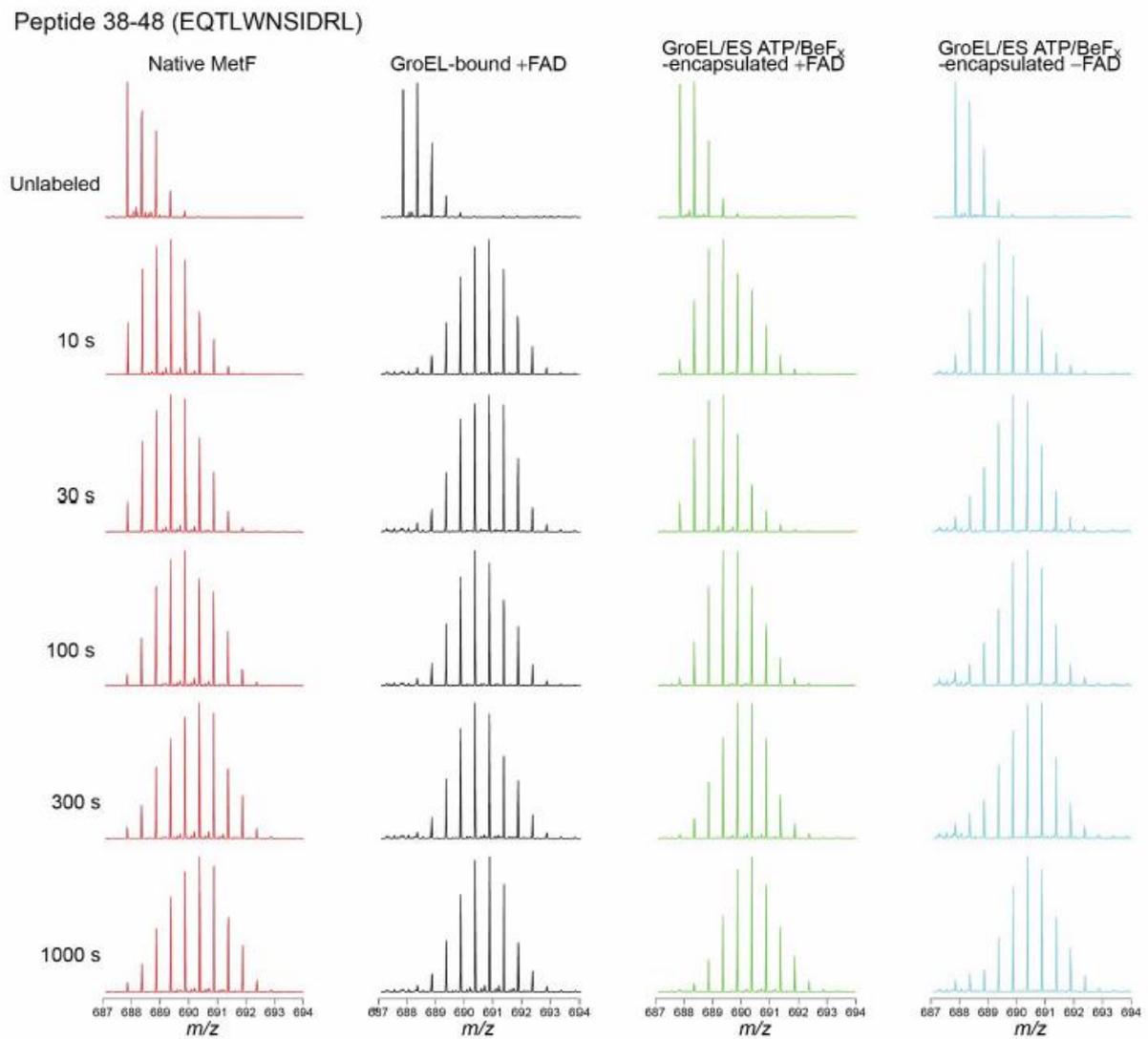
Figure 32: Deuterium incorporation into peptide 38-48 of MetF.

Mass spectra of peptide 38-48 after different times of deuterium exposure (10-1000 sec as in Figs. 28 and 29) of native homotetrameric MetF, GroEL-bound MetF (with FAD), GroEL: GroES$_2$ encapsulated MetF with FAD and GroEL: GroES$_2$ encapsulated MetF without FAD. Representative data from 3 independent experiments are shown.

Results

4.8 Functional role of negative charge clusters of the GroEL/ES cavity wall in protein folding

An effect of steric confinement and the negative charges of the GroEL/ES cavity wall have been suggested to be critical in the capacity of GroEL/ES to accelerate the folding of encapsulated protein by smoothing the folding energy landscape [20, 71, 79, 80, 89, 130]. Previous studies have shown that the negative net charge of the GroEL/ES cavity influences the substrates folding rate, since GroEL mutants with altered cavity charge were no longer able to accelerate folding [20, 79, 80]. The inner wall of the GroEL/ES *cis* cavity has a net charge of -42 (189 negatively and 147 positively charged amino acid residues). A number of negative charges (residues E252, D253, E255, D359, D361 and E363) are positioned in the apical domain in two circular layers (Fig. 33A). Most of these residues are highly conserved among GroEL homologs, although they do not play an apparent role in substrate and GroES binding [131, 132]. In our study, we explored the functional significance of these charges in promoting folding by analyzing a GroEL mutant, EL-KKK2, in which three residues (D359, D361, E363) are replaced by lysine so that the net charge of the GroEL cavity is zero [79]. EL-KKK2 has been shown to have similar substrate binding and encapsulation efficiency as wild-type GroEL for various substrate proteins [20, 79, 130]. Accordingly, EL-KKK2 prevented the aggregation of denatured MetF upon dilution from denaturant as analyzed by turbidity assay (Fig. 20B). However, EL-KKK2/ES was substantially less efficient in catalyzing MetF folding, resulting in an at least 13-times slower folding rate ($t_{1/2}$ ~ 3.5 min) compared to wild-type GroEL/ES (Fig. 33B). The folding yield was only somewhat reduced, presumably reflecting the higher number of folding cycles resulting in some loss of released MetF to aggregation. This finding demonstrates that the negative charge character of the GroEL cavity wall is critical for rapid folding of MetF [20, 79, 130]. Notably, similar to other obligate chaperonin substrates, such as Rubisco and DapA, that experience accelerated folding upon encapsulation, MetF carries a negative surface net charge (of -3.5). Thus, charge repulsion effects with the negatively charged cavity wall are important in providing a non-interacting environment for rapid folding of MetF.

Further experimental and theoretical studies will be required to describe the detailed mechanism by which the physicochemical properties of the chaperonin cavity allow highly efficient folding of obligate chaperonin substrate proteins.

(A)

(B)
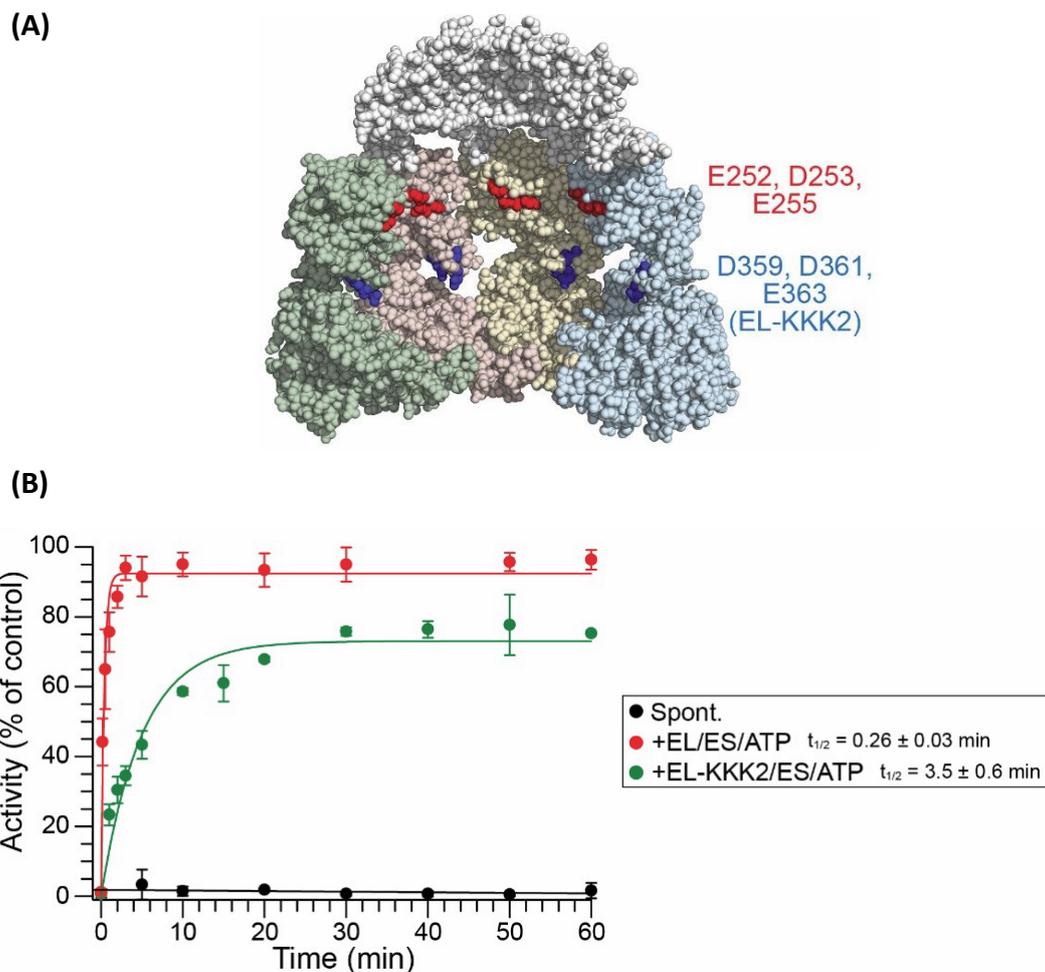
Figure 33: (A) Space-filling model of the GroEL/GroES-(ADP)$_7$ complex (PDB: 1AON) offering a view into the chaperonin cavity in the GroES-bound state. Clusters of negatively charged residues in the apical domain exposed toward the cavity are highlighted in red (E252, D253, E255) and blue (D359, D361, E363). (B) The net negative charge is required for rapid folding of MetF in the GroEL/ES cage. Spontaneous and assisted folding by GroEL/ES or EL-KKK2/ES was performed at 25°C as in Figure 20C. Averages ± SD from 3 independent experiments are shown.

5 Discussion

In the present study, we have investigated how the *E. coli* chaperonin system can actively promote folding beyond preventing aggregation of its endogenous substrate protein 5,10-methylenetetrahydrofolate reductase (MetF), a tetrameric TIM-barrel enzyme involved in methionine biosynthesis. Interestingly, several obligate substrates of GroEL share the $(\beta/\alpha)_8$TIM-barrel fold and are prone to populate kinetically trapped intermediate states, but how the GroEL/ES system mediates the folding of these proteins remained unclear. The key question is why are some proteins highly dependent on GroEL/ES for folding while others are not. MetF proved to be a suitable substrate to address this question as its pathway of spontaneous folding involves a deeply kinetically trapped intermediate. Here, we set out to differentiate between different models proposed for the mechanism of GroEL/ES-assisted folding. By comparing the spontaneous and GroEL/ES assisted folding of MetF, we found that GroEL/ES catalyzes rapid and efficient folding of denatured MetF upon encapsulation inside the chaperonin nanocage. During spontaneous folding, MetF populates a kinetically trapped folding intermediate(s) (MetF-I) that is associated with a high entropic barrier and is unable to proceed to the native state, even in the absence of aggregation.

How does GroEL/ES catalyze MetF folding? H/DX data suggests that GroEL binds MetF-I and stabilizes it in an unfolded state. Binding of ATP and GroES to the MetF-bound GroEL ring (the *cis*-ring) causes the displacement of MetF into the central GroEL/ES cavity. More than 50% of protein molecules fold in one round of encapsulation. After ATP hydrolysis in the *cis*-ring, binding of ATP to the opposite GroEL *trans*-ring then induces an allosteric signal that causes dissociation of GroES and ADP. Folded monomer is released for assembly while misfolded protein is rapidly recaptured for another folding cycle. These findings provide support for the "active cage" mechanism of chaperonin action.

Importantly, the environment of the GroEL/ES cage is highly conducive to folding. In contrast to refolding in bulk solution, MetF-I does not accumulate inside the cage and instead folding occurs with high efficiency, with only ~3 GroEL/ES reaction cycles being required for >90% folding yield. The remarkable efficiency of this reaction, revealed by studying the cognate substrate MetF, suggests that GroEL/ES and its substrates have co-evolved.

Discussion

Early studies with heterologous model substrates showed that GroEL/ES assisted folding requires a large number of protein encapsulation cycles (~25 cycles for 50% folding of Rubisco from the bacterium *Rhodospirillum rubrum*; ~40 cycles for mitochondrial Rhodanese) [79, 89, 115, 133], with only a few percent (2%–5%) of protein reaching the native state each round. These results suggested that GroEL/ES is a rather inefficient machine consuming high amounts of ATP [115]. That the physical properties of the chaperonin cage likely allow for additional roles in actively promoting folding beyond prevention of aggregation has now been seen for several endogenous and heterologous substrate proteins [71, 79, 80, 88, 89, 93]. Our data supports a mechanism in which the physical environment of the GroEL/ES cavity catalyzes the folding reaction by lowering the entropic component of the folding energy barrier. Thus, the GroEL/ES cage acts as a folding catalyst for a subset of proteins that otherwise fail to convert to the native state at a biologically relevant timescale. This kinetic effect is critical in adjusting the rate of folding to the speed of translation. Additional studies with different endogenous substrates will be useful in further establishing the general mechanistic principle underlying the function of GroEL/ES as a folding catalyst.

5.1 Catalysis of folding by the GroEL/ES chaperonin is biologically relevant

Previous studies have shown that the GroEL/ES complex can accelerate the rate of protein folding [20, 71, 88]. However, general conclusions as to the underlying mechanism remained difficult. Specifically, a clear distinction between effects of preventing aggregation from active promotion of folding could not be achieved. Here we used the *E. coli* protein-MetF, an endogenous obligate substrate of the chaperonin system, to carry out such an analysis.

A key finding was that MetF fails to fold spontaneously under standard folding conditions and forms a kinetically trapped intermediate state, even in the absence of aggregation. These intermediates are captured by GroEL/ES for rapid folding.

To exclude aggregation, we compared the spontaneous and assisted folding of MetF by FCS at 100pM. At this low concentration, MetF populates the monomeric intermediate state (s) (MetF-I), which remains competent for GroEL/ES assisted folding for at least 1 hr after dilution from denaturant (Fig. 26E). We characterized MetF-I by using intrinsic tryptophan (Trp) fluorescence. Native MetF shows a Trp fluorescence emission maximum (λ_{max}) at 345 nm, while the fully denatured MetF displays a redshift in fluorescence to a λ_{max} of 356 nm with a

79

Discussion

relatively high fluorescence intensity. MetF-I, formed upon dilution from denaturant has a λ_{max} of 345 nm with high fluorescence intensity (Fig. 24C). This suggests that the Trp residues of MetF-I are collapsed into a more hydrophobic environment, as observed for compact, molten globule-like folding intermediates [117].

Catalysis of protein folding by GroEL/ES is highly biologically relevant. About 45% of the obligate substrate proteins of GroEL, including MetF, share the TIM-barrel domain [12, 70] and many of these proteins were shown to either undergo degradation or to accumulate in aggregates in *E. coli* cells when GroEL is depleted [12, 35]. Among these, MetF also aggregates upon depletion of DnaK/DnaJ/GrpE in *E. coli* [12, 35]. The results obtained in this study provide mechanistic insight into the interplay between Hsp70 and the chaperonin system inside the cell. We suggest that after release from the ribosome, MetF, presumably as MetF-I is protected from aggregation and maintained in a folding-competent state by DnaK/DnaJ/GrpE, which then transfers it to the downstream chaperonin for subsequent folding [12, 39]. Under optimal growth conditions at 37°C, *E. coli* cells divide every ~20 minutes [134]. Thus, GroEL/ES assisted protein folding must be completed faster than the time of synthesis (~14 sec, assuming a translation rate of 20 amino acids/sec), to avoid the build-up of unfolded protein and making efficient use of available chaperonin capacity [71].

5.2 GroEL/ES modifies folding pathway of encapsulated protein

The nature of the protein folding pathway inside the GroEL/ES *cis*-cavity remains poorly characterized. Particularly, it is unclear whether polypeptide folding inside the chaperonin cage takes the same route to the native state as in free solution [135, 136].

Other unanswered questions concern the differences in chaperonin-assisted folding behavior of endogenous and heterologous substrate proteins. For example, why do only a few percent of ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) or mitochondrial malate dehydrogenase (MDH), fold to the native state in any given round of encapsulation by *E. coli* GroEL/ES? What are the conformational differences between proteins that utilize GroEL/ES productively and others that do not? And how does the non-native state look structurally after the release at the end of each GroEL/ES reaction cycle into the bulk solution?

To begin to address these questions, we analysed the GroEL/ES assisted folding of MetF using continuous deuterium exchange (H/DX-MS) at peptide resolution.

Discussion

Overall native, assembled MetF (with bound FAD) exhibits low deuterium uptake that indicates stable secondary and tertiary structure (Fig. 28B). In contrast, the GroEL-bound MetF was globally destabilized (Fig. 28C). Notably, the MetF subunit stably encapsulated in a non-cycling GroEL-ES$_2$ complex in the presence of FAD showed deuterium uptake properties equivalent to those of native MetF, except for the peptide regions at the strong dimer interface (peptides 239–255, 275–283, and 284–296) (Fig. 28D), which acquire exchange protection only upon subunit assembly. We also confirmed that the active site of MetF was correctly folded inside the GroEL/ES cage. In the absence of FAD, some peptides were destabilized, including the regions comprising peptides 38-48 and 150-166 (Fig. 29B, C), which are involved in the FAD binding site. Thus, after folding inside the GroEL/ES cage, the MetF subunit reaches a native-like conformation with a fully formed active site that can bind FAD. Finally, these folded MetF monomers are releases into solution for assembly to native homo-tetramer. The intermediate state MetF-I, populated under conditions of attempted refolding in bulk solution, could not be analyzed by H/DX-MS, as it aggregates at the concentrations required for these measurements. Conformational stretching upon binding to GroEL may prepare the MetF-I for optimal folding inside the cage, in which repartitioning to kinetically trapped intermediate is avoided and the folding reaction proceeds unimpaired by aggregation.

How does GroEL/ES catalyze protein folding? Theory predicts that steric confinement in a repulsive (net-negatively charged) chaperonin cage can accelerate folding by one to two orders of magnitude by smoothing the folding energy landscape [137-139].

Our previous studies have shown that the physical properties inside the chaperonin cage play an important role in substrate folding. Specifically, GroEL mutants with altered cavity net charge are unable to accelerate folding of several substrate proteins, indicating that the rate of protein folding in the GroEL/ES cage depends on the net negative charge of the cage wall [79]. The highly charged character of the GroEL/ES cavity may enhance the hydrophobic effect by ordering water molecules connected with the cavity wall [79, 140], although experimental evidence for the existence of a hydration shell is still missing [94]. In our study, we investigated folding using a GroEL mutant (EL-KKK2), in which the normal cavity wall net charge of -42 is removed to 0. We observed that EL-KKK2/ES was much less efficient in catalyzing MetF folding, with the folding rate at least 13 times slower ($t_{1/2}$ ~ 3.5 min) than that of wild-type GroEL/ES (Fig. 33B). These findings are consistent with previous reports [79, 94] that the net negative

81

charge of the GroEL cage provides a physical environment conducive to folding, probably associated with an effect of entropic confinement by the restricted space limiting the conformational flexibility of the enclosed folding intermediates [79, 80, 89, 137-139]. Conformationally dynamic folding intermediates of proteins with complex domain topologies such as the TIM-barrel- that have to cross a high entropic component of the folding energy barrier toward the native state would benefit the most from such constraining effects of confinement, resulting in a reduced search time for the formation of native contacts [39, 71, 89, 137-139]. Furthermore, during folding, the encapsulated substrate may also be remodeled by flexible C- terminal sequences emanating from the equatorial domains of GroEL [59, 67, 79, 88, 141]. Our findings are consistent with previous studies, which showed that the repulsive cage environment of the GroEL/ES results in acceleration of substrate folding, whereas potential interactions between the encapsulated protein Rhodanese and the GroEL cavity walls induces a strong reduction of the folding rates [138, 142, 143].

However, further studies are required to understand the exact mechanism of how the physical environment of the chaperonin cage influences protein folding.

5.3 Escape from chaperonin dependence

Interestingly, although MetF is an essential metabolic enzyme in *E. coli,* it is absent in certain bacteria of the genus *Mycoplasma* and *Ureaplasma* that lack GroEL/ES. MetF differs in this regard from the TIM-barrel protein N-acetylneuraminic acid aldolase (NanA). While *E. coli* NanA is an obligate GroEL-dependent substrate, its ortholog in *Mycoplasma synoviae* folds efficiently without chaperonin [71]. MsNanA is enriched with aromatic residues in the hydrophobic core and several α-helices of the TIM-barrel domain contain numerous solvent exposed lysines, which enhance α-helical propensity [71]. These structural features may explain the ability of MsNanA to fold without chaperonin. Similar adaptations in MetF may interfere with FAD binding.

Furthermore, highly homologous proteins (orthologs) may display a strong adaptation to their cognate chaperonin. A striking example is the large subunit (RbcL) of plant Rubisco, which cannot utilize the *E. coli* GroEL/ES system and requires the chloroplast chaperonin Cpn60 for folding [144], although plant RbcL is highly homologous to cyanobacterial RbcL, which is GroEL/ES-dependent [145, 146]. While these examples emphasize the importance of the

Discussion

chaperonin system in protein evolution [147, 148], the optimization of folding of a specific protein by substrate chaperonin coevolution is restricted by the need of GroEL/ES to maintain the capacity to fold numerous different proteins [149]. In *E. coli*, TIM-barrel proteins including MetF, DapA, and NanA form a subset of GroEL substrates [12, 39, 70, 71, 113], which may have succeeded in optimizing their chaperonin-mediated folding to a greater extent than proteins with other complex topologies that are less frequent among the obligate substrates.

6 Conclusion

The mechanism by which the GroEL/ES chaperonin system mediates protein folding has been subject to intensive investigation for more than two decades. While initial models suggested that the GroEL/ES nanocage acts purely as a passive anti-aggregation device, more recent reports indicate that the chaperonin can substantially accelerate folding reactions. How folding energy landscapes are modulated by GroEL/ES remains only partially understood. In the present study, we provided strong evidence for the "active cage" model of GroEL/ES using the obligate GroEL substrate protein MetF as an example. We show that MetF is unable to fold to the native state spontaneously under standard refolding conditions, even in the complete absence of aggregation. Thus, aggregation prevention by chaperonin cannot explain how GroEL/ES promotes MetF folding. In contrast to folding in bulk solution, which results in the population of an unproductive folding intermediate (MetF-I), encapsulation of non-native MetF subunit in the chaperonin cage results in rapid folding to a native-like state, as judged by its H/DX properties and the ability to bind the cofactor FAD.
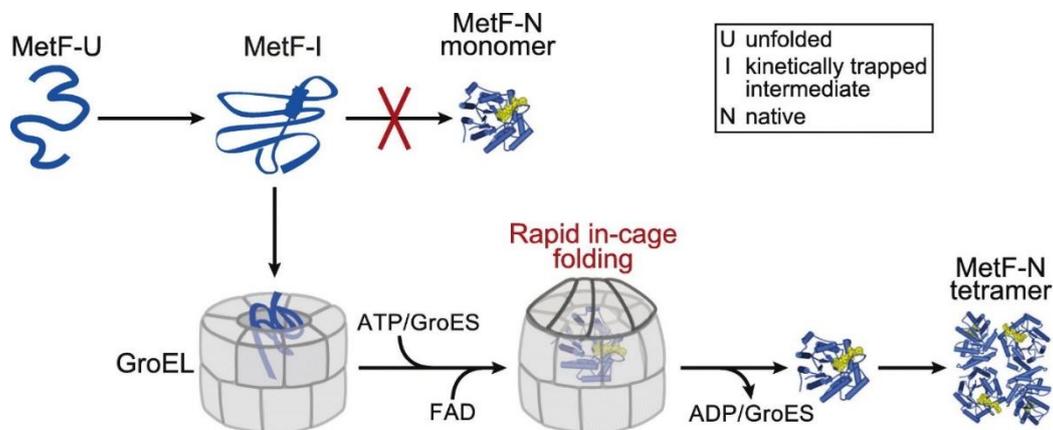
Figure 34: Proposed model of MetF folding in the GroEL/ES chaperonin cage.

The GroEL/ES reaction cycle is initiated by binding of non-native MetF, exposing hydrophobic residues, to the apical domains of an open GroEL ring. Binding of 7 ATP molecules and GroES results in displacement of substrate into the central GroEL cavity. Encapsulated substrate is now free to fold unimpaired by aggregation within the hydrophilic cage of GroEL/ES for the time required to hydrolyse 7 ATP in the GroEL ring that interacts with GroES (~5-10 sec at 25°C) [20]. The GroEL cage actively promotes MetF folding to the FAD-binding native state by reducing the entropic component of the energy barrier. Binding of 7 ATP and GroES to the opposing ring triggers release of ADP and GroES,

84

Conclusion

completing the cycle. Folded MetF monomer is released for assembly into the cytosol, while incompletely folded or misfolded protein is rapidly recaptured for another round of folding upon encapsulation.

Folding is efficiently promoted by the physical environment of the GroEL/ES cage, with the net-negative charge of the cavity wall having a critical role. In addition to the negative charges exposed on the wall of the folding cage, the spatial restriction experienced by encapsulated folding intermediates as well as the flexible C-terminal sequences protruding from GroEL into the cavity are likely to contribute to promoting folding. The high efficiency of folding observed in a single round of encapsulation - ~50% of MetF molecules fold in a single round of encapsulation excludes repetitive cycles of binding and unfolding of kinetically trapped states by GroEL as a mechanistic requirement of folding. Further studies are required to define the exact mechanism by which the chaperonin cage catalyzes the folding of proteins like MetF.

7 References

1. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(4096): p. 223-30.
2. Radford, S.E., *Protein folding: progress made and promises ahead.* Trends Biochem Sci, 2000. **25**(12): p. 611-8.
3. Nickson, A.A. and J. Clarke, *What lessons can be learned from studying the folding of homologous proteins?* Methods, 2010. **52**(1): p. 38-50.
4. Ivarsson, Y., et al., *Mechanisms of protein folding.* Eur Biophys J, 2008. **37**(6): p. 721-8.
5. Udgaonkar, J.B., *Multiple routes and structural heterogeneity in protein folding.* Annu Rev Biophys, 2008. **37**: p. 489-510.
6. Dill, K.A., et al., *The protein folding problem.* Annu Rev Biophys, 2008. **37**: p. 289-316.
7. Kim, Y.E., et al., *Molecular chaperone functions in protein folding and proteostasis.* Annu Rev Biochem, 2013. **82**: p. 323-55.
8. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nat Struct Biol, 1997. **4**(1): p. 10-9.
9. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *In vivo aspects of protein folding and quality control.* Science, 2016. **353**(6294): p. aac4354.
10. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884-90.
11. Jackson, S.E., *How do small single-domain proteins fold?* Fold Des, 1998. **3**(4): p. R81-91.
12. Kerner, M.J., et al., *Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli.* Cell, 2005. **122**(2): p. 209-20.
13. Jahn, T.R. and S.E. Radford, *Folding versus aggregation: polypeptide conformations on competing pathways.* Arch Biochem Biophys, 2008. **469**(1): p. 100-17.
14. Wales, T.E. and J.R. Engen, *Hydrogen exchange mass spectrometry for the analysis of protein dynamics.* Mass Spectrom Rev, 2006. **25**(1): p. 158-70.
15. Engen, J.R., *Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS.* Anal Chem, 2009. **81**(19): p. 7870-5.
16. Morgan, C.R. and J.R. Engen, *Investigating solution-phase protein structure and dynamics by hydrogen exchange mass spectrometry.* Curr Protoc Protein Sci, 2009. **Chapter 17**: p. Unit 17 6 1-17.
17. Marcsisin, S.R. and J.R. Engen, *Hydrogen exchange mass spectrometry: what is it and what can it tell us?* Anal Bioanal Chem, 2010. **397**(3): p. 967-72.
18. Pirrone, G.F., R.E. Iacob, and J.R. Engen, *Applications of hydrogen/deuterium exchange MS from 2012 to 2014.* Anal Chem, 2015. **87**(1): p. 99-118.
19. Mashaghi, A., et al., *Chaperone action at the single-molecule level.* Chem Rev, 2014. **114**(1): p. 660-76.
20. Gupta, A.J., et al., *Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level.* J Mol Biol, 2014. **426**(15): p. 2739-54.
21. Schuler, B. and H. Hofmann, *Single-molecule spectroscopy of protein folding dynamics--expanding scope and timescales.* Curr Opin Struct Biol, 2013. **23**(1): p. 36-47.
22. Ries, J. and P. Schwille, *Fluorescence correlation spectroscopy.* Bioessays, 2012. **34**(5): p. 361-8.
23. Fitzpatrick, J.A. and B.F. Lillemeier, *Fluorescence correlation spectroscopy: linking molecular dynamics to biological function in vitro and in situ.* Curr Opin Struct Biol, 2011. **21**(5): p. 650-60.
24. Bacia, K., S.A. Kim, and P. Schwille, *Fluorescence cross-correlation spectroscopy in living cells.* Nat Methods, 2006. **3**(2): p. 83-9.
25. Hartl, F.U., *Molecular chaperones in cellular protein folding.* Nature, 1996. **381**(6583): p. 571-9.

References

26. Hingorani, K.S. and L.M. Gierasch, *Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge.* Curr Opin Struct Biol, 2014. **24**: p. 81-90.

27. Ellis, J., *Proteins as molecular chaperones.* Nature, 1987. **328**(6129): p. 378-9.

28. Hartl, F.U., A. Bracher, and M. Hayer-Hartl, *Molecular chaperones in protein folding and proteostasis.* Nature, 2011. **475**(7356): p. 324-32.

29. Brockwell, D.J. and S.E. Radford, *Intermediates: ubiquitous species on folding energy landscapes?* Curr Opin Struct Biol, 2007. **17**(1): p. 30-7.

30. Batey, S., A.A. Nickson, and J. Clarke, *Studying the folding of multidomain proteins.* HFSP J, 2008. **2**(6): p. 365-77.

31. Braselmann, E., J.L. Chaney, and P.L. Clark, *Folding the proteome.* Trends Biochem Sci, 2013. **38**(7): p. 337-44.

32. Imamoglu, R., et al., *Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein.* Nat Commun, 2020. **11**(1): p. 365.

33. Deuerling, E. and B. Bukau, *Chaperone-assisted folding of newly synthesized proteins in the cytosol.* Crit Rev Biochem Mol Biol, 2004. **39**(5-6): p. 261-77.

34. Gershenson, A. and L.M. Gierasch, *Protein folding in the cell: challenges and progress.* Curr Opin Struct Biol, 2011. **21**(1): p. 32-41.

35. Calloni, G., et al., *DnaK functions as a central hub in the E. coli chaperone network.* Cell Rep, 2012. **1**(3): p. 251-64.

36. Moran Luengo, T., M.P. Mayer, and S.G.D. Rudiger, *The Hsp70-Hsp90 Chaperone Cascade in Protein Folding.* Trends Cell Biol, 2019. **29**(2): p. 164-177.

37. Cuellar, J., et al., *The structure of CCT-Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin.* Nat Struct Mol Biol, 2008. **15**(8): p. 858-64.

38. Langer, T., et al., *Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding.* Nature, 1992. **356**(6371): p. 683-9.

39. Singh, A.K., et al., *Efficient Catalysis of Protein Folding by GroEL/ES of the Obligate Chaperonin Substrate MetF.* J Mol Biol, 2020. **432**(7): p. 2304-2318.

40. Hartl, F.U. and M. Hayer-Hartl, *Molecular chaperones in the cytosol: from nascent chain to folded protein.* Science, 2002. **295**(5561): p. 1852-8.

41. Mogk, A., C. Ruger-Herreros, and B. Bukau, *Cellular Functions and Mechanisms of Action of Small Heat Shock Proteins.* Annu Rev Microbiol, 2019. **73**: p. 89-110.

42. Hayer-Hartl, M. and F.U. Hartl, *Chaperone Machineries of Rubisco - The Most Abundant Enzyme.* Trends Biochem Sci, 2020. **45**(9): p. 748-763.

43. Hipp, M.S., S.H. Park, and F.U. Hartl, *Proteostasis impairment in protein-misfolding and -aggregation diseases.* Trends Cell Biol, 2014. **24**(9): p. 506-14.

44. Maier, T., et al., *A cradle for new proteins: trigger factor at the ribosome.* Curr Opin Struct Biol, 2005. **15**(2): p. 204-12.

45. Preissler, S. and E. Deuerling, *Ribosome-associated chaperones as key players in proteostasis.* Trends Biochem Sci, 2012. **37**(7): p. 274-83.

46. Agashe, V.R., et al., *Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed.* Cell, 2004. **117**(2): p. 199-209.

47. Horowitz, S., et al., *Folding while bound to chaperones.* Curr Opin Struct Biol, 2018. **48**: p. 1-5.

48. Kramer, G., et al., *The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins.* Nat Struct Mol Biol, 2009. **16**(6): p. 589-97.

49. Deuerling, E., M. Gamerdinger, and S.G. Kreft, *Chaperone Interactions at the Ribosome.* Cold Spring Harb Perspect Biol, 2019. **11**(11).

50. Kramer, G., A. Shiber, and B. Bukau, *Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins.* Annu Rev Biochem, 2019. **88**: p. 337-364.

51. Rosenzweig, R., et al., *The Hsp70 chaperone network.* Nat Rev Mol Cell Biol, 2019. **20**(11): p. 665-680.

52. Clerico, E.M., et al., *Hsp70 molecular chaperones: multifunctional allosteric holding and unfolding machines.* Biochem J, 2019. **476**(11): p. 1653-1677.

References

53. Zhu, X., et al., *Structural analysis of substrate binding by the molecular chaperone DnaK.* Science, 1996. **272**(5268): p. 1606-14.

54. Mayer, M.P., *Hsp70 chaperone dynamics and molecular mechanism.* Trends Biochem Sci, 2013. **38**(10): p. 507-14.

55. Mayer, M.P., *Gymnastics of molecular chaperones.* Mol Cell, 2010. **39**(3): p. 321-31.

56. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *Recent advances in understanding catalysis of protein folding by molecular chaperones.* FEBS Lett, 2020. **594**(17): p. 2770-2781.

57. Horwich, A.L., et al., *Two families of chaperonin: physiology and mechanism.* Annu Rev Cell Dev Biol, 2007. **23**: p. 115-45.

58. Yebenes, H., et al., *Chaperonins: two rings for folding.* Trends Biochem Sci, 2011. **36**(8): p. 424-32.

59. Hayer-Hartl, M., A. Bracher, and F.U. Hartl, *The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding.* Trends Biochem Sci, 2016. **41**(1): p. 62-76.

60. Gestaut, D., et al., *The ATP-powered gymnastics of TRiC/CCT: an asymmetric protein folding machine with a symmetric origin story.* Curr Opin Struct Biol, 2019. **55**: p. 50-58.

61. Ditzel, L., et al., *Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT.* Cell, 1998. **93**(1): p. 125-38.

62. Xu, Z., A.L. Horwich, and P.B. Sigler, *The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex.* Nature, 1997. **388**(6644): p. 741-50.

63. Clare, D.K., et al., *ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin.* Cell, 2012. **149**(1): p. 113-23.

64. Saibil, H.R., et al., *Structure and allostery of the chaperonin GroEL.* J Mol Biol, 2013. **425**(9): p. 1476-87.

65. Braig, K., et al., *The crystal structure of the bacterial chaperonin GroEL at 2.8 A.* Nature, 1994. **371**(6498): p. 578-86.

66. Horwich, A.L. and W.A. Fenton, *Chaperonin-assisted protein folding: a chronologue.* Q Rev Biophys, 2020. **53**: p. e4.

67. Weaver, J. and H.S. Rye, *The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein.* J Biol Chem, 2014. **289**(33): p. 23219-23232.

68. Hunt, J.F., et al., *The crystal structure of the GroES co-chaperonin at 2.8 A resolution.* Nature, 1996. **379**(6560): p. 37-45.

69. Horwich, A.L., et al., *Folding in vivo of bacterial cytoplasmic proteins: role of GroEL.* Cell, 1993. **74**(5): p. 909-17.

70. Fujiwara, K., et al., *A systematic survey of in vivo obligate chaperonin-dependent substrates.* EMBO J, 2010. **29**(9): p. 1552-64.

71. Georgescauld, F., et al., *GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding.* Cell, 2014. **157**(4): p. 922-934.

72. Sharma, S., et al., *Monitoring protein conformation along the pathway of chaperonin-assisted folding.* Cell, 2008. **133**(1): p. 142-53.

73. Dobson, C.M., A. Sali, and M. Karplus, *Protein Folding: A Perspective from Theory and Experiment.* Angew Chem Int Ed Engl, 1998. **37**(7): p. 868-893.

74. Chaudhuri, T.K., et al., *GroEL/GroES-mediated folding of a protein too large to be encapsulated.* Cell, 2001. **107**(2): p. 235-46.

75. Haldar, S., et al., *Chaperonin-Assisted Protein Folding: Relative Population of Asymmetric and Symmetric GroEL:GroES Complexes.* J Mol Biol, 2015. **427**(12): p. 2244-55.

76. Taguchi, H., *Reaction Cycle of Chaperonin GroEL via Symmetric "Football" Intermediate.* J Mol Biol, 2015. **427**(18): p. 2912-8.

77. Yang, D., X. Ye, and G.H. Lorimer, *Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine.* Proc Natl Acad Sci U S A, 2013. **110**(46): p. E4298-305.

References

78. Yan, X., et al., *GroEL Ring Separation and Exchange in the Chaperonin Reaction.* Cell, 2018. **172**(3): p. 605-617 e11.

79. Tang, Y.C., et al., *Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein.* Cell, 2006. **125**(5): p. 903-14.

80. Chakraborty, K., et al., *Chaperonin-catalyzed rescue of kinetically trapped states in protein folding.* Cell, 2010. **142**(1): p. 112-22.

81. Farr, G.W., et al., *Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes.* EMBO J, 2003. **22**(13): p. 3220-30.

82. Sheppard, C.A., E.E. Trimmer, and R.G. Matthews, *Purification and properties of NADH-dependent 5, 10-methylenetetrahydrofolate reductase (MetF) from Escherichia coli.* J Bacteriol, 1999. **181**(3): p. 718-25.

83. Guenther, B.D., et al., *The structure and properties of methylenetetrahydrofolate reductase from Escherichia coli suggest how folate ameliorates human hyperhomocysteinemia.* Nat Struct Biol, 1999. **6**(4): p. 359-65.

84. Misra, S.K. and V. Bhakuni, *Unique holoenzyme dimers of the tetrameric enzyme Escherichia coli methylenetetrahydrofolate reductase: characterization of structural features associated with modulation of the enzyme's function.* Biochemistry, 2003. **42**(13): p. 3921-8.

85. Apetri, A.C. and A.L. Horwich, *Chaperonin chamber accelerates protein folding through passive action of preventing aggregation.* Proc Natl Acad Sci U S A, 2008. **105**(45): p. 17351-5.

86. Horwich, A.L., A.C. Apetri, and W.A. Fenton, *The GroEL/GroES cis cavity as a passive anti-aggregation device.* FEBS Lett, 2009. **583**(16): p. 2654-62.

87. Tyagi, N.K., et al., *Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented.* FEBS Lett, 2011. **585**(12): p. 1969-72.

88. Weaver, J., et al., *GroEL actively stimulates folding of the endogenous substrate protein PepQ.* Nat Commun, 2017. **8**: p. 15934.

89. Brinker, A., et al., *Dual function of protein confinement in chaperonin-assisted protein folding.* Cell, 2001. **107**(2): p. 223-33.

90. Mallam, A.L. and S.E. Jackson, *Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins.* Nat Chem Biol, 2011. **8**(2): p. 147-53.

91. Lin, Z., D. Madan, and H.S. Rye, *GroEL stimulates protein folding through forced unfolding.* Nat Struct Mol Biol, 2008. **15**(3): p. 303-11.

92. Lin, Z., et al., *Repetitive protein unfolding by the trans ring of the GroEL-GroES chaperonin complex stimulates folding.* J Biol Chem, 2013. **288**(43): p. 30944-55.

93. Ye, X., et al., *Folding of maltose binding protein outside of and in GroEL.* Proc Natl Acad Sci U S A, 2018. **115**(3): p. 519-524.

94. England, J.L., D. Lucent, and V.S. Pande, *A role for confined water in chaperonin function.* J Am Chem Soc, 2008. **130**(36): p. 11838-9.

95. Franck, J.M., et al., *Probing water density and dynamics in the chaperonin GroEL cavity.* J Am Chem Soc, 2014. **136**(26): p. 9396-403.

96. Naqvi, M.M., et al., *Protein chain collapse modulation and folding stimulation by GroEL-ES.* Sci Adv, 2022. **8**(9): p. eabl6293.

97. Thirumalai, D., G.H. Lorimer, and C. Hyeon, *Iterative annealing mechanism explains the functions of the GroEL and RNA chaperones.* Protein Sci, 2020. **29**(2): p. 360-377.

98. Korobko, I., et al., *Measuring protein stability in the GroEL chaperonin cage reveals massive destabilization.* Elife, 2020. **9**.

99. Hayer-Hartl, M.K., et al., *Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin.* EMBO J, 1994. **13**(13): p. 3192-202.

100. Poso, D., A.R. Clarke, and S.G. Burston, *A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL.* J Mol Biol, 2004. **338**(5): p. 969-77.

101. Traverso-Cori, A., H. Chaimovich, and O. Cori, *Kinetic Studies and Properties of Potato Apyrase.* Arch Biochem Biophys, 1965. **109**: p. 173-84.

References

102. Valenzuela, M.A., et al., *Effects of protein-modifying reagents on an isoenzyme of potato apyrase.* Biochem J, 1973. **133**(4): p. 755-63.

103. Geladopoulos, T.P., T.G. Sotiroudis, and A.E. Evangelopoulos, *A malachite green colorimetric assay for protein phosphatase activity.* Anal Biochem, 1991. **192**(1): p. 112-6.

104. Farr, G.W., et al., *Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms.* Cell, 1997. **89**(6): p. 927-37.

105. Muller, B.K., et al., *Pulsed interleaved excitation.* Biophys J, 2005. **89**(5): p. 3508-22.

106. Enderlein, J., et al., *Performance of fluorescence correlation spectroscopy for measuring diffusion and concentration.* Chemphyschem, 2005. **6**(11): p. 2324-36.

107. Guttman, M., et al., *Tuning a High Transmission Ion Guide to Prevent Gas-Phase Proton Exchange During H/D Exchange MS Analysis.* J Am Soc Mass Spectrom, 2016. **27**(4): p. 662-8.

108. Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data.* Nucleic Acids Res, 2019. **47**(D1): p. D442-D450.

109. Sadat, A., et al., *GROEL/ES Buffers Entropic Traps in Folding Pathway during Evolution of a Model Substrate.* J Mol Biol, 2020. **432**(20): p. 5649-5664.

110. Kellner, R., et al., *Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein.* Proc Natl Acad Sci U S A, 2014. **111**(37): p. 13355-60.

111. Sekhar, A., et al., *Mapping the conformation of a client protein through the Hsp70 functional cycle.* Proc Natl Acad Sci U S A, 2015. **112**(33): p. 10395-400.

112. Sekhar, A., et al., *Hsp70 biases the folding pathways of client proteins.* Proc Natl Acad Sci U S A, 2016. **113**(20): p. E2794-801.

113. Azia, A., R. Unger, and A. Horovitz, *What distinguishes GroEL substrates from other Escherichia coli proteins?* FEBS J, 2012. **279**(4): p. 543-50.

114. Tartaglia, G.G. and M. Vendruscolo, *Proteome-level interplay between folding and aggregation propensities of proteins.* J Mol Biol, 2010. **402**(5): p. 919-28.

115. Martin, J., et al., *Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate.* Nature, 1991. **352**(6330): p. 36-42.

116. Hawe, A., M. Sutter, and W. Jiskoot, *Extrinsic fluorescent dyes as tools for protein characterization.* Pharm Res, 2008. **25**(7): p. 1487-99.

117. Baldwin, R.L. and G.D. Rose, *Molten globules, entropy-driven conformational change and protein folding.* Curr Opin Struct Biol, 2013. **23**(1): p. 4-10.

118. Hayer-Hartl, M.K., F. Weber, and F.U. Hartl, *Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis.* EMBO J, 1996. **15**(22): p. 6111-21.

119. Horst, R., et al., *Direct NMR observation of a substrate protein bound to the chaperonin GroEL.* Proc Natl Acad Sci U S A, 2005. **102**(36): p. 12748-53.

120. Fei, X., et al., *Formation and structures of GroEL:GroES2 chaperonin footballs, the protein-folding functional form.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. 12775-80.

121. Koike-Takeshita, A., et al., *Crystal structure of a symmetric football-shaped GroEL:GroES2-ATP14 complex determined at 3.8A reveals rearrangement between two GroEL rings.* J Mol Biol, 2014. **426**(21): p. 3634-41.

122. Dong, D., et al., *The crystal structure of Cpf1 in complex with CRISPR RNA.* Nature, 2016. **532**(7600): p. 522-6.

123. Mukhopadhyay, S., et al., *A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2649-54.

124. Pack, C.G., et al., *Analysis of interaction between chaperonin GroEL and its substrate using fluorescence correlation spectroscopy.* Cytometry, 1999. **36**(3): p. 247-53.

125. Hodge, E.A., M.A. Benhaim, and K.K. Lee, *Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry.* Protein Sci, 2020. **29**(4): p. 843-855.

126. Balchin, D., et al., *Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC.* Cell, 2018. **174**(6): p. 1507-1521 e16.

References

127. Engen, J.R. and D.L. Smith, *Investigating protein structure and dynamics by hydrogen exchange MS.* Anal Chem, 2001. **73**(9): p. 256A-265A.

128. Robinson, C.V., et al., *Conformation of GroEL-bound alpha-lactalbumin probed by mass spectrometry.* Nature, 1994. **372**(6507): p. 646-51.

129. Natesh, R., et al., *A two-domain folding intermediate of RuBisCO in complex with the GroEL chaperonin.* Int J Biol Macromol, 2018. **118**(Pt A): p. 671-675.

130. Tang, Y.C., et al., *Essential role of the chaperonin folding compartment in vivo.* EMBO J, 2008. **27**(10): p. 1458-68.

131. Brocchieri, L. and S. Karlin, *Conservation among HSP60 sequences in relation to structure, function, and evolution.* Protein Sci, 2000. **9**(3): p. 476-86.

132. Stan, G., et al., *Annealing function of GroEL: structural and bioinformatic analysis.* Biophys Chem, 2003. **100**(1-3): p. 453-67.

133. Todd, M.J., G.H. Lorimer, and D. Thirumalai, *Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism.* Proc Natl Acad Sci U S A, 1996. **93**(9): p. 4030-5.

134. Pine, M.J., *Turnover of intracellular proteins.* Annu Rev Microbiol, 1972. **26**: p. 103-26.

135. Horst, R., et al., *Folding trajectories of human dihydrofolate reductase inside the GroEL GroES chaperonin cavity and free in solution.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20788-92.

136. Chen, J., et al., *Folding of malate dehydrogenase inside the GroEL-GroES cavity.* Nat Struct Biol, 2001. **8**(8): p. 721-8.

137. Baumketner, A., A. Jewett, and J.E. Shea, *Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape.* J Mol Biol, 2003. **332**(3): p. 701-13.

138. Sirur, A. and R.B. Best, *Effects of interactions with the GroEL cavity on protein folding rates.* Biophys J, 2013. **104**(5): p. 1098-106.

139. Hayer-Hartl, M. and A.P. Minton, *A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding.* Biochemistry, 2006. **45**(44): p. 13356-60.

140. England, J.L. and V.S. Pande, *Potential for modulation of the hydrophobic effect inside chaperonins.* Biophys J, 2008. **95**(7): p. 3391-9.

141. Chen, D.H., et al., *Visualizing GroEL/ES in the act of encapsulating a folding protein.* Cell, 2013. **153**(6): p. 1354-65.

142. Hofmann, H., et al., *Single-molecule spectroscopy of protein folding in a chaperonin cage.* Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11793-8.

143. Koculi, E. and D. Thirumalai, *Retardation of Folding Rates of Substrate Proteins in the Nanocage of GroEL.* Biochemistry, 2021. **60**(6): p. 460-464.

144. Aigner, H., et al., *Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2.* Science, 2017. **358**(6368): p. 1272-1278.

145. Goloubinoff, P., et al., *Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP.* Nature, 1989. **342**(6252): p. 884-9.

146. Liu, C., et al., *Coupled chaperone action in folding and assembly of hexadecameric Rubisco.* Nature, 2010. **463**(7278): p. 197-202.

147. Durao, P., et al., *Opposing effects of folding and assembly chaperones on evolvability of Rubisco.* Nat Chem Biol, 2015. **11**(2): p. 148-55.

148. Tokuriki, N. and D.S. Tawfik, *Chaperonin overexpression promotes genetic variation and enzyme evolution.* Nature, 2009. **459**(7247): p. 668-73.

149. Wang, J.D., et al., *Directed evolution of substrate-optimized GroEL/S chaperonins.* Cell, 2002. **111**(7): p. 1027-39.