Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural Features of the GroEL-GroES Nano-Cage

Required for Rapid Folding of Encapsulated Protein

Yun-Chi Tang

aus

Taipei

Taiwan, R.O.C.

2007

Acknowledgements

First of all, I would like to express my deepest gratitude to Prof. Dr. F. Ulrich Hartl for giving me the opportunity to study and learn the extremely interesting subject in his laboratory. I would like to thank him for the encouragement and the continual support throughout the entire period of my study.

Uncountable thanks go to my direct supervisor Dr. Manajit Hayer-Hartl for her invaluable advice and constant support. She is not only a good advisor of my work but also a good mentor in my personal life.

I would like to thank Prof. Dr. Jürgen Soll for his kindly help for correcting my dissertation and being the co-referee of my thesis committee.

I thank colleagues in the department of cellular biochemistry for providing accommodative environment to a foreigner like me and many helps. In particularly, I would like to thank Andrea, Silke, Elisabeth and Bernd Grampp for keeping the laboratory at good running. Special thanks to Nadine and Dirk for their excellent technical assistance.

Many great thanks to Sandra, Sarah, Shruti, Andreas, Christian, Gregor, José, Kausik, Martin and Michael for generously sharing their speciality opinions and many insightful discussions. Their friendships and the good working atmosphere became the main basis for the success of this work.

The deepest thanks go to my husband, Hung-Chun Chang, for his enormous support and patience and valuable scientific discussions. The same deep thanks belong to my parents and my family in Taiwan for their understanding and support.

Contents

1. Summary

The chaperonin GroEL and GroES form a nano-cage for proteins up to ~60 kDa to fold in isolation. The GroEL and GroES system has been thought of as an important but passive player in protein folding, providing an encapsulated and isolated environment that allows folding to proceed without impaired by aggregation. However, recent experiments showed that the folding of bacterial ribulose-bisphosphate carboxylase (RuBisCo) is accelerated in the GroEL/GroES folding cage, providing the first hint that the GroEL/GroES cavity could be more than just a passive folding container.

Here we explored the structural features of the chaperonin cage critical for modulating the folding of encapsulated substrates. We performed a series of experiments in which the volume and surface properties of the GroEL central cavity were altered, and the effects on the folding rate and yield of substrate proteins were measured. The substrate proteins of different molecular size selected for this study included the small (33 kDa) proteins rhodanese and MetF (33 kDa), 41 kDa maltose binding protein (MBP) and the larger, 50 kDa bacterial RuBisCo.

By deleting the GroEL C-terminal GGM repeats (13 amino acids) or replicating them two, three, or four times, the volume of the GroEL/GroES *cis* cavity was changed by -13% to +4%. Interestingly, modulating the volume of the GroEL cavity affected folding speed in accordance with confinement theory. For relatively small proteins of ~30 kDa, rhodanese and MetF, reducing cavity size first increased the rate of folding until a critical size limit, which, once exceeded, led to a significant decrease in folding rate. For the larger proteins of ~40-50 kDa, MBP and RuBisCo, either expanding or reducing the *cis*-cage volume decelerated folding.

The GroEL/GroES *cis* cavity wall exposes 189 negatively and 147 positively charged residues with a net negative charge of 42. This suggested that electrostatic interactions may

also influence the folding rate. By substituting one or more of the negative charged residues in each GroEL subunit with Asn, Gln, or Lys, we determined the importance of the charges on the folding of the model substrates. Strikingly, for many substrates either the refolding yields were reduced or folding rates were affected. The results revealed that the *cis*-cavity lining can have a profound influence on folding in general.

We suggest that the GroEL/GroES cage has a tripartite in folding by combining the following features: (1) encapsulation offers a safe environment for folding unimpaired by aggregation; (2) cavity volume presents a confinement effect which can speed up folding for some proteins; (3) by combining negatively charged wall properties with a mildly hydrophobic surface, the cage can facilitate rearrangement steps during folding. These properties allow GroEL to assist the folding of a wide range of cytosolic proteins.

2. Introduction

Proteins perform most biological processes in cells. Proteins not only provide the structural blocks (molecules of the cytoskeleton, epidermal keratin, viral coat proteins) to maintain the cell shape, but also execute nearly all cell functions. For instance, catalytic proteins (enzymes) mediate biochemical reactions, regulatory proteins (many hormones, receptors, kinases, phosphatases and DNA binding proteins) control cellular signal transduction and gene expression, transport proteins (hemoblobin, myoglobin, ferritin) deliver small molecules or ions to target cells, membrane proteins (channels and pumps) regulate the passage of molecules in and out of cells, and the immunoglobulin superfamily of proteins (antibodies and proteins involved in cell-cell recognition) dominate the immune system and signaling. To fulfill these biological activities, proteins must adopt precise three-dimensional structures. The process for acquiring the unique native structure of a polypeptide is called protein folding.

2.1. Protein folding

2.1.1. Protein structure

Structurally, proteins are polymers of amino acids, joined together by peptide bonds in a polypeptide chain. The amino acid sequence of a polypeptide chain is called its primary structure. Different regions of the sequence form local regular secondary structure, such as α-helices or β-sheets. The tertiary structure is formed by packing such secondary structure elements into one or several compact globular units called domains. As many proteins may contain several polypeptide chains, a protein's quaternary structure refers to the spatial arrangement of its subunits.

In the primary structure, the α carbons of adjacent amino acid residues are separated by three covalent bonds, arranged as Cα – C – N – Cα (Figure 1). The six atoms of the peptide group lie in a single plane, with the oxygen atom of the carboxyl group and the hydrogen of the amide nitrogen group. The peptide C – N bonds are unable to rotate freely because of their partial double bond character. Rotation is allowed of the N – Cα and Cα – C bonds. The bond angle resulting from rotations at Cα is labeled φ (phi) for the N – Cα bond and ψ (psi) for the Cα – C bond (Figure 1). In principle, φ and ψ can have any angle between -180° and 180°, but many angles are excluded by steric interference between atoms in the polypeptide backbone and amino acid side chains. G. N. Ramachandran calculated the energy contained in various pairs of ψ and φ angles and found two most stable pairs, the so called α and β conformations (Ramachandran and Sasisekharan, 1968). These two pairs of angles are found to almost exclusively occur in folded proteins, including the two most prominent examples of secondary structure: α-helix and β-strand.
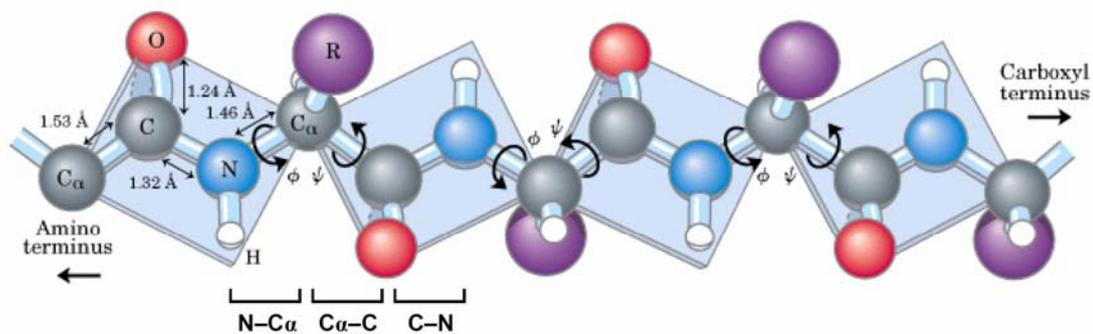
Figure 1. Rotation about bonds in a polypeptide chain

Three bonds separate sequential α carbons in a polypeptide chain. The N – Cα and Cα – C bonds can rotate, with bond angles designated φ and ψ, respectively. The peptide C – N is not free to rotate. Other signal bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. The peptide bond is planar as represent in blue shading. Adapted from (Lehninger *et al.*, 2000).

In spite of the conformational uniqueness of a protein, the secondary structural elements are simple, consisting of helices, sheets and turns. The α-helix and the β-sheet elements keep the main chain in an unstrained conformation, and allow hydrogen-bonding potential of the main-chain N-H and C=O groups. The α-helical structure was first described in 1951 by Pauling and Corey (Pauling and Corey, 1951a). In this structure the polypeptide backbone is tightly wound around a middle axis, and the R groups of amino acid residues protrude outward from the helical backbone. The first α-helix was described in the protein α-keratin, which is an abundant protein of the skin and its derivatives are found in hair, nails and horns (Pauling and Corey, 1951a). More generally, about one-fourth of all amino acid residues on polypeptides are found in α-helices. Pauling and Corey also predicted a second type of repetitive structural element, the β-sheet (Pauling and Corey, 1951b), which poses a more extensive conformation of a polypeptide chain. In the β-conformation, the polypeptide backbone is extended into a zigzag rather than a helical structure. The zigzag polypeptide chains can be arranged side by side as was typically found in the protein fibroin, the major constituent of silk (Pauling and Corey, 1951b). These two patterns are particularly common because they result from hydrogen bonding between the N-H and C=O groups in the polypeptide backbone, without involving the side chains of the amino acids. Thus, they can be formed by many different amino acid sequences. In each case, the protein chain adopts a regular, repeating conformation.

2.1.2. The complexity of protein folding

How does a polypeptide chain reach its native conformation? Pioneering experiments on protein folding were performed in the late 1950s by Christian Anfinsen. Purified, denatured Ribonuclease A was shown to fold spontaneously to its native state *in vitro* upon removal of the denaturant, as measured by regain of enzymatic activity (Anfinsen, 1973; Anfinsen *et*

al., 1955). These experiments demonstrated that the amino acid sequence of a polypeptide chain contains all the information required to fold the protein into its native, three-dimensional structure (Anfinsen, 1973).

Assuming the spontaneous folding process is a random process in which a polypeptide chain tries out all possible conformations around every single peptide bond until it eventually finds its native form, plus considering that each amino acid residues could on average have 10 different conformations, a protein containing 100 amino acid residues would result in 10^{100} different possible conformations. Since the interconversion between conformations needs $\sim 10^{-13}$ seconds, the 100-residue polypeptide would take about $\sim 10^{77}$ years (10^{100} x 10^{-13} s = 10^{87} s) to explore its conformational space. This is far beyond the time range of any biological process. In fact, proteins are assembled from amino acids at a very fast rate within cells. For instance, *E. coil* can produce a functional active protein containing 100 amino acid residues in about 5 second at 37°C. Thus protein folding cannot be a random, trial and error process. This argument was first made by Cyrus Levinthal in 1968, called Levinthal's paradox. Levinthal concluded that proteins must fold to their native conformation by specific folding pathways (Levinthal, 1968). The efficient folding must proceed through specific transient intermediates, in which local folded elements are stabilized to determine further folding of the polypeptide (Baldwin, 1996; Baldwin and Rose, 1999; Levinthal, 1968). These intermediates would significantly reduce the number of possible conformations during folding and thus allow protein folding to take place on a biologically relevant time scale.

2.1.3. Protein folding mechanism

How does a protein find the right pathway and aviod misfolding or aggregation? Serveral plausible models for the mechanism of folding have been proposed (Figure 2). One

model, the hydrophobic collapse model, suggests that a protein buries its hydrophobic side chains from solvent water, forming a collapsed intermediate or molten globule species, from which the native state develops by searching within this conformational state (Baldwin, 1989; Schellman, 1955). The framework model, on the other hand, suggests that local elements of secondary structure form first and then dock into the native tertiary structure of the protein (Kim and Baldwin, 1982; Shoemaker *et al*., 1985). By contrast, the nucleation growth model proposes that the amino acid residues adjacent in sequence form a nucleus from which the native structure then develops in a sequential manner. Lastly, the jigsaw model suggests that each protein molecule could fold by a different path (Kim and Baldwin, 1982).
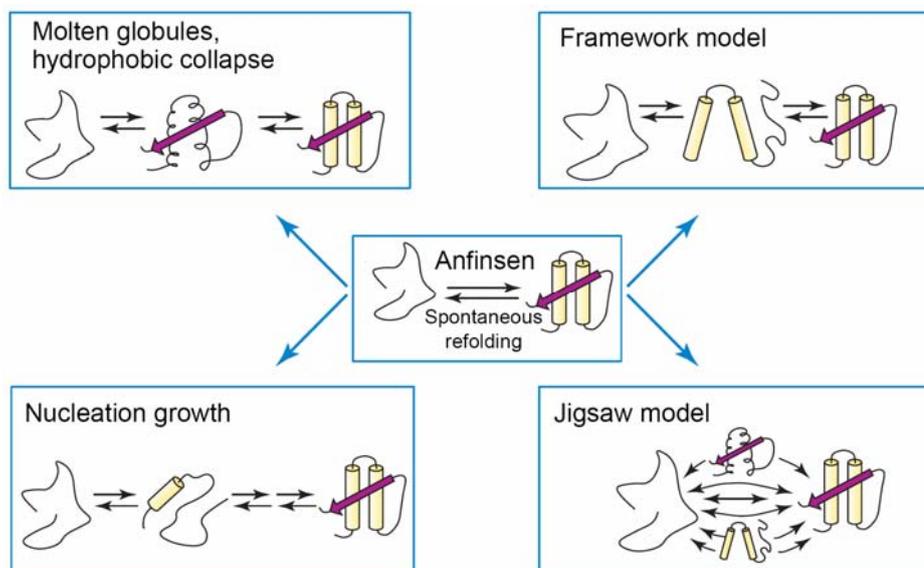
Figure 2. Plausible models for mechanisms of folding

At least four models were proposed to tackle the mechanisms of folding after Anfinsen's refolding work. As shown as molten globules/ hydrophobic collapse model, framework model, nucleation growth model and jigsaw model. The red arrows indicate β-sheet structure and the yellow cylinders represent α-helix. Adapted from (Radford, 2000).

Currently available evidences suggest that most proteins fold *via* a process that incorporates more than one of the models at the same time. Instead of following a single pathway, a population of polypeptide chains can take a variety of routes to the native state. The folding trajectory can be described in a three- or multi-dimensional energy landscape or folding funnel as shown in Figure 3 (Radford, 2000; Schultz, 2000).
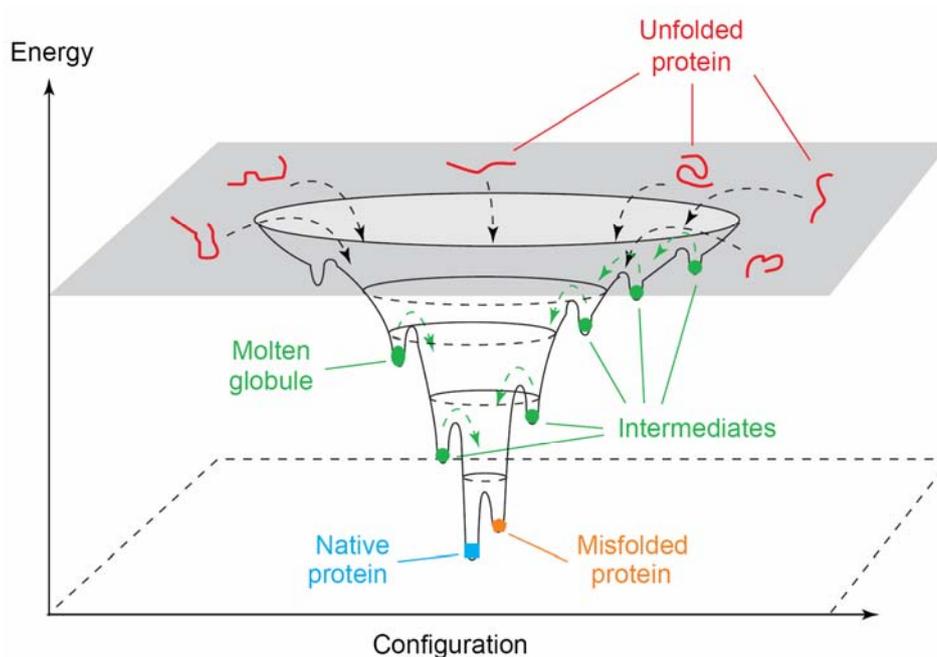
Figure 3. Scheme of the folding energy landscape

The multiple states of the unfolded protein located at the top of the folding funnel might fold to the native state by many of different routes. Most of these states represent transient intermediates in the folding process. Some of these intermediates retain a stable structure such as the molten globule, whereas others may fall into a kinetic trap resulting in the formation of misfolded proteins. For proteins that fold without populating intermediates, the surface of the funnel would be smooth. Adapted from Schultz, (2000) and Radford, (2000).

The energy landscape of folding is represented by a rough surface. Each point on the energy landscape presents a conformation of the polypeptide and the corresponding energy. The native state of a protein, defined as the conformation with the lowest free energy, is thus at the lowest point of the energy landscape or the bottom of the folding funnel. The multiple unfolded states are characterized by high conformational entropy and relative high free energy. Along the folding process, the protein follows a route from the rim of the funnel to the bottom of funnel. Populated intermediates on the way from unfolded to the native state are local minima in the energy landscape (Radford, 2000; Schultz, 2000; Troullier *et al.*, 2000). If the folding intermediate cannot escape a local minimum, it becomes kinetically trapped and results in the so-called misfolded protein.

2.1.4. Methods for studying protein folding

Major advances have been made in illuminating the folding mechanism of proteins since the original models described above. These have been derived from a wealth of new and elegant experimental approaches (Table 1), combined with theoretical methods. The key to discerning the nature of folding mechanisms is to combine the results from different techniques so that different aspects of folding can be probed and the results combined into a common picture of the folding process (Dobson *et al.*, 1994). Experimental methods commonly used include circular dichroism and fluorescence (Goldbeck *et al.*, 1997), which provide information about the extent of secondary and tertiary structure formation. When combined with stop-flow methods, they can measure folding events on the millisecond timescale (Radford, 2000).

Table 1. Experimental approaches to investigate protein folding

Technique	Time scale	Structural parameter probed
Fluorescence	ns-s	
1. Intrinsic fluorescence		Environment of Trp and Tyr
2. ANS binding		Exposure of hydrophobic surface area
3. Substrate binding		Formation of active site
4. FRET		Inter-residue distance
5. Anisotropy		Correlation time
Circular dichroism	ns-s	
1. Far UV		Secondary structure formation
2. Near UV		Tertiary structure formation
Small-angle X-ray scattering	≥ ms	Dimension and sharp of polypeptide
Absorbance (near UV)	ns-s	Environment of aromatic residues
FTIR	ns-s	Secondary structure formation
NMR		
1. Real time	ms-s	Environment of individual residues
2. Dynamic NMR	250 μs	Lineshape analysis provides folding-unfolding rate close to equilibrium
Hydrogen exchange (HX)		
1. Native state	min-months	Global stability and metastable states
2. Plused HX ESI MS	ms-s	Folding population
3. Plused HX NMR	ms-s	Hydrogen-bond formation in specific residues
Atomic force spectroscopy (AFM)	s	Unfolding forces and rate constants of single molecules

The table was modified from Brockwell, (2000) and Radford, (2000).

Abbreviations: ANS: 1-anilino napthalene sulphonic caid; ESI MS: elctrospray ionization mass spectrometry; FRET: fluorescence resonance energy transfer; FITR: fourier transform intra-red.

Information about folding at the level of individual residues can be obtained using hydrogen exchange (Englander, 2000) and protein engineering. Hydrogen exchange, when combined with detection by multidimensional NMR, can be used to determine the location and stability of individual hydrogen bonds at different stages of folding. Protein engineering, by contrast, provides site-specific information about the role of individual side chains in stabilizing populated intermediates and transient high-energy transition states. In this approach, an amino acid side chain is removed from the protein of interest and the effect (e.g. by mutating Val to Ala) of the mutation on the stability of the native protein (determined by equilibrium denaturation) and the intermediate or transition state (determining using kinetics) are measured and compared. The ratio of these stabilities is known as a Φ value. By determining many Φ values for residues spread through the native protein, the structure of intermediates (if they are populated) and the rate-limiting transition state can then be inferred (Fersht *et al.*, 1992).

A major developing area in folding is the establishment of techniques that can monitor the process on sub-millisecond timescales, so that important early events can be monitored (Brockwell *et al.*, 2000) (Table 1). Such methods include ultra-rapid mixing, temperature jump and pressure jump experiments (Eaton *et al.*, 2000; Roder and Shastry, 1999). When combined with detection methods such as fluorescence or circular dichroism, these can access folding events occurring on very fast timescales (nanosecond to microsecond). In parallel to these advances, a number of theoretical methods have been developed to simulate protein folding. These include molecular dynamics that can be used to monitor the unfolding of proteins at atomic resolution (usually at very high temperatures to speed up the process) and the folding of small proteins for up to milliseconds using the computer power that is now available (Duan and Kollman, 1998). In contrast, lattice simulations rely on very simple models for proteins based on polymer beads but have the advantage that the conformational space can be searched exhaustively (Dinner *et al.*, 2000). Together, these

approaches have provided new insights into the folding and unfolding processes of several proteins and agree well with experimental results (Brockwell *et al.*, 2000).

2.2. Protein folding in the cell

Although the protein folding mechanism has been extensively studied *in vitro* (Dobson and Karplus, 1999), folding of polypeptides in the cellular environment is much more complex, since the *in vivo* environment is extremely crowded and dynamic in comparison to the conditions of *in vitro* refolding.

2.2.1. Highly crowded milieu in the cell

The effective protein concentration in *E. coli* cells has been estimated to be as high as 300 to 400 mg/ml (Ellis, 2001a; Zimmerman and Minton, 1993) (Figure 4). Because of high concentration of proteins, nucleic acids and others macromolecules, the cytoplasm no longer serves as an ideal place for polypeptides to fold. Besides, this crowded environment gives rise to excluded volume effects, which can result in a significant increase in the affinities between interacting macromolecules by up to 10-100 fold (Minton, 2000; van den Berg *et al.*, 1999). As a consequence, the intermolecular binding constants between partially folded states are increased, leading to an increased probability of aggregation during folding (van den Berg *et al.*, 1999).

Another major difference between *in vitro* refolding and *de novo* folding is that folding in the cell must be accomplished in the context of the vectorial synthesis of polypeptide chains on ribosomes. The formation of a stable tertiary structure requires a complete protein domain sequence (usually 50 to 300 amino acid residues in length), and the unavailability of the C-terminal residues of a translating polypeptide within the ribosomal channel results in

prolonging the time for hydrophobic residues of a polypeptide to be exposed (Jaenicke, 1991). Nascent chains must therefore avoid forming misfolded intermediates and avoid aggregation with other nascent chains. Since translation occurs on polyribosomes, nascent chains may be especially prone to undergo inter-molecular clustering (Ellis and Hartl, 1999).
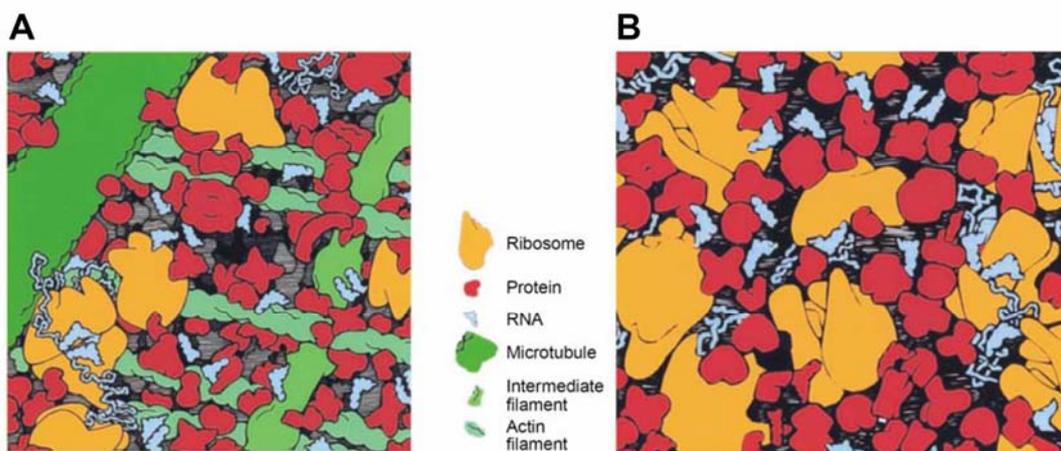
Figure 4. The crowded state of the cytoplasm

(A) Eukaryotic and (B) *E. coli* cells. Each square illustrates the face of a cube of cytoplasm with an edge 100 nm in length. The sizes, shapes and numbers of macromolecules are in the order of actual cytosolic concentration. Small molecules are not shown. Adapted from Ellis (2001).

Misfolding and aggregation in the cell leads not only to inactivation of the affected proteins, but can result in severe cellular dysfunction, causing a number of human diseases (Barral *et al.*, 2004). Protein misfolding diseases can be divided into two groups: in the first, diseases in caused by excessive amounts of misfolded proteins aggregating into the form of fibrils. A well known disease of amyloidoses is Alzheimer's disease (Dobson, 2003). One of the main characters of Alzheimer's disease is the accumulation of insoluble β-amyloid in the brain. The second group of protein misfolding diseases are those caused by smaller errors (point mutations) in the genetic blueprint leading to incomplete folding of a protein,

which affects its function. For instance, there are several point mutations found to constraint the folding of p53 with high correlation in association with different cancers, such as R249S in hepatocellular carcinoma as one of the most prominent case (Hsu *et al*., 1991).

Over the past two decades, a group of molecules have been identified that assist non-native proteins in folding under cellular conditions. This group of molecules is referred to as molecular chaperones (Martin and Hartl, 1993). Molecular chaperones interact with non-native polypeptides and prevent their misfolding and aggregation. They contribute towards successful folding to the native structure without providing specific conformational information to the folding process, in agreement with the conclusion reached by Anfinsen that the tertiary structure of a protein is solely determined by its own amino acid sequence (Anfinsen, 1973).

2.3. Molecular chaperone systems

In order to assist many diverse types of proteins to fold in living cells, molecular chaperones have evolved to perform versatile tasks by cooperating with each other (Young *et al*., 2004). In both prokaryotic and eukaryotic cells, the cytosol provides a well-developed network of chaperone pathways that help polypeptides from translation to the folded protein.

2.3.1. The chaperone network in the cytosol

As summarized in Figure 5 (Hartl and Hayer-Hartl, 2002), all three kingdoms of life are equipped with molecular chaperones for efficient protein folding. All nascent chains presumably interact with ribosome-associated chaperones. Small polypeptides chains (~65 to 80% of total) probably fold rapidly upon releasing from these chaperones. Longer nascent chains (10 to 20% of total) interact subsequently with Hsp70/Hsp40 system which assists

folding through ATP-dependent binding and release. About 10 to 15% of proteins which fail to fold with the upstream chaperones are further transferred to the chaperonin system which provides a sequestered environment for protein folding.
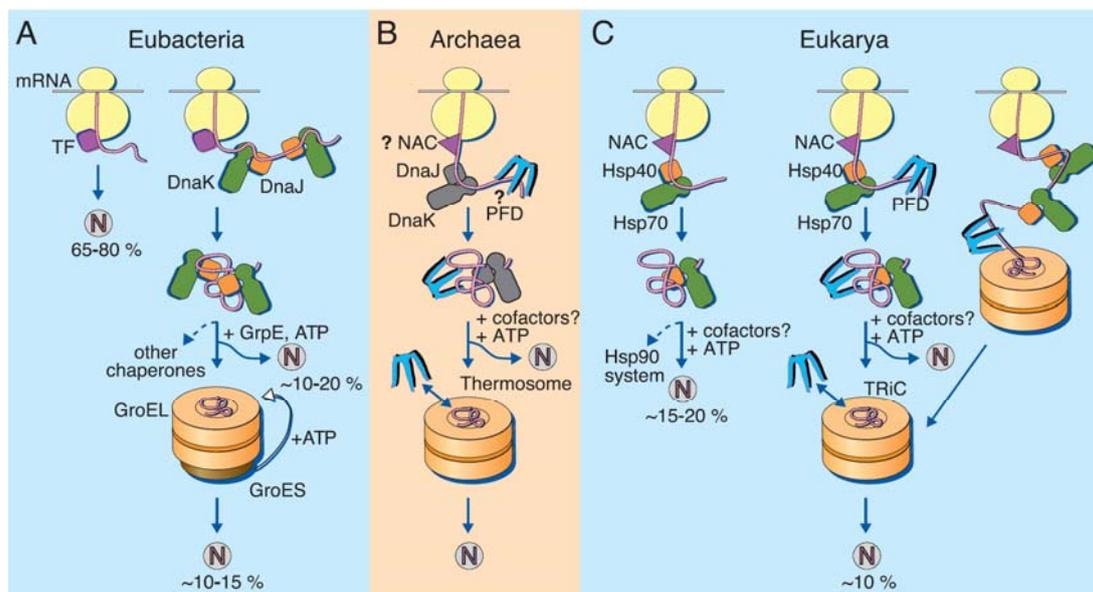
Figure 5. Models for the chaperone-assisted folding of *de novo* polypeptides

(A) In eubacteria, nascent chains probably first interact with trigger factor (TF), and most small proteins (~65 to 80%) fold to their native protein (N) without further chaperone interaction. Longer chains (10 to 20%) interact consequently with the DnaK/DnaJ/GrpE system and fold upon the ATP-dependent binding and release. About 10 to 15% of chains transit the chaperonin system (GroEL and GroES) for folding after their interaction with DnaK. (**B**) In archaea. Only some species contain DnaK/DnaJ. Interaction of prefoldin (PFD) with nascent chains and existence of nascent chain-associated complex (NAC) are not experimentally confirmed. (**C**) In eukarya, NAC probably interacts with nascent chains similar to TF in bacteria. The majority of small chains may fold upon release from ribosome. About 15 to 20% of chains reach their native states by Hsp70 and Hsp40. A subset of Hsp70 substrates needs further assistances by the Hsp90. About 10% of chains are co- or post-translationally passed on to the chaperonin TRiC in a reaction mediated by PFD. From Hartl and Hayer-Hartl (2002).

2.3.2. Ribosome–associated chaperones

Nascent chains emerging from the ribosomal exit tunnel are generally awaited by ribosome-associated chaperones, such as trigger factor (TF) in bacteria and nascent chain-associated complex (NAC) in eukaryotes (Figure 5) (Hartl and Hayer-Hartl, 2002).

In *E. coli*, TF was found to interact with most nascent chains by a cross-linking approach (Hesterkamp *et al*., 1996). TF is a 48 kDa protein, which binds at a 1:1 stoichiometry to the large subunit of the ribosome (Hesterkamp *et al*., 1996). TF docks specifically through separate domain onto the ribosomal L23 protein and also contacts the ribosomal protein L29. Although TF has peptidyl-prolyl *cis/trans* isomerase (PPIase) activity, the significance of PPIase activity of TF remains unclear (Genevaux *et al*., 2004; Kramer *et al*., 2004). Notably, TF is an ATP-independent chaperone, it does not assist folding in a nucleotide-regulated manner. TF is thought to function by scanning nascent chains and shielding hydrophobic regions to keep them from misfolding or aggregating.

In eukaryotes, the cytosol contains instead of TF a ribosome-associated heterodimeric complex of α (33 kDa) and β (22 kDa) subunits, called NAC (Shi *et al*., 1995; Wiedmann *et al*., 1994). NAC was shown to bind ribosomes and interact with nascent chains as they emerge from the ribosomal exit tunnel. NAC seems to influence the fidelity of co-translational targeting of nascent chains to the endoplasmic reticulum (Wiedmann *et al*., 1994). The actual function of NAC in protein folding remains to be established. However, NAC has no ATPase activity and probably functions by simply binding and protecting hydrophobic stretches of nascent chains in a manner similar to TF. In addition to NAC, the eukaryotic model organism *Saccharomyces cerevisiae* contains another hetero-dimeric chaperone, called ribosome-associated complex (RAC), which associates with ribosome nascent chain complex (RNC). RAC consists of Ssz1, a member of the Hsp70 family, and the DnaJ-related Hsp40 protein zoutin (Gautschi *et al*., 2001; Gautschi *et al*., 2002;

Michimoto *et al*., 2000). The J-domain of Zoutin was proposed to interact with Hsp70's, most likely serving as the Hsp40 partner for Ssb1 and Ssb2 (Bukau *et al*., 2000; Yan *et al*., 1998). Mutations in the J-domain of zoutin or deletion of the *ssz1* gene, but not the Ssz1 truncated mutant lacking the putative substrate binding domain, result in the same phenotype that is observed upon deletion of the *ssb* genes, indicating that RAC is required for the recruitment of the Ssb proteins to the ribosome (Gautschi *et al*., 2002; Hundley *et al*., 2002). Ssz does not seem to rely on ATPase activity (Huang *et al*., 2005), therefore it might act as a modulator of zoutin rather than as a chaperone itself (Huang *et al*., 2005).

2.3.3. The Hsp70 system

The Hsp70 chaperones are monomeric proteins of ~70 kDa in size, composed of a ~45 kDa amino-terminal ATPase domain and a ~25 kDa carboxyl-terminal polypeptide binding domain (Figure 6A). They are found in the cytosol of eubacteria, eukarya and some achaea (Figure 5), as well as within eukaryotic organelles, such as mitochondria and endoplasmic reticulum. They have important roles in protein metabolism both under stress and non-stress conditions, including in *de novo* protein folding, membrane translocation and the degradation of misfolded proteins (Bukau *et al*., 2006).

The crystal structures of individual domains of Hsp70s have been solved: the N-terminal ATPase domain structure of DnaK binding to its nucleotide exchange factor, GrpE (Harrison *et al*., 1997), and a C-terminal domain that binds substrate polypeptides (Figure 6 A) (Zhu *et al*., 1996). The substrate-binding domain can be further divided into a β-sandwich subdomain with a peptide-binding cleft and an α-helical latchlike segment (Zhu *et al*., 1996). Recently, a full length structure of bovine Hsc70 has been determined (Jiang *et al*., 2005), that allows us further understanding of the inter-domain interaction and how the ATPase and the substrate cycling proceed for Hsp70.
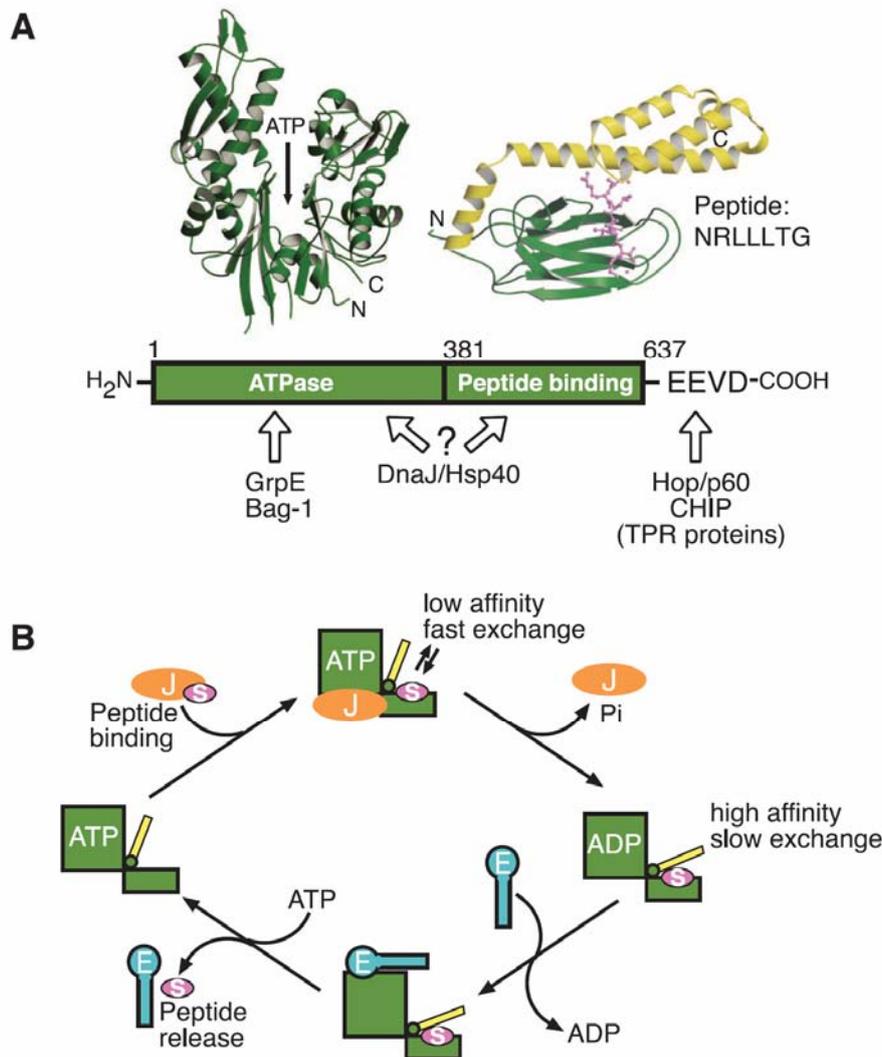
Figure 6. The structure and reaction cycle of the DnaK system

(A) (Top) Structures of the ATPase domain (Harrison *et al.*, 1997) and the peptide-binding domain (Zhu *et al.*, 1996) of DnaK. ATP indicates the nucleotide binding site of ATPase domain. The α-helical latch of peptide-binding domain is shown in yellow and the peptide substrate in pink. (Bottom) The interaction of prokaryotic and eukaryotic cofactors with Hsp70 is shown schematically. The C-terminal EEVD sequence is only present in eukaryotic Hsp70s and is involved in binding of TPR motifs. (B) K: DnaK; J: DnaJ; E: GrpE, S: substrate peptide. Non-native substrate polypeptides associate with either DnaJ (J) or DnaK (K) in the ATP-bound open state. DnaJ and substrate protein stimulate ATP hydrolysis of DnaK, leading to closure of its substrate binding pocket. GrpE is required for efficient release of ADP from the complex, and subsequent ATP binding results in opening of the substrate binding pocket and exchange of substrate polypeptides. The released substrate can either fold towards the native state or rebind to DnaJ or DnaK. Adapted from Hartl and Hayer-Hartl (2002).

The best studied Hsp70 is the *E. coli* Hsp70, DnaK. Its reaction cycle is shown in Figure 6B. The cochaperone DnaJ (41 kDa) first interacts with an unfolded polypeptide *via* its C-terminal domain, and then targets the polypeptide to DnaK by its N-terminal J domain. DnaK binds the polypeptide in the ATP-bound state. The interaction with DnaJ then stimulates the hydrolysis of ATP by DnaK and stabilizes its ADP-bound state (Mayer *et al*., 2000). In the ATP-bound state, DnaK binds and releases substrates rapidly; the α-helical latch over the peptide-binding cleft is in an open conformation (Figure 6B). In contrast, the ADP-bound state has slow exchange rate for peptides since the α-helical latch is closed over the substrate (Zhu *et al*., 1996). The nucleotide exchange factor GrpE promotes the release of ADP from DnaK, which is rate-limiting in this cycle. The substrate dissociates from DnaK upon subsequent ATP binding to DnaK (Schmid *et al*., 1994) and has the option of either folding, rebinding to DnaJ and DnaK or being transferred to another chaperone system, such as the chaperonin, for final folding (Figure 5).

Eukaryotic Hsp70s follow the similar ATPase cycle as DnaK. The DnaJ homologs (Hsp40s) also interact with polypeptides and stimulate the ATP hydrolysis of their partner Hsp70 proteins (Johnson and Craig, 2001; Lopez *et al*., 1998). Although the prokaryotic and eukaryotic Hsp70 systems have a similar mechanism, a GrpE homolog for nucleotide exchange has not been found in the eukaryotic cytosol. The nucleotide exchange reactions for Hsp70s in the eukaryotic cytosol are performed by structurally unrelated proteins, for instance, the mammalian co-chaperone BCL2-associated athanogene-1 (Bag1) (Hohfeld and Jentsch, 1997) and the mammalian Hsp70-binding protein (HspBP1) and its *S. cerevisiae* homolog Fes1(Kabani *et al*., 2002a; Kabani *et al*., 2002b). More recently Hsp110, a distant homolog of Hsp70 was found to also serve as a nucleotide exchange factor for Hsp70 (Dragovic *et al*., 2006; Raviol *et al*., 2006). Interestingly, these factors distort the ATPase domain to release ADP in different manners and efficiencies, suggesting the nucleotide exchange event is a regulatory tactic rather than an acceleration strategy for eukaryotic

Hsp70s. A possible explanation could derive from the general feature that eukaryotic Hsp70s have slower ATP hydrolysis rates than their prokaryotic counterparts, which implies that the rate-limiting step is the ATPase cycle of eukaryotic Hsp70s is not dissociation of ADP but the hydrolysis of ATP itself.

How does Hsp70 recognize unfolded polypeptides? By using phage display and synthetic peptide library approaches, Hsp70 was found to recognize linear polypeptide sequence enriched in hydrophobic amino acids (Flynn *et al*., 1991; Rudiger *et al*., 1997b). Because of its hydrophobic nature, this binding motif would typically be located in the interior of a folded protein; consequently, surface exposure of such a sequence may be a distinctive feature of nonnative conformations. Such hydrophobic regions are probably present in all unfolded polypeptides, and it has been predicted that an Hsp70-binding site occurs, on average, every 40 residues (Rudiger *et al*., 1997a). Association with Hsp70 results in the stabilization of the substrate protein in an extended conformation, thereby preventing its aggregation. *In vitro*, polypeptides can undergo multiple rounds of binding and release from Hsp70. This process is sufficient to promote folding of some model substrates, such as firefly luciferase (Szabo *et al*., 1994). However, in many cases, the Hsp70-bound substrates must be transferred to the chaperonin system for productive folding.

Besides ribosome-associated chaperones, Hsp70s are important in chaperoning nascent chains. As it has been shown that DnaK preferentially associates with elongating polypeptides larger than 20 to 30 kDa, thus DnaK interacts with nascent chains subsequent to TF (Teter *et al*., 1999) (Figure 5A). DnaK has an overlapping chaperone function with TF. Upon deletion of TF in *E. coli*, DnaK can substitute TF function in chaperoning nascent chains. However, the deletion of both TF and DnaK at 37°C leads to a severe loss of viability, and results in the accumulation of misfolded, aggregated proteins (Agashe *et al*., 2004; Teter *et al*., 1999). Intriguingly, while bacterial proteins efficiently apply TF and

DnaK for folding, their actions in concert were found to cause a delay of the folding process relative to translation for certain multi-domain proteins. The unexpected working paradigm indicates a fundamental difference of folding pathways utilized in bacterial and eukaryotic cells to prevent misfolding and aggregation during translation (Agashe *et al*., 2004). Similar to DnaK, mammalian Hsc70 also binds a wide range of nascent chains (> 15% of total), including many multi-domain protein >50 kDa (Thulasiraman *et al*., 1999). In cooperation with the Hsp40 homologs Hdj1 and Hdj2, Hsp70s can achieve the co-translational folding of polypeptides (Nagata *et al*., 1998; Terada *et al*., 1997). In addition, Hsc70 is probably recruited to the ribosome by the recently identified human zoutin ortholog Mpp11, which illustrates that ribosome-tethered chaperones have been conserved through evolution (Hundley *et al*., 2005).

2.3.4. The chaperonins: Hsp60 and Hsp10

Chaperonins are a conserved class of large double-ring complex of ~800 kDa with a central cavity (Braig *et al*., 1994; Hartl, 1996). There are two chaperonin subgroups that are similar in architecture but distantly related in amino acid sequence. Group I chaperonins, are generally found in eubacteria, such as GroEL in *E. coli* and the Hsp60 in mitochondria and chloroplasts. They cooperate with a ring-shaped co-chaperone, GroES or Hsp10, that forms the lid on a folding cage in which polypeptide substrates can be encapsulated during folding (Mayhew *et al*., 1996; Weissman *et al*., 1996). While Group II chaperonins exist in archaea and eukarya. They are independent of a co-chaperone but have helical protrusions in the apical domains of the chaperonin subunits which function as an inbuilt lid (Ditzel *et al*., 1998).

Bacterial GroEL is the most prominent example of Group I chaperonins (Bukau and Horwich, 1998; Hartl, 1996; Hartl and Hayer-Hartl, 2002). GroEL was first identified in the early 1970s by Costa Georgopoulos and colleagues; they observed that certain temperature-

sensitive mutations in the *GroE* operon were unable to support the growth of bacteriophage λ (Georgopoulos *et al*., 1973). It was subsequently demonstrated by Georgopoulos's group that the two proteins encoded in this operon, GroES and GroEL, are essential for the viability of *E. coli* at all temperatures (Fayet *et al*., 1989). In 1980s, John Ellis and colleagues identified a 60 kDa RuBisCo binding protein (RBP) that transiently associated with chloroplast RuBisCo Large-subunits during synthesis (Barraclough and Ellis, 1980) and RBP was then recognized as a homolog of GroEL. The *in vitro* refolding efficiency of bacterial RuBisCo was shown to be significantly improved by GroEL and GroES (Goloubinoff *et al*., 1989). Along with these findings, mitochondrial hsp60 was soon found to play a role in the folding of proteins imported into mitochondria (Ostermann *et al*., 1989).

GroEL contains 14 identical 57-kDa subunits arranged in two stacked rings of seven subunits each. And according to its crystal structure, each subunit of GroEL consists of three domains; the equatorial domain contains the ATP-binding site and is connected *via* a hinge-like intermediate domain to the apical domain (Figure 7) (Braig *et al*., 1994; Steinbacher and Ditzel, 2001; Xu *et al*., 1997). The equatorial domain contains the ATP-binding sites and mediates the most inter-subunit contacts within and between GroEL rings. The small intermediate domain has flexible hinge regions at the domain junctions that allow large structural rearrangements upon cooperative binding of seven ATP molecules and subsequent GroES binding. The apical domain forms the ring opening and exposes a number of hydrophobic residues toward the cavity for the binding of a nonnative substrate as well as for interacting with GroES (Fenton *et al*., 1994; Xu *et al*., 1997), a hepameric dome-shape ring of 10-kDa subunits. (Figure 7A) (Hunt *et al*., 1996). The details of the GroEL and GroES system will be discussed more thoroughly in the next section.

Group II chaperonins are more heterogeneous in sequence and structure than members of Group I (Leroux and Hartl, 2000). Although Group II chaperonins also have a double ring structure, they are generally hetero-oligomeric and the number of subunits varies

between eight and nine *per* ring (Figure 7B and 7D). For example, the eukaryotic chaperonin, TRiC (TCP-1 ring complex) or CCT (chaperonin containing TCP-1), contains eight different subunits *per* ring, ranging from 50 to 60 kDa (Frydman *et al.*, 1992; Gao *et al.*, 1992). On the other hand, the archaeal chaperonin, thermosome, consists of one to three different subunits, which are arranged in eight- or nine-fold symmetrical rings. The crystal structure of the archaeal thermosome complex revealed that individual subunits have a domain arrangement similar to those in GroEL (Ditzel *et al.*, 1998). In contrast to the equatorial (ATP-binding) domain which is relatively well conserved among all chaperonins, apical domain of Group II chaperonin is highly divergent in sequence from that in GroEL. Although the backbone trace of the apical domain of the thermosome is almost identical to that of GroEL, an additional α-helical insertion is found to extend as a large protrusion toward the central cavity (Klumpp *et al.*, 1997). Since the major difference between Group I and Group II chaperonins is the lack of a GroES-like cofactor for Group II chaperonins, this protrusion is thought to function as a built-in lid, as well as to be involved in substrate binding (Klumpp *et al.*, 1997; Llorca *et al.*, 1999).

Compared to the Group I chaperonins, the mechanism by which the Group II chaperonins mediate protein folding is less understood. The most prominent substrates of TRiC are cytoskeletal proteins actin and tubulin. Interestingly, the folding of these proteins cannot be mediated by the bacterial homolog, GroEL and GroES. Considering the high abundance of actin and tubulin in cells, and only a low concentration of TRiC exist in the cytosol, it strongly suggests that TRiC might have a more specific function in folding a subset of cytosolic proteins. Additionally, TRiC was shown to interact with nascent polypeptides like actin and firefly luciferase, by using a cross-linking approach (McCallum *et al.*, 2000), indicating that TRiC may act co- translationally in the folding of proteins that are too big to be encapsulated as a whole (Frydman *et al.*, 1994), unlike the post-translational folding catalyst GroEL. Consistent with that, a pulse-chase analysis in

mammalian cells revealed that TRiC interacts transiently with a wide range of newly synthesized chains of 30 to 120 kDa in size (Thulasiraman *et al.*, 1999).
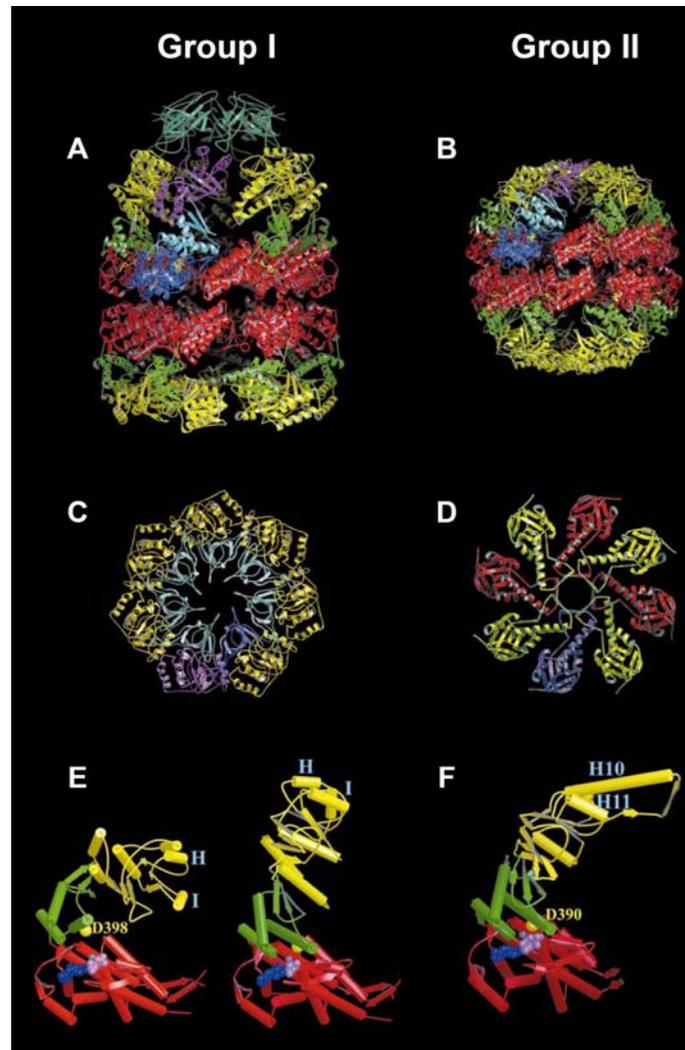
Figure 7. Architecture of Group I and II chaperonins.

(A) Side view of Group I chaperonin, GroEL/GroES complex in ADP-bound state (Xu *et al.*, 1997). The equatorial ATPase domain (red) is linked to the substrate-binding apical domain (yellow) by a flexible hinge or intermediate domain (green). GroES is shown in dark green. (B) Side view of the *T. acidophilium* thermosome (Ditzel *et al.*, 1998), color coded as in (A). (C) Top view of the *cis* ring with apical domains in yellow and GroES in gray. (D) Top view showing only the apical domain of thermosome. (E) Domain arrangements in GroEL: the left panel shows the nucleotide-free state of GroEL and the right panel presents that in the ADP-bound state. (F) Domain arrangement in the thermosome. Nucleotides are shown as ball models. Modified from (Steinbacher and Ditzel, 2001).

The nascent chain-associated chaperone, prefoldin (PFD), was found to cooperate with Group II chaperonins. PFD, also known as the GimC (genes involved in microtubule biogenesis complex), is a ~90 kDa complex of two α and four β subunits in the archaeal and eukaryotic cytosol (Siegert *et al*., 2000; Vainberg *et al*., 1998). Substrates binding and release by PFD is ATP independent, and *in vitro*, mammalian and archaeal PFD can stabilize nonnative proteins for subsequent transfer to a chaperonin (Leroux *et al*., 1999; Siegers *et al*., 1999). The substrate-binding site of PFD overlaps with its chaperonin-binding site, suggesting a mechanism for the transfer of PFD-bound substrates to Group II chaperonin (Martin-Benito *et al*., 2002; Okochi *et al*., 2004). The completion of folding of substrates is most likely achieved by Group II chaperonin after the delivery by PFD.

2.4. The *E. coli* chaperonin system: GroEL and GroES

2.4.1. Structure and function of GroEL and GroES

The structure of the GroEL/GroES complex has been extensively studied by electron microscopy (Braig *et al*., 1993; Langer *et al*., 1992; Saibil *et al*., 1991) and X-ray crystallography (Braig *et al*., 1994; Xu *et al*., 1997). Under physiological conditions, one GroES heptamer binds to one GroEL tetradecamer, thereby forming an asymmetric complex. The GroES bound ring is called *cis* ring, and the unliganded GroEL ring is called *trans* ring (Figure 8).

The domain rearrangement of GroEL between nucleotide-free and bound states is shown in Figure 9. The dramatic reshaping of the *cis* ring is due to the intermediate and apical domains rearrangements (Figure 9C) (Xu *et al*., 1997). First, the intermediate domain swings down towards the equatorial domain and the central cavity, pivoting approximately 25° around Pro137 and Gly 410. This movement locks the nucleotide binding sites and generates new interactions with the bound nucleotide and the equatorial domain. Second, the
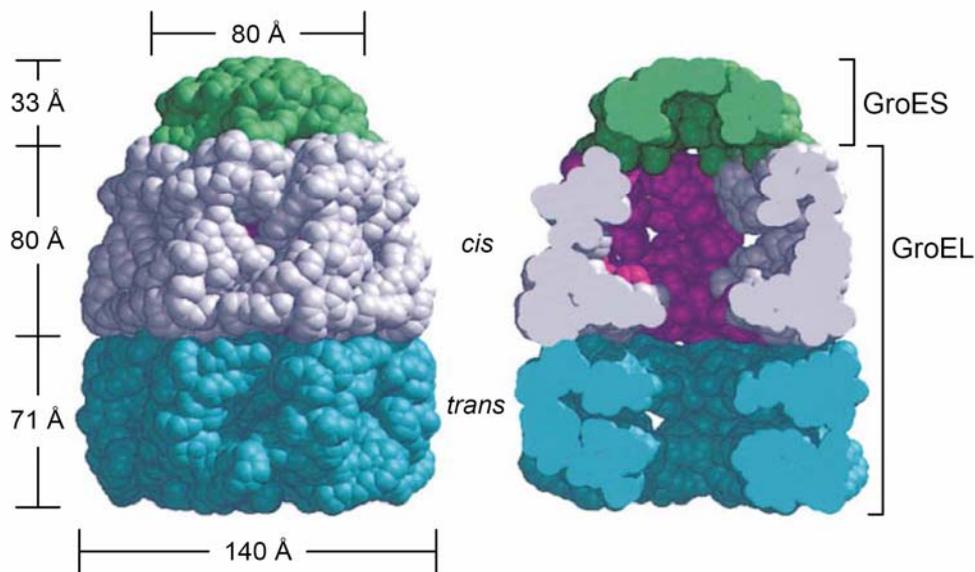
Figure 8. Architecture of GroEL-GroES-(ADP) $_7$ complex

Space-filling models of GroEL–GroES–(ADP) $_7$ with 6 Å Van der Waals spheres around Ca atoms. The *cis* ring of GroEL is gray and *trans* ring is blue. GroES is shown in green. The left panel is view from outside; the right panel is from the inside, generated by slicing the models with a vertical plane that contains the cylindrical axis. Figures are modified from Bukau and Horwich, (1998).

apical domain swings up 60° and twists around the long axis of the domain about 90°, forming new interactions with nearby apical domains and leading to an interaction with the mobile loop of GroES. Eventually, the domain rearrangements result in burying hydrophobic residues and hence changing the environment inside the GroEL-GroES cavity to hydrophilic (Figure 10). In addition the volume of the cavity is approximately enlarged by two-fold (Hayer-Hartl *et al*., 1996; Roseman *et al*., 1996; Weissman *et al*., 1994). This hydrophilic cage can accommodate single partially folded polypeptides up to ~60 kDa (Sigler *et al*., 1998; Viitanen *et al*., 1992).
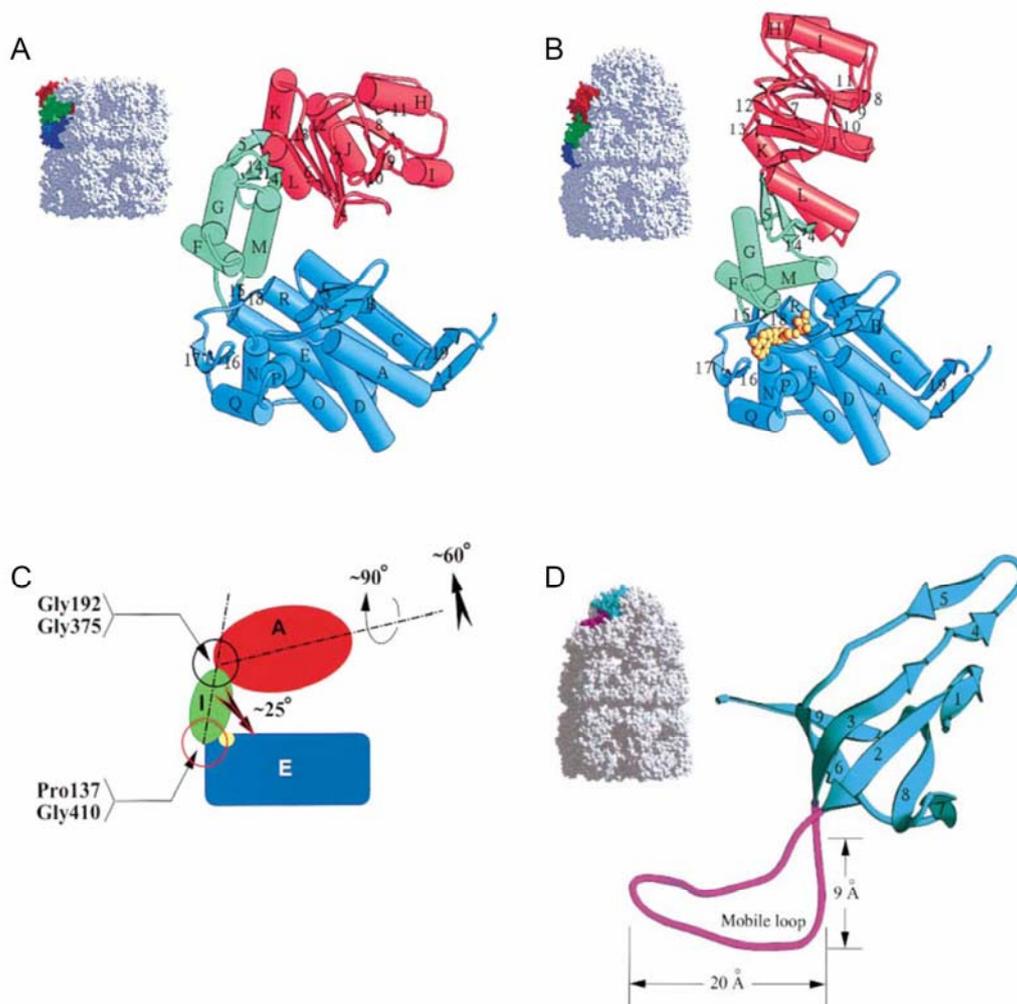
Figure 9. The domain movements within individual subunits of the *cis* GroEL ring

The upper panels show ribbon diagrams of an individual subunit of unliganded (A) and liganded (B) GroEL. The orientation of the representative subunit is the same as the colored subunit in the nearby space-filling model. The equatorial, intermediate, and apical domains are blue, green, and red, respectively. Note that GroES is only shown in B, to reveal more clearly the extent of motion of the apical domain. The nucleotide (ADP) in the right-hand structure is a yellow space-filling model. (C) Schematic representation of GroEL showing diagrammatically the en bloc movements that occur around the pivot points at the ends of the intermediate domain. Domains are colored as in the upper panels, and the small yellow circle on the top of the equatorial domain represents the nucleotide. (D) The left panel shows a space filling model of a side view of a GroEL-GroES complex. The GroES subunit is colored individually. The right panel shows a side view of a single GroES subunit in a ribbon drawing. Mobile loop as well as the GroEL interacting loop is indicated. Adapted from Xu *et al*. (1997).

Similar to GroEL, GroES forms a heptameric ring of ~10 kDa subunits. Each GroES subunit is folded into a single domain, which contains nine β-strands with one exceptionally long β-hairpin loop, the so-called mobile loop (Figure 9D). The GroEL-GroES contact is mediated through the mobile loop (Landry *et al*., 1993). This loop contains 16 amino acids which appear highly mobile in the uncomplexed GroES but become more structured upon interacting with GroEL. Because the hydrophobic binding regions of GroEL for polypeptide overlap with those for GroES (Fenton *et al*., 1994), GroES binding leads to the displacement of the bound polypeptide from the apical domains into the central cavity which is then permissive for folding (Martin *et al*., 1993; Weissman *et al*., 1996).
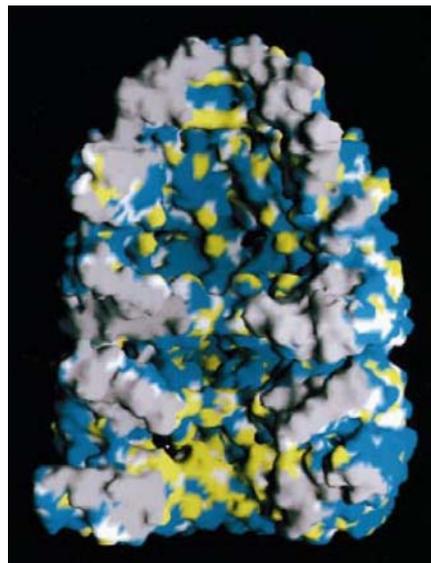
Figure 10. Hydrophobic surface at the interior of the GroEL cavity

An interior view of four subunits from each ring of the asymmetric structure, colored to reflect the relative hydrophobicity of the interior surface. Hydrophobic side-chain atoms are shown in yellow; polar and charge side-chain atoms are shown in blue. All solvent – excluded surfaces at the subunit interface are grey; and exposed backbone atoms are white. The surface of *trans* GroEL is considerably more hydrophobic than the surface of *cis* GroEL. From Xu *et al*. (1997)

2.4.2. Substrates of GroEL and GroES

Almost all unfolded proteins, including some heterologous substrates such as mitochondrial malate dehydrogenase, *R. rubrum* RuBisCo and bovine rhodanese, interact with GroEL *in vitro* (Goloubinoff *et al*., 1989; Hartman *et al*., 1993; Martin *et al*., 1991). *In vivo*, 10-15 % of cytosolic proteins of *E. coli* were observed to co-immunoprecipitate with GroEL (Ewalt *et al*., 1997; Houry *et al*., 1999). A recent extensive proteomic study revealed that approximately 250 different proteins of the ~2,400 cytosolic *E. coli* proteins interact with GroEL upon synthesis (Kerner *et al*., 2005). These newly identified GroEL substrates are classified into three classes. Class I proteins are largely independent of chaperone interaction but their folding yield can be increased by chaperones. Class II proteins do not refold efficiently in the absence of chaperones, but can utilize either the Hsp70 (DnaK/DnaJ/GrpE) or the GroEL/GroES system for folding. Class III substrates are fully dependent on the GroEL/GroES system for folding. The Hsp70 system can bind Class III proteins and prevent their aggregation, but folding is only achieved upon transfer to GroEL. Of these 250 substrates, ~85 are predicted to be obligate chaperonin substrates (Kerner *et al*., 2005). Importantly, the obligate GroEL substrates include at least 13 essential proteins, explaining why the GroEL/GroES system is indispensable for *E. coli* viability (Fayet *et al*., 1989).

Obligate GroEL substrates are typically 30-50 kDa in size and display complex α/β or α+β domain topologies, with (βα)$_8$ TIM barrel domains being significantly represented compared to the fold distribution of total cytosolic proteins (Kerner *et al*., 2005). These proteins appear to rely on GroEL for folding and to avoid or overcome kinetically trapped states at a biologically relevant time scale.

2.4.3. Mechanisms of GroEL and GroES mediated protein folding

GroEL and GroES assisted polypeptide folding involves two functional elements: (1) prevention of aggregation by binding non-native polypeptides; and (2) release of unfolded polypeptides into a sequestered compartment permissive for folding. Although the binding of unfolded polypeptide is independent of GroES, GroEL is critically dependent on GroES for providing a folding compartment. GroES cycles on and off either end of GroEL cylindrical rings in a manner regulated by the ATPase activity of GroEL. The chaperonin reaction begins by the binding of substrate polypeptide to the nucleotide-free state of GroEL (*trans* ring) (Figure 11). ATP and GroES then bind to the same ring, thereby resulting in the displacement of substrate into a newly formed *cis* ring and causing the dissociation of seven ADP molecules and GroES from the former *cis* complex. Upon binding to GroES, the apical domains undergo a massive rotation and upward movement (Roseman *et al*., 1996; Xu *et al*., 1997), generating an enlarged hydrophilic cavity (Figure 10). Nonnative proteins up to ~ 60 kDa can be encapsulated and are free to fold in the GroEL-GroES cage, the so-called Anfinsen cage (Brinker *et al*., 2001; Ellis, 2001b; Mayhew *et al*., 1996; Weissman *et al*., 1996). Folding is allowed to proceed for ~ 10 s, timed by the hydrolysis of seven ATPs in the *cis* ring. Binding of ATP and GroES to the opposite *trans* ring triggers the opening of the Anfinsen cage. Both native protein and nonnative intermediate exit at this point (Figure 11), but folding intermediates that still expose extensive hydrophobic surfaces are rapidly recaptured and the refolding cycle is repeated until the protein reaches its native state.

Although there is general agreement on the GroEL *cis* cycling process, two competing models, the "Anfinsen cage" model and the "iterative annealing" model, however, have been proposed to address how chaperonins can accelerate protein folding. (Betancourt and Thirumalai, 1999; Ellis and Hartl, 1996; Wang and Weissman, 1999).
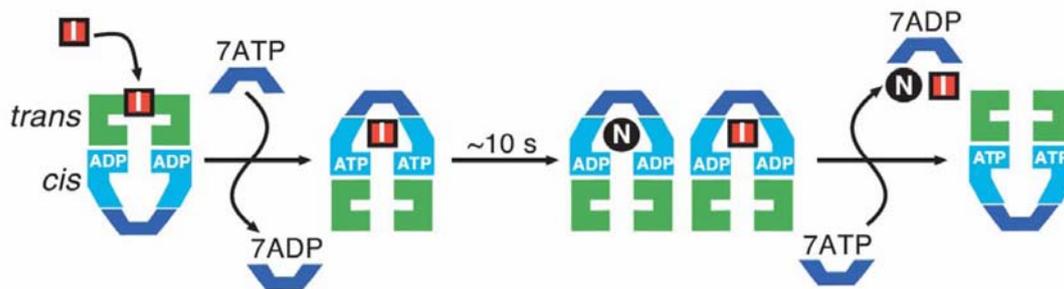
Figure 11. Simplified reaction of protein folding in the GroEL-GroES cavity.

A folding intermediate (I) binds to the apical domains of the *trans* ring (green) of GroEL. Binding of ATP and GroES (a dome-shape heptamer indicated in blue) displaces this polypeptide into the new forming *cis* cavity (cyan). Folding intermediate is given the chance to fold in this sequestered cage during one round of ATP hydrolysis (~10 s). Binding of other ATP and GroES to the opposite ring triggers the release of ADP, GroES and encapsulated polypeptide in either native (N) or intermediate state. If the polypeptide is still in intermediate state, it is then rapidly rebound by GroEL. Adapted from Hartl and Hayer-Hartl (2002).

The Anfinsen cage model is based on the hypothesis that a folding event is largely limited by intermolecular interactions that result in aggregation. This model proposes that the GroEL cavity provides a sequestered environment where folding of a substrate protein proceeds similar to infinite dilution (Ellis, 1996).

The iterative annealing model, on the other hand, proposes that the rate-limiting step in protein folding is the intramolecular reorganization of misfolded and kinetically trapped protein (Sosnick *et al*., 1994). This model suggests that the energy of ATP hydrolysis is used for the forceful unfolding of the misfolded substrate protein (Todd *et al*., 1996; Todd *et al*., 1994). Upon release from the apical domain into the bulk solution or the GroEL cavity, the actively unfolded substrate protein is then given a chance to fold. Incompletely folded proteins undergo further unfolding cycles until the native state is achieved. Since the time of encapsulation in the GroEL/GroES complex (~10 s) is relatively short compared to the time

required for folding of many GroEL substrates (minutes), this model also suggests that folding may continue to proceed outside the chaperonin cage upon release.

Although the iterative annealing model has been supported by an intriguing hydrogen-tritium exchange measurement that detected partial unfolding of bacterial ribulose bis-phosphate carboxylase (RuBisCo) during a single reaction cycle of GroEL (Shtilerman *et al*., 1999), similar unfolding was failed to be detected by NMR analysis of using malate dehydrogenase as substrate (Chen *et al*., 2001), neither with analogous experiment on human cyclophilin (Zahn *et al*., 1994), human DHFR (Gross *et al*., 1996), and β-lactamase (Gervasoni *et al*., 1998) is displaying such protection and hydrogen exchange. Indeed, substrate protein folding in the GroEL central cavity as described in the Anfinsen cage model has been repeatedly shown with several substrates including RuBisCo, mMDH, and rhodanese (Wang *et al*., 1998), as well as the use of the non-cycling single-ring mutant of GroEL for following GFP folding (Weissman *et al*., 1996), nicely addressed this *cis* cavity is central to allow folding to occur.

One of the most convincing results that supported the cage model over the active unfolding model came from the work of Brinker *et al* (Brinker *et al*., 2001), where they applied a clever method to prevent released substrate proteins from rebinding to GroEL by coupling biotin to cysteine residues located at the apical domain of GroEL, together with the addition of streptavidin to rapidly and irreversibly block the entry to the cage. Under buffer conditions in which partly folded chains of RuBisCo readily aggregate, the addition of streptavidin after several ATPase reaction cycles blocked RuBisCo folding completely and instantly. This observation is inconsistent with the iterative annealing model, which predicts that some released chains would fold correctly in free solution. It is therefore in accord with the cage model, which predicts that correct folding occurs only within cage.

2.5. Aim of the study

The basic mechanism of GroEL/GroES action involves encapsulation of a single molecule of non-native protein in a cage-like structure, thereby allowing folding to occur unimpaired by aggregation (Mayhew *et al*., 1996; Weissman *et al*., 1996). However, recent experimental findings together with theoretical analyses suggested that the physical environment of the chaperonin cage, in addition to providing a sequestrated folding space, may profoundly affect the energy landscape and the kinetic trajectories along which folding proceeds (Baumketner *et al*., 2003; Brinker *et al*., 2001; Jewett *et al*., 2004; Takagi *et al*., 2003; Zhou, 2004). Evidence for such effect was provided for bacterial RuBisCo, a TIM barrel protein and model GroEL substrate. Folding of RuBisCo inside the cage occurred with a considerably faster rate than spontaneous folding, even when aggregation in free solution was avoided by adjusting protein concentrations to very low levels (Brinker *et al*., 2001). This discovery offered the prospect of using the chaperonin system as an experimental tool to address a series of questions fundamental to our understanding of protein folding in general.

Although the theoretical analysis suggested that this rate enhancement is due to the spatial confinement of the folding protein, which would entropically destabilize unfolded conformations and reduce the search time for the energy basin of the compact native state (Baumketner *et al*., 2003; Takagi *et al*., 2003; Zhou, 2004), actual biochemical support illustrating this mechanism is still missing. To provide such information, here we performed a mutational analysis of the GroEL cavity to examine the structural features that play a critical role in accelerating folding. In probing the geometric confinement as a major contributor, we systematically reduced or increased the volume of the chaperonin cage and followed the consequences on folding speeds of different-sized model substrates. In addition, a number of conserved negative charges exposed on the cavity wall were also investigated for its impact in facilitating protein folding.

3. Materials and Methods

3.1. Materials

3.1.1. Chemicals

L-Amino acids	Sigma-Aldrich
Acetic acid	Merck
Adenosine 5′-(β,γ-imido)triphosphate tetralithium salt (AMP-PNP)	Sigma-Aldrich
Adenosine triphosphate, disodium salt (ATP)	Sigma-Aldrich
Agarose (SeaKem LE)	Cambrex Bio Science
Alexa Fluor 488 C_5 maleimide	Molecular Probes
Ammonium persulfate (APS)	Sigma-Aldrich
Ampicillin	Merck
Amylose resin	New England Biolabs
Arabinose	Sigma-Aldrich
Bacto agar	Difco
Bacto trypton	Difco
Bacto yeast extract	Difco
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Calcium chloride	Merck
CDTA (trans-1,2-diaminocyylohexane-N,N,N',N'-tetracetic acid)	Sigma-Aldrich
Chloramphenicol	Sigma-Aldrich
Complete EDTA-free protease inhibitor	Roche
Coomassie brilliant blue R-250	Roth
Dextran 40	Sigma-Aldrich
Dimethylsulfoxide (DMSO)	Merck
DTSSP [3,3'-dithiobis(sulfo-succinimidylpropionate)]	PIERCE

Dithiothreitol (DTT)	Roche
ECL™ detection kit	Amersham Pharmacia Biotech
Ethanol	Merck
Ethidium bromide	BioRad
Ethylenediaminetetraaceticacid –sodium salt (EDTA)	Merck
Ferric nitrate	Sigma-Aldrich
Flavine adenine dinucleotide (FAD)	Sigma-Aldrich
Ficoll 70	Sigma-Aldrich
Formaldehyde	Sigma-Aldrich
Glucose	Sigma-Aldrich
Glycerol	Merck
Glycine	Roth
Guanidium hydrochloride (GuHCl)	Sigma-Aldrich
HEPES	Sigma-Aldrich
Hydrochloric acid (37%)	Merck
IANBD (*N*-((2-(iodoacetoxy) ethyl)-*N*-methyl) amino-7-nitrobenz-2-oxa-1, 3-diazole ester	Molecular Probes
Imidazol	Merck
Isopropyl-β-D-thiogalactopyranoside (IPTG)	BioMol
Kanamycin	Sigma-Aldrich
Magnesium chloride	Merck
Malachite green	Sigma-Aldrich
Maltose	Sigma-Aldrich
Menadione	Sigma-Aldrich
β-mercaptoethanol	Sigma-Aldrich
Methanol	Merck
Nickel-NTA agarose beads	Qiagen
Nicotinamide adenine dinucleotide (NADH)	Roche
PIPES	Sigma-Aldrich
Phenyl-methyl-sulfonyl-fluoride (PMSF)	Sigma-Aldrich
Polyacrylamide/bisacrylamide solution 30 % (30 : 0.8)	Roth

Polyethylene glycol 2000 (PEG 2000)	Merck
Potassium Cyanide	Sigma-Aldrich
Potassium hydroxide	Sigma-Aldrich
D-ribulose 1,5-diphosphate (RuDP)	Sigma-Aldrich
Silver nitrate	Sigma-Aldrich
Sodium chloride	Merck
Sodium [^{14}C]bicarbonate	Amersham Pharmacia Biotech
Sodium dodecylsulfate (SDS)	Sigma-Aldrich
Sodium hydroxide	Sigma-Aldrich
Sodium thiosulfate	Merck
Spectinomycin	Sigma-Aldrich
Sucrose	Merck
N, N, N', N'-Tetramethylethylendiamine (TEMED)	Sigma-Aldrich
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl)	PIERCE
Tris-base	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Tween-20	Calbiochem

3.1.2. Enzymes

Apyrase	Sigma-Aldrich
Benzonase	Merck
Lysozyme	Sigma-Aldrich
Pfu DNA polymerase	Stratagene
Restriction enzymes	New England Biolabs
Rhodanese	Sigma-Aldrich
Shrimp Alkaline Phosphatase	Roche
T4 DNA ligase	New England Biolabs
Vent DNA polymerase	New England Biolabs

3.1.3. Materials

Centricon 10 kDa cut-off	Amicon
Centricon 30 kDa cut-off	Amicon
High performance chemiluminescence film	Amersham Pharmacia Biotech
Microcon 10 kDa cut-off	Amicon
Microcon 30 kDa cut-off	Amicon
Nitrocellulose transfer membrane	Whatman Schleicher & Schuell
Sterile filter 0.22 μm	Millipore
Sterile filter 0.45 μm	Millipore

3.1.4. Instruments

AIDA gel imaging software version 2.31	Raytest
ÄKTA Explorer 100	Amersham Pharmacia Biotech
Balance AG285, PB602	Mettler Toledo
Centrifuges: Avanti J-25, Avanti J20 XP, J-6B, GS-6R	Beckmann
Centrifuges 5415C and 5417R	Eppendorf
Chromatography columns	Amersham Pharmacia Biotech
(HiPrep Desalting, MonoQ, HiTrap Heparin, Sephacryl S200/S300, Superdex 200, Superose 6, Sephadex G25 (NAP-5, NAP-10); chromatography resins: Q-Sepharose, DE52, Source 30 Q, Source 30 S)	
Deionization system MilliQ plus PF	Millipore
Electrophoresis chambers MiniProtean 3	Bio-Rad
Electrophoresis power supply Power PAC 300	Bio-Rad
Fluorescence spectrometer Fluorolog 3	HORIBA Jobin Yvon
FPLC systems	Amersham Pharmacia Biotech
EmulsiFlex high pressure homogenizer	Avestin
Gene Pulser II electroporation system	Bio-Rad
Gilson Pipetman (2, 10, 20, 100, 200, 1000 μl)	Abimed
Incubators Innova 4430	New Brunswick Scientific

Luminescent Image Analyzer LAS-3000	FUJIFILM
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad
PCR-Thermocycler T3	Biometra
pH meter Accumet Basic	Fisher Scientific
SMART system	Amersham Pharmacia Biotech
Sonicator Ultrasonic Processor XL	Misonix Inc.
Spectrophotometer DU 640 UV/VIS	Beckmann
Spectrophotometer LS50	Perkin-Elmer
Synergy HT UV/VIS/fluorescence/luminescence plate reader	Bio-Tek
UV/VIS Spectrometer V-560	Jasco
Thermomixer Comfort	Eppendorf
Vortex	Ikamag
Water bath	Bioblock Scientific

3.1.5. Media

LB medium:	10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, (+ 15 g/l agar for solid medium). Adjusted to pH 7.0 with NaOH (Sambrook *et al*., 1989).
SOC medium	20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l NaCl, 0.186 g/l KCl, 0.95 g/l $MgCl_2$. After autoclave, add 20 ml of filter sterilized 1M glucose (Sambrook *et al*., 1989).

3.1.6. Antibiotic stock solutions

Antibiotic additives to growth media were prepared as 1000x stock solutions and filter-sterilized before usage: ampicilin: 100 g/l, kanamycin: 25 g/l, Chloramphenicol 25 g/l, Spectinomycin 50 g/l.

3.2. Bacterial strains and plasmids

3.2.1. *E. coli* strains

DH5α F'

F'/ *endA1 hsdR17*(r_k^-, m_k^+) *glnV44 thi-1 recA1 gyrA* (Nal^r) *relA1Δ(lacIZYA-argF) U169 deoR (φ80dlacΔ(lacZ)M15)*

BL21(DE3) Gold
(Stratagene)

B strain, F- *dcm+* Hte *ompT hsdS*(r_B- m_B-) *gal l* (DE3) *endA* Tet^r

MC4100
(Genevaux *et al.*, 2004)

F- *araD139 Δ(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR*

MC4100 SC3
(Kerner *et al.*, 2005)

F- *araD139 Δ(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR PgroE::Para* Kan^R

3.2.2. Plasmids

GroEL and all chaperonin mutants were constructed in a pCH vector backbone (Chang *et al.*, 2005) inserted *via* the *Nde*I and *Nhe*I sites. Synthetic oligonucleotides encoding wild-type or mutant C-terminal extensions of GroEL were introduced into the pCH-ELΔC or SR-ELΔC plasmid between the *Nhe*I and *Hind*III sites. The SR-EL charge-mutants (SR-QNQ, SR-NNQ, SR-3N3Q, SR-KKK(1), SR-KKK(2), SR-D253N, SR-D253K, SR-D359N and SR-D359K) and MBP mutants (SM-MBP (Y283D), DM-MBP (V8G,Y283D), WT-MBP (D95C), SM-MBP (D95C), DM-MBP (D95C)) were generated by site-directed mutagenesis.

GroES was constructed in a pET11a vector inserted *via* the *Nde*I and *Bam*HI sites (Brinker *et al.*, 2001).

Constructs for chaperonin *in vivo* co-expression experiments were prepared by cutting out GroEL-mutant DNA fragments between *Cla*I, *Bsa*AI sites from pCH-EL plasmid series, and ligated into *Cla*I, *Ecl*136II cutted pOFXtacSL2 vector (Castanie *et al*., 1997). The resulting constructs are *tac* promoter driven, IPTG inducible with a p15A origin, which are compatible with ColEI constructs for co-expression.

Constructs of WT and mutant MBP and MetF for *in vivo* experiments were generated by subcloning from the vectors described above into the arabinose promoter controlled vector pBAD18 (Guzman *et al*., 1995) *via Xba*I and *Hind*III sites.

The sequences of all final constructs were confirmed by DNA sequencing (Sequiserve and Medigenomix GmbH).

All DNA constructs used in this work are list in Table 2.

Table 2. The DNA consructs applied in this work

Plasmid	Promoter/ Origin	Selection marker	Purpose	Reference
(A) *In Vitro*				
GroEL and GroES				
pCH-EL	T7/ColEI	Amp^R	protein purification	this work
pCH-SR	T7/ColEI	Amp^R	protein purification	this work
pET11a-ES	T7/ColEI	Amp^R	protein purification	this work
GroEL size-mutant				
pCH-ELΔC	T7/ColEI	Amp^R	protein purification	this work
pCH-EL2[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL3[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL4[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL[AAA]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL[GGA]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL2[GGA]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-EL[GGA]$_4$ [GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-SR	T7/ColEI	Amp^R	protein purification	this work
pCH-SRΔC	T7/ColEI	Amp^R	protein purification	this work
pCH-SR2[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-SR3[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-SR4[GGM]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-SR[AAA]$_4$	T7/ColEI	Amp^R	protein purification	this work
pCH-SR[GGA]$_4$	T7/ColEI	Amp^R	protein purification	this work
GroEL charge-mutant				
pCH-SRD253N	T7/ColEI	Amp^R	protein purification	this work
pCH-SRQNQ	T7/ColEI	Amp^R	protein purification	this work
pCH-SRD359N	T7/ColEI	Amp^R	protein purification	this work
pCH-SRNNQ	T7/ColEI	Amp^R	protein purification	this work
pCH-SR3N3Q	T7/ColEI	Amp^R	protein purification	this work
pCH-SRD253K	T7/ColEI	Amp^R	protein purification	this work
pCH-SRKKK(1)	T7/ColEI	Amp^R	protein purification	this work
pCH-SRD359K	T7/ColEI	Amp^R	protein purification	this work
pCH-SRKKK(2)	T7/ColEI	Amp^R	protein purification	this work
GroEL substrates				
pCH-WTMBP	T7/ColEI	Amp^R	protein purification	this work
pCH-SMMBP	T7/ColEI	Amp^R	protein purification	this work
pCH-DMMBP	T7/ColEI	Amp^R	protein purification	this work
pCH-WTMBP D95C	T7/ColEI	Amp^R	protein purification	this work
pCH-SMMBP D95C	T7/ColEI	Amp^R	protein purification	this work
pCH-DMMBP D95C	T7/ColEI	Amp^R	protein purification	this work

(B) *In Vivo*

GroEL and EL mutants

pOFXtacSL2	tac/ p15A	Cm^R, Spc^R	Co-expression	(Castanie *et al.*, 1997)
pOFXtacL2	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSLΔC	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL2[GGM]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL3[GGM]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL4[GGM]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL[AAA]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL[GGA]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL2[GGA]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL[GGA]$_4$ [GGM]$_4$	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSLNNQ	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSL3N3Q	tac/ p15A	Cm^R, Spc^R	Co-expression	this work
pOFXtacSLKKK(2)	tac/ p15A	Cm^R, Spc^R	Co-expression	this work

GroEL substrates

pBAD18-WTMBP	ara/ ColEI	Amp^R	Co-expression	this work
pBAD18-SMMBP	ara/ ColEI	Amp^R	Co-expression	this work
pBAD18-DMMBP	ara/ ColEI	Amp^R	Co-expression	this work
pBAD18-MetF	ara/ ColEI	Amp^R	Co-expression	this work

3.3. Molecular cloning methods

3.3.1. Preparation and transformation of *E. coli* competent cells

For preparation of chemically-competent *E. coli* cells, a single colony was used to inoculate 500 ml LB medium (including antibiotic, if applicable) and grown to an optical density (OD$_{600}$) of 0.25 - 0.5 at 37°C. The cells were then chilled on ice for 15 min and harvested at 5000 x *g* for 10 min at 4°C. The cell pellet was washed with 80 ml ice-cold Ca/glycerol buffer (10 mM PIPES, 60 mM CaCl$_2$, 15 % glycerol; pH 7.0, adjusted with NaOH, and filter-sterilized) once and incubated with additional 80 ml Ca/glycerol buffer on ice for 30 min. Finally, the cells were pelleted and resuspended in 6 ml of Ca/glycerol buffer. 100 µl aliquots were frozen in liquid nitrogen and stored at –80°C.

For transformation, ~50 µl competent cells were mixed with 0.05 - 0.2 µg plasmid DNA or 1-5 µl ligation reaction and incubated on ice for 15 min. The cells were heat-shocked at $42^{o}C$ for 45-90 s and subsequently placed on ice for 2 min. 1 ml of LB medium was added and the cells were shaken at $37^{o}C$ for 1 h. The cell suspension was then plated on selective plates and incubated at $37^{o}C$, until colonies had developed (typically 10-16 h).

Alternatively, electroporation was applied to improve the transformation efficiency. Electrocompetent cells were prepared as follows: 500 ml bacterial culture was grown to an optical density (OD_{600}) of 0.8 in LB medium at $37^{o}C$. The cells were washed carefully with 250 ml ice-cold sterilized water for two times and finally the cells were pelleted and resuspended in 2 ml of ice-cold 10% glycerol. 40 µl aliquots were frozen in liquid nitrogen and stored at $-80^{o}C$. For electroporation transformation, competent cells (40 µl) were mixed with 1-2 µl plasmid DNA (or ligation product) and transferred into a 0.2 cm Gene Pulser cuvette. The electroporation was done at 2.5 kV, 25 µFD and 200Ω settings with a Gene Puser II elecporation device. The transformed cells were allowed to recover in 1 ml of SOC medium with 225 rpm shaking at $37^{o}C$ for 1 h. The cell suspension was then plated on selective plates and incubated until colonies had developed (Dower *et al*., 1988).

3.3.2. Plasmid purification

LB medium containing the appropriate antibiotic was inoculated with a single *E. coli* colony harboring the DNA plasmid of interest and shaken 8 – 16 h at $37^{o}C$. Plasmids were isolated using the QIAprep Spin Miniprep Kit or QIAGEN Plasmid Midi Kit (Qiagen) according to the manufacturer's instructions.

3.3.3. PCR amplification

PCR (polymerase chain reaction)-mediated amplification of DNA was performed according to a standard protocol with minor modifications:

DNA Template:	10-20 ng (plasmid DNA)
	250 ng or less (bacterial genomic DNA)
Primers:	20 pmol each
dNTPs:	200 µM each
Pfu DNA Polymerase:	2.5 U
Polymerase buffer:	1 x
Additives:	3-6 % DMSO if GC content was >50 %,
Final volume:	50 µl

Cycling conditions (35 cycles):

Initial denaturation:	94°C, 5 min
Cycle denaturation:	94°C, 30-60 s
Annealing:	~55°C, 30-60 s
Extension:	72°C, duration dependent on template length:
	1 kbp/min.
Final Extension:	72°C, 10 min.

Stored at 4°C or -20°C.

PCR products were further purified using the QIAquick PCR purification and gel extraction kits (Qiagen) according to the manufacturer's instructions.

3.3.4. DNA restriction and ligation

DNA restriction was performed according to the manufacturer's instructions (New England Biolabs) of the respective enzymes. Typically, a 50 µl reaction contained 1-2 µl of each restriction enzyme and 0.5-2 µg purified PCR product or 1-5 µg plasmid DNA in the appropriate reaction buffer. Digested vector DNA was dephosphorylated with shrimp alkaline phosphatase.

For ligation, 100-200 ng (~1-2 µl) dephosphorylated vector DNA, 100-200 ng (~5-10 µl) DNA insert and 1 µl (100 U) T4 ligase were incubated in ligase buffer at 25°C for 1 h or, for increased efficiency, at 16°C overnight. The ligation product was transformed into competent *E. coli* DH5α cells as described.

3.3.5. DNA analytical methods

DNA concentrations were measured by UV absorption spectroscopy at $\lambda = 260$ nm. A solution of 50 µg/ml of double stranded DNA in H_2O exhibits approximately A_{260} nm = 1.

Agarose gel electrophoresis was performed in TAE buffer (40 mM Tris, 1 mM EDTA, 20 mM acetic acid) and 1 – 2 % TAE-agarose gels, supplemented with 1 µg/ml ethidium bromide, at 4 – 6 V/cm.

DNA sequencing was performed by Medigenomix GmbH (Martinsried, Germany) or Sequiserve (Vaterstetten, Germany).

3.4.   Protein purification

Proteins DnaK, DnaJ, GrpE (Hayer-Hartl *et al.*, 1996) and RuBisCo from *R. rubrum* were obtained from the Hartl laboratory collection of purified proteins.

3.4.1.   GroEL expression and purification

GroEL was purified with modifications to the protocol described by Hayer-Hartl *et al.* (1996). *E. coli* BL21 (DE3) Gold cells harboring the plasmid pCH-GroEL were grown in 6 l LB medium containing 100 mg/l ampicillin at 37°C to OD_{600} of ~0.5. The induction was then proceeded by the addition of 1 mM (final concentration) of IPTG to the culture for 5-6 h. After harvesting the cultures by centrifugation for 30 min at 2500 x g, cells were resuspended in 100 ml lysis buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA) and Complete protease inhibitor (1 tablet/25 ml). The suspension was further treated with lysozyme (~0.5 mg/ml) and benzonase (~500 units) for 60 min at 4°C. Lysis was achieved by homogenization of the cell suspension in an EmulsiFlex C5 device kept on ice. Cell debris was removed by ultracentrifugation for 60 min at 40,000 x g, 4°C and the supernatant subsequently passed through 0.2 μm filter. The supernatant was applied to a 400 ml DE52 column attached to an ÄKTA Explorer chromatography system. After washing with two column volumes of the lysis buffer, the protein was eluted using a NaCl gradient from 50 mM to 600 mM in five column volumes. The GroEL containing fractions were collected and dialyzed in 5 l lysis buffer overnight at 4°C. The deslated supernatant pool was applied onto a 20 ml MonoQ column, equilibrated in 30 mM Tris-HCl pH 7.5, 1 mM DTT, 1 mM EDTA and GroEL eluted with a NaCl gradient from 0 to 0.5 M. GroEL containing fractions were collected and dialyzed against 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM DTT and 1 mM EDTA. The sample was then applied to a 4 x 5 ml Heparin Sepharose column (HiTrap Heparin) and eluted with 30 mM Tris-HCl pH 7.5 with a NaCl gradient

from 0 to 0.5 M NaCl. GroEL-eluted flowthrough was collected and concentrated to less than 5 ml in 50 kDa cut-off Centriprep concentrators. Finally the concentrated sample was applied to a Sephacryl S 300 (XK 26/60) size exclusion column equilibrated in 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM DTT, 1 mM EDTA and 10 % glycerol. GroEL oligomer (approximate size 800 kDa) fractions were collected and concentrated to ~ 35 mg/ml (equivalent to 44 μM of GroEL oligomer). Protein concentration was determined based on extinction coefficient of GroEL ($\varepsilon 280=126,800$ M^{-1}cm^{-1}). And aliquots were flash frozen in liquid N$_2$ and stored at -80°C. Total yield of GroEL was typically ~ 600 mg.

3.4.2. GroES expression and purification

The expression and purification of GroES was similar as GroEL as described above, including the induction, lysis and centrifugation procedures. The supernatant was applied to a 400 ml DE52 column attached to an ÄKTA Explorer chromatography system. After washing with two column volumes of the above buffer, the protein was eluted using a NaCl gradient from 50 mM to 500 mM in five column volumes. GroES containing fractions were collected and dialyzed in 5 l lysis buffer overnight at 4°C. The deslated pool was applied into a 20 ml MonoQ column. Proteins were eluted in 30 mM Tris-HCl pH 7.5, 1 mM DTT, 1 mM EDTA and a NaCl gradient from 0 to 0.5 M. GroES containing fractions were collected and concentrated to less than 5 ml in 10 kDa cut-off Centriprep concentrators. The concentrated sample was finally applied to a Sephacryl S 200 (XK 26/60) size exclusion column equilibrated in 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM DTT, 1 mM EDTA and 10 % glycerol. GroES oligomer (approximate size 70 kDa) fractions were collected and concentrated to ~ 15 mg/ml based on extinction coefficient of GroES ($\varepsilon 280=8,943$ M^{-1}cm^{-1}). Aliquots were frozen in liquid N2 and stored at -80°C. Total yield of GroES was typically ~ 400 mg.

3.4.3. MBP and MBP mutants expression and purification

MBP and MBP mutants were purified using an amylose affinity column (New England Biolab). *E. coli* BL21 (DE3) Gold cells harboring the plasmid pCH-MBP wild-type and pCH-MBP mutants were grown in 6 l LB medium containing 100 mg/l ampicillin at 37°C to OD_{600} of ~0.1. The induction was then proceeded by adding 0.1 mM final concentration of IPTG to the culture for 12-16 h at 25°C. After harvesting the cultures by centrifugation for 30 min at 2500 x g, cells were resuspended in 100 ml amylose buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM DTT, and 1 mM EDTA) containing Complete protease inhibitor (1 tablet/25 ml). The same lysis and centrifugation conditions as used for GroEL purification were applied here. The supernatant was next dialyzed in amylose buffer to remove cellular maltose and slowly loaded on to a 100 ml amylose column. After washing with 12 column volumes of amylose buffer, MBP was eluted with amylose buffer containing 10 mM maltose. Fractions containing MBP were collected and dialyzed in 5 l amylose low salt buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM DTT, and 1mM EDTA) overnight at 4°C. MBP was concentrated to 10 mg/ml, and aliquots were frozen in liquid N2 and stored at -80°C. Typical yield of WT-MBP from 6 l culture was ~ 500 mg, while the SM-MBP, DM-MBP gave yields of ~200 mg and ~100 mg respectively, due to largely partition of the expressed protein in the insoluble fractions. Protein concentration was determined using extinction coefficient of $\varepsilon280=64,720$ $M^{-1}cm^{-1}$ of WT-MBP, $\varepsilon280=63,440$ $M^{-1}cm^{-1}$ of SM-MBP and DM-MBP.

3.4.4. MetF expression and purification

MetF was expressed at 30°C in *E. coli* cells harboring elevated level of GroEL and GroES (Kerner *et al.*, 2005). Cells were harvested and resuspended in running buffer (50 mM Tris-HCl pH 7.3, 300 mM NaCl) containing Complete EDTA-free protease inhibitor (1

tablet/ 25 ml). Lysis was achieved by homogenization of the cell suspension in an EmulsiFlex high pressure homogenizer device and the lysate was cleared by centrifugation at 40,000 x g for 1 h, 4°C. The supernatant was applied to a 5 ml HiTrap metal chelating column pre-charged with Ni^{2+}. The column was washed with a gradient of 10 to 50 mM imidazole in running buffer (for over 10 column volumes) and the proteins were eluted with 250 mM imidazole in running buffer. Fractions containing MetF were dialyzed in 20 mM MOPS-KOH, pH 7.4, 200 mM NaCl. MetF concentration was determined based on the absorption of bound FAD at 447 nm (ε_{280}=14,300 $M^{-1}cm^{-1}$) (Sheppard et $al.$, 1999). Proteins were aliquoted, flash-frozen in liquid N_2 and stored at -80 °C.

3.4.5. Rhodanese preparation

Bovine mitochondrial rhodanese was purchased from Sigma (R-1756). 11.5 mg rhodanese was dissolved in 600 μl rhodanese buffer (20 mM MOPS-NaOH, pH 7.4, 50 mM NaCl). Impurities were removed by ultracentrifugation for 20 min at 4°C and ~20,000 x g. The supernatant was transferred to a new vial and the protein concentration was determined using extinction coefficient (ε_{280}=59,840 $M^{-1}cm^{-1}$). Proteins were aliquoted and stored at -80°C.

3.5. Protein analytical methods

3.5.1. Determination of protein concentration

Protein concentrations were determined spectrophotometrically by A_{280} (in 6 M GuHCl), based on the theoretical extinction coefficient of the respective protein at λ=280 nm (Gill and von Hippel, 1989) as calculated by the ProtParam tool at the ExPASy proteomics server

(http://www.expasy.org). Molar concentrations of chaperones are expressed for the native state oligomer, while the GroEL substrates are presented as monomer.

3.5.2. SDS-PAGE

SDS-Polyacrylamide gels were prepared as follows:

4 % stacking gels and 10 %, 12 % or 15 % sepqrating gels were prepared according the following recipe:

Chemicals	Stacking gel	Separating gel		
	4 %	10 %	12 %	15 %
30 % Acryalmide (0.8% bis)	6.5 ml	16.7 ml	20 ml	25 ml
0.5 M Tris, pH 6.8	12.5 ml	_	_	_
1.5 M Tris, pH 8.8	_	12.5 ml	12.5 ml	12.5 ml
10 % SDS	0.5 ml	0.5 ml	0.5 ml	0.5 ml
2M Sucrose	_	12.5 ml	12.5 ml	12 ml
H_2O (up to 50 ml)	30.5 ml	7.8 ml	4.5 ml	_
TEMED	50 µl	25 µl	25 µl	25 µl
10% APS	500 µl	500 µl	500 µl	500 µl

SDS-PAGE was performed using a discontinuous buffer system (Laemmli, 1970) in BioRad Mini-Protean 3 electrophoresis chambers employing a constant current of 30 mA/gel in 50 mM Tris-Base, 380 mM glycine, 0.1 % SDS (pH 8.3). Protein samples were prepared for SDS-PAGE by mixing with 5x Laemmli buffer (Laemmli, 1970) (final concentration of 1x Laemmli buffer: 60 mM Tris-HCl, pH6.8, 1% SDS, 10 % glycerol, 0,01% Bromophenol blue, 0,1 mM β-mercaptoethanol) and boiling samples at 95°C for 3-5 min before loading onto a gel. After electrophoresis, gels were stained with Coomassie blue

staining solution (0.1 % Coomassie brilliant blue R-250, 40 % ethanol, 7 % acetic acid) for

1 h or longer and destained in 20 % ethanol, 7 % acetic acid.

3.5.3. Western-blotting

Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes in a

Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) in 25 mM Tris, 192 mM glycine,

20 % methanol, pH 8.4 at constant current of 150 mA/gel for 1 h (Towbin *et al*., 1979).

Nitrocellulose membranes were blocked in 5 % skim milk powder in TBST (50 mM Tris-

HCl, pH8.0, 150 mM NaCl, 0.05% Tween 20) for 1 h. The membranes were then incubated

with a 1:2000 – 1:10,000 dilution of primary antibody serum in TBST for 2 h and

extensively washed in TBST before incubation with a 1:5000 (for anti-mouse IgG) or

1:10,000 (for anti-rabbit IgG) dilution of secondary antibody in TBST (anti-rabbit IgG and

anti-mouse IgG, whole molecule – horseradish peroxidase conjugate, Sigma-Aldrich). After

extensive washing, protein bands were detected by incubating the membranes with ECL

chemiluminescence solution and exposure to X-ray film (High performance

chemiluminescence film) or a Luminescent Image Analyzer LAS-3000 system.

3.5.4. Sliver staining

Polyacryamide gels were fixed in fix solution (50 % methanol, 5% acetic acid) for at

least 1 h and briefly soaked into incubation solution (0.02% sodium thiosulfate) for 1 min.

After rinsing in water twice, gels were incubated in silver solution (0.2 % silver nitrate with

0.06% v/v formaldehyde) for 20 min and staining initiated by the developing solution (6%

sodium carbonate, 0.0002% sodium thiosulfate with 0.04% v/v formaldehyde) until the

desired signal intensity appeared. Further color development was stopped by incubation in

10% acetic acid for 10 min.

3.6. GroEL functional activity assays

3.6.1. ATPase assay

GroEL (0.2 µM oligomer) was incubated in buffer (20 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM $MgCl_2$) for 5 min at 37^oC, and GroES was present at a 2-fold molar excess over GroEL. The reaction was initiated at 37^oC by the addition of 2 mM ATP. The kinetics of the ATPase activities of GroEL and GroEL/GroES were followed for 0- 20 min with time points taken every 2.5 min. ATPase activity at the various time points was stopped by the addition of 15 mM CDTA. Quantification of liberated inorganic phosphate was performed by the malachite green assay (Lanzetta *et al.*, 1979). After 30 min incubation at 25^oC, the absorption was followed at 637 nm. A calibration curve of inorganic phosphate (0–20 nmol) was measured as linear range control in parallel. Spontaneous ATP hydrolysis was not observed under the conditions used.

3.6.2. Aggregation prevention assay of denatured rhodanese

Rhodanese (25 µM) was denatured for 1 h at 25^oC in denaturing buffer (6 M GuHCl, 20 mM Tris-HCl, pH 7.5, 20 mM KCl, 5 mM $MgCl_2$, 5 mM DTT) and diluted 100-fold into buffer A (20 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM $Mg(OAc)_2$) alone or into buffer A containing 0.25 µM GroEL (oligomers). Aggregation was monitored by light scattering at 320 nm (Weber *et al.*, 1998). Values indicating aggregation in buffer A after 10 min is set to 100%.

3.6.3. Surface plasmon resonance (SPR)

SPR measurements were performed on a BIAcore 2000 instrument (Brinker *et al.*, 2001; Hayer-Hartl *et al.*, 1995). All buffers used for SPR were filtered (0.2 µm) and degassed. All

SPR analyses in this work were in buffer A at a flow rate of 20 μl/min at 25°C. To immobilized GroES-His$_6$ on a chelating NTA biosensor chip, NTA was first loaded with Ni^{2+} by injection of 20 μl 0.5 mM NiCl$_2$, GroES was then immobilized on flow cell 1 and 3 (resulting in 50-100 RU), leaving flow cell 2 and 4 as buffer reference. 250 nM GroEL oligomer was injected in buffer (20 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM Mg(OAc)$_2$) plus the indicated nucleotides, followed by its dissociation kinetics for 10 min.

3.7. *In vitro* protein refolding and activity assays

Protein refolding reactions containing chaperones (when present) were carried out with the following molar concentration ratios of chaperones to substrate:

1 substrate (monomer) : 2 GroEL (tetradecamer) : 4 GroES (heptamer) .

3.7.1. MBP Refolding

WT-MBP and mutants (25 μM) were denatured in 6 M GuHCl, 20 mM Tris-HCl, pH 7.5, 20 mM KCl for 1 h at 25°C and refolded upon 100-fold dilution into high salt buffer A (20 mM Tris-HCl, pH 7.5, 200 mM KCl, 5 mM Mg(OAc)$_2$) or low salt buffer B (20 mM Tris-HCl, pH 7.5, 20 mM KCl, 5 mM Mg(OAc)$_2$) in the absence or presence of chaperones. GroEL/GroES assisted refolding was initiated at 25°C by the addition of 5 mM ATP. Intrinsic tryptophan fluorescence was monitored on a Fluorolog 3 Spectrofluorometer (Spex) with an excitation wavelength of 295 nm (slit width 2 nm) and an emission wavelength of 345 nm (slit width 5 nm)(Chun *et al*., 1993).

3.7.2. MetF Refolding

MetF (50 µM) was denatured with 6 M GuHCl in buffer A containing 10 mM DTT for 1 h at 25°C and refolded upon 100-fold dilution into buffer A containing 50 µM FAD and 1 g/L BSA at 25°C in the absence or presence of chaperones. A 2-fold molar excess of GroES over GroEL was added and refolding initiated with 5 mM ATP. At the times indicated, aliquots were taken and reactions stopped with 40 mM CDTA. MetF activity was measured at 25°C using an NADH-menadione oxidoreductase assay as described previously (Sheppard *et al.*, 1999).

3.7.3. Rhodanese Refolding

Rhodanese (50 µM) was denatured for 1 h at 25°C in denaturing buffer (6 M GuHCl, 20 mM Tris-HCl, pH 7.5, 20 mM KCl, 5 mM MgCl$_2$, 5 mM DTT) and refolded upon 100-fold dilution into buffer A or B supplemented with chaperones as indicated in the figure legends. A 2-fold molar excess of GroES over GroEL was added and refolding initiated with 5 mM ATP. At the times indicated, further refolding were stopped with 50 mM CDTA and rhodanese activity measured at 460 nm at 25°C, as previously described (Hayer-Hartl *et al.*, 1996; Martin *et al.*, 1991).

3.7.4. RuBisCo Refolding

RuBisCo (50 µM) was denatured in 6 M GuHCl, 50 mM Tris, pH 7.8, 1 mM EDTA, 10 mM DTT for 1 h at 25°C and diluted 100-fold into buffer A or B containing 1 mg/ml BSA, 5 mM DTT, supplemented with chaperones as indicated in the figure legends. Assisted refolding was initiated by adding 2-fold molar excess of GroES over GroEL and 5 mM ATP. At the times indicated, aliquots were withdrawn and rapidly mixed with solution

containing 7.5 mM CDTA and apyrase to stop GroEL action and RuBisCo refolding. Enzymatic activity was determined after incubation at 25°C for 1 h as described in Brinker *et al* (2001).

3.8. Biochemical and biophysical methods

3.8.1. Thermal denaturation of MBP

Temperature denaturation of MBP was monitored by circular dichroism spectroscopy (Jasco J-715) at 222 nm in buffer A. Thermal denaturation was carried out from 20- 90°C with a heating rate of 0.5°C *per* min and monitored at 222 nm. After incubation at 90°C for 30 min the sample was cooled to 20°C. The midpoint of thermal denaturation (Tm) for 2.5 μM MBP wild-type and mutants were determined by fitting the circular dichroism signal to a two state transition model.

3.8.2. Equilibrium unfolding of MBP

The equilibrium unfolding of MBP (0.25 μM) was performed at 25°C in buffer A containing 0−1.5 M GuHCl and followed by intrinsic tryptophan fluorescence at 345 nm. Equilibrium unfolding curves fit a two state transition.

3.8.3. Fluorescence assay of maltose binding to MBP

MBP D95C mutants (50 μM) (Marvin *et al*., 1997) were labeled in buffer C (100 mM HEPES, pH 7.8, 1 mM TCEP, 2 mM EDTA) for 4 h on ice in the presence of a 20-fold excess of the fluorophore IANBD (*N*-((2-(iodoacetoxy) ethyl)-*N*-methyl) amino-7-nitrobenz-2-oxa-1, 3-diazole ester, Molecular Probes, Inc.). Unbound fluorophore was

removed using micro Bio-Gel P6 columns (BIO-RAD) equilibrated in buffer B. The coupling efficiency measured by the absorption of MBP ($\varepsilon_{280} = 69,000$ M^{-1}cm^{-1}) and IANBD ($\varepsilon_{472} = 23,000$ M^{-1}cm^{-1}) was > 90%. IANBD fluorescence was monitored at 538 nm (slit width 8 nm) with an excitation wavelength at 470 nm (slit width 2 nm) at 25°C. During a typical MBP refolding experiment as described in section 3.7.1., 5 mM maltose were either present through the fluorescence measurement, or added at the end of refolding to confirm the native status of refolded MBP.

3.8.4. Fluorescence anisotropy

MBP D95C mutants (50 µM) were labeled in buffer C for 12 h on ice with a 2.5-fold excess of Alexa Fluor 488 C$_5$ maleimide (Molecular Probes). Unbound fluorophore was removed as above. The coupling efficiency measured by the absorption of MBP ($\varepsilon_{280} = 69,000$ M^{-1}cm^{-1}) and Alexa 488 C$_5$ maleimide ($\varepsilon_{493} = 72,000$ M^{-1}cm^{-1}) was > 90%. Steady state fluorescence anisotropy was monitored at the emission wavelength of 518 nm (slit width 7 nm) with an excitation wavelength at 495 nm (slit width 5 nm) at 25°C using a LS50 spectroflurometer (Perkin-Elmer).

3.8.5. Proteinase K protection of GroEL-GroES substrate complex

Rhodanese, DM-MBP or RuBisCo (25 µM each) was denatured in GuHCl as described and diluted 100-fold into buffer A or B in the presence of a 2- or 4-fold molar excess of GroEL or SR-EL, respectively, at 25°C. Treatment with proteinase K (2 µg/ml) was followed for 0-20 min . At the indicated times, sample was removed and further proteolysis was stopped with 1.5 mM PMSF. Protease protection of substrate protein by GroEL-GroES

complex was determined by immunoblotting and quantification by Aida Image Analyzer v. 3.52.

3.8.6. Intermolecular crosslinking of MBP

Crosslinking with DTSSP [3,3'-dithiobis(sulfo-succinimidylpropionate)] was performed according to the manufacturer's instruction (PIERCE). Native or GuHCl denatured MBP (25, 50, 100, 200 μM) was diluted 100-fold into 20 mM MOPS pH 7.5, 200 mM KCl and 5 mM $MgCl_2$ containing 0.5 mM DTSSP. After incubation at 25^oC for 60 min crosslinking was quenched by addition of 1M Tris, pH 7.4 to a final concentration of 50 mM with further incubation for 15 min. Samples were analyzed by SDS-PAGE under non-reducing conditions followed by sensitive silver staining. A small amount of crosslinked MBP was only observed with the denatured protein during refolding at a final MBP concentration greater than 1 μM.

3.9. *In vivo* assays

3.9.1. Solubility of MBP and MetF *in vivo*

E. coli MC4100 strain containing the plasmid pOFXtac-SL2 , expressing GroEL/GroES or EL-mutants/GroES, was transformed with the arabinose-controlled expression plasmid (pBAD 18) for MBP or MetF. Cells were grown in LB medium at 37^oC to an $OD_{600}=0.8$, and chaperonins were induced with 0.1 mM IPTG for 1 hour before induction of substrate protein with 0.2% arabinose for 1 hour. Cells were harvested by centrifugation (5000 x *g*, 5 min) and spheroplasts were prepared as following (Ausubel *et al*., 2003): 3 OD_{600} units cell pellet was resuspended in 100 μl of 50 mM Tris-Cl, pH 8.0, 20% sucrose, followed by adding 20 μl of 5 mg/ml lysozyme solution and incubating the cells on ice for 5 min. 40 ul

of 20 mM of EDTA were further added and the cells were incubated on ice for 5 min. Finally, the cell suspension was mixed with 40 μl of 50 mM Tris-Cl, pH 8.0 and the incubation temperature was shifted to 37°C for 5 min. Spheroplasts were lysed by dilution into an equal volume (200 μl) of 2x lysis buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM MgSO$_4$, 0.2% Triton X-100, Complete EDTA-free protease inhibitors, 100 U/ml Benzonase). Aliquots were fractionated into supernatant and pellet by centrifugation at 20,000 x g for 30 min and analyzed by 12% SDS-PAGE and Coomassie blue staining or immunoblotting, as indicated.

3.9.2. Complementation assay of GroEL/GroES depletion strain

A GroEL depletion strain, MC4100 SC3 where the chromosomal groE promoter is replaced with the araC gene and the pBAD promoter, and a kanamycin resistance (KanR) cassette is immediately upstream of groE (transcribed in the reverse orientation) was a generous gift from C. Gerogopoulos (University of Geneva, Switzerland).

E. coli MC4100 SC3 cells harboring plasmids pOFXtac-SL2 series, expressing GroEL/GroES or EL-mutants/GroES were grown in LB- 0.05 % arabinose medium at 37°C to OD$_{600}$ = 1 (~7 h), where cell numbers were predetermined to be similar among all strains tested (~5 x 10^8 cells/ml) based on colony counting results. After confirming the OD$_{600}$ values of the cultures, serial dilutions were made to give 5 x 10^4, 5 x 10^3, 5 x 10^2, 5 x 10^1 cells/ 5 μl suspension, and spotted them on either LB-Spe50 plate containing 0.05 % arabinose or 1 mM IPTG. Plates were scanned after 16 h incubation.

4. Results

4.1. GroEL/GroES can accelerate MBP folding more than ten-fold

4.1.1. MBP as a suitable model substrate to study the rate of chaperonin assisted folding

To explore how encapsulation actively accelerates the folding reaction, a suitable substrate is required for further study. Since the obligate GroEL dependent substrate proteins were shown to aggregate upon *in vitro* refolding (Kerner *et al*., 2005), it is difficult to compare their spontaneous and chaperonin-assisted folding rates. To avoid this complication, we explored the suitability of MBP as a model substrate, based on previous reports that GroEL/GroES can increase the folding speed of a mutant form of MBP (Beissinger *et al*., 1999; Sparrer *et al*., 1997).

MBP is a periplasmic *E. coli* protein that is involved in sugar transport as a primary acceptor for maltose and maltodextrin (Spurlino *et al*., 1991). It is synthesized in the cytosol as a precursor containing a 26-residue signal peptide at the N-terminus; however it folds robustly in the cytosol when expressed without its cleavable N-terminal export sequence. MBP is composed of two globular domains formed by discontinuous sequence elements consisting of secondary structural βαβ units, with the binding site for maltose located in a cleft between the domains (Figure 12) (Spurlino *et al*., 1991).

MBP has several properties that make it desirable as a folding substrate: firstly, MBP is a monomeric protein (~41 kDa) which does not require an assembly step upon release from the chaperone. Secondly, *in vitro* refolding of MBP has been extensively characterized (Chun *et al*., 1993) and it folds spontaneously up to a concentration of 1 to 2 μM (Ganesh *et al*., 2001). Thirdly, MBP possesses eight tryptophans (10, 62, 94, 129, 158, 230, 232 and

340) distributed over both domains (Figure 12); thereby the formation of its native state results in a dramatic enhancement in intrinsic tryptophan fluorescence, the tryptophan fluorescence signal is reduced 5-fold upon unfolding, and the recovery of fluorescence can be used as a measure of folding (Chun *et al*., 1993).. Importantly, it is possible to continuously monitor the formation of native MBP protein both in the presence and absence of GroEL/GroES, which lack tryptophans in resulting significant fluorescence background.
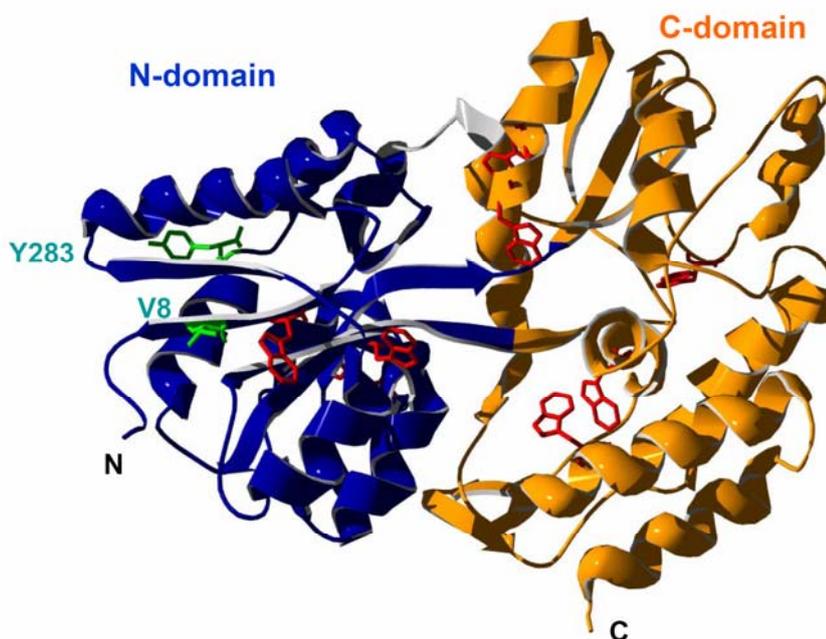
Figure 12. Ribbon diagram of the tertiary structure of MBP.

The two discontinuous domains are shown in blue and orange (N-domain: 1-109 and 264-309, C-domain: 114-258 and 316-370), respectively, the eight tryptophans in red. Positions of mutated amino acids (V8G and Y283) are indicated in green. The structure was adapted from Spurlino *et al*., (1991) (PDB code: 1OMP) and figure was generated with SwissPDB-Viewer v3.7 and POV-Ray v3.6.

The spontaneous rate of folding of wild-type MBP (WT-MBP) is very fast; therefore WT-MBP is not a suitable substrate for following folding rate enhanced by GroEL. However, several slow folding mutants of MBP are known and we analyzed two of these,

the single mutant Y283D (SM-MBP) and the double mutant V8G/Y283D (DM-MBP) (Chun *et al*., 1993; Wang *et al*., 1998). Both mutations, V8G and Y283D, are located in the N-domain, in close proximity in a strand and loop segment, respectively (Figure 12). Formation of native contacts within the N-domain is rate-limiting for MBP folding and is therefore slowed by these mutations (Chun *et al*., 1993).

As monitored by tryptophan fluorescence, at 25°C, WT-MBP rapidly refolded with an apparent rate of ~0.03 s^{-1} (t$_{1/2}$ ~25 s), upon dilution from 6 M guanidine-HCl (GuHCl) into the refolding buffer. In addition, SM-MBP and DM-MBP refolded to full yield but with ~7-fold (t$_{1/2}$ ~175 s) and ~75-fold (t$_{1/2}$ ~1900 s) slower rates, respectively (Figure 13 and Supplemental Table S1).
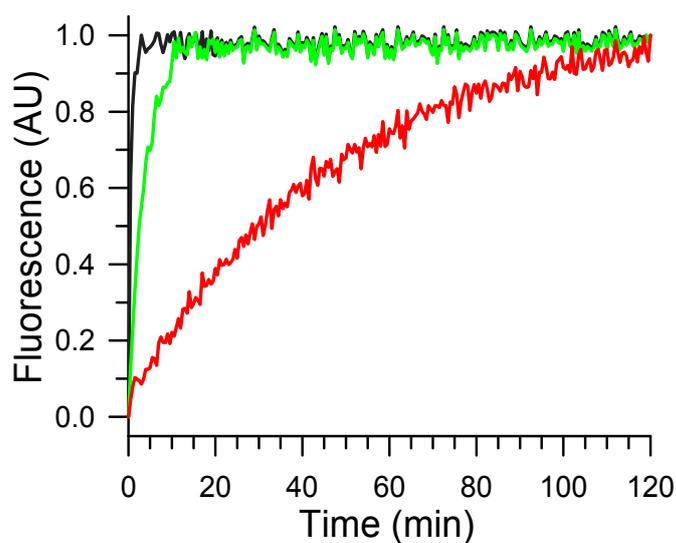
Figure 13. Spontaneous refolding of wild-type and mutant MBP

Refolding of GuHCl-denatured MBP (25 μM) at 25°C upon 100-fold dilution into buffer A. Wild-type MBP was indicated in black, SM-MBP in green and DM-MBP in red. The maximum recovery of tryptophan fluorescence was set to 1 (~100% of native MBP control).

Importantly, the slower folding rates of MBP mutants are not due to lower stabilities of their final states. We applied temperature and denaturant-dependent unfolding experiments and demonstrated that the native states of SM-MBP and DM-MBP were only moderately destabilized relative to WT-MBP (Figure 14 and Supplemental Table S1A). This indicates that these three proteins are equivalently stable in the further experimental conditions.
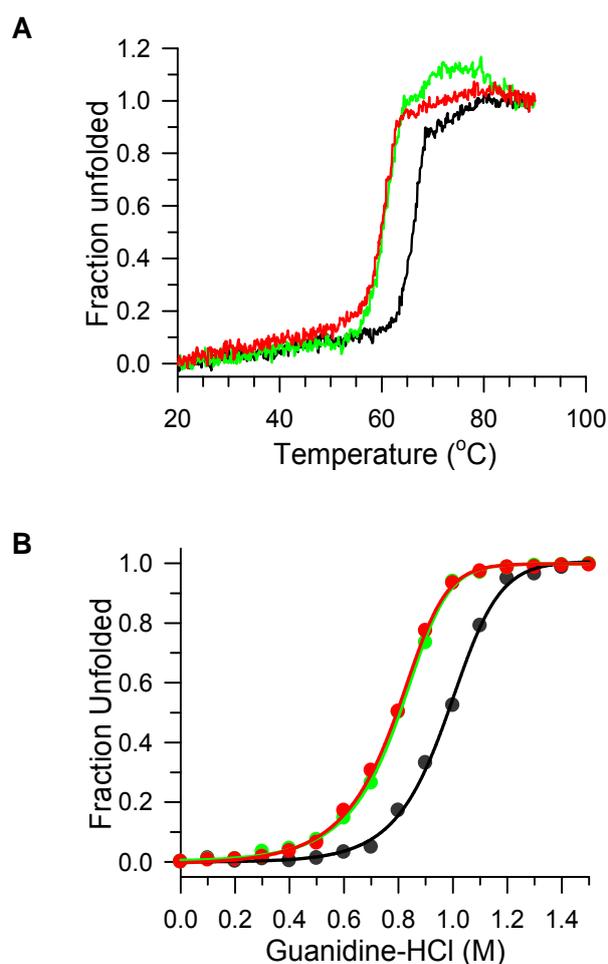
Figure 14. Folding properties of WT-MBP and MBP mutants

(A) Thermal denaturation of MBP monitored by circular dichroism (Jasco J-715) at 222 nm in buffer A. WT-MBP is indicated in black, SM-MBP in green and DM-MBP in red. (B) The equilibrium unfolding of WT-MBP and MBP mutants (0.25 μM) at 25°C in buffer A containing 0-1.5 M GuHCl, followed by intrinsic tryptophan fluorescence at 345 nm. Both the thermal denaturation and equilibrium unfolding curves fit a two state transition. Thermal denaturation was fully reversible.

4.1.2. Folding acceleration of MBP is GroEL dependent in a noncycling manner

To test whether GroEL can bind denatured MBP, we first incubated the denatured substrates in absence or presence of nucleotide together with only GroEL, we found that in the absence of ATP, GroEL inhibited the spontaneous folding of WT-MBP and two mutant proteins, indicating GroEL can efficiently recognize all three unfolded MBP. Proteins in the presence of ATP, slow refolding was only observed with SM-MBP and DM-MBP, whereas WT-MBP refolded with kinetics similar to spontaneous folding, suggesting that the SM-MBP and DM-MBP bury hydrophobic residues more slowly, and allowing efficient GroEL rebinding. Strikingly, substantial acceleration of SM-MBP and DM-MBP folding was observed upon addition of the complete GroEL/GroES system and ATP, in which the folding rate of SM-MBP was accelerated ~3-fold compared to spontaneous folding and that of DM-MBP was accelerated ~13-fold (Figure 15A-C and Supplemental Table S1). Spontaneous folding of WT-MBP was rapid and efficient and no further enhancement by the GroEL/GroES system was observed.

In contrast to GroEL/GroES, the bacterial Hsp70 chaperone system, consisting of DnaK, DnaJ, GrpE and ATP, can bind unfolded MBP but strongly retarded the folding of SM-MBP and DM-MBP. On the other hand, the Hsp70 system maintained both SM-MBP and DM-MBP competent for further folding by GroEL/GroES (Figure 15B and C). This suggests that the acceleration in folding of the mutant MBPs can only occur with the GroEL/GroES system. These properties are similar to those recently described for several highly aggregation sensitive, authentic GroEL substrates (Kerner *et al*., 2005).

To demonstrate that encapsulation of mutant MBP in the GroEL-GroES cage is sufficient for accelerated folding, refolding experiments were carried out with the non-cycling single-ring mutant of GroEL (SR-EL). SR-EL contains four point mutations in
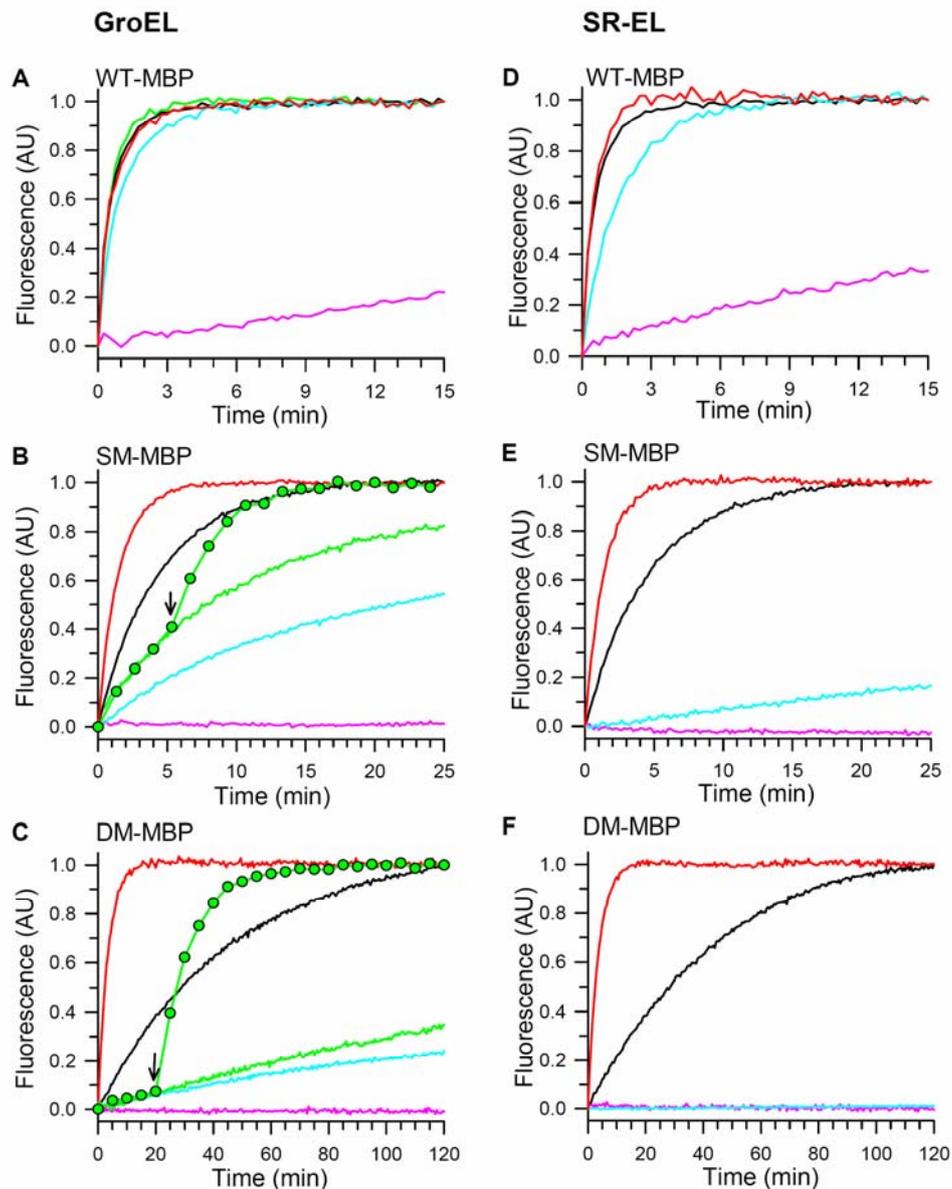
Figure 15. Effects of GroEL/GroES on wild-type and mutant MBP refolding.

(A-F) Refolding of GuHCl-denatured MBP (25 μM) at 25°C upon 100-fold dilution into reactions containing either buffer A alone (spontaneous; black), buffer with 0.5 μM GroEL or 1.0 μM SR-EL (pink); 0.5 μM GroEL/5 mM ATP or 1.0 μM SR-EL/5 mM ATP (blue); 0.5 μM GroEL/1 μM GroES/5 mM ATP or 1.0 μM GroEL/2 μM GroES/5 mM ATP (red); 1.25 μM DnaK/0.625 μM DnaJ/1.25 μM GrpE /5 mM ATP (green); or 1.25 μM DnaK/ 0.625 μM DnaJ/1.25 μM GrpE /5 mM ATP followed by addition of 0.5 μM GroEL/1 μM GroES/5 mM ATP (green circles) at the time indicated by the arrow. The maximum recovery of tryptophan fluorescence in the presence of GroEL/GroES/ATP was set to 1 (~100% of native MBP control).

amino acid residues Arg452, Glu461, Ser463 and Val464 that are responsible for the major contacts between two rings of GroEL (Weissman *et al.*, 1995). Residue 452 is replaced by Glu and the other three residues by Ala. These mutations cause the loss of electrostatic interaction between two rings and result in the formation of single heptameric ring (Weissman *et al.*, 1996). SR-EL has previously been shown to fold rhodanese upon binding of GroES in the presence of ATP; however, the folded rhodanese was not released from the ring cavity since GroES remained associated with SR-EL (Hayer-Hartl *et al.*, 1996; Weissman *et al.*, 1996).

In the presence of GroES and ATP, SR-EL fully reproduced the rate acceleration of SM-MBP and DM-MBP folding observed with the cycling GroEL/GroES system, while the rate of WT-MBP folding remained unchanged (Figure 15D-F and Supplemental Table S1A). Thus, the physical environment of the GroEL-GroES cage is likely responsible for the observed increase in folding speed.

To be certain that the folding rates of MBP measured by intrinsic tryptophan fluorescence indeed reflected acquisition of the native state competent in binding maltose; we introduced a unique cysteine at position D95 in the N-domain in MBP and MBP mutants and labeled them with the fluorophore IANBD. The D95C residue is buried in the absence of maltose. Upon maltose binding, MBP undergoes a large conformational change thereby D95C residue is exposed and results in a ~2-3 fold fluorescence increase at 538 nm (Marvin *et al.*, 1997). Introduction of the D95C mutation slowed the spontaneous refolding of wild-type and mutant MBP, but very similar folding rates were determined by monitoring intrinsic tryptophan or IANBD fluorescence (Supplemental Table S1). As shown in Figure 16, the folding of MBP followed by D95C-IANBD fluorescence at 538 nm reflects maltose binding and hence the native state.
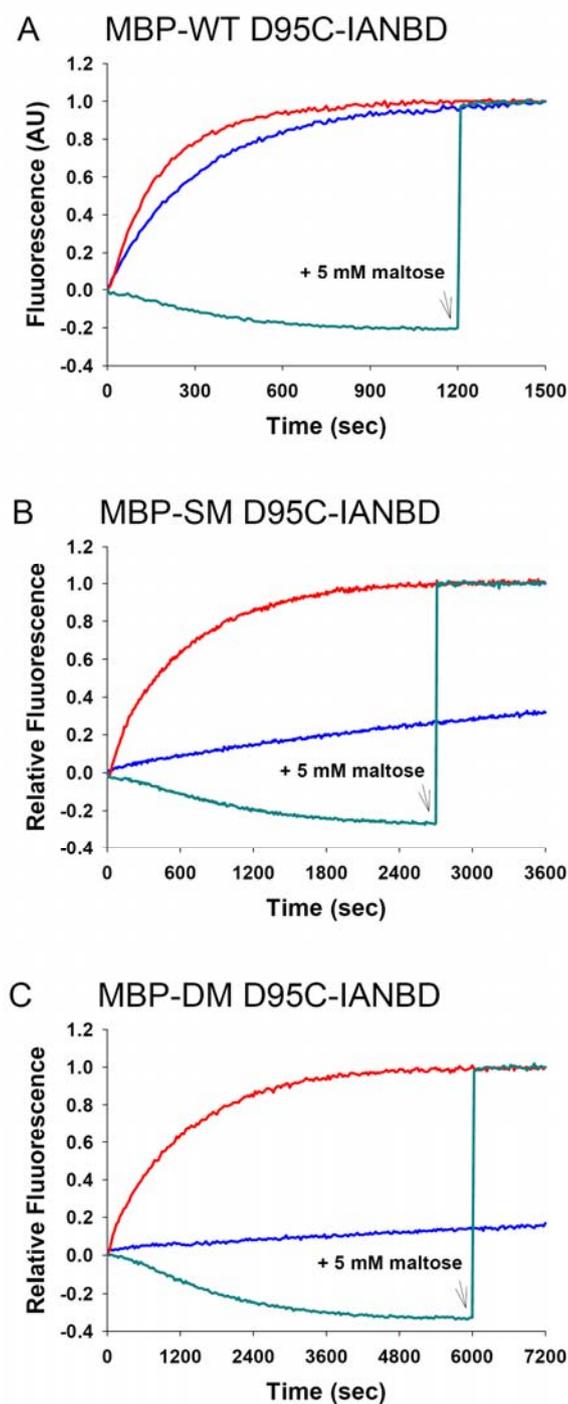
Figure 16. Maltose binding of wild-type and mutant MBP.

Maltose binding of GuHCl-denatured MBP (25 μM) at 25°C upon 100-fold dilution into buffer A containing 5 mM maltose alone (blue), buffer with 0.5 μM GroEL/1 μM GroES/5 mM ATP/ 5 mM maltose (red) or 0.5 μM GroEL/1 μM GroES/5 mM ATP followed by addition of 5 mM maltose (green) at the time indicated by the arrow. The maximum recovery of IANBD fluorescence in the presence of GroEL/GroES/ATP was set to 1 (~100% of native MBP control).

One other plausible explanation for the accelerated folding rate observed with the GroEL and GroES system may be due to the disassembly of small MBP aggregates by GroEL in the refolding reactions. We performed refolding experiments with increased concentration of the three MBPs. The folding rates and yields were essentially independent of concentration between 50 nM and 1μM for WT-MBP (Ganesh *et al*., 2001) and the two mutant MBPs (Figure 17), arguing against reversible aggregation as the cause of slow spontaneous folding of mutant MBP.
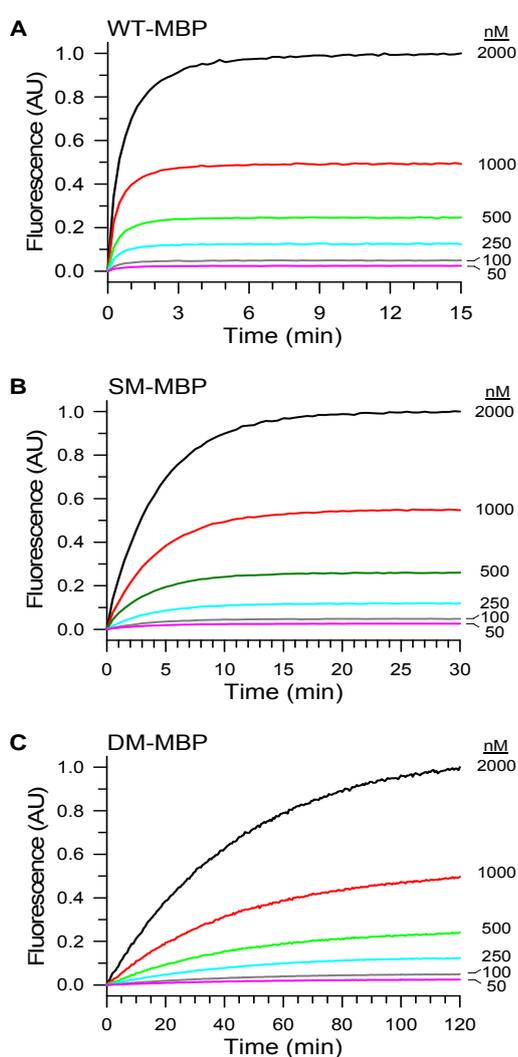
Figure 17. Folding properties of WT-MBP and MBP mutants

The concentration titration (50 nM- 2μM) of WT-MBP, SM-MBP and DM-MBP in buffer A at 25°C followed by the intrinsic tryptophan fluorescence at 345 nm.

Similarly, a chemical crosslinking result also excluded possibility of aggregates formed during MBP refolding. Crosslinking reaction by applying DTSSP (3, 3′-dithiobis [sulfosuccinimidyl-propionate]) revealed association of MBP monomers during refolding can only occurred at protein concentrations above 2 μM (Figure 18), while a control reaction carried out in an aggregation-favored environment with 5% Ficoll as crowding agent demonstrated the crosslinking procedure is sufficiently sensitive to detect even small amount of dimers of MBP. Finally, aggregation of MBP is not observed by ultracentrifugation, confirmed again the absence of MBP aggregates in the relevant concentration range applied in this work.
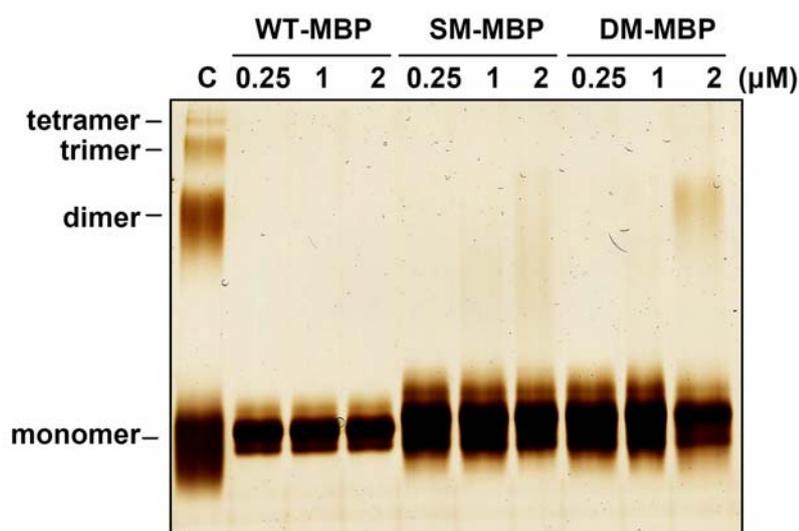
Figure 18. Folding properties of WT-MBP and MBP mutants.

GuHCl-denatured WT-MBP was diluted 100-fold into MOPS buffer containing 0.5 mM DTSSP. Final concentrations of MBP varied between 0.25- 2 μM. After incubation at 25°C for 60 min crosslinking was quenched and samples were analyzed by SDS-PAGE under non-reducing conditions followed by silver staining. Control experiment (C) was performed with crosslinking of 0.25 μM denatured DM-MBP in 5% Ficoll buffer.

4.2.   Effects of GroEL cavity size on folding

As mentioned above, the rate acceleration is GroEL dependent and occurred inside the GroEL/GroES cavity. To further understand what the driving forces for accelerated folding are, we investigated two major physical properties of the *cis* chaperonin cage. Firstly, the limited volume of cavity and secondly, the hydrophilic properties of the cavity wall.

4.2.1.   Properties of GroEL cavity size

Molecular simulations of the enhancement of the rate of folding of eight small proteins inside a cage of various sizes indicated that the ratio of size of protein to the size of the cage is important (Takagi *et al*., 2003). For a 50 kDa polypeptide chain inside the GroEL-GroES cage, the predicted rate enhancement is about six-fold, falling to about two-fold for 30 kDa protein. These theoretical estimates are in rough agreement with the experimentally observed approximate four-fold enhancement for 50 kDa RuBisCo and no enhancement for 33 kDa Rhodanese folding within the GroEL-GroES cage (Brinker *et al*., 2001). So, if considering that RuBisCo (50 kDa) closely approaches the size limit of the cage (~60 kDa), the effect of topological confinement may happen in the GroEL-GroES cage and contribute to accelerate folding by increasing the compaction of polypeptide and blocking the formation of certain kinetically trapped conformers.

To explore this possibility, we engineered a series of GroEL mutants with varying cavity size. The GroEL subunits contain flexible, C-terminal sequences of 13 residues, consisting of 4 Gly-Gly-Met (GGM) repeats and ending with an additional Met residue (Figure 19). These [GGM]$_4$M sequences protrude from the equatorial domains into the GroEL cavity but are not resolved in the crystal structure (Braig *et al*., 1994). Deletion or extension of these segments afforded the possibility to vary the size of the GroEL-GroES cage (Figure 19). Taking the 7-fold symmetry of the structure into account, we estimated

that deletion of [GGM]$_4$M, resulting in ELΔC, would increase the volume capacity of GroEL for folding intermediates by ~4.4%. In contrast, duplication of the C-terminal segment (EL-2[GGM]$_4$) will reduce the volume by ~4.4% compared to WT-GroEL, and the mutants EL-3[GGM]$_4$ and EL-4[GGM]$_4$ are expected to have ~91% and 87% of WT-GroEL volume, respectively (Figure 19). Additionally, by mutating [GGM]$_4$ to [AAA]$_4$, [GGA]$_4$ or 2[GGA]$_4$, we changed the size of the cavity in small increments in a manner independent of the specific GGM sequence.

	cis-cavity Volume: Å^3 x10^3	% change
ELΔC	168.2	+4.4
WT-GroEL	161.1	0.0
EL-2[GGM]$_4$	154.1	−4.4
EL-3[GGM]$_4$	147.1	−8.7
EL-4[GGM]$_4$	140.0	−13.1
EL-[AAA]$_4$	162.1	+0.6
EL-[GGA]$_4$	163.1	+1.2
EL-2[GGA]$_4$	158.1	−1.9
EL-[GGA]$_4$[GGM]$_4$	156.1	−3.1
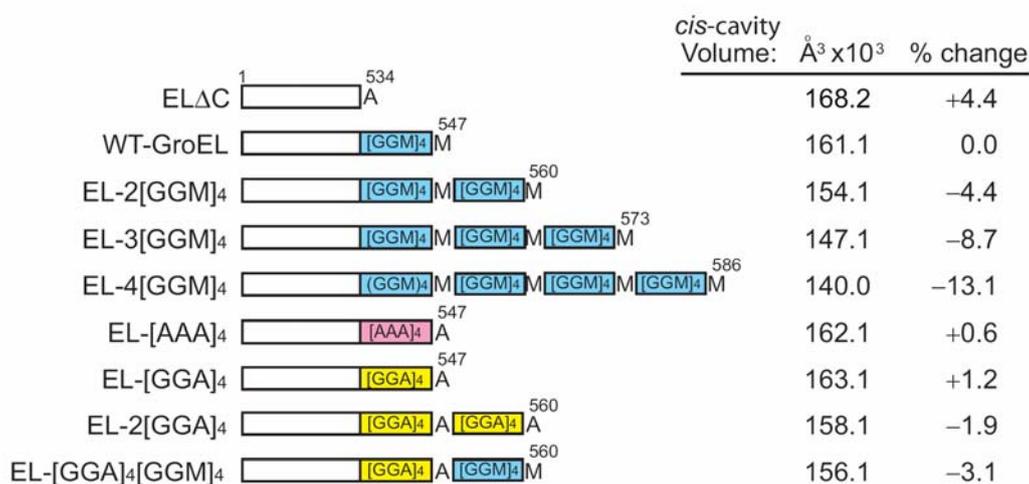
Figure 19. Properties of GroEL cavity size-mutants.

Schematic representation of a series of GroEL constructs with deletions, mutations or extensions in the C-terminal [GGM]$_4$ repeat sequences. The *cis*-cavity volume of wild-type chaperonin was calculated as 161,100 Å^3 from the structure of the GroEL-GroES complex , taking into account that the N-terminal Met and C-terminal 23 amino acids of GroEL (~14,000 Å^3) were not resolved in the crystal structure. Volume changes resulting from modification of C-terminal segments were estimated based on the known volume of specific amino acid residues.

These cavity size mutant chaperonins were generated both for GroEL and SR-EL. Upon overexpression and purification of these chaperonin mutants in the bacterial strain BL21 (DE3), we estimate the incorporation of endogenous GroEL is less than 5%, meaning there is maximally a single endogenous GroEL subunit assembled in the mutant chaperonin complex. Biochemical assays revealed that these purified chaperonins bound unfolded protein with similar affinity as WT-GroEL, as evidenced by their ability to inhibit the aggregation of rhodanese in the absence of ATP (Figure 20). Surface plasmon resonance experiments demonstrated efficient ATP-dependent GroES cycling and stable GroES binding in the presence of the non-hydrolysable ATP analog AMP-PNP (or ATP in case of non-cycling SR-EL), which indicates that these cavity size mutants are all functional chaperonins with regard to their to their substrate binding and ATP/GroES cycling abilities .

To determine the functional volume capacity of the GroEL mutants, here we applied an encapsulation assay which by measuring the degree of protease protection conferred to the GroEL-bound substrate protein under GroES in the presence of AMP-PNP. As shown for rhodanese (33 kDa), DM-MBP (41 kDa) and bacterial RuBisCo (50 kDa), GroEL-bound protein was rapidly degraded in the absence of GroES (Figure 21A). Addition of GroES to WT-GroEL resulted in ~50% protection of substrate, as expected due to the asymmetrical binding of GroES (Figure 21B) (Hayer-Hartl *et al*., 1996). While a similar degree of protection was observed with ELΔC, EL-[AAA]$_4$ and EL-[GGA]$_4$. However, the stepwise extension of the [GGM]$_4$M segment resulted in a reduced capacity of protein encapsulation. This effect was most pronounced with the largest protein RuBisCo. For example, while EL-4[GGM]$_4$, having a ~13% reduced cavity volume, allowed efficient encapsulation of rhodanese, encapsulation of DM-MBP and RuBisCo was reduced by 40% and 90%, respectively (Figure 21A and B). Similar results were obtained with the cavity size-mutants of SR-EL (Figure 22).
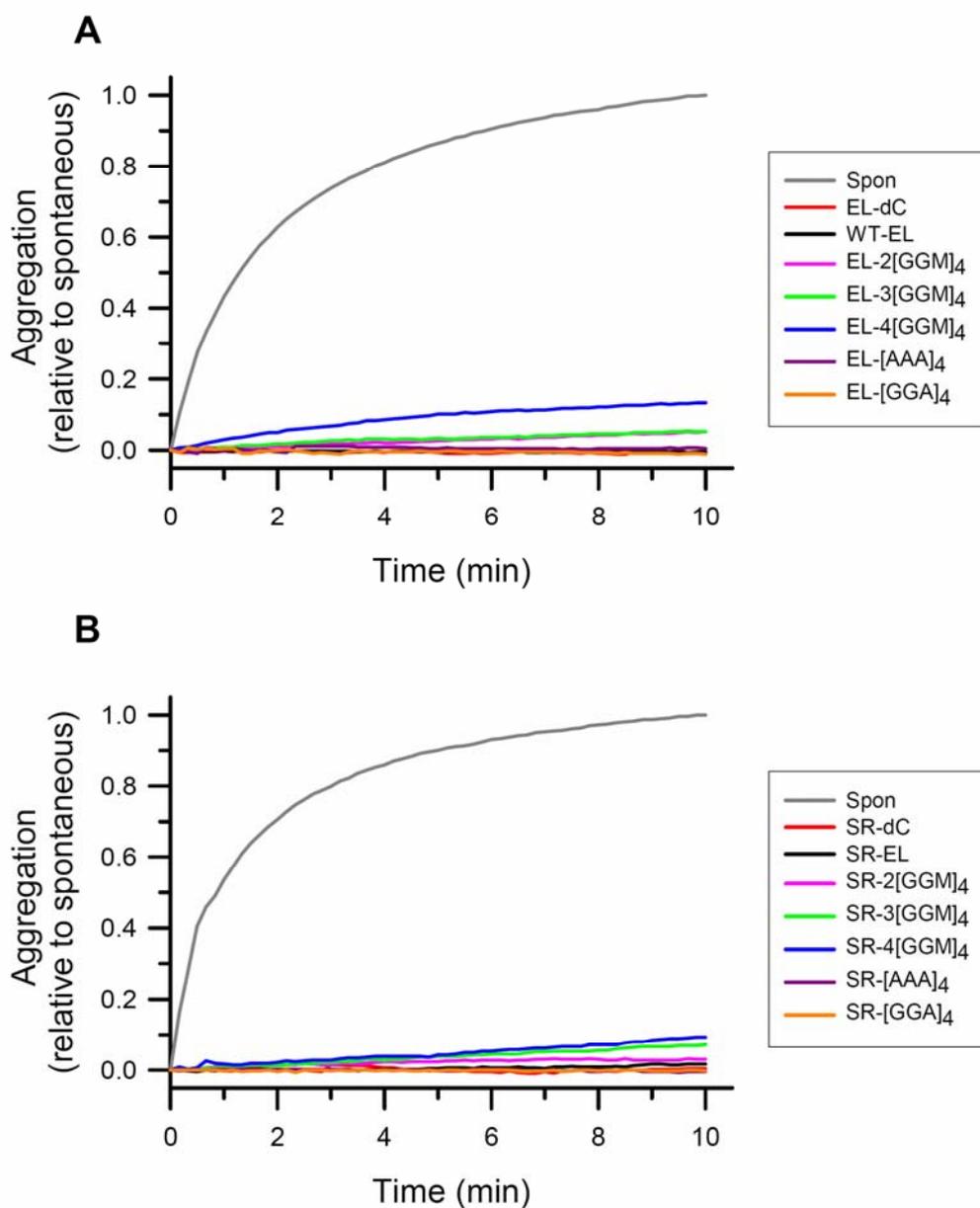
Figure 20. Inhibition of rhodanese aggregation by GroEL and SR cavity size mutants.

Rhodanese (25 µM) was denatured in 6M GuHCl and 100-fold dilution into buffer without (Spon) or with presence of chaperones as indicated for either GroEL (0.5 µM, A) or SR-EL (1µM, B). Aggregation was monitored by spectrophotometer at 320nM and the absorbance value of spontaneous experiment after 10min was set to 1.
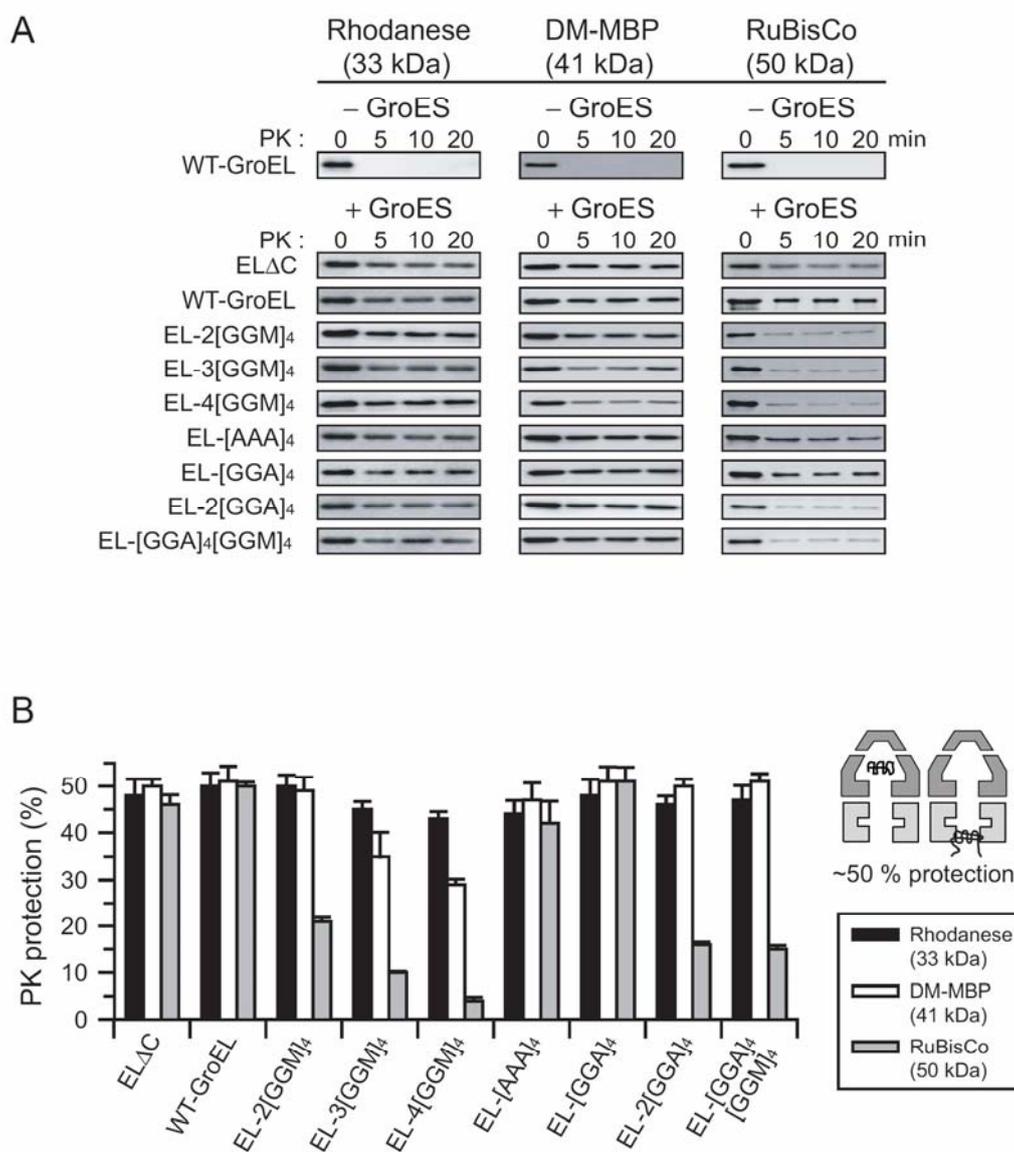
Figure 21. Proteinase K protection of GroEL- GroES- substrate complex.

(A) Proteinase K (PK) protection of rhodanese (250 nM), DM-MBP (250 nM) or RuBisCo (250 nM) bound to WT-GroEL or GroEL cavity size mutants (0.5 μM). GroEL-substrate complexes were incubated with PK (2 μg/ml) in buffer A containing 4 mM AMP-PNP in the absence or presence of GroES (1 μM) at 25°C. (B) Protected substrate protein was quantified by immunoblotting and densitometry. Amounts in non-protease treated reactions correspond to 100%.
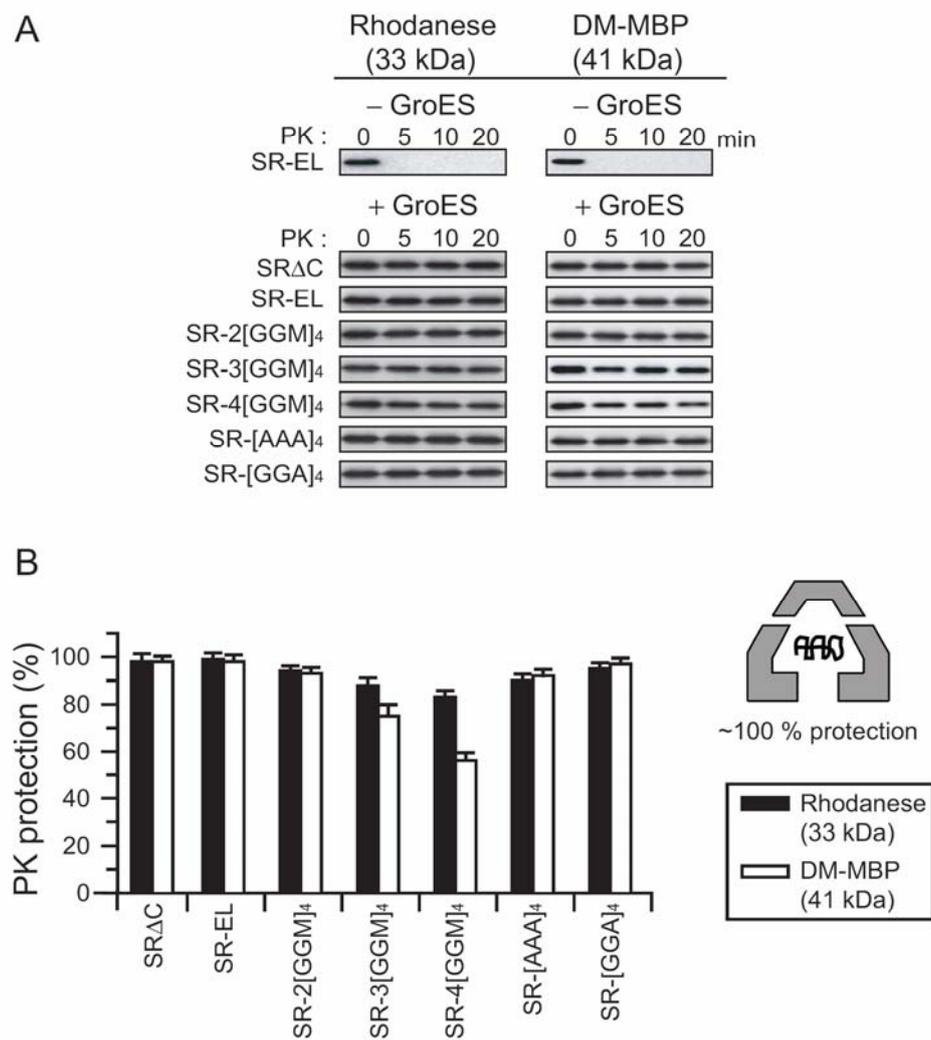
Figure 22. Proteinase K protection of SREL- GroES- substrate complex.

(A) Proteinase K (PK) protection of rhodanese (250 nM) or DM-MBP (250 nM) bound to SREL or SR cavity size mutants (1 μM). SREL-substrate complexes were incubated with PK (2 μg/ml) in buffer B containing 4 mM AMP-PNP in the absence or presence of GroES (2 μM) at 25°C. (B) Protected substrate protein was quantified by immunoblotting and densitometry. Amounts in non-protease treated reactions correspond to 100%.

4.2.2. Effects of GroEL cavity size on folding

We next investigated how changing the size of the GroEL cavity affected the folding rates of proteins differing in molar mass, including mutant MBP (41 kDa) and the GroEL-dependent substrates rhodanese (33 kDa), MetF (33 kDa) and RuBisCo (50 kDa). Importantly, these experiments were performed with 500 nM substrates refolding at 25°C, which is non-permissive, namely, no spontaneous refolding detectable for rhodanese, MetF and RuBisCo due to aggregation, except for MBP.

Dependent on protein size, optimal folding rates were either observed with normal sized cavity or upon reduction of cage volume in both double ring and single ring chaperonin (Figure 23 and Figure 24). Deletion of [GGM]$_4$M (ELΔC), expanding the GroEL *cis* cavity by ~4.4%, generally reduced folding speed without changing the folding yields (Figure 23 and Figure 25). Strikingly, reducing wild-type cavity size by ~1.9, 3.1 and 4.4% in constructs EL-2[GGA]$_4$, EL-[GGA]$_4$ [GGM]$_4$ and EL-2[GGM]$_4$, respectively, resulted in a highly reproducible, stepwise enhancement of folding rate for rhodanese and MetF (Figure 23A and B). This effect correlated well with the decrease in available cage volume of the GroEL mutants. Notably, it was independent of the hydrophobic Met residues in the C-terminal extensions, and was therefore attributed to spatial confinement rather than to specific interactions with the extended GroEL sequences. Further reduction of cavity size (EL-3[GGM]$_4$) reversed the rate acceleration without affecting the folding yield. Finally, very slow folding rate was observed with EL-4[GGM]$_4$ (Figure 23A and B), accompanied by a 40-70% reduction in folding yield (Figure 25). Because encapsulation of rhodanese by GroES was still fully efficient (Figure 21), this indicates that the restriction in space may have limited critical rearrangement steps during folding.
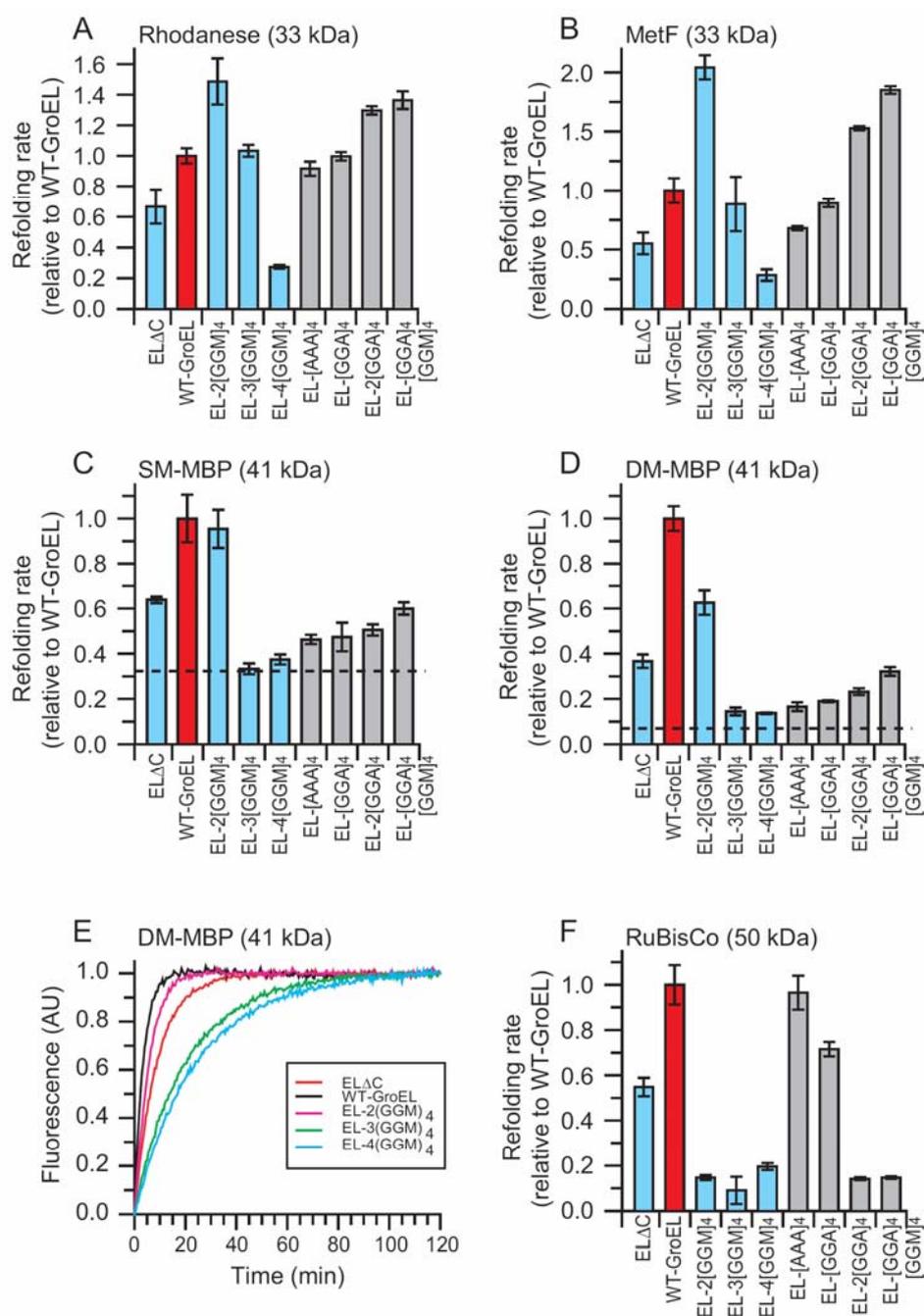
Figure 23. Effect of GroEL cavity size on folding rates.

GroEL/GroES assisted refolding of rhodanese (A), MetF (B), SM-MBP (C), DM-MBP (D-E) and RuBisCo (F) at 25°C with the GroEL mutants indicated. Blue bars, cavity size mutants (decreasing cavity size from left to right); light gray bars, mutants with reduced hydrophobic character of the C-terminal repeat sequences. The refolding rate obtained with WT-GroEL (red bar) was set to 1. The dashed line represents the rate of spontaneous folding for MBP. Representative tryptophan fluorescence folding traces for DM-MBP are shown in (E). Standard deviations of at least three independent experiments are shown.

The accelerated folding of rhodanese found for EL-2[GGM]$_4$ can be recapitulated with the non-cycling single ring chaperonin SR-2[GGM]$_4$. And interestingly, other cavity size mutants in SR followed essentially the same trend of change for folding speed, indicating the space confinement contributed evidently to the enhancement of folding rate, thus implied the ATP-driven cycling for substrates is dispensable for this increase-of-rate finding (Figure 24).

In contrast to rhodanese and MetF, reducing cavity size did not accelerate folding for the larger protein MBP. While EL-2[GGM]$_4$ still supported folding of SM-MBP at the rate similar to WT-GroEL, interestingly, the folding speed of DM-MBP was ~40% reduced, suggesting that the folding pathways of the mutant proteins differ (Figure 23C-E). Further reduction in cavity size (EL-3[GGM]$_4$) slowed the folding of both proteins without reducing the folding yield (Figure 23C, D and Figure 25), although encapsulation by GroES was still ~70% efficient (Figure 21). Again, comparable folding rates and yields of DM-MBP were observed in SR cavity size mutants suggested the result is independent of multiple rounds of folding cycles (Figure 24 and Figure 25).

Additionally, changing the [GGM]$_4$ sequence to [AAA]$_4$, [GGA]$_4$ and [GGA]$_4$[GGM]$_4$ all affected the folding speed of MBP, implying a sequence specific effect of [GGM]$_4$M may contribute to the organization of MBP folding intermediates in the *cis* cavity. A sequence-specific effect of [GGM]$_4$M on folding will be discussed below in 4.2.3.
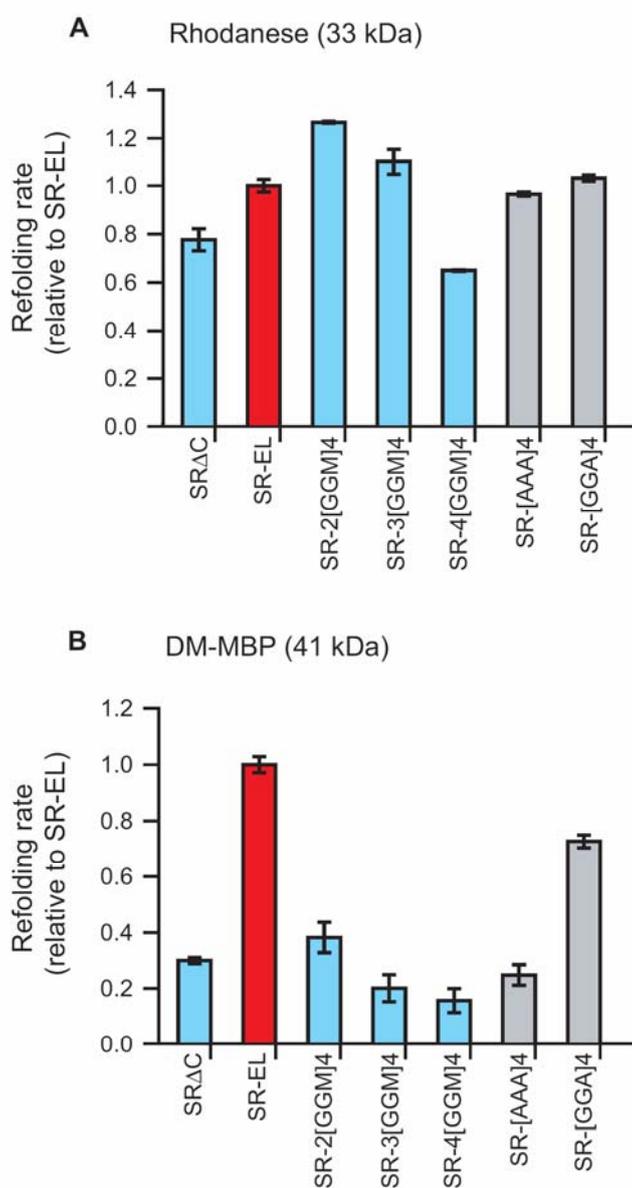
Figure 24. Effect of SR-EL cavity size on folding rates.

SR-EL/GroES assisted refolding of rhodanese (A), and DM-MBP (B) at 25°C with the SR-EL mutants indicated. Blue bars, cavity size mutants (decreasing cavity size from left to right); light gray bars, mutants with reduced hydrophobic character of the C-terminal repeat sequences. The refolding rate obtained with SR-EL (red bar) was set to 1. Standard deviations of at least three independent experiments are shown.
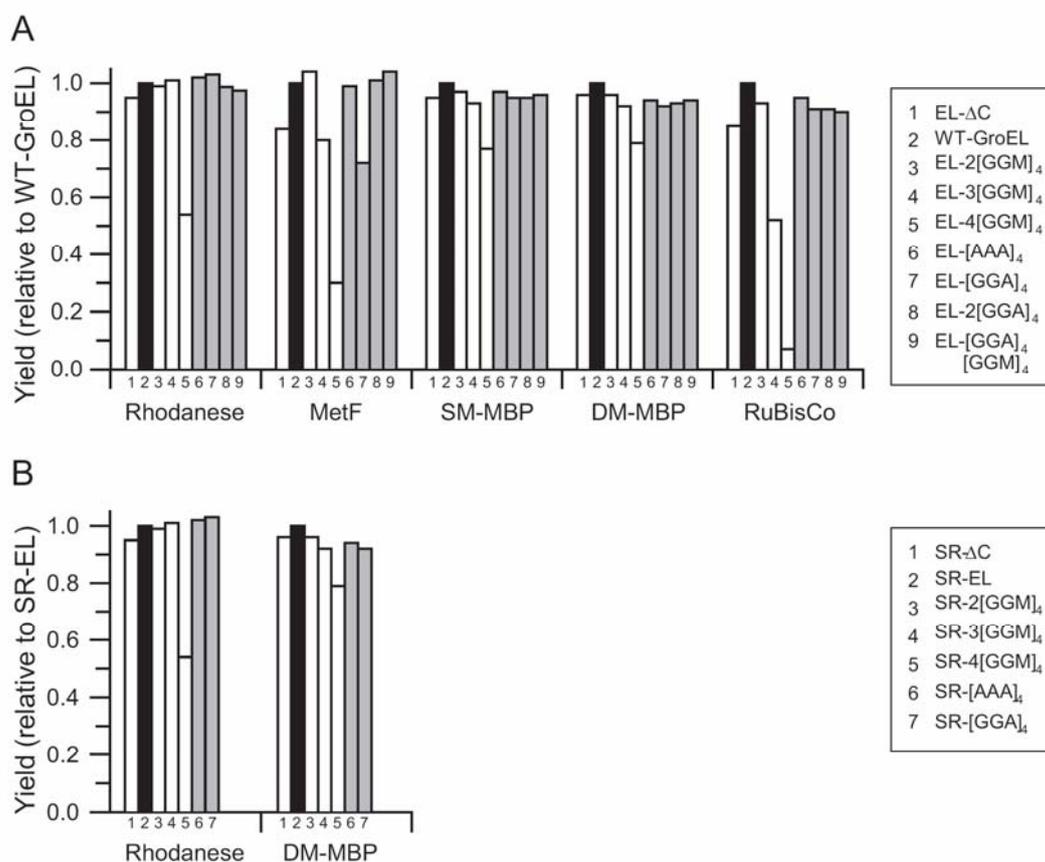
Figure 25. Refolding yields observed for the various GroEL cavity size constructs.

Refolding yields of rhodanese, MetF, SM-MBP, DM-MBP and RuBisCo in the refolding assays from Figure 23 (Panel A, double ring chaperonin) and Figure 24 (Panel B, single ring chaperonin). White bars indicate cavity size mutants (decreasing cavity size from left to right) and the light gray bars indicate constructs that modulate the mildly hydrophobic character of the C-terminal repeat sequence. The refolding yield obtained with WT-GroEL or SR-EL (black bar) was set to 1.

Consistent with its larger size, the folding of RuBisCo was even more strongly affected by decreasing the volume of the chaperonin cage (Figure 23F). This effect was independent of the specific sequence of the C-terminal extension, because both EL-[AAA]$_4$ and EL- shown similar capacities in folding RuBisCo as GroEL wild-type. And, along with observation, EL-2[GGA]$_4$ and EL-2[GGM]$_4$ equally slowed RuBisCo folding to ~15% without affecting the yield (Figure 23 and Figure 25). In contrast, the folding yield was

reduced by ~50% with EL-3[GGM]$_4$ and by ~95% with EL-4[GGM]$_4$, correlating with the loss of encapsulation (see Figure 21).

Fluorescence anisotropy measurements using the D95C variants of mutant MBP labeled with Alexa 488 confirmed the idea that reducing cavity size restricted the mobility of enclosed protein. To ensure the result is obtained with single binding step the chaperonin instead of binding-release cycle between the bulk and the chaperonin, these experiments were performed with the non-cycling SR-EL cavity size-mutants. When bound denatured MBP to the apical domains of SR-EL, the anisotropy value of the unfolded protein was high, reflecting the low rotational dynamics of the large SR-EL-substrate complex (Figure 26A and B). Upon GroES binding triggered by ATP addition, a rapid drop in anisotropy occurred, indicating increased dynamics resulting from displacement of the bound protein into the cage (Rye *et al*., 1997). For SM-MBP in SR-EL or SR-2[GGM]$_4$, this step was followed by a time-dependent increase in mobility occurring with kinetics corresponding to folding, as measured by tryptophan fluorescence (Figure 26A). In contrast, the folding protein was increasingly restricted in mobility in SR-3[GGM]$_4$ and SR-4[GGM]$_4$ (Figure 26A). DM-MBP generally experienced a more pronounced restriction in mobility (Figure 26B), suggesting that this protein populates more extended folding intermediates.

These results are consistent with theoretical simulations of the effects of steric confinement on protein folding (Baumketner *et al*., 2003; Takagi *et al*., 2003; Zhou, 2004), which predict that proteins will experience a rate acceleration of folding with increasing confinement up to a point where further restriction in space would limit necessary reconfiguration steps.
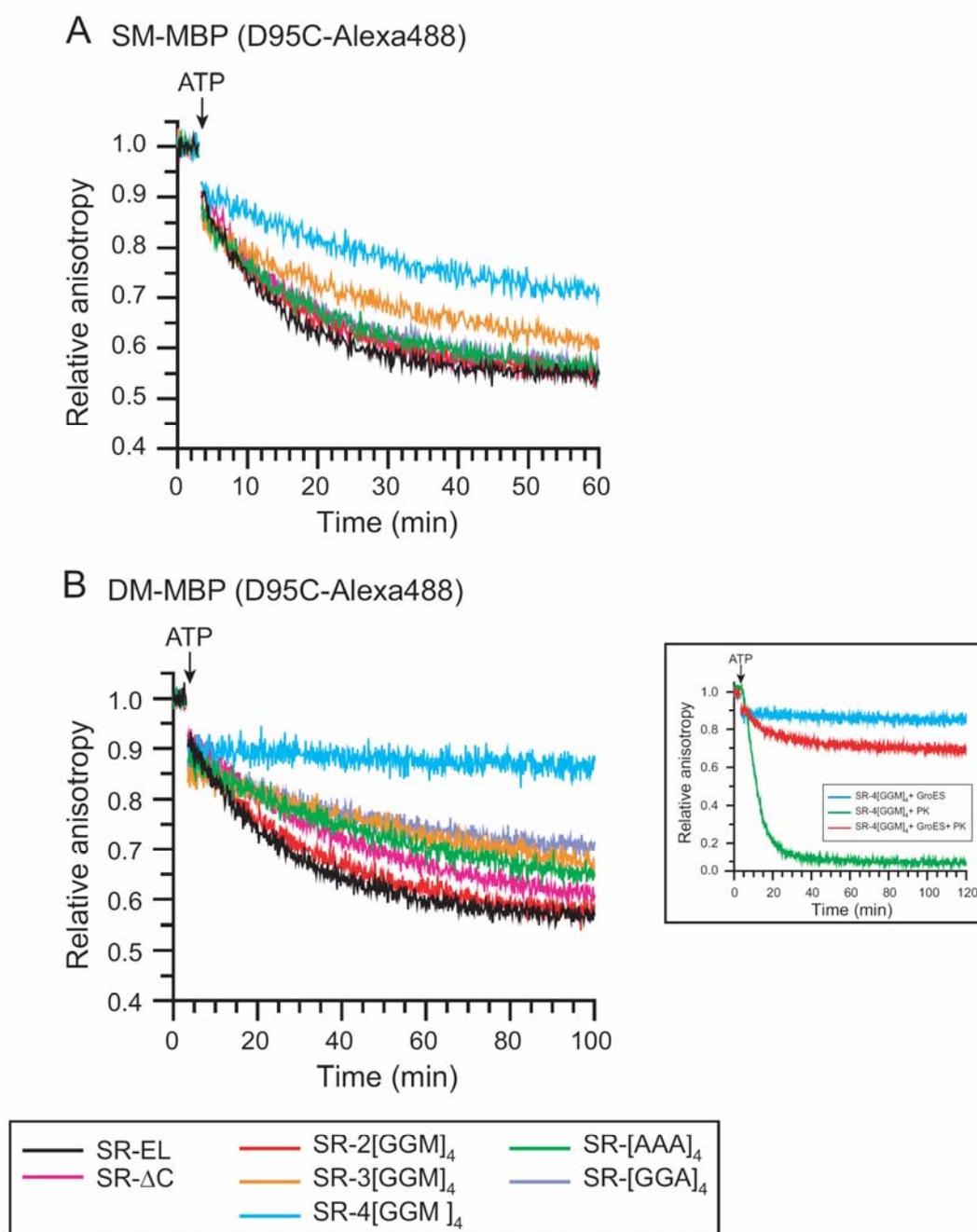
Figure 26. Restriction in substrate protein mobility upon encapsulation in SR-EL and in SR-EL with mutated C-terminal sequences.

Kinetics of steady state fluorescence anisotropy of SM-MBP (A) and DM-MBP (B) upon encapsulation by GroES in the SR-EL cavity size-mutants indicated. D95C versions of MBP were labeled with Alexa 488 (see Experimental Procedures). GroES binding was initiated by addition of ATP (arrow). Note that removal of non-encapsulated DM-MBP by proteinase K in the reaction with SR-4[GGM]$_4$ in (B) had only a small effect on anisotropy, indicating that largely encapsulated protein was measured (see inset).

4.2.3. Function of GGM repeats in folding

The flexible, mildly hydrophobic GGM repeats of GroEL are highly conserved among GroEL homologs from different species (Brocchieri and Karlin, 2000). GroEL lacking this sequence was found to support the growth of *E. coli* but in contrast to WT-GroEL was unable to suppress temperature sensitive mutations of chromosomal replication initiator protein dnaA (McLennan *et al.*, 1993). We found that changing $[GGM]_4M$ to $[AAA]_4A$ or $[GGA]_4A$ decelerated the folding of mutant MBP to a greater extent than deleting $[GGM]_4M$ altogether (Figure 23C and D). Anisotropy measurements revealed that, contrary to expectations, increasing cavity size by deleting $[GGM]_4M$ (SRΔC) did not increase protein mobility during folding (Figure 26A and B). Moreover, replacement of $[GGM]_4M$ by $[AAA]_4A$ caused a substantial restriction in mobility of the folding protein, an effect that was again most pronounced with DM-MBP and was not seen with WT-MBP (Figure 26B). Notably, these mutations had only a small effect on the folding of rhodanese, MetF and RuBisCo (Figure 23A, B, F). Similarly, mutants EL-2$[GGA]_4$ and EL-$[GGA]_4[GGM]_4$ were less effective than EL-2$[GGM]_4$ in the folding of mutant MBP (Figure 23C and D). These results argue for a specific role of $[GGM]_4$ in facilitating the rearrangement of certain folding intermediates by providing a mildly hydrophobic, interactive surface. This function may be particularly important for proteins which have acquired mutations that result in highly energetically frustrated folding pathways, such as the mutant versions of MBP.

4.3. Role of negative charge clusters on the cavity wall in GroEL assisted folding

4.3.1. GroEL mutants with altered cavity charge

The wall of the GroEL *cis* cavity has a net charge of minus 42 (189 negatively and 147 positively charged amino acid residues). A number of negative charges (residues E252, D253, E255, D359, D361, and E363), which are all positioned in the apical domain, cluster in two circular layers (Figure 27). Most of these residues (E252, D253, E255, E363) are highly conserved among GroEL homologs, although they have no apparent role in the basic GroEL functions of substrate and GroES binding (Brocchieri and Karlin, 2000; Stan *et al.*, 2003). To explore their possible significance in promoting folding, we replaced individual or multiple residues by either asparagine or glutamine (neutral) or lysine (positive) in SR-EL, in attempt to demonstrate the effect in a non-substrate-cycling manner. As a consequence of the 7-fold symmetry of GroEL, these mutations dramatically change the electrostatic character of the cavity wall (Figure 27). The mutant proteins were efficiently overexpressed and purified in soluble form and have been verified on blue native gels as heptamer. All the SR-EL charge-mutants bound GroES stably in the presence of ATP, reflecting the inability of SR-EL to cycle GroES. Binding of GroES to preformed complexes of mutant SR-EL and unfolded DM-MBP resulted in 90-100% protease protection (Figure 28). In contrast, several of the SR-EL charge mutants, such as SR-QNQ, SR-3N3Q and SR-D359K had a 40-50% reduced capacity to support RuBisCo encapsulation, suggesting an interference with the compaction of the molecule normally occurring upon its displacement into the cage by GroES (Lin and Rye, 2004). A ~75% reduced encapsulation efficiency was observed with mutation D253K (Figure 28). This mutant was not analyzed further with regard to RuBisCo folding.

Figure 27. Properties of GroEL cavity charge-mutants.

Space-filling model of GroEL/GroES-(ADP)$_7$ complex (, pdb 1AON, DS ViewerPro) offering a view into the *cis*-cavity with 4 subunits of GroEL and GroES shown. Clusters of negatively charged residues exposed towards the *cis*-cavity are highlighted in red (E252, D253, E255) and blue (D359, D361, E363). The net charge of the *cis*-cavity wall formed by 7 GroEL subunits is indicated for the different mutants.
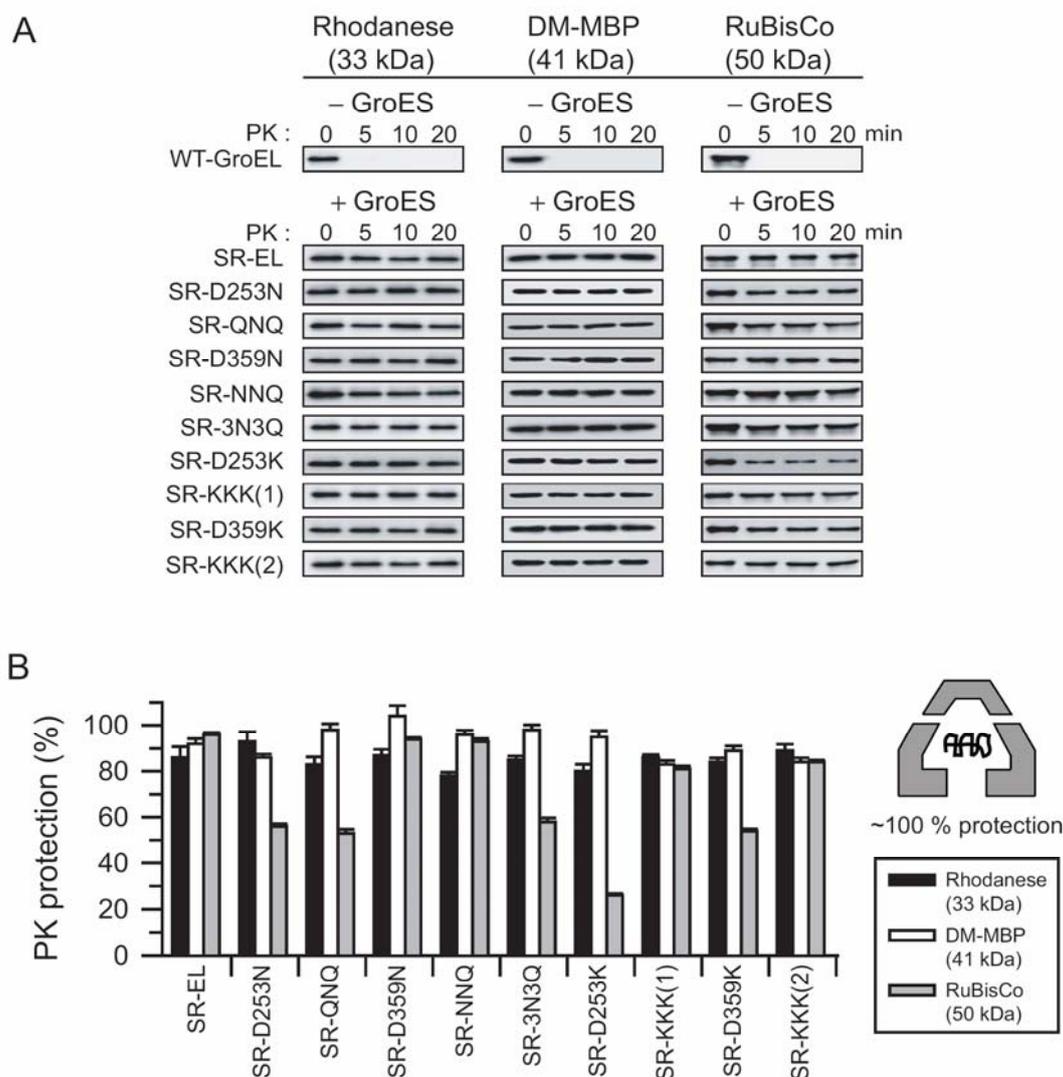
Figure 28. Proteinase K protection of SREL- GroES- substrate complex.

(A) PK protection of rhodanese, DM-MBP and RuBisCo in complexes with the various mutant forms of SR-EL and GroES. PK treatment was performed as in Figure 21. SREL-substrate complexes were incubated with PK (2 μg/ml) in buffer A/4 mM AMP-PNP in the absence or presence of GroES at 25°C. (B) Protected substrate protein was quantified by immunoblotting and densitometry. Amounts in non-protease treated reactions correspond to 100%.

4.3.2. Effects of GroEL cavity charge on folding

The charge mutations were without effect on the rate or yield of WT-MBP folding, but moderately reduced the folding speed of SM-MBP and markedly decelerated DM-MBP folding (Figure 29A-C and Figure 30). Changing single or multiple negative charges to neutral residues slowed DM-MBP folding by 30 to 80%, with multiple mutations generally having a more severe effect (Figure 29C). The effects of replacing negative with positively charged residues varied considerably dependent on the specific protein tested. For example, the single charge reversal of SR-D359K, while strongly decelerating the folding of DM-MBP, caused a moderate acceleration of rhodanese folding and was without effect on the folding rate of RuBisCo (Figure 29 C and D). In the case of RuBisCo, some of the charge mutants strongly diminished the folding yield. An interesting example is SR-NNQ, which caused an 80% reduction in yield (Figure 30), although the protein was efficiently encapsulated by GroES (Figure 28). However, the subpopulation of molecules that reached native state did so with almost the normal rate as for SR-EL (Figure 29D). This indicated that a large fraction of RuBisCo was trapped inside the SR-NNQ-GroES cage in a non-native state. A virtually complete folding arrest of encapsulated RuBisCo was observed with SR-KKK(2), containing positive charges at positions D359, D361 and E363. Indeed, upon dissociation of GroES at low temperature in the presence of EDTA, most of the RuBisCo was released from the cavity in a PK-sensitive, non-native state. In contrast, PK-resistant, folded protein was detected when the same experiment was carried out with wild-type SR-EL. The complete removal of cavity net charge in SR-KKK(2) also strongly decelerated the folding of DM-MBP, but caused a moderate increase in folding speed for rhodanese (Figure 29C and D). It is noteworthy in this context that WT-MBP and RuBisCo have a negative net charge of -8 and -11, respectively, whereas rhodanese, the protein least affected by the charge mutations, has a net charge of only -1 (Supplemental Table S2). The correlation of

interaction strength between substrate net charge and GroEL cage charge composes an interesting subject that remains to be studied further.
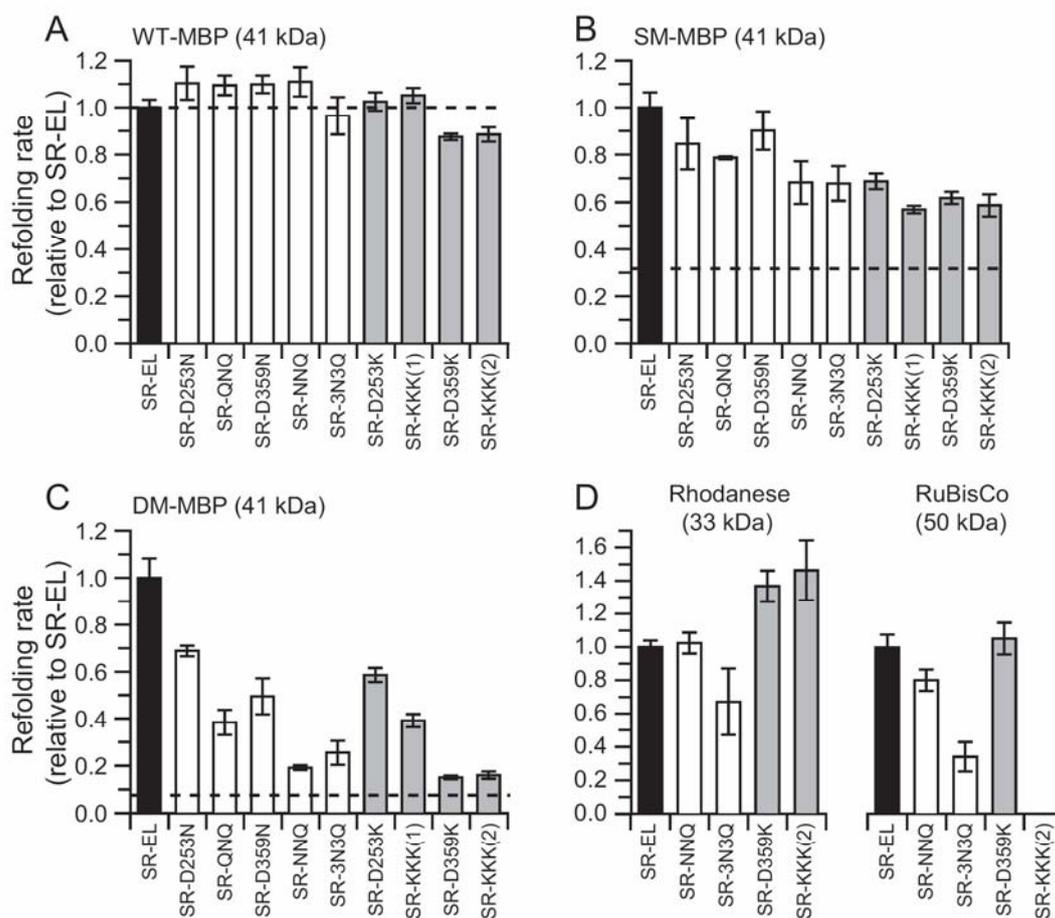
Figure 29. Effect of GroEL cavity charge on folding rates.

Refolding of WT-MBP (A), SM-MBP (B), DM-MBP (C), rhodanese and RuBisCo (D) with the indicated SR-EL charge mutants and GroES was analyzed in buffer B/5 mM ATP at 25°C as described in Experimental Procedures. White bars indicate amino acid changes from negative to neutral and light gray bars changes from negative to positive. The refolding rate obtained with SR-EL was set to 1 (black bar). Dashed line represents the rate of spontaneous folding for the respective proteins.
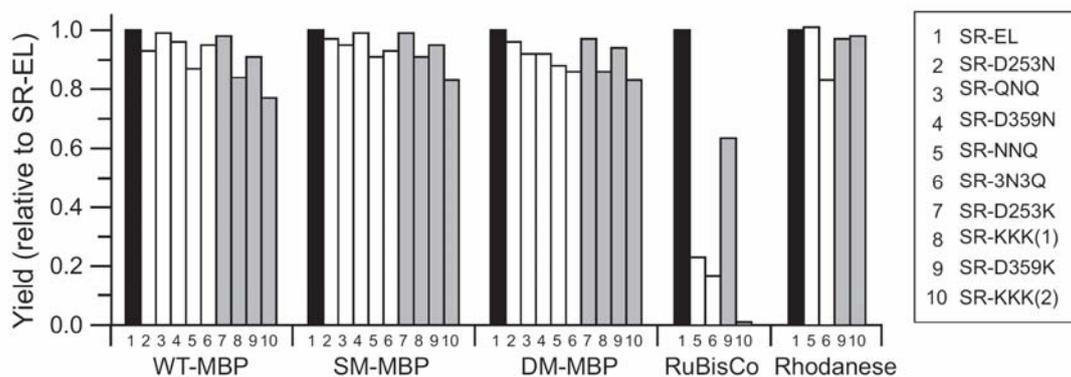
Figure 30. Refolding yields observed for the various SR-EL charge constructs

Refolding yields of WT-MBP, SM-MBP, DM-MBP, RuBisCo and rhodanese of the refolding assays from Figure 29. The white bars indicate negative to neutral amino acid changes, and the light gray bars are negative to positive amino acid changes. The refolding yield obtained with SR-EL was set to 1 (black bar).

Reducing the negative net charge of the cavity wall strongly impaired the mobility of MBP in the chaperonin cage. This effect was already apparent with the D95C version of WT-MBP (Figure 31A). The protein interacted substantially with the less negatively charged cavity wall, both during folding and after reaching native state. Further, mobility was even more restricted with the slower folding SM-MBP and DM-MBP (Figure 31B and C). Mutants which caused complete loss of cavity wall net charge, such as SR-3N3Q and SR-KKK(2) (see Figure 27), significantly slowed the rapid mobilization of SM-MBP and DM-MBP normally occurring immediately upon GroES binding (Figure 31B and C). This suggests that the non-native states of these proteins interact with the cavity wall immediately after release from the apical GroEL domains, presumably resulting in delayed burial of hydrophobic residues.

These findings indicate that the charge properties of the GroEL cavity wall are of profound significance in the ability of the chaperonin to promote the folding of certain substrate proteins. While the charge effects on specific proteins may vary, the overall

negative surface charge of the cavity wall of the apical domains appears to provide a non-interactive surface optimized to accomplish the efficient folding of many different proteins.
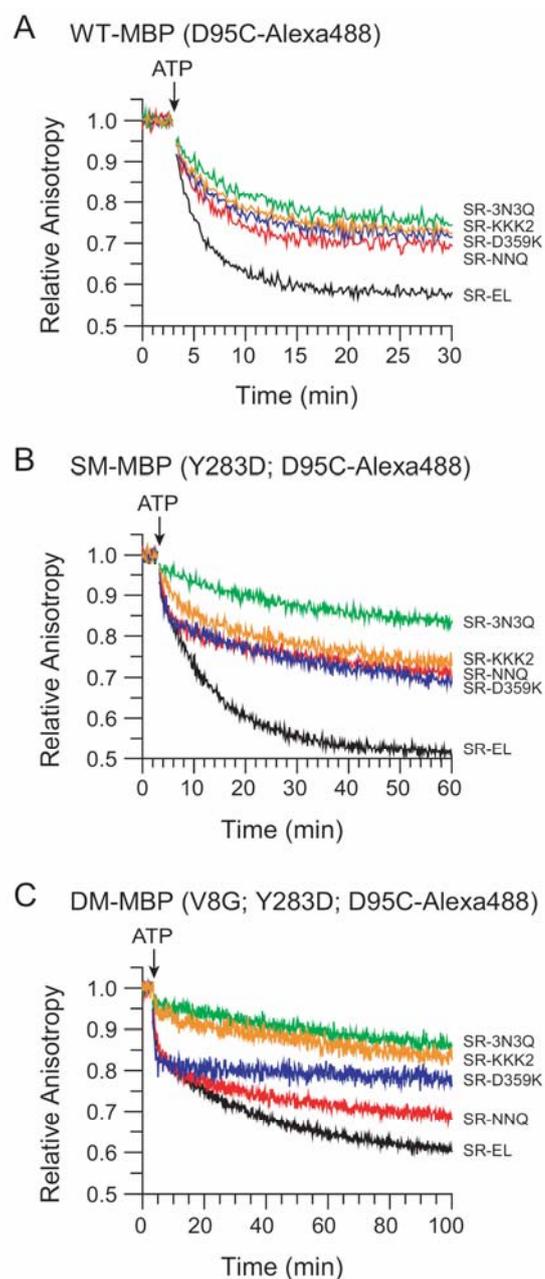
Figure 31. Restriction in substrate protein mobility upon encapsulation in SR-EL and SR-EL cavity charge mutants.

Kinetics of steady state fluorescence anisotropy of WT-MBP (**A**), SM-MBP (**B**) and DM-MBP (**C**) upon encapsulation by GroES in the SR-EL cavity charge mutants indicated. D95C versions of MBP were labeled with Alexa 488 (see Experimental Procedures). GroES binding was initiated by addition of ATP (arrow).

4.4. Study GroEL/GroES assisted folding *in vivo*

4.4.1. Significance of accelerated folding by GroEL/GroES *in vivo*

The requirement of GroEL/GroES for efficient protein folding *in vivo* is well established, but it is unclear whether the capacity of the chaperonin to accelerate folding is biologically relevant. We addressed this question by using our model substrate MBP and a recently identified *in vivo* substrate of GroEL, MetF. Overexpression of WT-MBP from an arabinose controlled expression plasmid in *E. coli* resulted in the production of fully soluble protein. In contrast, expression of SM-MBP, DM-MBP and MetF produced largely insoluble protein (Figure 32A). Additional overexpression of GroEL/GroES, but not of GroEL alone, dramatically reduced the formation of aggregates and allowed the production of soluble protein (Figure 32B and C). Overexpression of ELΔC suppressed the aggregation of SM-MBP, DM-MBP and MetF only partially (Figure 32D), consistent with the reduced folding rates observed with ELΔC *in vitro* (Figure 23B-D). Expression of the GroEL variant with reduced cavity size, EL-2[GGM]$_4$, resulted in a similar effect in the case of mutant MBP, but allowed the production of soluble MetF with close to 100% yield (Figure 32E). This enhancement of solubility corresponds with the accelerated folding of MetF by EL-2[GGM]$_4$ observed *in vitro* (Figure 23B-D). As expected, EL-4[GGM]$_4$ was unable to support the folding of mutant MBP or MetF (Figure 32F). Changing the repeat motif from [GGM]$_4$M to [GGA]$_4$A failed to produce significant amounts of native mutant MBP but only partially suppressed the aggregation of MetF (Figure 32G), confirming the sequence-specific contribution of [GGM]$_4$M to mutant MBP folding (Figure 23C and D). Similarly, the charge mutants EL-NNQ, EL-3N3Q and EL-KKK(2), while partially efficient in MetF folding, markedly reduced the amount of soluble mutant MBP (Figure 32H-J), again consistent with the observations *in vitro* (Figure 29B-C). Collectively, these results demonstrate the biological relevance of accelerated folding achieved by the chaperonin

system. Reducing the ability of GroEL to accelerate folding diminishes its capacity to handle recalcitrant proteins such as the mutant versions of MBP. On the other hand, decreasing the size of the GroEL cavity is beneficial for the folding of the smaller protein, MetF.

4.4.2. GroEL depletion strain

To further address the functional relevance of these GroEL variants *in vivo*, we assayed for growth complementation of WT-GroEL. Since GroEL is essential under all conditions for the viability of *E. coli* (Fayet *et al*., 1989), we used a special *E. coli* strain, MC4100 SC3, in which the groE operon is under the control of the arabinose promoter. Endogenous GroEL and GroES expression is therefore critically dependent on the presence of arabinose in the growth medium (Kerner et al., 2005 and Figure 30). Control tests with empty vector or only GroEL overexpression failed to complement for the growth of this strain, supporting previous findings that both GroES and GroEL are non-dispensable for growth. Interestingly, gradually shrinking the cage size resulted in a loss of complementation ability of ~10 times and ~1,000 times for EL-3[GGM]$_4$ and EL-4[GGM]$_4$, respectively (Figure 33). This is probably due to the consequence of failure to fold obligate substrates of GroEL in the small cage, correlating well with the *in vitro* refolding and *in vivo* co-expression experiments on MetF (Figure 23 and Figure 32). Similarly, strong impairment in growth with the cavity charge mutants like EL-3N3Q and EL-KKK(2) also suggested the significance of surface property of the GroEL/GroES *cis*-cage can contribute to the protein folding.
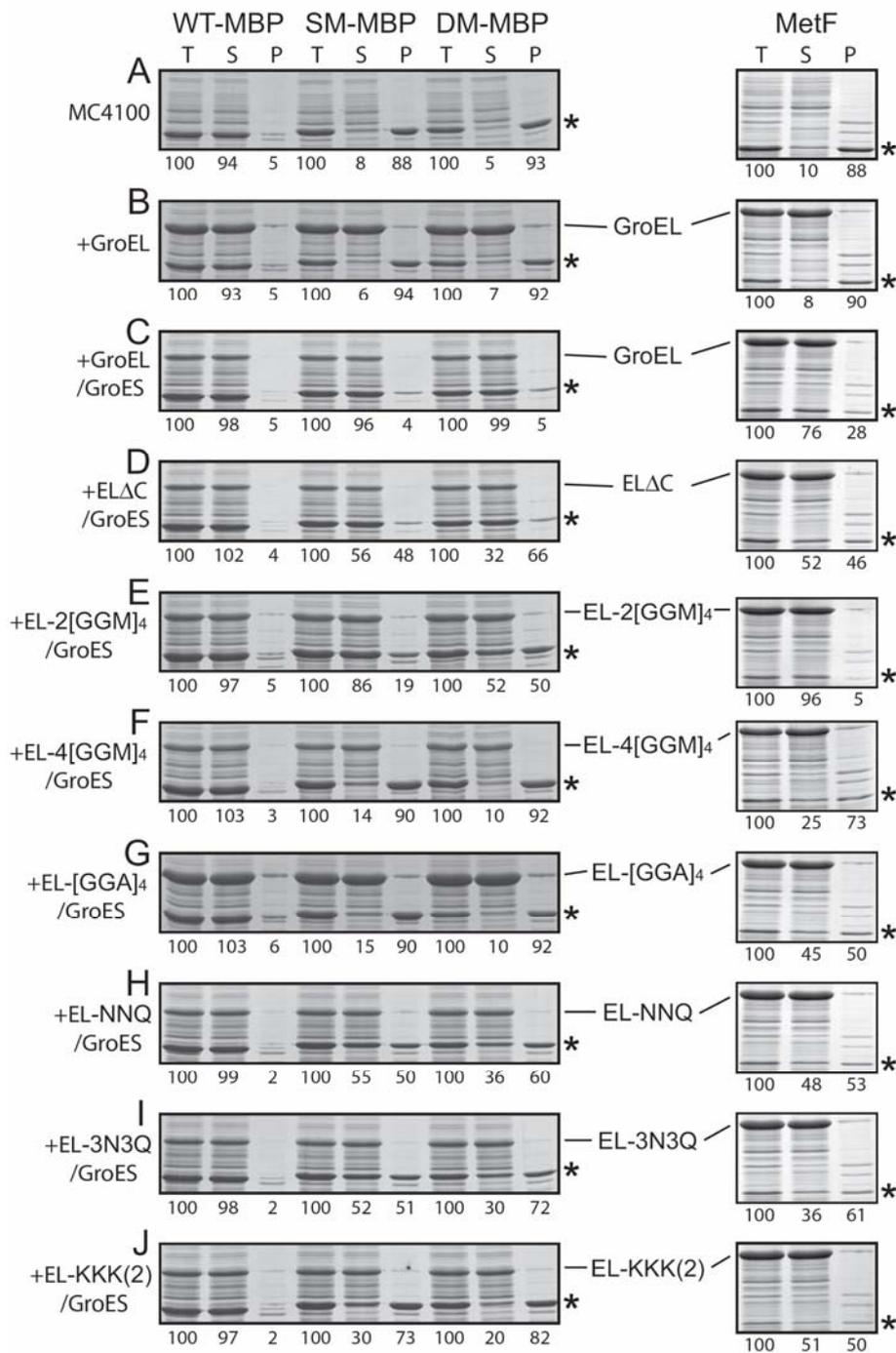
Figure 32. Effect of WT-GroEL and GroEL mutants on folding *in vivo*.

WT-MBP, SM-MBP, DM-MBP or MetF were overexpressed in *E. coli* cells of strain MC4100C either without (A) or with additional overexpression of GroES and the GroEL mutants indicated (C-J). GroES expression was omitted in (B). Total (T), supernatant (S) and pellet (P) fractions were analyzed by SDS-PAGE and Coomassie staining. Amounts of MBP or MetF protein in S and P fractions, determined by densitometry, are given in % with total protein (T) set to 100%. The asterisk indicates the position of MBP or MetF.
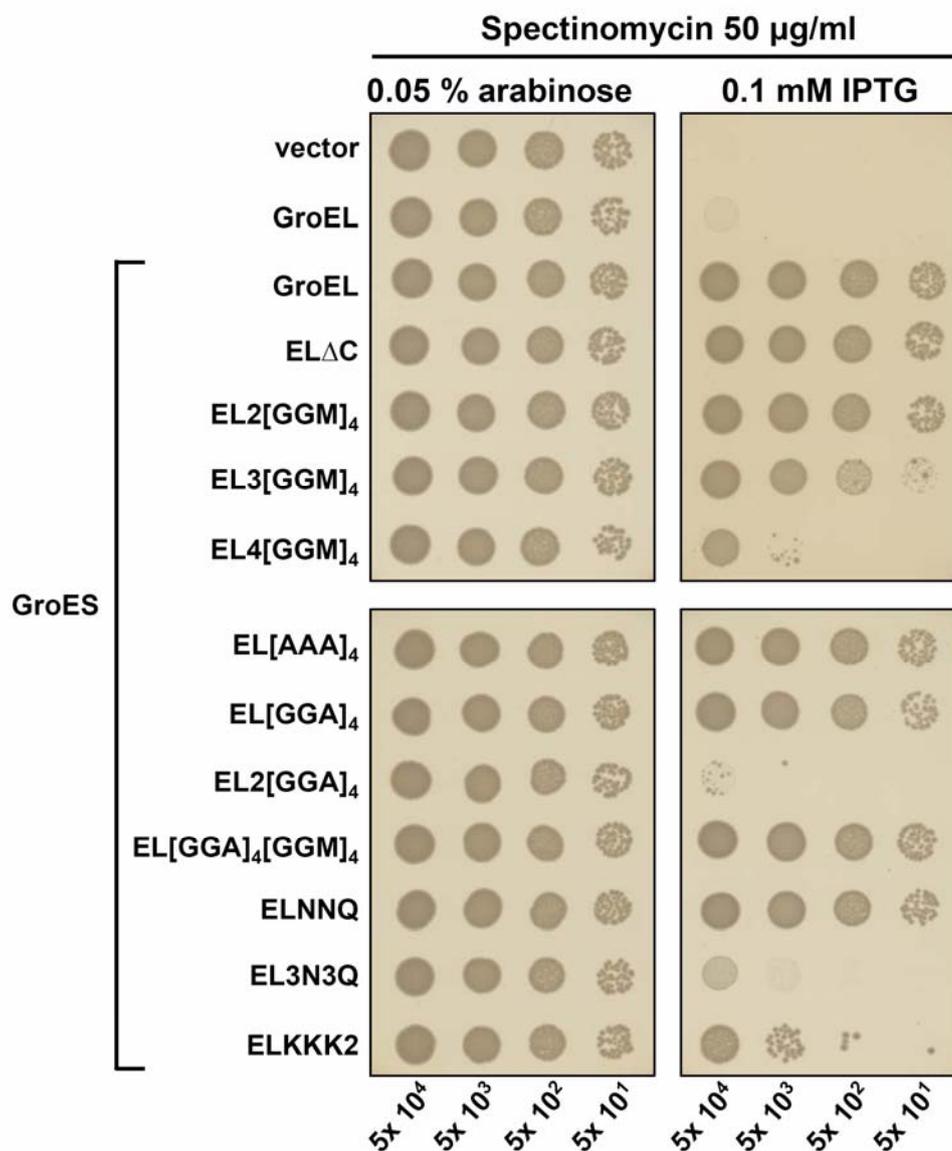
Figure 33. Complementation effect of GroEL size and charge mutants in EL/ES-depletion stain.

Spectinomycin-resistant constructs encoding denoted genes were transformed into *E. coli* MC4100 SC3 cells (Kerner *et al.*, 2005) and grew with either wild-type GroEL/GroES induction (left, 0.05% arabinose) or engineered GroEL induction (right, 0.1 mM IPTG) on LB plates. Serial dilutions were performed according to Experimental Procedures.

5. Discussion

The GroEL/GroES nano-cage allows a single protein molecule to fold in isolation. This reaction has been compared to spontaneous folding at infinite dilution. However, recent theoretical studies indicated that the physical environment of the chaperonin cage can alter the folding energy landscape, resulting in accelerated folding for some proteins (Takagi *et al*., 2003; Zhou, 2004). This suggested that GroEL/GroES has a role in folding beyond providing an infinite dilution box.

By performing an extensive mutational analysis of GroEL, we have identified three structural features of the chaperonin cage as major contributors to this capacity: i) geometric confinement exerted on the folding protein inside the limited volume of the cage, ii) a mildly hydrophobic, interactive surface at the bottom of the cage, and iii) clusters of negatively charged amino acid residues exposed on the cavity wall. We suggest that these features in combination provide a physical environment that has been optimized in evolution to catalyze the structural annealing of proteins with kinetically complex folding pathways. Thus, the chaperonin system and its mutant versions may prove as useful tools in understanding how proteins navigate their energy landscape of folding.

5.1. Effect of spatial confinement on folding rate

Initial experimental observations by Brinker *et al*., (2001) demonstrated an acceleration effect of GroEL on the folding rate. The conditions for an efficient renaturation of *R. rubrum* RuBisCo *in vitro* were elaborated, and a comparative study was performed. This result showed that the folding of RuBisCo inside the GroEL-GroES cavity was four times faster than spontaneous folding in bulk solutions, suggesting a catalytic role of the GroEL-GroES system. Since the non-cycling SR-EL mutant could reproduce the rate acceleration

of the folding of RuBisCo, this indicated that the rate acceleration in the GroEL cavity is unrelated to repeated cycles of binding and release. The fact that similar experiments carried out with another substrate protein, rhodanese, whose molecular weight (~33 kDa) is considerably lower than RuBisCo (~50 kDa), revealed no acceleration effect by GroEL, suggested that this acceleration effect may be dependent on the size of the substrate protein relative to the dimensions of the central cavity.

In the crystal structure, the GroEL-GroES cage has a volume of ~175,000 Å^3, in principle large enough to accommodate proteins of >70 kDa (Xu *et al*., 1997). However, the functionally relevant volume is smaller, due to the C-terminal 23 amino acids of the GroEL subunits, which protrude into the cavity but are not resolved in the structure. Because of their flexible character, these segments are likely to occupy more than their nominal volume of ~14,000 Å^3 per GroEL ring. Moreover, since the geometry of the cage resembles a truncated cone, part of the volume may be unavailable to certain substrate proteins. Consistent with these considerations, most GroEL-dependent proteins are smaller than 50 kDa (Kerner *et al*., 2005) and the 56 kDa phage T4 capsid protein, Gp23, already requires an enlarged, phage-encoded version of GroES (Gp31) for encapsulation (Bakkes *et al*., 2005; Hunt *et al*., 1997). It follows that a typical GroEL substrate would undergo considerable compaction upon displacement into the cage from a loosely packed bound state (Horst *et al*., 2005). This step is mediated by ATP and GroES binding, which drive large allosteric domain movements in GroEL (Figure 11).

Confinement of a protein within a cavity might prevent formation of non-native conformations possessing radii of gyration larger than the radius of the enclosing cavity. And an enclosed substrate protein should demonstrate a significantly higher stability compared to bulk solution, because unfolded states with expanded volume will be disallowed in bulk, which increased the risk of incorrect compaction or aggregation (Thirumalai *et al*., 2003). In support of this hypothesis, a range of computational studies

(Takagi *et al*., 2003; Zhou, 2004), as well as experimental observations on protein confined within silica glass nanopores (Eggers and Valentine, 2001), have shown that spatial confinement of a protein can, in fact, lead to a large increase in protein stability. More significantly, however, the spatial restriction of a non-native folding intermediate could have a profound effect on the kinetics and mechanism of folding. The experimental observation of accelerated RuBisCo folding inside the GroEL-GroES cavity suggested that the folding landscape of a protein can be fundamentally altered by encapsulation (Brinker *et al*., 2001). Several computational studies using lattice and molecular dynamics simulation models have also provided support for this conclusion (Baumketner *et al*., 2003; Takagi *et al*., 2003; Xu *et al*., 2005; Zhou, 2004).

The geometric confinement exerted by the cage would result in a destabilization of unfolded conformers relative to bulk solution and in the preferential population of compact intermediates, thus potentially smoothing rugged folding energy landscapes and enhancing the folding rate (Figure 34) (Baumketner *et al*., 2003; Brinker *et al*., 2001; Takagi *et al*., 2003; Zhou, 2004). Because the entropic penalty for establishing long-range interactions is large, the acceleration of folding is predicted to be more pronounced for proteins with a high proportion of long-range tertiary contacts (Takagi *et al*., 2003), such as the GroEL-dependent proteins with complex α/β or α+β domain topologies (Kerner *et al*., 2005).

We have performed the first systematic test of these ideas by gradually reducing or increasing the volume of the chaperonin cage. The results of these experiments are remarkably consistent with prediction. Relatively small proteins such as rhodanese and MetF (33 kDa) experienced a rate acceleration of folding upon reducing cage size to a point where further restriction in space slowed folding dramatically. For MBP (41 kDa) and RuBisCo (50 kDa), on the other hand, either reducing or increasing cage volume decelerated folding, indicating that WT-GroEL provides an optimal level of spatial confinement for

these proteins. The optimum for productive confinement proved to be remarkably narrow, with as little as 2-5% change in cage volume affecting folding rates by 2-fold or more.

On the other hand, "over-confinement" may stabilize misfolded states that require substantial expansion in order to return to a productive folding trajectory. Taking the geometries of cage and substrate proteins into consideration, the extent of conformational movement possible during folding is indeed very limited. For rhodanese, MetF and MBP, the longest axes of the native proteins are between 60-73 Å, compared to 85 Å as the longest dimension of the cage (Supplemental Table S2). Remarkably, in the case of RuBisCo, the long axis of the native monomer is 95 Å, suggesting either that the GroEL-GroES complex is conformationally plastic or the product of RuBisCo folding is a compressed monomer. The latter possibility would be consistent with recent FRET measurements for this protein when enclosed in the GroEL-GroES cage (Lin and Rye, 2004).
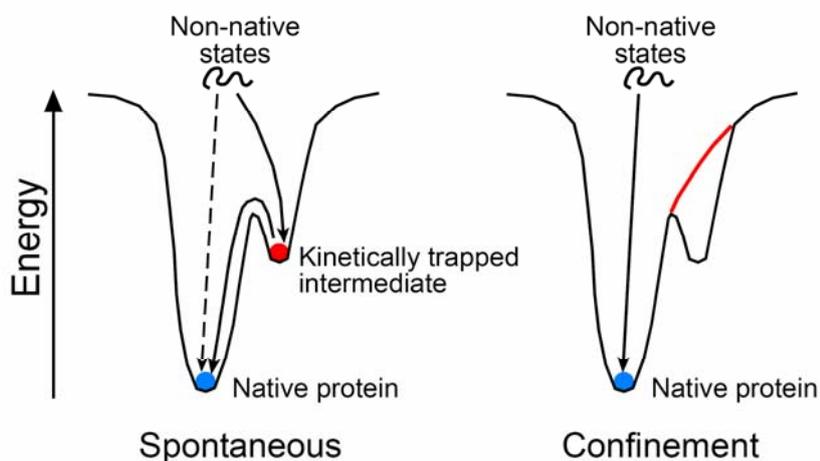
Figure 34. The mechanism of GroEL-GroES accelerated folding .

Simple energy diagrams are shown for (Left) Spontaneous folding. Unfolded protein partitions between a fast pathway to the native state and preferred formation of a trapped intermediate whose spontaneous conversion to the native state is slow. (Right) Accelerated chaperonin-assisted folding as a result of confinement. Formation (or accumulation) of trapped intermediate is thought to be avoided in the confined environment of the chaperonin cage. This is reflected in a smoothing of the energy surface (red line). In this model, the energy of ATP is utilized to discharge unfolded protein into the narrow hydrophilic space of the chaperonin cage. From Brinker *et al*. (2001)

5.2. Effect of the mildly hydrophobic C-terminal GGM repeat on folding rate

The amino acid sequences of all members of the Group I chaperonin, including GroEL, share a striking motif located at the C-terminus. This motif, variable in length and exact sequence, consists mainly of glycine and methionine residue. The GGM repeats are also observed in other ATPase protein families, like Hsp70 (Karlin and Brocchieri, 1998) or RecA (Brendel *et al*., 1997). The significance of these repetitive elements is unknown. This motif in Hsp70 was suggested to react with the Hsp70 peptide binding site, leading to Hsp70 oligomer formation (Rippmann *et al*., 1991). In other glycine-rich protein, keratins, the glycine-rich motifs are organized into loops in which the hydrophobic residues (*i.e.* methionines) stack and the flexible structure is akin to a molecular spring. However its function in the GroEL *cis* cage may be different and profound. Molecular dynamics simulations of the folding of a highly energetically frustrated protein inside the chaperonin cage suggested that a moderately hydrophobic wall would accelerate folding substantially (Jewett *et al*., 2004). The mildly hydrophobic GGM repeats may fulfill such a role, perhaps by intercalating between hydrophobic regions of folding intermediates, thereby preventing the formation of kinetically stable, misfolded states. This GGM repeat-substrate interaction maybe in particularly important for certain substrates, as revealed from the folding experiments. When changing the GGM repeat to AAA or GGA repeats, with little change in the relative size of the cage, both SM-MBP and DM-MBP displayed impaired folding rates, and RuBisCo was mildly affected, in the mutated cavities (Figure 23). The anisotropy measurement for MBP mobility upon encapsulation implied that such a slight hydrophobicity change of the cage bottom can affect the rearrangement of folding intermediates as much as the over-confined environment of EL-3[GGM]$_4$ (Figure 26).

Taken together with the confinement effect results, our data support a model in which the bimodal character of the cavity wall facilitates the re-configuration of folding intermediates within the confined cage. Notably, the resulting annealing mechanism is independent of repeated cycles of active GroES and ATP-dependent unfolding, in contrast to the "iterative annealing" model (Thirumalai and Lorimer, 2001). Instead, "cage-mediated annealing" would achieve a smoothing of the folding energy landscape in a single encapsulation cycle by sequestrating the protein in a confined space with an optimized mixture of a hydrophobic foundation property. Consistent with this proposal, changes in these properties had the most pronounced effect on the folding of DM-MBP and RuBisCo, those proteins in the test set which experienced the highest enhancement in folding rate by GroEL.

5.3.   Physical properties of the GroEL cavity wall

In theoretical models of confinement, proteins are generally assumed to be enclosed in a volume limited by an inert, repulsive wall. Our mutational analysis demonstrates that polar and hydrophobic wall properties of the chaperonin cage, acting in conjunction with geometric confinement, contribute critically to the ability of the system to accelerate folding. As known charge allows versatile ways of interactions, from weak to strong between molecules, the highly conserved charged residues lining the central cavity of the GroEL is likely to be necessary for the reconfiguration of the misfolded or partially unfolded proteins into their native conformations, by providing a strong hydrophilic environment that compels formation of the hydrophobic core. This electrostatic repulsion effects is a mechanism frequently happens for rapid and long-range effects that help orient the surface, as well as facilitating hydrophobic interactions. Additionally, given that most GroEL substrates have a negative net charge (Kerner *et al*., 2005), the electrostatic repulsion

force may also contribute to repel the substrate from the GroEL cavity wall upon succeeding with its folding (Hunt *et al.*, 1997; Wang *et al.*, 2002).

In fact, substrate intermediate and the cavity wall would most likely accommodate no more than a thin hydration layer around the exterior of the substrate protein. As even demonstrated for native and well folded proteins of modest size, like GFP (~27 kDa), confinement within the GroEL-GroES cavity appears to impose serious constrains on the rotational freedom of the protein. The rotational correlation time of native GFP enclosed within the SR-EL/GroES cavity was found to be 4-fold slower than in free solution (Weissman *et al.*, 1996). This suggests either significant, direct interactions between native GFP protein and the walls of the GroEL cavity, or a considerable increase in the effective viscosity of the hydration layer between the GFP and SR-GroES cavity wall that all because of the charge-mediated electrostatic force. For much larger substrate protein, like RuBisCo, the extent of physical constriction is likely to be far more significant. Indeed, the average hydration layer between a marginally expanded and large folding intermediate and the cavity wall might be no thicker than a single water molecule (Thirumalai and Lorimer, 2001).

5.4. Biological relevance of cage-mediate annealing

Based on the recent analysis of the GroEL substrate proteome, ~ 85 *E. coli* cytoplasmic proteins are predicted to be strictly dependent on GroEL/GroES for folding, including 13 proteins with essential functions (Kerner *et al.*, 2005). It would appear that the chaperonin cage has been optimized to accomplish the folding of these proteins at a biologically relevant time scale. As noted previously, the properties of the cage must therefore represent an evolutionary compromise to support a variety of folding pathways (Wang *et al.*, 2002), and this would explain why mutating certain features may improve the folding of a specific

protein while potentially being detrimental to the folding of others. However, significant structural deviations may be tolerated when additional, specialized forms of GroEL are expressed to allow adaptation of an organism to specific growth conditions. Interestingly, Mycobacteria express two forms of GroEL, of which GroEL1 lacks the C-terminal GGM repeat and instead has an 18 amino acid, histidine-rich sequence. This C-terminal sequence appears to be critical for GroEL1 to support the folding of proteins required for bacterial biofilm formation (Ojha *et al.*, 2005).

An additional important role of cage-mediated annealing is to preserve the foldability of a protein despite the presence of mutations, as shown for mutant MBP. This capacity would explain the recent finding that overproduction of GroEL/GroES reduces the phenotypic penetrance of deleterious mutations in bacterial cell lineages (Maisnier-Patin *et al.*, 2005), in a manner comparable to the conformational buffering effects proposed for other chaperone systems (Rutherford and Lindquist, 1998).

Overall, the size and surface properties of the cage represent an evolutionary compromise that helps the bacterial cell to produce functional proteins fast enough to survive in a competitive microbial world.

5.5.   Perspectives

With this new view of chaperonin-assisted folding we may need to reconsider several issues in protein folding. Are the intermediates of folding within the GroEL/GroES cavity different from the intermediates formed during spontaneous folding? Are new, more efficient routes for folding opened up by the special cavity environment? How can we exploit knowledge of chaperonin structure and function to enhance the folding of "difficult" proteins *in vitro* or to create new expression systems for the production of a particular protein without compromising host survival? Over the last 30 years, theoretical and experimental methods have been extensively applied to study the folding of small, rapidly folding domains. However, delineating the folding mechanism of larger, slower folding proteins poses an enormous experimental challenge, due to the complex structure of these proteins that are stabilized by long-range interactions. Understanding their folding within the constraints of the chaperonin nanocage may allow us to evolve enzymes of therapeutic interest with the use of the chaperonin system. For example, variants of GroEL described in this study, such as GroELΔC with a larger cavity size, may expand the spectrum of proteins that can utilize the chaperonin to substrates greater than 50 kDa. GroEL variants with smaller cavity size may be used to improve the folding rate of other proteins.

6. References

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

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

Anfinsen, C. B., Harrington, W. F., Hvidt, A., Linderstrom-Lang, K., Ottesen, M., and Schellman, J. (1955). Studies on the structural basis of ribonuclease activity. Biochim Biophys Acta *17*, 141-142.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (2003). Current Protocols in Molecular Biology (New York: John Wiley & Sons, Inc.).

Bakkes, P. J., Faber, B. W., van Heerikhuizen, H., and van der Vies, S. M. (2005). The T4-encoded cochaperonin, gp31, has unique properties that explain its requirement for the folding of the T4 major capsid protein. Proc Natl Acad Sci U S A *102*, 8144-8149.

Baldwin, R. L. (1989). How does protein folding get started? Trends Biochem Sci *14*, 291-294.

Baldwin, R. L. (1996). Why is protein folding so fast? Proc Natl Acad Sci U S A *93*, 2627-2628.

Baldwin, R. L., and Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci *24*, 77-83.

Barraclough, R., and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. Biochim Biophys Acta *608*, 18-31.

Barral, J. M., Broadley, S. A., Schaffar, G., and Hartl, F. U. (2004). Roles of molecular chaperones in protein misfolding diseases. Semin Cell Dev Biol *15*, 17-29.

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

Beissinger, M., Rutkat, K., and Buchner, J. (1999). Catalysis, commitment and encapsulation during GroE-mediated folding. J Mol Biol *289*, 1075-1092.

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

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature *371*, 578-586.

Braig, K., Simon, M., Furuya, F., Hainfeld, J. F., and Horwich, A. L. (1993). A polypeptide bound by the chaperonin groEL is localized within a central cavity. Proc Natl Acad Sci U S A *90*, 3978-3982.

Brendel, V., Brocchieri, L., Sandler, S. J., Clark, A. J., and Karlin, S. (1997). Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. J Mol Evol *44*, 528-541.

Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U., and Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. Cell *107*, 223-233.

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

Brockwell, D. J., Smith, D. A., and Radford, S. E. (2000). Protein folding mechanisms: new methods and emerging ideas. Curr Opin Struct Biol *10*, 16-25.

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

Bukau, B., and Horwich, A. L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell *92*, 351-366.

Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. Cell *125*, 443-451.

Castanie, M. P., Berges, H., Oreglia, J., Prere, M. F., and Fayet, O. (1997). A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in Escherichia coli. Anal Biochem *254*, 150-152.

Chang, H. C., Kaiser, C. M., Hartl, F. U., and Barral, J. M. (2005). De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. J Mol Biol *353*, 397-409.

Chen, J., Walter, S., Horwich, A. L., and Smith, D. L. (2001). Folding of malate dehydrogenase inside the GroEL-GroES cavity. Nat Struct Biol *8*, 721-728.

Chun, S. Y., Strobel, S., Bassford, P., Jr., and Randall, L. L. (1993). Folding of maltose-binding protein. Evidence for the identity of the rate-determining step in vivo and in vitro. J Biol Chem *268*, 20855-20862.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci *25*, 331-339.

Ditzel, L., Lowe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R., and Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. Cell *93*, 125-138.

Dobson, C. M. (2003). Protein folding and disease: a view from the first Horizon Symposium. Nat Rev Drug Discov *2*, 154-160.

Dobson, C. M., Evans, P. A., and Radford, S. E. (1994). Understanding how proteins fold: the lysozyme story so far. Trends Biochem Sci *19*, 31-37.

Dobson, C. M., and Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol *9*, 92-101.

Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988). High efficiency transformation of E. coli by high voltage electroporation. Nucleic Acids Res *16*, 6127-6145.

Dragovic, Z., Broadley, S. A., Shomura, Y., Bracher, A., and Hartl, F. U. (2006). Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s. Embo J *25*, 2519-2528.

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

Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000). Fast kinetics and mechanisms in protein folding. Annu Rev Biophys Biomol Struct *29*, 327-359.

Eggers, D. K., and Valentine, J. S. (2001). Crowding and hydration effects on protein conformation: a study with sol-gel encapsulated proteins. J Mol Biol *314*, 911-922.

Ellis, R. J. (1996). Revisiting the Anfinsen cage. Fold Des *1*, R9-15.

Ellis, R. J. (2001a). Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol *11*, 114-119.

Ellis, R. J. (2001b). Molecular chaperones: inside and outside the Anfinsen cage. Curr Biol *11*, R1038-1040.

Ellis, R. J., and Hartl, F. U. (1996). Protein folding in the cell: competing models of chaperonin function. Faseb J *10*, 20-26.

Ellis, R. J., and Hartl, F. U. (1999). Principles of protein folding in the cellular environment. Curr Opin Struct Biol *9*, 102-110.

Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. Annu Rev Biophys Biomol Struct *29*, 213-238.

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

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. J Bacteriol *171*, 1379-1385.

Fenton, W. A., Kashi, Y., Furtak, K., and Horwich, A. L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. Nature *371*, 614-619.

Fersht, A. R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol *224*, 771-782.

Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. Nature *353*, 726-730.

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

Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F. U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. Nature *370*, 111-117.

Ganesh, C., Zaidi, F. N., Udgaonkar, J. B., and Varadarajan, R. (2001). Reversible formation of on-pathway macroscopic aggregates during the folding of maltose binding protein. Protein Sci *10*, 1635-1644.

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

Gautschi, M., Lilie, H., Funfschilling, U., Mun, A., Ross, S., Lithgow, T., Rucknagel, P., and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. Proc Natl Acad Sci U S A *98*, 3762-3767.

Gautschi, M., Mun, A., Ross, S., and Rospert, S. (2002). A functional chaperone triad on the yeast ribosome. Proc Natl Acad Sci U S A *99*, 4209-4214.

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

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. J Mol Biol *76*, 45-60.

Gervasoni, P., Gehrig, P., and Pluckthun, A. (1998). Two conformational states of beta-lactamase bound to GroEL: a biophysical characterization. J Mol Biol *275*, 663-675.

Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem *182*, 319-326.

Goldbeck, R. A., Kim-Shapiro, D. B., and Kliger, D. S. (1997). Fast natural and magnetic circular dichroism spectroscopy. Annu Rev Phys Chem *48*, 453-479.

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

Gross, M., Robinson, C. V., Mayhew, M., Hartl, F. U., and Radford, S. E. (1996). Significant hydrogen exchange protection in GroEL-bound DHFR is maintained during iterative rounds of substrate cycling. Protein Sci *5*, 2506-2513.

Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol *177*, 4121-4130.

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

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. Nature *381*, 571-579.

Hartl, F. U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852-1858.

Hartman, D. J., Surin, B. P., Dixon, N. E., Hoogenraad, N. J., and Hoj, P. B. (1993). Substoichiometric amounts of the molecular chaperones GroEL and GroES prevent thermal denaturation and aggregation of mammalian mitochondrial malate dehydrogenase in vitro. Proc Natl Acad Sci U S A *90*, 2276-2280.

Hayer-Hartl, M. K., Martin, J., and Hartl, F. U. (1995). Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. Science *269*, 836-841.

Hayer-Hartl, M. K., Weber, F., and Hartl, F. U. (1996). Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis. Embo J *15*, 6111-6121.

Hesterkamp, T., Hauser, S., Lutcke, H., and Bukau, B. (1996). Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proc Natl Acad Sci U S A *93*, 4437-4441.

Hohfeld, J., and Jentsch, S. (1997). GrpE-like regulation of the Hsc70 chaperone by the anti-apoptotic protein BAG-1. EMBO J *16*, 6209-6216.

Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L., and Wuthrich, K. (2005). Direct NMR observation of a substrate protein bound to the chaperonin GroEL. Proc Natl Acad Sci U S A *102*, 12748-12753.

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

Hsu, I. C., Metcalf, R. A., Sun, T., Welsh, J. A., Wang, N. J., and Harris, C. C. (1991). Mutational hotspot in the p53 gene in human hepatocellular carcinomas. Nature *350*, 427-428.

Huang, P., Gautschi, M., Walter, W., Rospert, S., and Craig, E. A. (2005). The Hsp70 Ssz1 modulates the function of the ribosome-associated J-protein Zuo1. Nat Struct Mol Biol *12*, 497-504.

Hundley, H., Eisenman, H., Walter, W., Evans, T., Hotokezaka, Y., Wiedmann, M., and Craig, E. (2002). The in vivo function of the ribosome-associated Hsp70, Ssz1, does not require its putative peptide-binding domain. Proc Natl Acad Sci U S A *99*, 4203-4208.

Hundley, H. A., Walter, W., Bairstow, S., and Craig, E. A. (2005). Human Mpp11 J protein: ribosome-tethered molecular chaperones are ubiquitous. Science *308*, 1032-1034.

Hunt, J. F., van der Vies, S. M., Henry, L., and Deisenhofer, J. (1997). Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. Cell *90*, 361-371.

Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L., and Deisenhofer, J. (1996). The crystal structure of the GroES co-chaperonin at 2.8 A resolution. Nature *379*, 37-45.

Jaenicke, R. (1991). Protein stability and protein folding. Ciba Found Symp *161*, 206-216; discussion 217-221.

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

Jiang, J., Prasad, K., Lafer, E. M., and Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. Mol Cell *20*, 513-524.

Johnson, J. L., and Craig, E. A. (2001). An essential role for the substrate-binding region of hsp40s in Saccharomyces cerevisiae. Journal of Cell Biology *152*, 851-856.

Kabani, M., Beckerich, J. M., and Brodsky, J. L. (2002a). Nucleotide exchange factor for the yeast Hsp70 molecular chaperone Ssa1p. Mol Cell Biol *22*, 4677-4689.

Kabani, M., McLellan, C., Raynes, D. A., Guerriero, V., and Brodsky, J. L. (2002b). HspBP1, a homologue of the yeast Fes1 and Sls1 proteins, is an Hsc70 nucleotide exchange factor. FEBS Lett *531*, 339-342.

Karlin, S., and Brocchieri, L. (1998). Heat shock protein 70 family: multiple sequence comparisons, function, and evolution. J Mol Evol *47*, 565-577.

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

Kim, P. S., and Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem *51*, 459-489.

Klumpp, M., Baumeister, W., and Essen, L. O. (1997). Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. Cell *91*, 263-270.

Kramer, G., Patzelt, H., Rauch, T., Kurz, T. A., Vorderwulbecke, S., Bukau, B., and Deuerling, E. (2004). Trigger factor peptidyl-prolyl cis/trans isomerase activity is not essential for the folding of cytosolic proteins in Escherichia coli. J Biol Chem *279*, 14165-14170. Epub 12004 Jan 14116.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., and Gierasch, L. M. (1993). Characterization of a functionally important mobile domain of GroES. Nature *364*, 255-258.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F. U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. Embo J *11*, 4757-4765.

Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., and Candia, O. A. (1979). An improved assay for nanomole amounts of inorganic phosphate. Anal Biochem *100*, 95-97.

Lehninger, A. L., Nelson, D. L., and Cox, M. M. (2000). Lehninger Principles of Biochemistry, W.H. Freeman & Company.

Leroux, M. R., Fandrich, M., Klunker, D., Siegers, K., Lupas, A. N., Brown, J. R., Schiebel, E., Dobson, C. M., and Hartl, F. U. (1999). MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. Embo J *18*, 6730-6743.

Leroux, M. R., and Hartl, F. U. (2000). Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. Curr Biol *10*, R260-264.

Levinthal, C. (1968). Are there pathways for protein folding? J Chim Phys *65*, 44-45.

Lin, Z., and Rye, H. S. (2004). Expansion and compression of a protein folding intermediate by GroEL. Mol Cell *16*, 23-34.

Llorca, O., McCormack, E. A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. L., Willison, K. R., Fernandez, J. J., and Valpuesta, J. M. (1999). Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. Nature *402*, 693-696.

Lopez, B. P., Pfund, C., and Craig, E. A. (1998). The biochemical properties of the ATPase activity of a 70-kDa heat shock protein (Hsp70) are governed by the C-terminal domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 15253-15258.

Maisnier-Patin, S., Roth, J. R., Fredriksson, A., Nystrom, T., Berg, O. G., and Andersson, D. I. (2005). Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet *37*, 1376-1379.

Martin, J., and Hartl, F. U. (1993). Protein folding in the cell: molecular chaperones pave the way. Structure *1*, 161-164.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F. U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. Nature *352*, 36-42.

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

Martin-Benito, J., Boskovic, J., Gomez-Puertas, P., Carrascosa, J. L., Simons, C. T., Lewis, S. A., Bartolini, F., Cowan, N. J., and Valpuesta, J. M. (2002). Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. Embo J *21*, 6377-6386.

Marvin, J. S., Corcoran, E. E., Hattangadi, N. A., Zhang, J. V., Gere, S. A., and Hellinga, H. W. (1997). The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors. Proc Natl Acad Sci U S A *94*, 4366-4371.

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

Mayhew, M., da Silva, A. C., Martin, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1996). Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature *379*, 420-426.

McCallum, C. D., Do, H., Johnson, A. E., and Frydman, J. (2000). The interaction of the chaperonin tailless complex polypeptide 1 (TCP1) ring complex (TRiC) with ribosome-bound nascent chains examined using photo-cross-linking. J Cell Biol *149*, 591-602.

McLennan, N. F., Girshovich, A. S., Lissin, N. M., Charters, Y., and Masters, M. (1993). The strongly conserved carboxyl-terminus glycine-methionine motif of the Escherichia coli GroEL chaperonin is dispensable. Mol Microbiol *7*, 49-58.

Michimoto, T., Aoki, T., Toh-e, A., and Kikuchi, Y. (2000). Yeast Pdr13p and Zuo1p molecular chaperones are new functional Hsp70 and Hsp40 partners. Gene *257*, 131-137.

Minton, A. P. (2000). Protein folding: Thickening the broth. Curr Biol *10*, R97-99.

Nagata, H., Hansen, W. J., Freeman, B., and Welch, W. J. (1998). Mammalian cytosolic DnaJ homologues affect the hsp70 chaperone-substrate reaction cycle, but do not interact directly with nascent or newly synthesized proteins. Biochemistry *37*, 6924-6938.

Ojha, A., Anand, M., Bhatt, A., Kremer, L., Jacobs, W. R., Jr., and Hatfull, G. F. (2005). GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell *123*, 861-873.

Okochi, M., Nomura, T., Zako, T., Arakawa, T., Iizuka, R., Ueda, H., Funatsu, T., Leroux, M., and Yohda, M. (2004). Kinetics and binding sites for interaction of the prefoldin with a group II chaperonin: contiguous non-native substrate and chaperonin binding sites in the archaeal prefoldin. J Biol Chem *279*, 31788-31795.

Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F. U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. Nature *341*, 125-130.

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

Pauling, L., and Corey, R. B. (1951b). The pleated sheet, a new layer configuration of polypeptide chains. Proc Natl Acad Sci U S A *37*, 251-256.

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

Ramachandran, G. N., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. Adv Protein Chem *23*, 283-438.

Raviol, H., Sadlish, H., Rodriguez, F., Mayer, M. P., and Bukau, B. (2006). Chaperone network in the yeast cytosol: Hsp110 is revealed as an Hsp70 nucleotide exchange factor. Embo J *25*, 2510-2518.

Rippmann, F., Taylor, W. R., Rothbard, J. B., and Green, N. M. (1991). A hypothetical model for the peptide binding domain of hsp70 based on the peptide binding domain of HLA. Embo J *10*, 1053-1059.

Roder, H., and Shastry, M. R. (1999). Methods for exploring early events in protein folding. Curr Opin Struct Biol *9*, 620-626.

Roseman, A. M., Chen, S., White, H., Braig, K., and Saibil, H. R. (1996). The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. Cell *87*, 241-251.

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

Rudiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997b). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. Embo J *16*, 1501-1507.

Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. Nature *396*, 336-342.

Rye, H. S., Burston, S. G., Fenton, W. A., Beechem, J. M., Xu, Z., Sigler, P. B., and Horwich, A. L. (1997). Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. Nature *388*, 792-798.

Saibil, H., Dong, Z., Wood, S., and auf der Mauer, A. (1991). Binding of chaperonins. Nature *353*, 25-26.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Second Edition.Cold Spring Harbor Press, Cold Spring Harbor, NY.

Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solution. C R Trav Lab Carlsberg [Chim] *29*, 230-259.

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

Schultz, C. P. (2000). Illuminating folding intermediates. Nat Struct Biol *7*, 7-10.

Sheppard, C. A., Trimmer, E. E., and Matthews, R. G. (1999). Purification and properties of NADH-dependent 5, 10-methylenetetrahydrofolate reductase (MetF) from Escherichia coli. J Bacteriol *181*, 718-725.

Shi, X., Parthun, M. R., and Jaehning, J. A. (1995). The yeast EGD2 gene encodes a homologue of the alpha NAC subunit of the human nascent-polypeptide-associated complex. Gene *165*, 199-202.

Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., and Baldwin, R. L. (1985). Nature of the charged-group effect on the stability of the C-peptide helix. Proc Natl Acad Sci U S A *82*, 2349-2353.

Shtilerman, M., Lorimer, G. H., and Englander, S. W. (1999). Chaperonin function: folding by forced unfolding. Science *284*, 822-825.

Siegers, K., Waldmann, T., Leroux, M. R., Grein, K., Shevchenko, A., Schiebel, E., and Hartl, F. U. (1999). Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system. Embo J *18*, 75-84.

Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U., and Moarefi, I. (2000). Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell *103*, 621-632.

Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A., and Horwich, A. L. (1998). Structure and function in GroEL-mediated protein folding. Annu Rev Biochem *67*, 581-608.

Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994). The barriers in protein folding. Nat Struct Biol *1*, 149-156.

Sparrer, H., Rutkat, K., and Buchner, J. (1997). Catalysis of protein folding by symmetric chaperone complexes. Proc Natl Acad Sci U S A *94*, 1096-1100.

Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991). The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem *266*, 5202-5219.

Stan, G., Thirumalai, D., Lorimer, G. H., and Brooks, B. R. (2003). Annealing function of GroEL: structural and bioinformatic analysis. Biophys Chem *100*, 453-467.

Steinbacher, S., and Ditzel, L. (2001). Review: nucleotide binding to the thermoplasma thermosome: implications for the functional cycle of group II chaperonins. J Struct Biol *135*, 147-156.

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

Takagi, F., Koga, N., and Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. Proc Natl Acad Sci U S A *100*, 11367-11372.

Terada, K., Kanazawa, M., Bukau, B., and Mori, M. (1997). The human DnaJ homologue dj2 facilitates mitochondrial protein import and luciferase refolding. J Cell Biol *139*, 1089-1095.

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

Thirumalai, D., Klimov, D. K., and Lorimer, G. H. (2003). Caging helps proteins fold. Proc Natl Acad Sci U S A *100*, 11195-11197.

Thirumalai, D., and Lorimer, G. H. (2001). Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct *30*, 245-269.

Thulasiraman, V., Yang, C. F., and Frydman, J. (1999). In vivo newly translated polypeptides are sequestered in a protected folding environment. Embo J *18*, 85-95.

Todd, M. J., Lorimer, G. H., and Thirumalai, D. (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. Proc Natl Acad Sci U S A *93*, 4030-4035.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. Science *265*, 659-666.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A *76*, 4350-4354.

Troullier, A., Reinstadler, D., Dupont, Y., Naumann, D., and Forge, V. (2000). Transient non-native secondary structures during the refolding of alpha-lactalbumin detected by infrared spectroscopy. Nat Struct Biol *7*, 78-86.

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

van den Berg, B., Ellis, R. J., and Dobson, C. M. (1999). Effects of macromolecular crowding on protein folding and aggregation. Embo J *18*, 6927-6933.

Vendruscolo, M., Paci, E., Karplus, M., and Dobson, C. M. (2003). Structures and relative free energies of partially folded states of proteins. Proc Natl Acad Sci U S A *100*, 14817-14821.

Viitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. Protein Sci *1*, 363-369.

Wang, J. D., Herman, C., Tipton, K. A., Gross, C. A., and Weissman, J. S. (2002). Directed evolution of substrate-optimized GroEL/S chaperonins. Cell *111*, 1027-1039.

Wang, J. D., Michelitsch, M. D., and Weissman, J. S. (1998). GroEL-GroES-mediated protein folding requires an intact central cavity. Proc Natl Acad Sci U S A *95*, 12163-12168.

Wang, J. D., and Weissman, J. S. (1999). Thinking outside the box: new insights into the mechanism of GroEL-mediated protein folding. Nat Struct Biol *6*, 597-600.

Weber, F., Keppel, F., Georgopoulos, C., Hayer-Hartl, M. K., and Hartl, F. U. (1998). The oligomeric structure of GroEL/GroES is required for biologically significant chaperonin function in protein folding. Nat Struct Biol *5*, 977-985.

Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A., and Horwich, A. L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. Cell *83*, 577-587.

Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. Cell *78*, 693-702.

Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M., and Horwich, A. L. (1996). Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell *84*, 481-490.

Wiedmann, B., Sakai, H., Davis, T. A., and Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. Nature *370*, 434-440.

Xu, W. X., Wang, J., and Wang, W. (2005). Folding behavior of chaperonin-mediated substrate protein. Proteins *61*, 777-794.

Xu, Z., Horwich, A. L., and Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature *388*, 741-750.

Yan, W., Schilke, B., Pfund, C., Walter, W., Kim, S., and Craig, E. A. (1998). Zuotin, a ribosome-associated DnaJ molecular chaperone. Embo J *17*, 4809-4817.

Young, J. C., Agashe, V. R., Siegers, K., and Hartl, F. U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol *5*, 781-791.

Zahn, R., Spitzfaden, C., Ottiger, M., Wuthrich, K., and Pluckthun, A. (1994). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. Nature *368*, 261-265.

Zhou, H. X. (2004). Protein folding and binding in confined spaces and in crowded solutions. J Mol Recognit *17*, 368-375.

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

Zimmerman, S. B., and Minton, A. P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct *22*, 27-65.

7. Appendices

7.1. Supplementary Tables

Table S1. Apparent rates of MBP refolding

MBP Protein		Tm °C	Rate of refolding e^{-3}s^{-1}				
			Spontaneous	GroEL ATP	GroEL/ES ATP	DnaK/DnaJ/GrpE ATP	SR-EL/ES ATP
(A)	WT-MBP	66.3	29.3 ± 0.8	14.90 ± 1.1	31.2 ± 1.0	26.5 ± 1.0	29.2 ± 0.9
	SM-MBP	61.5	4.0 ± 0.3	1.2 ± 0.1	12.45 ± 1.05	1.4 ± 0.2	13.4 ± 0.8
	DM-MBP	60.5	0.37 ± 0.009	0.15 ± 0.02	4.89 ± 0.26	0.06 ± 0.02	4.5 ± 0.3
(B)	WT-MBP(D95C)	--	3.85 ± 0.12	--	5.80 ± 0.11	--	--
	SM-MBP(D95C)	--	0.35 ± 0.004	--	2.60 ± 0.034	--	--
	DM-MBP(D95C)	--	0.07 ± 0.000	--	0.83 ± 0.002	--	--
(C)	WT-MBP(D95C-IANBD)	--	3.15 ± 0.11	--	5.28 ± 0.12	--	--
	SM-MBP(D95C-IANBD)	--	0.22 ± 0.007	--	1.59 ± 0.04	--	--
	DM-MBP(D95C-IANBD)	--	0.07 ± 0.002	--	0.80 ± 0.01	--	--
(D)	WT-MBP(D95C-Alexa488)	--	3.69 ± 0.11	--	4.08 ± 0.13	--	--
	SM-MBP(D95C-Alexa488)	--	0.33 ± 0.007	--	1.02 ± 0.09	--	--
	DM-MBP(D95C-Alexa488)	--	0.04 ± 0.004	--	0.40 ± 0.01	--	--

Refolding assays were performed at 25°C as described in Experiment procedures and followed by tryptophan fluorescence in (A, B, D) and IANBD fluorescence in (C). The values for Tm°C are midpoints (fitted to a two state transition model) obtained from the temperature denaturation curves for the unfolding of the protein (2.5 μM) followed by circular dichroism spectroscopy at 222 nm (Jasco J-715). (see Figure 14).

Table S2. Parameters of GroEL substrates

Substrate	MW (kDa)	Dimensions (Å)[a]	Radius of gyration (Å)[a]	Volume (Å³)[b]	PI	Net charge at pH 7.5[c]
MetF	33.1	58.9*51.4*49.0	18.3	27960	6.23	–3.90
Rhodanese	33.1	64.1*48.8*50.6	18.8	30280	7.00	–0.87
WT-MBP	41.0	73.3*58.3*45.7	21.6	37780	5.24	–8.12
SM-MBP	41.0	n.d	--	--	5.12	–9.12
DM-MBP	41.0	n.d.	--	--	5.12	–9.12
RuBisCo	50.0	93.3*60.4*44.3	25.0	45830	5.72	–10.67

[a] Moleman2: Stat, all atoms xyz align, van der Waals [G.J. Kleywegt (1999) Acta Cryst. D55, 1878-1857]

[b] VOIDOO, 1.0 Å grid including van der Waals radii [G.J. Kleywegt and T.A. Jones (1994) Acta Cryst. D50, 178-185]

[c] J. Kyte and R.F. Doolittle (1982) J. Mol. Biol. 157, 105-132

7.2. Abbreviations

ADP	adenosine 5'-diphosphate
Amp	ampicillin
AMP-PNP	Adenosine 5′-(β,γ-imido)triphosphate tetralithium salt
APS	ammonium peroxodisulfate
ATP	adenosine 5'-triphosphate
BSA	albumin bovine serum
CAM	chloramphenicol
CDTA	*trans*-1,2-diaminocyclohexane-*N,N,N',N'*-tetraacetic acid
DNA	deoxyribonucleic acid
DnaJ	bacterial Hsp40 chaperone
DnaK	bacterial Hsp70 chaperone
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
g	acceleration of gravity, 9.81 m/s^2
GdnHCl	guanidinium hydrochloride
GFP	Green fluorescent protein
GroEL	bacterial Hsp60 chaperonin
GroES	bacterial Hsp10 cochaperonin
GrpE	bacterial nucleotide exchange factor of DnaK
h	hour
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	Luria Bertani
MBP	Maltose binding protein
METF	5,10-methylenetetrahydrofolate reductase
MOPS	3-(N-morpholino)propanesulfonic acid
NAC	nascent chain-associate complex
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate

OAc	acetate
OD	optical density
PAGE	PolyAcrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank. Repository for processing and distribution of 3-D structure data of proteins and nucleic acids. http://www.rcsb.org/pdb/
PPIase	prolyl- *cis/trans* isomerase
RAC	ribosome-associate complex
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodiumdodecylsulfate
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TF	trigger factor
TRiC	Tailless complex polypeptide ring complex
Tris HCl	tris(hydroxymethyl)aminomethane hydrochloride

7.3. Publications

Journal articles

Tang, Y. C., Chang, H. C., Hayer-Hartl, M., and Hartl, F. U. (2007). SnapShot: Molecular Chaperones, Part II. Cell *128*, 412.

Chang, H. C., **Tang, Y. C**., Hayer-Hartl, M., and Hartl, F. U. (2007). SnapShot: Molecular Chaperones, Part I. Cell *128*, 212.

Tang, Y. C., Chang, H. C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U., and Hayer-Hartl, M. (2006). Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell *125*, 903-914.

Lee, Y. C*., **Tang, Y. C***., Chen, Y. H*., Wang, C. M., and Tsou, A. P. (2003). Selenite-induced survival of HuH7 hepatoma cells involves activation of focal adhesion kinase-phosphatidylinositol 3-kinase-Akt pathway and Rac1. J. Biol. Chem. *278*, 39615-39624.

*: Equal contribution

Poster

Tang, Y.C., Roeben, A., Hartl, F. U., and Hayer-Harrtl, M. 2003. Acceleration of protein folding by confinement in the GroEL-GroES cage. [Poster Presentation in EURESCO and FEBS conference "Biology of Molecular Chaperones" in Tomar/ Portugal (August 2003)].

7.4. *Curriculum vitae*

Name	Yun-Chi Tang
Birth date	11, February, 1973
Birth place	Taipei, Taiwan
Nationality	Taiwan, Republic of China

Dissertation

2002-present	Ph.D. dissertation directed by Prof. Dr. F. Ulrich Hartl in Department of Cellular Biochemistry, Max-Planck Institute of Biochemistry, Martinsried: *Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein.*

Education

1997-1999	Master of Science.
	Master thesis directed by Prof. Dr. A. P. Tsou in Institute of Biotechnology in Medicine, National Yang-Ming University, Taipei, Taiwan: *Studies on the selenium-induced survival pathway in a human hepatocellular carcinoma cell line, Huh7.*
1991-1995	Bachelor of Science.
	Department of Microbiology, Soochow University, Taipei, Taiwan.
1988-1991	School Leaving Certificate.
	Chung-San Senior High School, Taipei, Taiwan.
1985-1988	School Leaving Certificate.
	Sing-Ming Junior High School, Taipei, Taiwan.

Experience

1999- 2001	Research Assistant
1995-1997	Institute of Botany, Academia Sinica, Taipei, Taiwan.

Honors

2007	Junior research award of Max-planck institute of Biochemistry, Munich, Germany.