Max-Planck-Institut für Biochemie, Martinsried
Abteilung Zelluläre Biochemie

The *Escherichia coli* GroEL Interaction Proteome:

Identification and Classification

Michael Johannes Kerner

Danksagung

Diese Arbeit wurde in der Zeit von Dezember 1999 bis Dezember 2004 in der Abteilung Zelluläre Biochemie des Max-Planck-Instituts für Biochemie in Martinsried angefertigt.

Mein besonderer Dank gilt **Prof. Dr. F. Ulrich Hartl** für seine hervorragende Betreuung und stete Unterstützung, die Bereitstellung des spannenden Themas und die ausgezeichneten Arbeitsbedingungen in seiner Abteilung. Weiterhin möchte ich **Dr. Manajit Hayer-Hartl** für zahlreiche hilfreiche Diskussionen und Ratschläge danken.

Prof. Dr. Wolfgang Baumeister danke ich herzlich für seine Bereitschaft zur Vertretung dieser Dissertation vor der Technischen Universität München.

Ganz besonders danken möchte ich **Dr. Dean Naylor** für seine ausgezeichnete Betreuung und hervorragende Zusammenarbeit, ohne die diese Arbeit so nicht möglich gewesen wäre. Ebenfalls sehr herzlich danken möchte ich **Tobias Maier** für die erfolgreiche und sehr gute Zusammenarbeit an diesem Projekt. Beide sind mir nicht nur gute Kollegen, sondern auch sehr gute Freunde. Auch **Dr. Anna Stines** und **Hung-Chun Chang** bin ich sehr dankbar für die fruchtbare Zusammenarbeit im Rahmen des vorliegenden Projektes.

Dr. Yasushi Ishihama, **Prof. Dr. Matthias Mann**, **Morten Kirkegaard** (Syddansk Universitet, Odense, Dänemark) und **Prof. Dr. Dmitrij Frishman** (Technische Universität München und GSF Neuherberg) bin ich sehr dankbar für die äußerst erfolgreiche Kooperation an diesem Projekt und ihre stete Bereitschaft zu weiteren zusätzlichen Analysen.

Ebenfalls bedanken möchte ich mich bei **Prof. Dr. Costa Georgopoulos** (Université de Genève), **Dr. Luis Serrano** (EMBL Heidelberg), **Prof. Dr. Andrei Lupas** und **Dr. Johannes Soeding** (MPI für Entwicklungsbiologie, Tübingen) für ihre Hilfe und Unterstützung während dieser Arbeit.

Herzlich danken möchte ich auch besonders **Dr. José Manuel Barral** für seine Unterstützung, seine zahlreichen Hilfestellungen und guten Ratschläge und außerdem für unsere sehr gute Freundschaft. Weiterhin danke ich **allen Mitarbeitern** der Abteilung Zelluläre Biochemie für ihre Hilfe und die gute Atmosphäre. Ich habe hier viele gute Freunde gewonnen.

Von ganzem Herzen danke ich auch meiner Familie. Meine Eltern **Rotraud und Wolfgang Kerner** und meine Geschwister **Daniela** und **Andrea** waren mir immer eine große Unterstützung. Ganz besonders herzlich danke ich meiner Frau **Sabine Kerner**, die mir mit ihrer Hilfe und Liebe und ihrem großen Verständnis immer zur Seite steht und ohne die mir diese Arbeit sicher um ein Vielfaches schwerer gefallen wäre.

Contents

1 Summary

To become biologically active, proteins must reliably and rapidly acquire their correct three-dimensional structures. It has been firmly established that all the folding information is contained within the amino acid sequence of a protein. However, the mechanisms utilized by proteins to avoid sampling of the extraordinarily large number of possible conformations during their folding process are just beginning to be understood. In the cell, efficient folding of many newly synthesized proteins relies on the assistance of molecular chaperones, which act to prevent aggregation and promote folding. In *Escherichia coli* (*E. coli*), nascent chain-binding chaperones, such as DnaK (a member of the Hsp70 family), stabilize elongating polypeptides on ribosomes and prevent them from aggregating. Folding of the complete nascent chain in the cytosol is achieved either upon controlled release from these factors and/or following transfer to downstream chaperones, such as the GroEL/GroES chaperonin. This complex is formed by the barrel-shaped GroEL oligomer (also called Hsp60) and its GroES cofactor (Hsp10), which functions as a detachable lid. Newly synthesized proteins encapsulated by the chaperonin can attain their native structure unimpaired by aggregation, during repeated cycles of ATP-dependent binding and release.

Remarkably, in spite of our detailed knowledge of the chaperonin reaction mechanism, only little is known about the natural substrates of GroEL. Most mechanistic studies on the *E. coli* chaperonin have been performed with heterologous substrates. Our laboratory previously identified ~50 GroEL-interacting proteins (Houry *et al.*, 1999), but their dependence on GroEL during the process of protein folding remained to be determined. In an attempt to define the requirement of a polypeptide for GroEL-assisted folding, the folding properties of ten GroEL-interacting proteins were studied both *in vitro* and *in vivo*.

The development of *in vitro* refolding assays allowed categorization of GroEL interacting proteins into three distinct classes. Class I proteins are largely independent of chaperones but can improve their folding yield by general chaperone interactions. Class II proteins do not refold efficiently in the absence of chaperones, but can utilize both the DnaK and the GroEL/GroES systems. Class III substrates, on the other hand, are absolutely dependent on GroEL. DnaK can bind class III proteins and thereby prevent their aggregation, but folding is only achieved upon transfer to GroEL. These behaviors were confirmed *in vivo* by evaluating substrate folding during GroEL/GroES depletion, as well as by examining substrate solubility upon recombinant co-expression with chaperones.

In order to obtain a more global understanding of GroEL's importance *in vivo*, a set of ~250 GroEL interacting proteins in *E. coli* was then identified through a large scale proteomic approach, probably representing the nearly complete 'GroEL interactome'. Substrate classification was extended to these newly identified GroEL interactors by comparing the relative abundance of these new substrates in GroEL complexes to those of the previously established substrates, utilizing quantitative mass spectrometry. Among the newly identified GroEL substrates, 84 are predicted to be class III, 126 class II and 42 class I. This extended classification was confirmed by biochemical analysis of representative examples. Despite a generally low cellular abundance of class III proteins, ~80 % of all polypeptide chains bound

to GroEL are members of this class. Class II substrates occupy nearly 20 % of GroEL, and class I proteins use up only 1 – 2 % of the chaperonin capacity.

The set of class III proteins includes thirteen essential proteins, explaining why GroEL is uniquely essential for viability among the molecular chaperones of *E. coli* under all conditions examined. Class III preferentially contains proteins with sizes compatible with the available space inside the GroEL/GroES cavity, with an upper limit of approximately 55 – 60 kDa. Bioinformatic analysis of the GroEL interactome revealed a significant enrichment of a particular αβ fold (the TIM $(\beta\alpha)_8$ barrel) among the stringently GroEL-dependent class III substrates. This suggests that certain TIM barrel proteins preferentially populate kinetically trapped folding intermediates that depend on encapsulation by the chaperonin for acquisition of their native state.

In summary, this study has defined the GroEL interacting proteome and characterized its dependence on GroEL during the process of protein folding both *in vivo* and *in vitro*. Among the achievements presented here is the first report of authentic GroEL substrates, the characterization of a chaperone network for coordinated protein folding and the explanation why GroEL is essential to *E. coli* viability.

2 Introduction

Proteins are central to most processes of a living cell. They constitute the majority of the dry mass of a cell and not only provide the structural building blocks, but also execute nearly all cellular functions: enzymatic catalysis allows the many and diverse chemical reactions inside a cell; proteins serve as transport and storage devices; embedded in cellular membranes they form pumps and channels allowing selective passage of molecules into and out of the cell or between different compartments. Proteins also serve to deliver information among different cells, to just mention a few of the multitude of functions provided by this group of molecules.

2.1 Protein folding

2.1.1 Protein structure

In order to cover such a variety of functions, most proteins have to adopt specific and unique three-dimensional structures. To discuss the architecture of proteins, four levels of structure are generally referred to: The primary structure of a protein is the consecutive sequence of amino acids in a polypeptide chain, thus describing the covalent connections of a protein. Secondary structure is referred to as the steric relationship of amino acids of a protein, which are close to each other in the primary structure. The secondary structure thus describes local three-dimensional structure, usually restricted to only parts of a polypeptide chain. Examples are the α-helix and the β-sheet. The tertiary structure of a protein describes the arrangement of secondary structures within the whole protein chain, thus defining the overall fold and shape of the protein. In addition to the tertiary structure, multiple-subunit protein complexes feature a quaternary structure, describing the packing of the distinct polypeptide chains in the complex.

2.1.2 The protein folding problem

The folding process of a protein involves a complex molecular recognition phenomenon, dependent on many weak, non-covalent interactions among the amino acid residues within the folding polypeptide.

In vitro, this process can be experimentally followed by first unfolding a protein at, for instance, high temperature (usually >40 °C), extreme pH values or in highly concentrated solutions of chaotropic agents like guanidinium-hydrochloride (6M) or urea (8M). To start the folding process, the proteins are diluted from the denaturing condition into an environment allowing the native state to exist. The folding reaction itself can be monitored by a variety of different methods: circular dichroism, fluorescence or absorption spectroscopies, light scattering measurements, hydrogen-deuterium exchange and subsequent nuclear magnetic resonance spectroscopy or mass spectrometry, and many more. Another easily accessible method to monitor the folding state of enzymes is to follow the development of enzymatic activity upon refolding.

In pioneering experiments on protein folding, Christian Anfinsen showed in the

1960's that correct refolding of unfolded Ribonuclease A into its native and enzymatically active structure occurs spontaneously in free solution (Anfinsen, 1973). All information determining the native structure, which is the polypeptide conformation with the lowest Gibbs free energy of the full system, is fully contained in the amino acid sequence of a protein. Furthermore, Anfinsen showed that this spontaneous folding process occurs with biologically relevant rates (Schechter *et al.*, 1970) – an astonishing observation and not easily predicted.

When considering a carbon single bond, for instance, three rotation conformers with minimal energy exist. For rotation around the C2-C3 bond of butane, the *anti* conformation of the methyl groups is most favorable, and is stabilized by 18.8 kJ/mol against the most unfavorable totally eclipsed conformation. This energy barrier is easily overcome by thermal energy at room temperature, and thus the C2-C3 carbon bond of butane is nearly freely rotatable. Proteins, which frequently contain 300 or more amino acids, have an incomparably higher number of possible conformations. Disregarding the side chain residues of a polypeptide chain and just considering the backbone, the conformation of a single amino acid in a protein can be described by a pair of angles, Ψ and Φ (Figure 1). The peptide bond itself is planar, a consequence of its partial double bond character. Ψ describes the angle formed by rotation around the axis through C_α and the carboxyl carbon; Φ describes rotation around C_α and the amino group. Because of steric collisions between atoms within each amino acid,
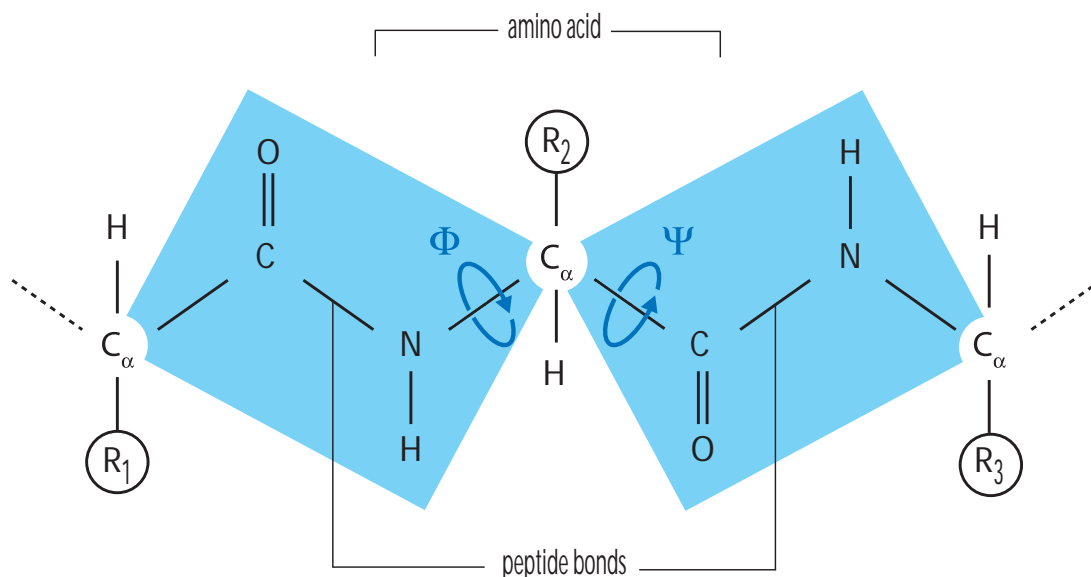
Figure 1: Steric limitations of the peptide bond angles in a protein.

Each amino acid contributes three bonds to the polypeptide backbone. The peptide bond is planar (blue shading) and does not allow rotation. The N-C_α and C_α-C bonds, however, allow rotation - their angles are called Φ and Ψ. R is used to indicate the side chain residues of the corresponding amino acid.

most pairs of Ψ and Φ angles do not actually occur. 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 naturally in folded proteins, including the two most prominent examples of secondary structure: α-helix and β-strand.

In order to assess the likelihood of correct folding of a protein, Cyrus Levinthal estimated the number of theoretically possible conformations of a polypeptide containing 150 amino acids (Levinthal, 1969). His calculations resulted in $\sim 10^{300}$ possible conformations of this theoretical protein. Even when considering only the two lowest energy regions of Ψ and Φ found by Ramachandran, there are still on the order of 2^{100} or $\sim 10^{30}$ main-chain conformations for a small protein with 100 amino acid residues. If one assumes that interconversion of main-chain conformations can occur at the maximum rate physics allows ($\sim 10^{11}$ s^{-1}), it would take about 10^{11} years to search the full amount of potential conformations (Dinner *et al.*, 2000). This so called 'Levinthal paradox' led Levinthal to the conclusion that a protein cannot sample all possible conformations during the process of folding, but folding "is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide. This suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process" (Levinthal, 1969).

2.1.3 Protein folding mechanisms

Protein folding thus seems to occur along certain pathways, thereby simplifying the folding process by splitting it up into sequential steps. Stabilized folding intermediates were proposed, defining the individual steps of such a pathway (Baldwin, 1996; Baldwin and Rose, 1999; Privalov, 1996). Folding intermediates possess partially native and stabilized structural elements, in combination with unstructured regions. A pathway folding mechanism drastically reduces the amount of possible conformations during the folding process, thus allowing effective protein folding during biologically relevant timescales.

There have been a variety of proposals for general folding pathways (Daggett and Fersht, 2003). The so-called 'nucleation-growth' mechanism (Wetlaufer, 1973) proposed tertiary structures propagating from an initially formed nucleus of stable secondary structure. However, since this model predicts the absence of folding intermediates, it rapidly lost interest. Two different models dominated the protein folding field, one being the 'framework model' (Kim and Baldwin, 1982; Kim and Baldwin, 1990; Ptitsyn and Rashin, 1975) in combination with its close relative, the 'diffusion-collision model' (Karplus and Weaver, 1976). Here, secondary structure is proposed to form initially, followed by diffusion of these stable intermediate structures until collision with others to form the final native tertiary structure. The second model was the one of 'hydrophobic collapse' (Baldwin, 1989; Kauzmann, 1959; Schellman, 1955; Tanford, 1962), where a rapid collapse of the hydrophobic polypeptide chain precedes folding, which can then proceed in a more confined environment.

More recently, some proteins have been observed to fold without intermediate structures in simple two-state reactions, for instance chymotrypsin inhibitor 2 (CI2) (Jackson and Fersht, 1991). Furthermore, proteins have been shown to form secondary and tertiary struc-

tures in parallel during a general collapse of the polypeptide chain (Otzen *et al.*, 1994), leading to formulation of the 'nucleation-condensation' mechanism (Fersht, 1997). This mechanism combines features of both the framework mechanism and the hydrophobic collapse. Long range and other native hydrophobic contacts are formed in the transition state, which is stabilized by the formation of secondary structural elements in combination with tertiary interactions. Without stabilization by tertiary interactions with the rest of the protein, secondary structure elements like α-helices or β-hairpins are often unstable and would not form when isolated.

Currently, simulations of protein folding by molecular dynamics computations, together with experimental data, are beginning to describe unfolding-folding pathways of proteins at atomic resolution (Fersht and Daggett, 2002; Mayor *et al.*, 2003). However, such simulations are still limited to very small proteins and are not yet applicable to the large majority of proteins in the cell.

2.1.4   *Folding energy landscapes*

The protein folding process is often illustrated by 'folding energy landscapes'. An example for the folding of hen lysozyme (Kiefhaber, 1995; Matagne *et al.*, 1997; Radford *et al.*, 1992) is given in Figure 2. An energy landscape, or hypersurface, is a three- or multidimensional representation of the free energy of a molecule dependent on certain variables. In the case of hen lysozyme (Figure 2), these variables are Q_α and Q_β, the number of native contacts in the α domain (which consists of mainly α-helices) and the β domain (which contains a considerable amount of β-sheets) (Dinner *et al.*, 2000). Lysozyme is known to have at least two different predominant folding pathways. Thus, it requires at least two variables for an adequate description of its folding behavior. The starting point of folding is the unfolded state, which is actually not a single conformation, but an entire ensemble of different random coil conformations of the polypeptide, each one without any native contacts formed yet.

Following the yellow trajectory, lysozyme follows a fast folding pathway and does not experience the formation of considerably stable folding intermediates. The protein follows the folding funnel along a pathway of constantly declining free energy into the native state without interruption by local minima. The red trajectory follows a slightly different track along the free energy surface in the beginning of the folding process. This pathway leads into a stable and long-lived intermediate with persistent structure in only the α but not the β domain. Further folding from this intermediate involves either a transition over a high energy barrier, or partial unfolding and change to the fast folding track. The red track thus represents the slow folding track of hen lysozyme.

2.2   **Protein folding in the cell**

In contrast to the situation *in vitro*, where many proteins have been shown to fold spontaneously, protein folding in the cell is not generally a spontaneous process. Upon synthesis at the ribosome, an elongating polypeptide chain encounters conditions that do not necessarily favor folding. The cytosolic environment of the cell is thought to promote aggregation, an adverse side-reaction competing with the process of protein folding. Moreover, the
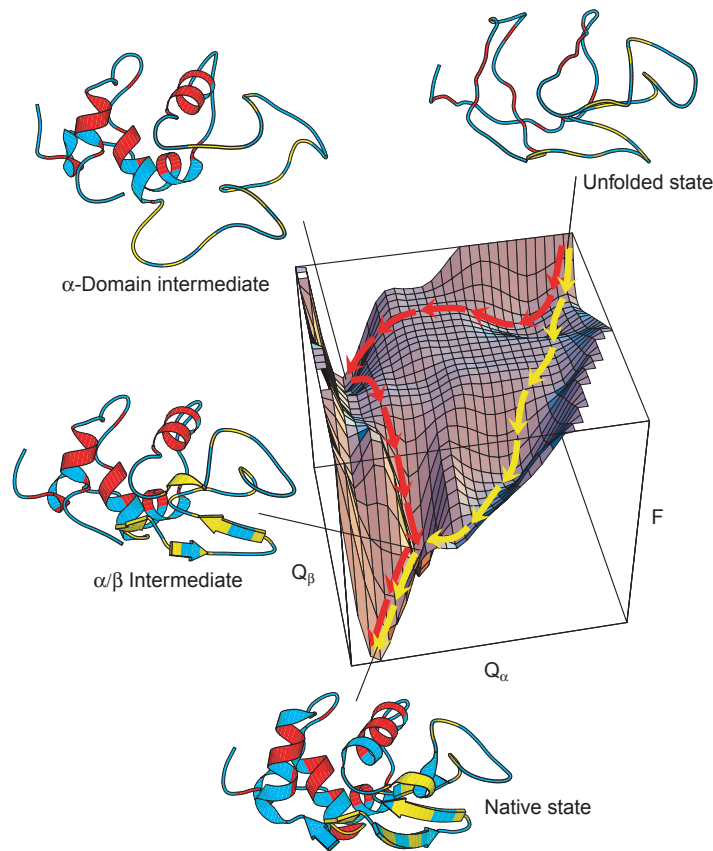
Figure 2: Folding energy landscape of hen lysozyme.

Schematic free-energy (F) surface dependent on the number of native contacts in the α and β domain of hen lysozyme (Q_α, Q_β). The yellow trajectory represents a fast track of folding, the red the red trajectory represents a slow track of folding in which a kinetically stable intermediate is populated, thus slowing down the folding process. From Dinner *et al.* (2000).

entire polypeptide chain is not available for folding all at once, since synthesis at the ribosome is a vectorial process and is slower than the rate of protein folding. Thus, incomplete polypeptide chains, which can not fold cotranslationally, have to be prevented from intra- and intermolecular misfolding and aggregation.

Evidence has become available since the late 1980's that a certain group of proteins, known as molecular chaperones, assists newly synthesized and unfolded proteins during their folding inside the cell. Molecular chaperones protect unfolded polypeptides from aggregation and misfolding and promote their folding towards the native structure. Notably, they do not usually carry specific conformational information required for successful folding of their substrates.

2.2.1   *Protein aggregation and the excluded volume effect*

Unfolded and partially folded polypeptides often expose hydrophobic regions which

are energetically unfavorable in the hydrophilic environment of aqueous solutions, such as the cytosol. Native proteins usually bury these hydrophobic regions inside their globular structure during folding. However, interaction of these hydrophobic patches with hydrophobic material from other unfolded or partially unstructured polypeptide chains leads to formation of unstructured aggregation that can heavily compete with the folding process (Figure 3). As aggregation is usually irreversible, aggregated proteins, which can not proceed down a successful folding pathway, are usually degraded.

In the cellular cytosol, aggregation is strongly enhanced by the high concentrations of proteins, nucleic acids and other macromolecules (> 500 g/l in *E. coli*). The theory of excluded volume, or macromolecular crowding, postulates that macromolecules are mutually impenetrable and sterically exclude their volume for other macromolecules in the solution (Ellis, 2001; Minton, 2001). Much of the solution volume is therefore inaccessible for a polymer in a crowded solution, leading to an effective increase in its concentration, or better thermodynamic activity, as well as an increased diffusion coefficient of the solution. The excluded volume effect predicts that small structures are favored above large structures in the presence of molecular crowding. Thus, the concept predicts that both the globular native state of a protein as well as compact aggregated structures are favored in crowded solutions. Whether aggregation or folding to the native state prevails in a crowded environment de-
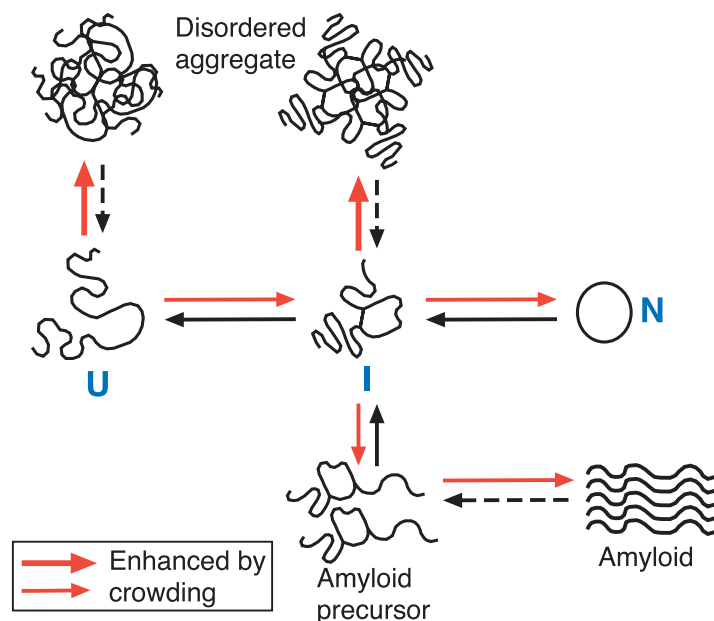
Figure 3: Aggregation of unfolded polypeptide chains is a side-reaction of protein folding.

Molecular crowding (red arrows) in the cytosol can amplify these side reactions, as well as folding to the native state. U: Unfolded polypeptide chain, I: partially structured folding intermediate, N: natively structured protein. From Hartl and Hayer-Hartl (2002).

pends on the specific protein involved. Generally, it is expected that aggregation is more enhanced for small polypeptides, because their diffusion is not as limited by the increased diffusion coefficient of a crowded solution, as it is for large polypeptides. Large polypeptides would thus rather react intramolecularly and bury their hydrophobic sites inside themselves by folding to the native state. However, the crowding agents of the complex cellular cytosol are not only steric agents but often expose hydrophobic surfaces themselves, and thus aggregation is most often favored under crowding situations of the living cell.

A special case of protein aggregation in live cells is amyloid formation, which is associated with certain so-called protein folding diseases, such as Huntington's, Parkinson's or Alzheimer's diseases, and the spongiform encephalopathies such as Creutzfeldt-Jacob's disease (Dobson, 1999). Amyloids are highly ordered, fibrillar protein aggregates with a characteristic pattern by X-ray diffraction and a typical structure by electron microscopy. Amyloid is thought to be a generic structural form of all proteins, usually favored under extreme conditions of pH or temperature. The proteins involved in the above diseases, however, usually assemble into amyloid structures under physiological conditions.

2.2.2 De novo protein folding: necessity of molecular chaperones

The problem of protein aggregation in the crowded environment of the cellular cytosol is aggravated by the fact that formation of tertiary structure, at least on the level of domains, is a cooperative process, dependent on the presence of the whole polypeptide chain. Polypeptides are generated and released into the cytosol vectorially from the ribosome and therefore offer large unstructured and hydrophobic regions during their synthesis. In order to prevent aggregation of these partly completed polypeptides, ribosome associated chaperones are necessary, which reversibly bind to aggregation-prone nascent polypeptide chains at the ribosomal exit tunnel (Hartl and Hayer-Hartl, 2002; Walter and Buchner, 2002).

Many chaperones promote folding of their substrates by repeated cycles of binding and release, often in an ATP hydrolysis dependent process. Interaction with chaperones may not only prevent intermolecular aggregation of polypeptides, but can help to diminish non-native intramolecular contacts and avoid the formation of kinetically trapped folding intermediates. Some chaperones, like the Clp (Hsp100) family, are even able to resolubilize aggregated proteins (Ben-Zvi and Goloubinoff, 2001).

Molecular chaperones generally bind to unfolded newly synthesized polypeptides and proteins denatured by stress conditions, like elevated temperature. Chaperones recognize structural elements usually not present in natively folded proteins, such as exposed hydrophobic residues and unstructured backbone regions. Many molecular chaperones are constitutively expressed and fulfill essential functions within the cell while the levels of other chaperones are greatly increased under stress conditions. Therefore molecular chaperones are also known as 'stress proteins' or 'heat shock proteins' (Hsp's).

Chaperones do not actively fold their substrate proteins, they rather create a local environment favoring productive protein folding above functionally non-productive side reactions. Molecular chaperones also do not usually catalyze folding in a classical sense. On the contrary, the folding process is often slowed down by chaperone interaction. On the other

hand, there have been recent experiments providing evidence for accelerated folding by encapsulation inside the chaperonin GroEL/GroES (see below). However, intermolecular folding catalysis with direct transient integration of tertiary structure has so far only been observed for peptide bond *cis-trans* isomerization and disulfide bond formation (Schiene and Fischer, 2000; Schiene-Fischer *et al.*, 2002) and more recently for the *E. coli* periplasmic chaperone FimC (Vetsch *et al.*, 2004). Furthermore, the folding of the protein subtilisin represents a particular case of folding catalysis: the subtilisin propeptide acts as an intramolecular chaperone for the folding of its own protease domain by serving as a folding template. This intramolecular chaperone does not form a part of the final structure of the protease domain and is removed by autolytic or exogenous proteolytic activity upon folding (Shinde *et al.*, 1997).

2.3   Molecular chaperone systems for *de novo* protein folding

2.3.1   *Overview of the substrate flux through chaperone networks in the cytosol*

The variety of molecular chaperones in a cell is thought to be organized in chaperone 'networks' (Figure 4). The interplay among the distinct chaperones in such networks is becoming increasingly more understood (Young *et al.*, 2004).

Most small proteins probably fold spontaneously upon release from the ribosome and ribosome bound factors. In addition to ribosome bound chaperones, Hsp70 proteins also bind emerging polypeptide chains as they are still associated with the ribosome, and further promote folding upon complete release of the proteins from the ribosome. Usually, however, Hsp70 proteins are not associated with the ribosome themselves.

Hsp70s are able to deliver substrates to the chaperonin system which is located further downstream and continues to mediate folding of proteins that previously failed to fold (see below). In eukaryotes, but not in bacteria, an additional factor exists (termed prefoldin or alternatively GimC) which can directly bind to nascent chains and cooperate with the chaperonin system to promote folding.

A subset of Hsp70 substrates in the eukaryotic cytosol is delivered to the Hsp90 system and both Hsp70 and Hsp90 cooperate in their folding. Several cofactors are involved in this process, which can modulate the action of these chaperones towards other processes, such as protein targeting to the mitochondrion.

2.3.2   *Ribosome-associated chaperones*

The first chaperones that interact with a nascent chain at the ribosome are trigger factor (TF) in bacteria and in the eukaryotic cytosol the nascent chain-associated complex (NAC), accompanied by the ribosome-associated complex (RAC) and Ssb, a member of the Hsp70 family, in *Saccharomyces cerevisiae* (*S. cerevisiae*). They all bind nascent polypeptides during their synthesis and are associated with the ribosome themselves.

The 48 kDa *E. coli* protein TF binds to a docking site at protein L23 of the large ribosomal subunit (Kramer *et al.*, 2002). TF is thought to scan the nascent polypeptide as it emerges from the ribosomal exit tunnel for hydrophobic regions and binds to these as they
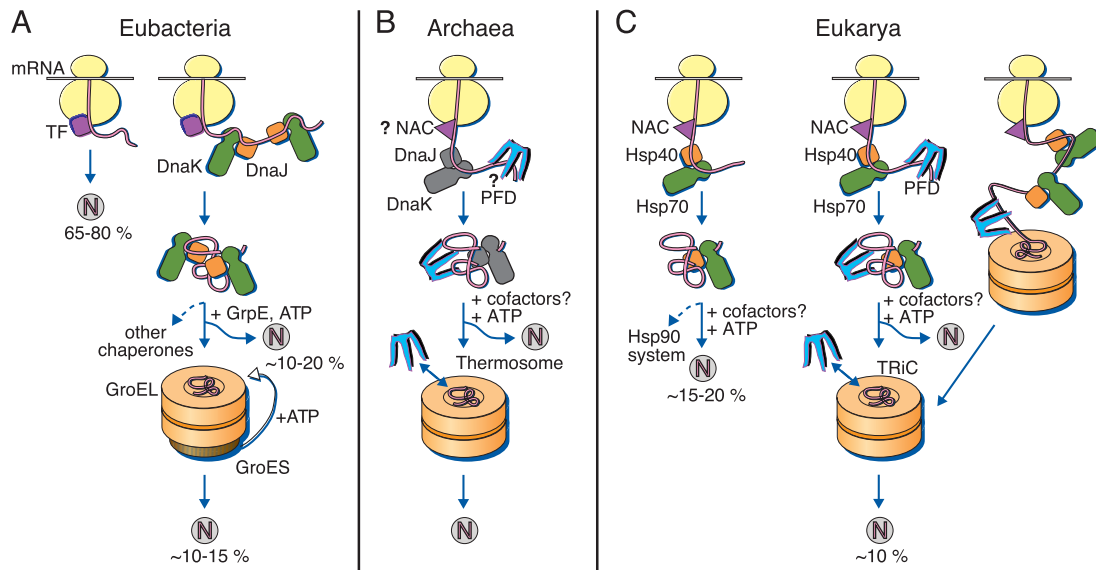
Figure 4: Model for de novo protein folding assisted by a network of molecular chaperones in the cytosol of bacteria, archaea and eukarya.

N: Natively folded protein, TF: trigger factor, NAC: nascent chain-associated complex, PFD: prefoldin. **(A)** Many proteins in the bacterial cytosol fold without further assistance upon release from the ribosome and ribosome-bound TF. DnaK assists the remainder of proteins in folding, and can transfer substrates to the chaperonin system (GroEL/GroES). **(B)** Only some archaea contain DnaK/DnaJ. Interaction of PFD with nascent chains and existence of NAC is not experimentally confirmed. **(C)** In the example of the mammalian cytosol, NAC probably interacts with nascent polypeptide chains, together with Hsp70 and Hsp40. The majority of proteins can fold upon release from these factors. A subset of Hsp70 substrates is transferred to the Hsp90 system. Furthermore, PFD interacts with nascent chains and transfers these to TRiC, the eukaryotic chaperonin. From Hartl and Hayer-Hartl (2002).

are encountered. The TF reaction is not ATPase driven, rather nascent polypeptides dissociate from TF spontaneously upon release from the ribosome (Hesterkamp *et al.*, 1996). TF exhibits peptidyl prolyl *cis-trans* isomerase activity, but the biological relevance of this activity for protein folding is still unclear, since it is not essential for the function of TF *in vivo* (Genevaux *et al.*, 2004; Kramer *et al.*, 2004).

TF and the DnaK system (the bacterial Hsp70 system; for a more detailed description of DnaK/Hsp70, see below) have overlapping functions in stabilization of nascent polypeptides and maintaining them in a competent state for subsequent folding (Deuerling *et al.*, 1999; Teter *et al.*, 1999). DnaK does not bind directly to the ribosome, but is known to bind emerging polypeptides, while these are still being synthesized (Teter *et al.*, 1999). *E. coli* cells lacking the *tig* gene encoding TF recruit DnaK as effective replacement for TF in chaperoning nascent polypeptides at the ribosome, as shown by an increased number of DnaK molecules associated with ribosome bound nascent chains in such a strain. *E. coli* cells can also tolerate the loss of DnaK without detectable growth defects, as long as they contain a

suppressor mutation in the heat shock transcription factor σ^{32}, preventing the otherwise un-controlled overexpression of further heat shock proteins in the absence of DnaK (Bukau and Walker, 1990). However, a combined deletion of TF and DnaK results in cells with severe protein misfolding and aggregation effects, leading to loss of viability at temperatures >30 °C (Deuerling *et al.*, 1999; Genevaux *et al.*, 2004; Teter *et al.*, 1999). It was recently shown that suppressor mutations can reverse this effect to some degree, when the cells are grown at low temperature and are slowly adapted to higher temperatures. Furthermore, GroEL/GroES overproduction can also counteract this phenomenon (Genevaux *et al.*, 2004; Vorderwülbecke *et al.*, 2004).

For folding of certain multidomain proteins in the bacterial cytosol, DnaK and TF have recently been observed to cooperate in preventing co-translational folding of single domains and rather allow post-translational folding of the completed polypeptide chain upon release from the ribosome. This is in contrast to the situation in the eukaryotic cytosol, where rapid co-translational folding is supported by the Hsp70/Hsp40 system (Agashe *et al.*, 2004).

The eukaryotic cytosol lacks a TF homolog, but NAC, found in *S. cerevisiae*, mammals and probably most other eukaryotes, is also a ribosome bound complex that interacts with substrate polypeptides emerging from the ribosomal exit tunnel (Wiedmann *et al.*, 1994). A chaperone activity, however, has not been directly shown so far. Notably, *S. cerevisiae* features another ribosome bound chaperone, RAC (ribosome associated complex), which is constituted from Ssz (a Hsp70 homolog) and zuotin (a Hsp40 homolog, see below) (Gautschi *et al.*, 2001).

2.3.3   *The Hsp70 system*

The Hsp70 system constitutes a central part of molecular chaperone networks. Homologs of Hsp70 are present in all bacteria, and also in the cytosol and in most organelles of eukarya. Furthermore, Hsp70 proteins are found in some archaea. They are involved in a variety of cellular processes, ranging from protein folding to protein disaggregation and degradation. The common mode of Hsp70 action appears to be binding to short, extended hydrophobic peptide sequences in the substrate proteins. This process inhibits further local folding of partially folded substrates for the time they are bound, and in addition prevents their aggregation. Native proteins do not usually expose such hydrophobic fragments and are thus not recognized by Hsp70.

Most Hsp70s cooperate with partner proteins during their functional cycle: the Hsp40 proteins and the nucleotide exchange factors. Additional interaction partners, especially in eukaryotes, link Hsp70 to other chaperone systems, in particular to Hsp90 (Walter and Buchner, 2002; Young *et al.*, 2003). Hsp70 also cooperates with the Hsp104/Clp family in resolubilization of protein aggregates. The exact role of Hsp70 in this mechanism, however, is still unclear (Mogk and Bukau, 2004).

Hsp70 chaperones are monomeric, ~70 kDa proteins comprised of two functional parts, a ~45 kDa amino-terminal ATPase domain and a ~25 kDa carboxy-terminal polypeptide-binding domain. Crystal structures of the individual domains of DnaK (the bacterial Hsp70) but not of the entire molecule, have been determined (Harrison *et al.*, 1997; Zhu *et*

al., 1996). Communication between the two domains in the functional cycle results in efficient binding and release of substrate polypeptides.

The reaction cycle, or 'ATPase cycle', of Hsp70 proteins is best understood for the DnaK system (Figure 5) (Bukau and Horwich, 1998; Naylor and Hartl, 2001). It involves large structural rearrangements of the DnaK molecule, driven by binding and release of ATP. Interplay with the Hsp40 co-chaperone DnaJ and the nucleotide exchange factor GrpE further controls the cycle. DnaK exists in mainly two structural states. When ATP is bound, the substrate binding site is in its open conformation, affinity for substrate polypeptides is low and characterized by fast association and dissociation rates. In the ADP bound state, the substrate binding pocket is closed, affinity for bound substrates is high and association/dissociation rates are low. Substrate binding thus occurs in the ATP bound state.

ATP hydrolysis, with concurrent structural rearrangements in the peptide binding domain, is the rate limiting step in the reaction cycle of Hsp70 proteins. For DnaK, this step is dynamically controlled by DnaJ (in particular by its J-domain) and GrpE, as they are able to (also stimulated by presence of substrate protein) dramatically increase the ATP hydrolysis rate of DnaK. Thus, upon substrate binding of DnaK in the ATP bound state, interaction with DnaJ triggers ATP hydrolysis and closure of the substrate binding pocket of DnaK with bound substrate. Furthermore, DnaJ is able to bind unfolded polypeptide proteins itself and can deliver them to DnaK.

After a suitable binding time of substrate proteins by DnaK, substrate dissociation is induced by release of ADP and binding of new ATP. This process requires interaction with the nucleotide exchange factor GrpE, which greatly accelerates release of ADP even if present only in catalytic amounts. Opening of the substrate binding pocket of DnaK, induced by ATP binding, then allows release and exchange of substrate polypeptides.

Eukaryotic Hsp70 proteins (*e.g.* Hsc70 in mammals, Ssa in *S. cerevisiae*) follow a similar reaction cycle. The Hsp40 homologs of DnaJ (*e.g.* Hdj1 and Hdj2 in mammals, Ydj1 and Sis1 in *S. cerevisiae*) are also able to bind unfolded polypeptides themselves and stimulate the ATP hydrolysis of their partner Hsp70 proteins (Young *et al.*, 2004). GrpE orthologs do not exist in the eukaryotic cytosol, nucleotide exchange of cytosolic eukaryotic Hsp70s is achieved by additional factors, such as BAG1 and HspBP1 in mammals or Fes1 in *S. cerevisiae*, although these are structurally unrelated to GrpE (Shomura *et al.*, 2005). The ADP bound state of Hsc70 can be stabilized by HIP (Hsp70 Interacting Protein), which counteracts BAG1 activity.

2.3.4   The chaperonins

The chaperonins constitute a highly conserved class of essential gene products (Fayet *et al.*, 1989) present in all three domains of life (Eukarya, Eubacteria and Archaea) and in almost every organism sequenced to date. The mitochondrial chaperonin Hsp60 of *Neurospora crassa* was the first chaperone identified to promote polypeptide folding (Ostermann *et al.*, 1989). Chaperonins are large, multimeric, nearly 1 million Da complexes with a double-ring structure, enclosing two central cavities. They are divided into two
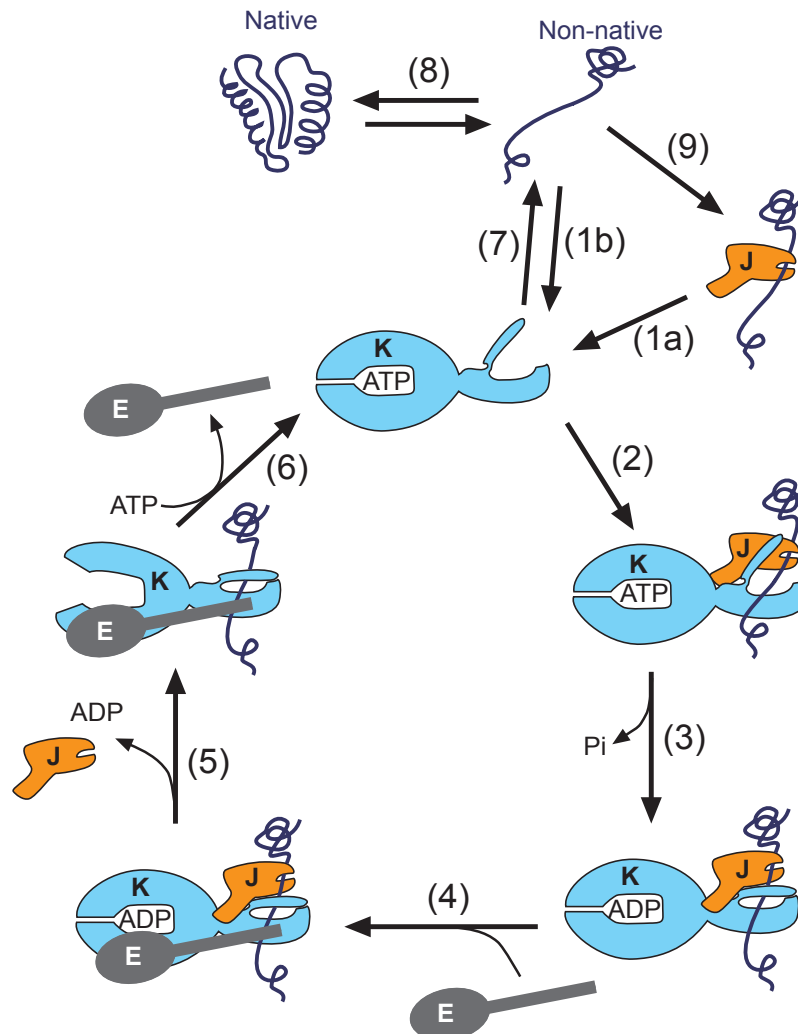
Figure 5: The reaction mechanism of DnaK in cooperation with DnaJ and GrpE.

Non-native substrate polypeptides associate with either DnaJ (J) **(1a)** or DnaK (K) in the ATP bound open state **(1b)**. DnaJ and substrate protein **(2)** stimulate ATP hydrolysis by DnaK **(3)**, leading to closure of the substrate binding pocket of DnaK. GrpE (E) interaction **(4)** is necessary for efficient release of ADP from the complex **(5)**, and subsequent ATP binding **(6)** results in opening of the substrate binding channel and exchange of substrate polypeptides **(7)**. The released substrate can either fold towards the native state **(8)** or re-bind to DnaJ **(9)** or DnaK **(1b)**. Adapted from Naylor and Hartl (2001).

groups, which are related in topology, but do not share close sequence similarity. Group I chaperonins occur in the bacterial cytosol (GroEL) and in eukaryotic organelles of bacterial endosymbiotic origin (Cpn60 in chloroplasts, and Hsp60 or Cpn60 in mitochondria). Group I chaperonins function in cooperation with cofactors of the Hsp10 family (GroES in bacteria, Hsp10 or Cpn10 in mitochondria and chloroplasts). Group II chaperonins occur in archaea (thermosome) and the eukaryotic cytosol (TRiC, the TCP1 Ring Complex, also called CCT for Chaperonin Containing TCP1). They do not utilize a Hsp10-like cofactor (see below), instead they contain an in-built lid-like appendage that is suggested to provide a similar function.

Unlike Hsp70 proteins, which promote polypeptide folding by binding and thereby preventing aggregation, chaperonins mediate protein folding by encapsulation in a central cavity. Sequestation allows folding to proceed unimpaired by interaction with other macromolecules in the cellular environment that would usually lead to misfolding and aggregation.

In *E. coli*, unfolded polypeptides have been shown to require both the DnaK and the GroEL systems (Ewalt *et al.*, 1997; Horwich *et al.*, 1993; Langer *et al.*, 1992b; Teter *et al.*, 1999). The chaperonin mechanism of substrate encapsulation makes direct folding of polypeptides at the ribosome unlikely, and GroEL has indeed not yet been found associated to ribosome attached nascent chains. Notably, a cooperation and subsequent interaction with substrate protein by first DnaK and then GroEL has already been shown (Langer *et al.*, 1992a). The group II chaperonin TRiC, on the other hand, can directly associate with nascent chains at the ribosome (Frydman *et al.*, 1994). Whether TRiC can promote folding cotranslationally or prefers an encapsulation mechanism upon completion of polypeptide synthesis is not yet fully resolved.

The substrate set of chaperonins is narrower than that of the upstream chaperones which bind to almost all nascent polypeptide chains, particularly in case of the ribosome associated chaperones. About 10 – 15 % of all synthesized proteins in the bacterial cytosol transit through GroEL (Ewalt *et al.*, 1997; Houry *et al.*, 1999). Among an identified subset of GroEL interacting polypeptides, a structural preference for multiple αβ motifs has been observed (Houry *et al.*, 1999), but GroEL does not exclusively fold such proteins. *In vitro*, GroEL binds to almost any unfolded polypeptide (Viitanen *et al.*, 1992) *via* hydrophobic interactions. However, in the cell GroEL is located at a relatively late stage of the chaperone network. Many proteins have already folded or are in late folding states when they would reach the chaperonin, and do not expose hydrophobic patches anymore. GroEL thus mostly binds to the remaining proteins that failed to complete their folding with the assistance of upstream chaperones.

The set of identified substrates of group II chaperonins is much smaller when compared to that of GroEL. TRiC was initially found to be responsible for folding of actin and tubulin (Sternlicht *et al.*, 1993; Yaffe *et al.*, 1992). These two proteins constitute its major substrates, although more substrates have been found subsequently. More recently, TRiC was shown to participate in the folding of several WD40-repeat proteins, which constitutes the first instance that a common structural class for group II chaperonin substrates has been identified (Siegers *et al.*, 2003).

2.4   The *E. coli* chaperonin GroEL/GroES

2.4.1   *Structure and function*

GroEL and GroES constitute one of the most intensively studied chaperone systems to date (reviewed in Bukau and Horwich, 1998; Fenton and Horwich, 2003; Hartl, 1996; Naylor and Hartl, 2001; Sigler *et al.*, 1998; and Walter and Buchner, 2002), which has allowed a detailed understanding of the structural principles and mechanics of this molecular machine. Crystallographic (Boisvert *et al.*, 1996; Braig *et al.*, 1994; Xu *et al.*, 1997) and electron microscopic studies (Ranson *et al.*, 2001; Roseman *et al.*, 1996; Saibil *et al.*, 1991) showed that GroEL is a tetradecamer of identical 57 kDa subunits, built from two heptameric rings stacked back to back. The rings of GroEL form two separated cavities (Figure 6).

Each GroEL subunit is comprised of three domains (Figure 7): (i) an equatorial domain, which binds ATP or ADP and forms the inter-subunit contacts between the two GroEL rings, thereby holding them together; (ii) an apical domain, which exposes hydrophobic surfaces at the entrance of the cavity promoting substrate binding, as well as subsequent binding
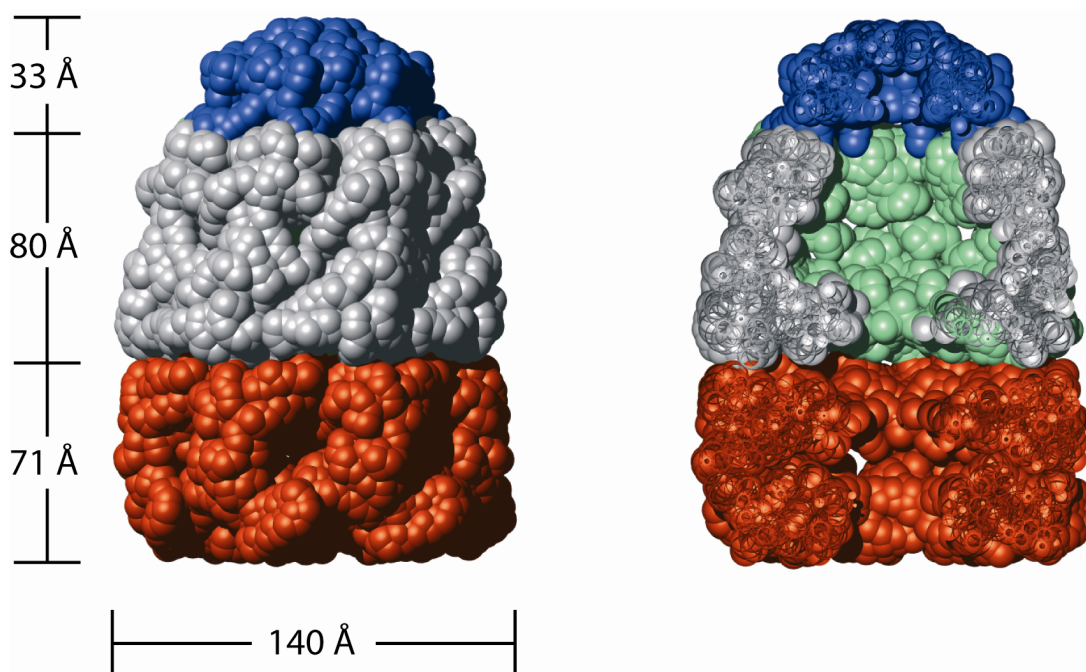
Figure 6: Asymmetric structure of the GroEL/GroES complex.

Space-filling models of GroEL/GroES (PDB 1AON; Xu *et al.*, 1997) with 6 Å Van der Waals spheres around C_α atoms. The two rings of GroEL are red and grey, GroES is shown in blue. Outside view (left) and inside view (right) of GroEL/GroES, generated by slicing the structure with a vertical plane through the heptameric symmetry axis. To indicate the interior of the cis-cavity, subunits located at the back of GroEL are colored in green. Figure created with MOLMOL (Koradi et al., 1996) and rendered with POV-Ray v3.6.
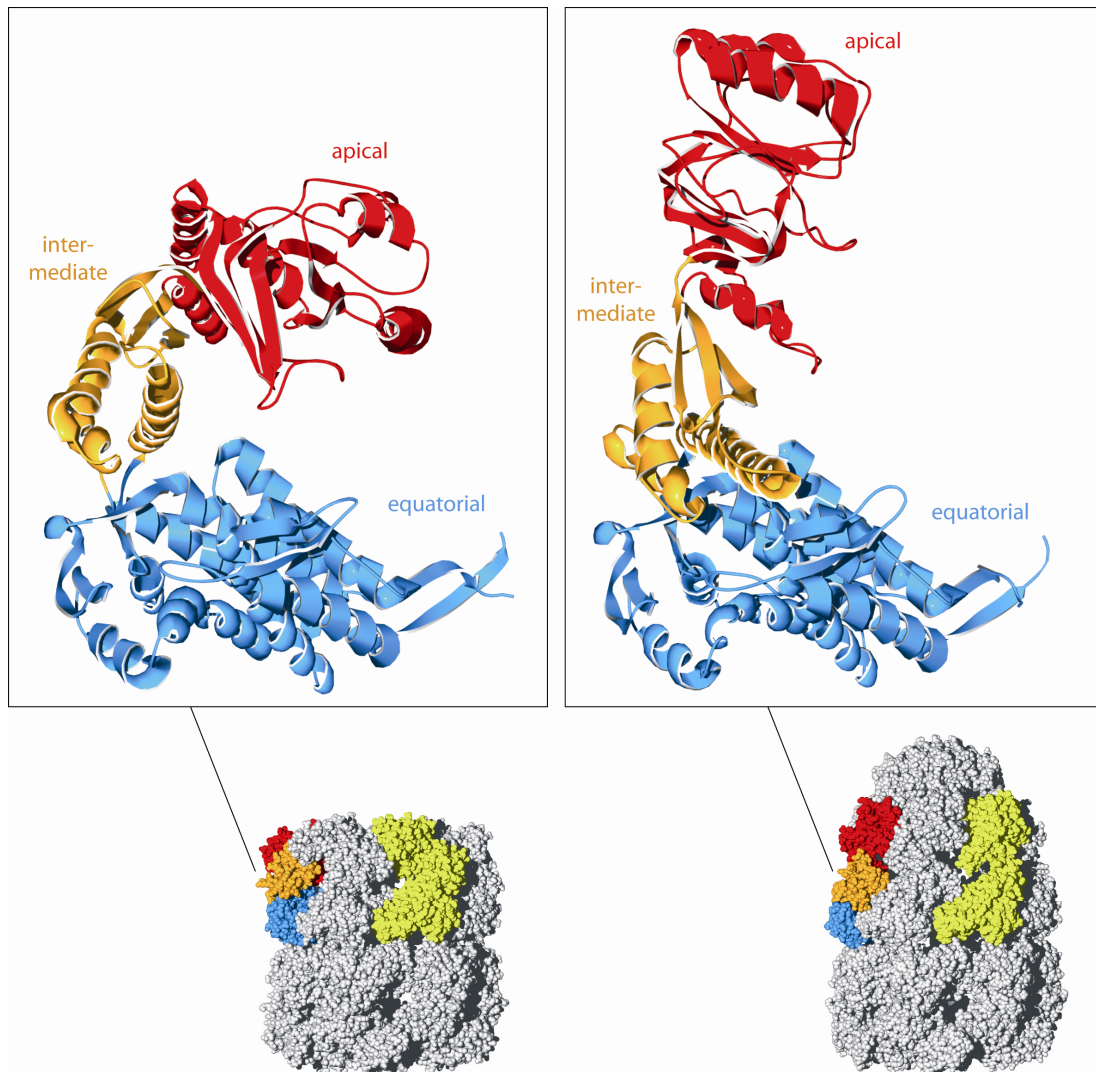
Figure 7: Structural rearrangements in GroEL upon binding of GroES.

The top panel shows ribbon diagrams of single GroEL subunits, oriented as indicated in the space filling models (lower panels) of GroEL (PDB code 1DER; Boisvert *et al.*, 1996) and GroEL/GroES (PDB code 1AON; Xu *et al.*, 1997) (models include Van der Waals spheres around all backbone and side-chain atoms of the complexes). GroEL monomers consist of three domains: the equatorial (blue), intermediate (yellow) and apical (red) domains. To indicate the location of individual subunits of GroEL in the complex, another subunit is colored in light green in the space filling models. Figure created with SwissPDB-Viewer v3.7 and POV-Ray v3.6.

to the co-factor GroES; and (iii) an intermediate domain, which connects the equatorial and the apical domain and acts like a hinge during large structural rearrangements in the GroEL ring when binding to the co-factor GroES (Figure 7, compare left with right, and see below).

The co-factor GroES is a single-ring, dome-shaped heptamer of identical 10 kDa subunits. GroES binds to the apical domains of either one of the two GroEL rings depending on the ATP or ADP bound state of GroEL (Figures 6 and 7). This creates an asymmetric GroEL/GroES complex with an enclosed cage below the GroES lid – the so called *cis* cavity. The GroEL ring opposite to bound GroES is described as being in the *trans* position. GroES binding is dynamic during the reaction mechanism of GroEL, and GroES subsequently binds to and releases from each ring of GroEL during the reaction cycle. GroES binding to GroEL is mediated by the so-called 'mobile loop' at the base of the GroES dome (Landry *et al.*, 1993; Richardson *et al.*, 2001). This loop contains 16 amino acids which fold into a β-hairpin structure upon association with their hydrophobic binding groove on GroEL. Affinity measurements of various GroEL mutants to GroES by surface plasmon resonance, conducted as a side project of this thesis, suggest that the free energy of the mobile loop order-disorder transition modulates the speed of chaperonin cycling (Shewmaker *et al.*, 2004).

The GroES binding site of GroEL has been repeatedly observed to partially overlap with the substrate binding region of GroEL (Chen and Sigler, 1999; Chen *et al.*, 1994; Fenton *et al.*, 1994; Xu *et al.*, 1997). GroES binding of GroEL thus impairs substrate binding to the apical domains and bound substrate is practically dislodged inside the central cage when GroES associates.

Binding of GroES is preceded by cooperative binding of seven ATP molecules to the *cis* GroEL equatorial subunits. This promotes large structural rearrangements of the *cis* GroEL ring, leading to an approximately twofold increase of the cavity volume (the cavity under GroES is then ~80 Å wide and ~85 Å high). The hydrophobic apical domains open up by twisting outward and upward, thereby releasing substrate into the central cavity (Chen *et al.*, 1994; Hayer-Hartl *et al.*, 1996; Langer *et al.*, 1992b; Martin *et al.*, 1993; Mayhew *et al.*, 1996; Roseman *et al.*, 1996; Weissman *et al.*, 1995). Recent experiments (Motojima *et al.*, 2004) indicate that binding of ATP to GroEL is necessary for this substrate release from the apical domains, while ADP binding to GroEL is sufficient for forming a *cis* complex with GroES, but substrate polypeptide cannot be efficiently released into the cavity. Bound substrate polypeptide thus presents a load on the apical domains of GroEL.

A consequence of the structural rearrangements in GroEL, induced by ATP binding, is the burial of hydrophobic surfaces on the GroEL apical domains and along the GroEL cavity. The inner lining of the GroEL cavity changes from a mostly hydrophobic surface in the unliganded (or *trans*) complex to a mostly hydrophilic surface in the *cis* cavity, enclosed by GroES (Figure 8). This hydrophilic cage can accommodate unfolded proteins up to ~60 kDa (Sigler *et al.*, 1998).
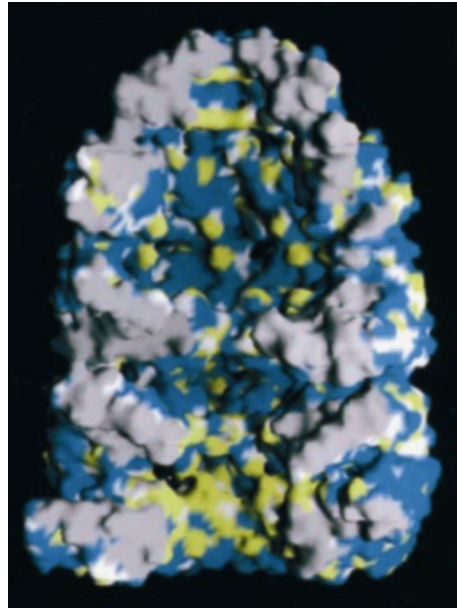
Figure 8: Hydrophobic surfaces at the interior of the GroEL cavity.

View of four subunits of each ring of the asymmetric GroEL/GroES structure (PDB 1AON). Hydrophobic side chains are colored yellow, polar and charged side-chains are blue, solvent-excluded surfaces at the interfaces with missing subunits are grey and exposed polypeptide backbone atoms are white. From Xu *et al.* (1997).

2.4.2   *Mechanisms of GroEL-mediated protein folding*

GroEL/GroES-mediated protein folding is a process which involves an alternating action of the two rings of GroEL (for a schematic depiction of the GroEL/GroES reaction cycle, see Figure 9). Encapsulation of unfolded proteins promotes folding to their native state probably because of a combination of different reasons. First, one GroEL ring usually binds and encapsulates only a single unfolded polypeptide molecule. Inside the GroEL/GroES cage, the protein is then free of interactions with other aggregation prone substances. It is then practically in a solution of infinite dilution, also called an 'Anfinsen cage' (Ellis, 1996). This name originates from the 'Anfinsen cage model', which implies that the folding protein has all the information for correct folding embedded into its primary amino acid sequence. The protein only needs to be prevented from aggregation during the process of its folding. In this model GroEL simply functions as a passive box where unhindered folding can proceed.

A different, but not mutually exclusive view of how substrate folding is achieved by GroEL interaction is described by the 'iterative annealing model' (Todd *et al.*, 1996). Here, GroEL association of substrate polypeptides is thought to partially unfold or rearrange them, before release into the GroEL/GroES cage. Thus, by encountering multiple rounds of GroEL binding and release, polypeptides repeatedly start in an unfolded state and have the chance to
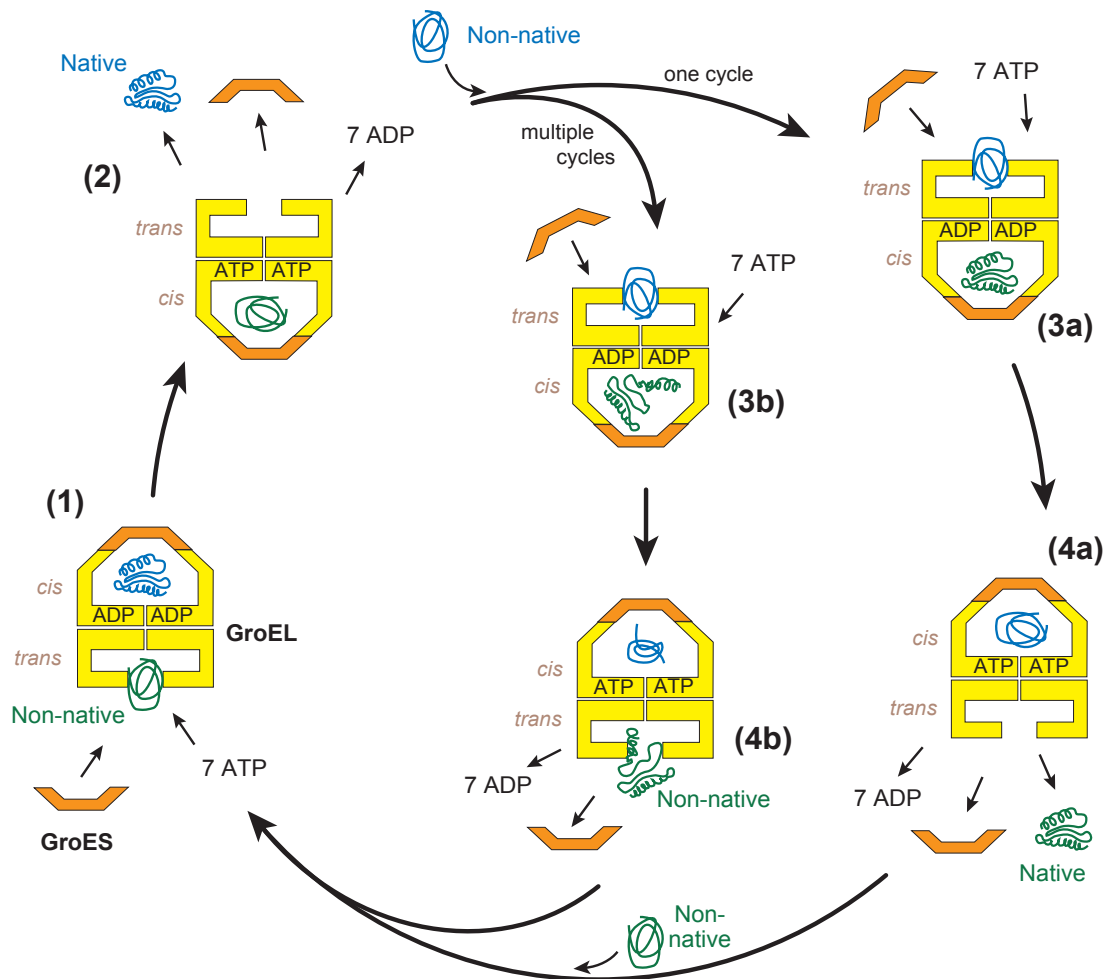
Figure 9: Schematic model of the GroEL/GroES reaction cycle.

Protein folding is mediated in an alternating fashion by the two rings of GroEL in combination with GroES. **(1)** Unfolded protein (green) associates with the hydrophobic apical domains of the unoccupied (lower) ring of the asymmetric GroEL/GroES complex. **(2)** Binding of ATP to this lower ring induces large structural rearrangements, leading to an up and outward twist of the GroEL apical domains. This allows GroES to bind to the apical domains, while concurrently substrate is released into the central cavity with a now greatly increased volume. At the same time, ADP and GroES are released from the opposite (upper) GroEL ring, allowing dissociation of another substrate protein (blue) previously encapsulated in that opposite cavity. **(3a and 3b)** Unfolded protein is given the chance fold in the newly formed (lower) cis cavity during ATP hydrolysis (~10-20 s) before substrate release. **(4a)** Binding of new unfolded substrate (blue), ATP and GroES to the (upper) trans ring induces release of ADP, GroES and now native folded substrate from the (lower) GroEL cis cavity, **(4b)** while substrate that could not reach the native state in this particular cycle can rebind to the apical domains of GroEL or be released into free solution (not shown). Adapted from Naylor and Hartl (2001).

reach the native conformation in the ~10 – 20 s of a GroEL cycle. If they do not complete their folding during one round of GroEL interaction, they can rebind to the GroEL apical domains, thereby partly rearranging their misfolded structure and subsequently having a new chance to fold correctly.

It has been proposed that binding to the apical domains of multiple subunits of GroEL can unfold a substrate protein, based on a study with the protein rubisco from *Rhodospirillum rubrum* (*R. rubrum*, Shtilerman *et al.*, 1999), a 51 kDa protein which is absolutely dependent on GroEL/GroES for folding under normal conditions (Goloubinoff *et al.*, 1989). GroEL-bound rubisco was suggested to be unfolded in a forced manner during the structural rearrangements of GroEL. However, this mechanism of GroEL action was not observed by experiments with different substrates (Chen *et al.*, 2001). A more recent study then confirmed rearrangements in unfolded rubisco during initial association with GroEL, but failed to produce evidence for forced unfolding coupled to the GroEL apical domain movements (Lin and Rye, 2004).

Encapsulation of substrate protein can also influence its folding pathway in additional manners. The space inside the GroEL/GroES cavity is limited, particularly for large proteins (~40 – 60 kDa). The theory of molecular crowding and confinement predicts smaller states to be energetically favored in confined spaces like the chaperonin cage (Ellis, 2001; Ellis, 2003; Minton, 2001), which would promote folding to the more compact native state. The limited range of conformations accessible in such a confined environment probably also contributes to the correct folding process by reducing the amount of different structures actually being sampled during this process. Additionally, the hydrophilic and slightly negatively charged surface of the GroEL/GroES cavity probably also favors formation of the native state, as this is accompanied by the burial of hydrophobic amino acid residues inside the substrate protein.

Notably, work performed early during the preparation of this doctoral thesis, which included investigation of the spontaneous and GroEL assisted (re-)folding and assembly behavior of *R. rubrum* rubisco, contributed to the finding that folding of this protein inside the GroEL/GroES cavity is considerably accelerated when compared to unassisted folding under permissive conditions (Brinker *et al.*, 2001). This acceleration was not observed for the smaller mitochondrial protein rhodanese (~33 kDa). A mechanism was proposed for GroEL/GroES-mediated protein folding of the large protein rubisco which included a smoothing of the otherwise rugged folding energy landscape with a high population of intermediate folding states (see section 2.1.4). GroEL destabilizes these intermediate states by confinement in the hydrophilic cage, and thus accelerates the folding rate of rubisco. This mechanism was subsequently confirmed by theoretical considerations of multiple groups (Baumketner *et al.*, 2003; Jewett *et al.*, 2004; Takagi *et al.*, 2003).

2.4.3   *Substrates of GroEL*

In spite of the detailed knowledge about the molecular mechanisms of the *E. coli* chaperonin machinery, little is known about the natural substrates of GroEL. All of the substrates used in *in vitro* experiments for elucidation of the GroEL mechanism have been het-

erologous proteins. Since even small changes in an otherwise highly homologous protein lead to different folding pathways (Ferguson *et al.*, 1999), clearly no direct parallels can be drawn from heterologous substrates.

In *E. coli* 10 – 15 % of cytosolic proteins transit through GroEL (Ewalt *et al.*, 1997; Houry *et al.*, 1999). Despite this relatively low number, which is based on the nature of the molecular chaperone network inside the cell (see above), GroEL is able to bind to a large and diverse variety of non-native proteins *in vitro* (Viitanen *et al.*, 1992). Employing a GroEL molecule comprised of two rings, each produced as a single continuous polypeptide (Farr *et al.*, 2000), it was possible to determine more closely the effects of mutations that had previously been observed to abolish polypeptide binding (Fenton *et al.*, 1994). Three consecutive wild-type apical domains were found to be necessary to facilitate stable binding of the substrates rhodanese and malate dehydrogenase *in vitro*. This arrangement was also found to be necessary to allow cell viability. Thus, a hydrophobic binding site of ~150° around the inside of a ring appears to be required for efficient binding of non-native peptides.

The main promiscuity of substrate binding by GroEL has been attributed to the plasticity of its hydrophobic binding sites, which results in their association with a multitude of substrates (Chen and Sigler, 1999). A high affinity peptide ('strongly binding peptide', SBP) was identified and crystallized in combination with GroEL apical domains and the full tetradecameric complex. The binding site for this peptide was found to be a flexible hydrophobic groove on the apical domains of GroEL, lining the openings of the cavity. The site was overlapping with the region that accommodates the mobile loop of GroES when bound to GroEL (Chen and Sigler, 1999). Other peptides which have been characterized in their binding to GroEL correspond *e.g.* to amphiphilic α-helical regions of rhodanese (Hlodan *et al.*, 1995). A crystal structure of GroEL with a complete bound substrate protein bound remains to be obtained. Problems associated with this are the large conformational variety of even a single type of bound protein, and the heptameric rotational symmetry of GroEL, which would decrease the abundance of even a single state of a bound substrate to ~15 % in a crystal.

Co-immunoprecipitation of *E. coli* GroEL with bound polypeptides and subsequent 2D-gel-mass spectrometry allowed the previous identification of a subset of potential GroEL substrates (Houry *et al.*, 1999). This study was performed under nucleotide free conditions, with polypeptides bound to the apical domains of GroEL, and without associated GroES. GroEL was also seen to be involved in interaction with a considerable amount of pre-existing proteins that presumably had become unfolded. This action of GroEL is thought to be particularly necessary under heat shock or other stress conditions that lead to protein unfolding in the cell. Among the identified GroEL interactors, there was a structural preference for proteins with multiple αβ domains. However, no more detailed structural information could be extracted from this limited dataset. Furthermore, only about half of the proteins had available structures or homologous structures at the time of analysis.

2.5   Aim of the study

It would be highly informative to study the *in vitro* and *in vivo* folding behavior of several of the natural GroEL interactors so far identified (Houry *et al.*, 1999). It has already been proposed that GroEL dependence of folding proteins could vary for different proteins (Ewalt *et al.*, 1997). The observed differences in molecular chaperone content of *E. coli* and other cells also suggests distinct chaperone requirements, in spite of the observed promiscuous *in vitro* substrate binding of GroEL. Thus far, *in vitro* folding experiments with GroEL have been exclusively conducted with heterologous proteins, which have not allowed direct conclusions for the folding behavior of its natural substrates.

The set of ~50 previously identified GroEL interactors (Houry et al., 1999) contains a number of proteins for which functional enzyme assays are available. These proteins were selected to study their GroEL requirement for refolding *in vitro*. The selected proteins could be tentatively grouped into three classes, based on the variation of their requirement for GroEL. A comprehensive identification and characterization of the complete set of GroEL interactors was undertaken to extend the classification of GroEL substrates to the entire *E. coli* proteome. No quantitative data were available from previous studies and new developments in mass spectrometry allowed identification and relative quantification of the nearly complete interaction proteome of *E. coli* GroEL. Bioinformatic analysis of the GroEL substrate proteome and the distinct classes was then performed to identify common properties of GroEL substrate classes.

3   Materials and Methods

3.1   Materials

3.1.1   Chemicals

Chemicals and biochemicals were of *pro analysi* grade and purchased from **Sigma-Aldrich** (Steinheim, Germany) unless stated otherwise.

Amersham Pharmacia Biotech (Freiburg, Germany): ECL™ detection kit.

BIAcore (Freiburg, Germany): biosensor chips CM5, NTA; Amine Coupling Kit; Maintenance Kit; Surfactant P20.

BioMol (Hamburg, Germany): IPTG; HEPES.

BioRad (Munich, Germany): ethidiumbromide; AffiGel 15 columns.

Cambridge Isotope Laboratories (Saarbrücken, Germany): Arg-^{13}C6; Leu-D3.

Difco (Heidelberg, Germany): Bacto tryptone, Bacto yeast extract, Bacto agar.

Fluka (Deisenhofen, Germany): ADP.

Merck (Darmstadt, Germany): ampicillin; PEI cellulose F TLC plates; [^{32}P]-PP$_i$ [NEX019].

New England Biolabs (Frankfurt am Main, Germany): restriction enzymes, T4 DNA ligase.

Pierce / Perbio Science (Bonn, Germany): Sulfo-GMBS.

Serva (Heidelberg, Germany): Pefabloc Protease inhibitor (4-(2-aminoethyl)benzene-sulfonylfluoride HCl).

Qiagen (Hilden, Germany): Plasmid Midi kit, QIAprep Spin Mini prep kit, QIAquick PCR purification and gel extraction kits.

Roche (Basel, Switzerland): ATP; benzonase; Complete protease inhibitor; hexokinase; Pefabloc; proteinase K (PK, *Tritriachium album*); shrimp alkaline phosphatase.

Roth (Karlsruhe, Germany): scintillation fluid; polyacrylamide/bisacrylamide solution 40 % (37.5 : 1).

Schleicher & Schuell: protran nitrocellulose transfer membrane.

3.1.2   Instruments

Abimed (Langenfeld, Germany): Gilson Pipetman (2, 10, 20, 100, 200, 1000 µl).

Amersham Pharmacia Biotech (Freiburg, Germany): electrophoresis power supply EPS 3500; FPLC systems; ÄKTA Explorer 100; SMART-System; prepacked chromatography columns: 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.

Amicon (Beverly, MA, USA): vacuum filtration unit (0.2 µm); concentration chambers (Centriprep, Centricon).

Avestin (Mannheim, Germany): EmulsiFlex C5 homogenizer.

Beckmann (Munich, Germany): DU 640 UV/VIS Spectrophotometer; centrifuges: Avanti J-25, Avanti J20 XP, J-6B, GS-6R, Optima LE-80k ultracentrifuge.

BIAcore (Freiburg, Germany): BIAcore 2000.

Biometra (Göttingen, Germany): T3 PCR-Thermocycler.

Bio-Rad (München, Germany): electrophoresis chambers MiniProtean 2 and 3; electrophoresis power supply Power PAC 300; Gene Pulser II electroporation system.

Bio-Tek (Bad Friedrichshall, Germany): Synergy HT UV/VIS/fluorescence/luminescence plate reader.

Clontech (Heidelberg, Germany): Talon IMAC resin.

Eppendorf (Hamburg, Germany): centrifuges 5415C and 5417R, Thermomixer Comfort.

Fisher Scientific (Schwerte, Germany): pH meter Accumet Basic.

Fuji (Tokyo, Japan): FLA 2000 Phosphoimager.

Hoefer Scientific Instruments (San Francisco, USA): SemiPhore blotting transfer unit.

Jasco (Groß-Umstadt, Germany): UV/VIS Spectrometer V-560.

Mettler Toledo (Gießen, Germany): balances AG285, PB602.

Millipore (Eschborn, Germany): deionization system MilliQ plus PF, Millex-HA filters 0.22 μm.

Misonix Inc. (New York, USA): sonicator Ultrasonic Processor XL.

New Brunswick Scientific (Nürtingen, Germany): orbital shaker and incubator Innova 4430.

Packard (Dreieich, Germany): liquid scintillation analyzer Tri-carb 1500.

Raytest (Straubenhardt, Germany): AIDA gel imaging software version 2.31.

3.1.3   *Media and buffers*

3.1.3.1   Media

Media were prepared with demineralized H_2O and autoclaved after preparation, unless stated otherwise.

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).

NZY^+ broth: 10 g/l NZ amine (casein hydrolysate), 5 g/l yeast extract, 5 g/l NaCl. pH adjusted to 7.5 with NaOH. 125 μl $MgCl_2$ (1M), 125 μl $MgSO_4$ (1M) and 200 μl glucose (20 % (w/v)) were added to 10 ml NZY broth before use and filter sterilized (0.2 μm).

M63 medium: 2 g/l $(NH_4)_2SO_4$, 13.6 g/l KH_2PO_4, 0.5 mg/l $FeSO_4$ x 7 H_2O. Before use, 1 ml $MgSO_4$ (1 M), 10 ml glucose (20 % w/v), L-amino acid mix (to 0.5 mM of each amino acid final) were added per 1 l medium and filter sterilized

(Sambrook *et al.*, 1989).

SILAC medium: as M63 medium, except Leu was exchanged for Leu-D3 or Arg for Arg-^{13}C6, respectively, in the amino acids mix.

3.1.3.2 Buffers

Buffer A: 20 mM MOPS-KOH, 100 mM KCl, 10 mM MgCl$_2$, pH 7.4.

HBS: 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005 % surfactant P20, pH 7.4.

PBS: 137 mM NaCl, 2.7 mM KCl, 20 mM KH$_2$PO$_4$/K$_2$HPO$_4$, pH 7.4.

TAE: 40 mM Tris-Acetate, 1 mM EDTA, pH 8.3.

TBS: 25 mM Tris-HCl, 140 mM NaCl, 3 mM KCl, pH 8.0.

TBST: TBS + 0.05 % (v/v) Tween 20.

3.1.4   Bacterial strains and plasmids

3.1.4.1   Bacterial strains

The following *E. coli* strains were used throughout this study:

BL21 (DE3) Gold (Stratagene),

DH5α (Novagen),

XL1-Blue (Stratagene),

MC4100 (Teter *et al.*, 1999, from Dr. E. Bremer *via* Dr. S. Raina),

MC4100 Δ*dnaK*Δ*dnaJ* (Teter *et al.*, 1999),

MG1655 (American Type Culture Collection – ATCC 47076),

SC3 (P. A. Lund, University of Birmingham, UK),

MC4100 GroE PBAD (C. Georgopoulos, this study).

3.1.4.2   Plasmids

The following plasmids were generated for recombinant protein expression and *in vivo* experiments:

pET11a Amp^R *E. coli* GroES inserted at the *Nde* I and *Bam* HI restriction endonuclease sites (pT7-ES, Brinker *et al.*, 2001). pET11a Amp^R *E. coli* GroEL inserted at the *Nde* I and *Bam* HI restriction endonuclease sites (pT7-EL, Brinker *et al.*, 2001). pET22b Amp^R *Methanosarcina mazei* GroES inserted at the *Nde* I and *Eco* RI restriction endonuclease sites (pT7-MmGroES, Klunker *et al.*, 2003). pBAD33-ESL Cam^R expressing *E. coli* GroEL/ GroES (Ewalt *et al.*, 1997).

Construction of pT7 and pT7-N(His)$_6$-substrate plasmids: the coding region of each GroEL substrate (ADD, ALR2, CRP, DAPA, DCEA, END4, ENO, G3P1, GATD, GATY, HEM2, LLDD, LTAE, METF, METK, NANA, SYT, TDH, TYPH, XYLA, YAJO, YHBJ) was amplified by PCR from MG1655 genomic DNA and inserted into pET22b Amp^R (No-

vagen) (for wild type proteins) and pET28b Kan^R (Novagen) (for amino-terminally hexahis-tidine tagged proteins) at *Nde* I and *Hind* III or *Eco* RI restriction endonuclease sites.

Construction of pT7-ES-C(His)$_6$ and pT7-*Mm*ES-C(His)$_6$ (carboxy-terminally hexa-histidine tagged *E. coli* GroES / *M. mazei* GroES): the coding regions, including a carboxy-terminal (His)$_6$-tag, were amplified from pT7-ES and pT7-*Mm*ES and inserted into pET22b Amp^R (Novagen) at the *Nde* I and *Hind* III or *Eco* RI restriction endonuclease sites.

Construction of pBAD18-ES, pBAD18-ES-cHis, pBAD18-*Mm*ES and pBAD18-*Mm*ES-cHis: the ribosomal binding site and coding region of the corresponding pT7-plasmid was excised with XbaI and HindIII restriction endonucleases. Each fragment was inserted into the same sites of pBAD18 (Guzman *et al.*, 1995).

Construction of pBAD33-EL and pBAD33-*Mm*ES: the ribosomal binding site and coding region of the pT7-EL and pT7-*Mm*ES plasmids was excised with *Xba* I and *Hind* III restriction endonucleases. Each fragment was inserted into the same sites of pBAD33 Cam^R (Guzman *et al.*, 1995).

Each construct was verified by DNA sequencing on both strands.

3.1.5 *Proteins*

The following purified proteins were obtained from the laboratory stock of the De-partment of Cellular Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Ger-many:

DnaJ (Zylicz *et al.*, 1985),

GrpE (Zylicz *et al.*, 1987), GrpE-(His)$_6$,

GroES (Hayer-Hartl *et al.*, 1996).

3.2 Molecular biological methods

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

Competence buffer I: 100 mM KCl, 30 mM KOAc, 60 mM CaCl$_2$, 15 % glycerol; pH 5.8, adjusted with acetic acid. Filter sterilized and stored at 4 °C.

Competence buffer II: 10 mM MOPS, 10 mM KCl, 75 mM CaCl2, 15 % glycerol; pH 6.8, adjusted with NaOH. Filter sterilized and stored at 4 °C.

For preparation of chemically competent *E. coli* cells, a single colony was used to inoculate LB medium (including antibiotic, if applicable) and grown to an optical density (OD$_{600nm}$) of 0.25 - 0.5. The cells were then chilled on ice for 15 min and centrifuged at 1500 g for 15 min at 4 °C. The supernatant was discarded; the cells were resuspended in one third of the original volume competence buffer I without vortexing, and incubated on ice for 1 h. Next, the cells were pelleted, resuspended in 1/25 of the original volume competence buffer II, incubated on ice for 15 min and shock-frozen in liquid N$_2$ in 20-100 µl aliquots.

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 30 min. The cells were heat

shocked at 42 °C for 30-45 s and subsequently placed on ice for 2 min. 1 ml of NZY$^+$ or LB medium was added and shaken at 37 °C for 1 h. The cell suspension was then plated on selective plates and incubated at 37 °C, until colonies had developed (typically 10-16 h).

3.2.2   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$_2$O exhibits approximately A$_{260nm} = 1$.

Agarose gel electrophoresis was performed in TAE buffer 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).

3.2.3   PCR amplification

PCR (polymerase chain reaction) mediated amplification of DNA was performed according to the following generic protocol (with minor modifications to it, when necessary):

DNA Template:	20 ng (plasmid DNA)
	250 ng or less (bacterial genomic DNA)
Primers:	20 pmol each
dNTPs:	200 µM each
Polymerase:	2.5 U
Polymerase buffer:	1 x
Additives:	4 % DMSO if GC content was >50 %,
	7 % DMSO if GC content was >60 %.
Final volume:	50 µl

Cycling conditions (30 cycles):

Initial denaturation:	95 °C, 5 min
Cycle denaturation:	95 °C, 30-60 s
Annealing:	~55 °C, 30-60 s
Extension:	~72 °C, duration dependent on template length: <400 bp, 30 s; 400-1000 bp, 45 s; 1-2 kbp, 1 min; >2 kbp, 2 min.
Final Extension:	72 °C, 10 min.
Stored at 4 °C or -20 °C.	

3.2.4   DNA restriction and ligation

DNA restriction was performed according to product instructions of the respective enzymes. Typically, a 50 µl reaction contained 1-2 µl of each restriction enzyme and 30 µl 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, 50-100 ng (~1-2 µl) dephosphorylated vector DNA, 200-300 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 and transformed into competent *E. coli* DH5α cells.

LB medium containing the appropriate antibiotic was inoculated with a single *E. coli* colony harboring the DNA plasmid of interest and shaken 8 – 14 h at 37 °C. Plasmids were isolated using the QIAprep Spin Miniprep Kit or QIAGEN Plasmid Midi Kit according to the instructions.

3.3 Protein biochemical methods

3.3.1 Protein expression and purification

All protein purifications steps were performed at 4 – 8 °C unless stated otherwise.

GroEL was purified with modifications to the protocol described by Hayer-Hartl *et al.* (1994). *E. coli* BL21 (DE3) Gold cells harboring the plasmid pT7-GroEL, grown at 37 °C in 6 l LB medium containing 100 mg/l ampicillin were induced with 1 mM IPTG at an OD_{600} ~0.6 for 5 – 6 h and harvested by centrifugation for 30 min at 2500 g. Cells were resuspended in 50 mM Tris-HCl pH 7.8, 50 mM NaCl, 1mM EDTA and Complete protease inhibitor (1 tablet/25 ml). The suspension was frozen in liquid N_2 and thawed before addition of lysozyme (~0.5 mg/ml) and benzonase (~200 units). Lysis was achieved by homogenization of the cell suspension in an EmulsiFlex C5 device kept on ice. Cell debris was removed by ultracentrifugation for 30 min at 4 °C and ~100 000 g and subsequent filtration (0.2 μm). Upon twofold dilution with 50 mM Tris-HCl pH 7.5, 1 mM EDTA, the supernatant was applied to a 200 ml Source 30Q column attached to an ÄKTA Explorer chromatography system. After washing with two column volumes of the above buffer, a gradient was introduced, elevating the NaCl concentration to 1 M in ~5 – 6 column volumes and resulting in elution of proteins and DNA. Collected fractions were analyzed for protein content by SDS-PAGE and GroEL containing fractions were pooled. This pool was diluted 4-fold with 25 mM histidine-HCl, 1 mM EDTA, pH 5.8 and applied to a 20 ml MonoQ column. Proteins were eluted in 25 mM His-HCl pH 5.8 1 mM EDTA and a NaCl gradient from 0 to 0.5 M. Fractions were collected in tubes already containing 1/10 of the final volume 1 M Tris-HCl pH 7.8 to immediately increase the pH and prevent protein precipitation caused by long exposure to low pH. The pooled fractions containing GroEL were dialyzed against 50 mM Tris-HCl pH 7.3. The sample was then applied to a 5 ml Heparin Sepharose column (HiTrap Heparin) and eluted with 50 mM Tris-HCl pH 7.1 and a NaCl gradient from 0 to 250 mM NaCl. GroEL-containing fractions were pooled and concentrated to ~7 ml in Centriprep concentrators. The concentrated sample was applied to a Sephacryl S 300 (XK 26/60) size exclusion column equilibrated in 20 mM MOPS-KOH pH 7.4, 100 mM NaCl and 10 % glycerol. Fractions containing GroEL oligomer (approximate size 800 kDa) were collected, concentrated using a Centriprep concentrator, frozen in aliquots in liquid N_2 and stored at -80 °C. Total yield of GroEL was typically ~400 – 600 mg.

GroEL-D87K (GroEL-Trap) (Fenton *et al.*, 1994; Weissman *et al.*, 1994) was purified with modifications to the purification protocol of wild type GroEL. Additional MonoQ anion exchange chromatography steps were introduced between heparin and final size exclusion chromatography. GroEL was eluted from MonoQ in 50 mM Tris pH 7.8 and 7.1, using

very shallow NaCl gradients at GroEL elution concentrations. Total yield was ~300 mg.

GroES-(His)$_6$ cell lysate from cells harboring the plasmid pT7-GroES-C(His)$_6$ was prepared as described for GroEL. The lysate was diluted with running buffer (50 mM Tris pH 7.4, 200 mM NaCl, 10 mM MgCl$_2$, and 10 mM KCl) and applied to a 10 ml Talon column using fresh or regenerated resin. The column was washed with 100 ml running buffer, followed by 10 ml running buffer including 1 mM ATP to allow release of potentially bound GroEL. Elution was achieved by applying a shallow imidazole gradient from 0 to 300 mM. Fractions containing GroES-(His)$_6$ were pooled and applied to a MonoQ ion exchange chromatography column (50 mM Tris, pH 7.2, NaCl gradient from 0 to 0.5 M NaCl) followed by size exclusion chromatography in 20 mM MOPS, 200 mM NaCl and 10 % glycerol and a final concentration step.

DnaK was purified with modifications to the method used by Jordan & McMacken (1995). *E. coli* BL21 (DE3) Gold cell lysate containing DnaK expressed from the plasmid pET11a-DnaK (laboratory stock, Hartl Dept) was prepared in 50 mM MOPS pH 7.6, 20 mM KCl, 1mM DTT, 1 mM EDTA and Complete protease inhibitor as described for GroEL from 4 l of cell culture. The lysate was applied to a 200 ml DE52 anion exchange column equilibrated in 20 mM MOPS pH 7.6, 1 mM DTT and 10 mM MgCl$_2$. Elution was achieved by applying a shallow KCl gradient from 0 – 0.6 M KCl. Fractions containing DnaK were pooled and further applied to a 25 ml ATP-agarose (C-8, 9 atom spacer, Sigma) column in 20 mM MOPS pH 7.6, 1 mM DTT, 10 mM MgCl$_2$, and 50 mM KCl running buffer. The column was washed with 150 ml buffer, followed by 90 ml high salt buffer (running buffer plus 1 M KCl), 70 ml GTP buffer (running buffer plus 1 mM GTP), 20 ml running buffer and eluted in running buffer supplemented with 5 mM ATP. DnaK containing fractions were pooled and applied to MonoQ anion exchange chromatography in 20 mM MOPS pH 7.6, 10 mM MgCl$_2$, and a gradient of 50 – 500 mM KCl. Final size exclusion chromatography was performed on a Sephacryl S200 (26/60) column, using storage buffer (20 mM MOPS pH 7.6, 200 mM KCl and 10 % glycerol). DnaK containing fractions were concentrated, frozen in liquid N$_2$ and stored in aliquots at -80 °C. DnaK activity was confirmed (also in combination with DnaJ and GrpE-(His)$_6$) by measurement of its ATPase activity as described by Liberek *et al.* (1991).

METK was purified from cells expressing untagged METK from pET22b-METK as described by Markham *et al.* (1983).

SYT was purified in a non-tagged form from pET22b-SYT with modifications to the protocol described by Brunel *et al.* (1993). Lysate was prepared as described for GroEL in 50 mM Tris-HCl, 20 mM NaCl, 2 mM EDTA, 2 mM DTT and Complete protease inhibitor. The lysate was applied to a DE52 anion chromatography column (150 ml resin), and elution performed in 50 mM Tris-HCl pH 8.0 , 5 mM MgCl$_2$, 1 mM DTT, 5 % glycerol and a NaCl gradient from 0 – 0.7 M. SYT containing fractions were then subjected to anion chromatography on MonoQ (20 ml resin) using the above buffer and gradient. SYT containing fractions were dialyzed against 20 mM MES-KOH pH 6.8, 10 mM KCl, 5 mM MgCl$_2$, 5 % glycerol, 1 mM DTT and applied to Source 30 S cation chromatography using dialysis buffer and a 0 – 200 mM KCl gradient. Next, SYT containing fractions were diluted 3-fold and applied to

Heparin chromatography (2 x 5 ml resin) in the same buffer. Following elution with a 0 – 1 M KCl gradient, combined SYT containing fractions were concentrated and subjected to Sephacryl S200 size exclusion chromatography. SYT dimer fractions were combined, concentrated in a Centricon concentrator, frozen in liquid N_2 and stored at -80 °C. Total yield was ~800 mg.

All other GroEL substrates (**ADD, ALR2, DAPA, DCEA, END4, ENO, G3P1, GATD, GATY, HEM2, LLDD, METF, NANA, TDH, XYLA, YAJO**) were expressed at 30 °C or 37 °C from the corresponding pET28-plasmids encoding amino-terminally $(His)_6$-tagged proteins. Highly insoluble GroEL substrates (METF, GATY) were expressed at 30 °C in cells harboring elevated levels of GroEL/GroES, as described below for *in vivo* coexpression experiments. Purification of hexahistidine tagged proteins was performed utilizing the following general protocol: Cell lysate was prepared as described above in running buffer (50 mM Tris-HCl pH 7.3, 300 mM NaCl, Complete protease inhibitor (EDTA free). Diluted lysate was applied to ~10 – 15 ml Talon resin column and washed with ~100 ml of 50 mM Tris-HCl, 300 mM NaCl. Potentially bound chaperones were eluted by washing with 30 ml running buffer plus 10 mM KCl, 5 mM $MgCl_2$ and 5 mM ATP. Elution was achieved by an imidazole gradient from 10 – 250 mM. Fractions containing the protein of interest were combined and, dependent on their purity, either subjected to MonoQ anion exchange chromatography (50 mM Tris-HCl pH 7.0 – 8.0, NaCl gradient) or directly to size exclusion chromatography (Sephacryl S200, S300 or Superdex 200) in 20 mM MOPS-KOH pH 7.4, 200 mM NaCl, 10 % glycerol. Following concentration in Centriprep concentrators, protein solutions were aliquoted, frozen in liquid N_2 and stored at -80 °C.

3.3.2 *Protein analytical methods*

3.3.2.1 Determination of protein concentrations

Protein concentrations were determined spectrophotometrically, based on the theoretical excitation 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), unless otherwise stated. Molar concentrations of chaperones are expressed for the native state oligomers. However, GroEL substrates are expressed as monomers unless otherwise stated, due to presumed monomeric binding of the substrates to chaperones in the denatured state.

3.3.2.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed using a discontinuous buffer system (Laemmli, 1970) in BioRad Mini-Protean II or 3 electrophoresis chambers employing a constant voltage of 160-220 V in 50 mM Tris-Base, 380 mM glycine, 0.1 % SDS (pH 8.3). SDS loading buffer was added to the protein samples (final concentration: 60 mM Tris-HCl, pH 6.8, 100 mM DTT, 2 % SDS, 10 % glycerol, 0.005 % bromphenol blue) prior to denaturation at 95 °C for 5 min and gel loading. Gels (0.75 and 1.0 mm width) had varying concentrations between 6 and 16 % polyacrylamide, dependent on the required resolution.

Stacking Gel:	5.1 % acrylamide/bisacrylamide (37.5:1), 125 mM Tris-HCl, pH 6.8, 0.1 % SDS, 0.1 % TEMED, 0.1 % APS.
Separating Gel:	6 - 16 % acrylamide/bisacrylamide (37.5:1), 0.75 M Tris-HCl, pH 8.8, 0.1 % SDS, 0.075 % TEMED, 0.05 % APS.

3.3.2.3 Coomassie blue staining

Polyacrylamide gels were fixed and stained in 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 for removal of unspecific background stain.

3.3.2.4 Silver staining

Polyacrylamide gels were fixed for at least 1 h in 12 % TCA, 50 % methanol, reduced for 15 min in 12 % TCA, 50 % methanol, 60 mM CuCl2, rinsed in H_2O and washed twice for 15 min in 5 % acetic acid, 10 % ethanol. Following oxidation for 15 min in 250 μM $KMnO_4$, gels were rinsed in H_2O, washed twice for 15 min in 5 % acetic acid, 10 % ethanol and rinsed in H_2O again. Staining occurred after 20 min incubation in 6 mM $AgNO_3$, a rinse in H_2O, and development in 145 mM K_2CO_3, 0.02 % HCHO until the desired staining degree was reached. Further color development was stopped by final incubation in 50 mM EDTA.

3.3.2.5 Western Blotting

Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes in a semi-dry western blotting unit (SemiPhore) in 25 mM Tris, 192 mM glycine, 20 % methanol, pH 8.4 at constant current of ~0.5 – 0.8 mA/cm^2 gel size for 1.5 h (Towbin *et al.*, 1979).

Nitrocellulose membranes were blocked in 5 % skim milk powder in TBST for 1 h. The membranes were then incubated with a 1:2000 – 1:10000 dilution of primary antibody serum in TBST and extensively washed in TBST before incubation with a 1:5000 dilution of secondary antibody in TBST (Anti-rabbit IgG, whole molecule – horseradish peroxidase conjugate. Antibody produced in goat. Sigma). After extensive washing, protein bands were detected by incubating the membranes with ECL chemiluminescence solution and exposure to X-ray film.

3.3.2.6 Generation of polyclonal antibodies

Rabbit polyclonal antibodies were generated at the animal facilities of the MPI for Biochemistry according to Harlow and Lane (1988). Purified proteins were injected subcutaneously as water in oil emulsion formed out of 1 volume of protein solution (~0.2 – 1 mg) in PBS and 1 volume Freund's Adjuvant (Freund and McDermot, 1942). Complete Freund's adjuvant was used for the initial immunization and incomplete Freunds adjuvant for 4 – 6 succeeding boosts, which were injected at intervals of 4 – 7 weeks. Serum was taken ~10 days post injection. When necessary, antibodies were affinity purified with the corresponding antigen immobilized to AffiGel 15 (Harlow and Lane, 1988).

3.4 Biochemical and biophysical methods

3.4.1 *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) : 5 DnaK (monomer) : 2.5 DnaJ (monomer) : 2.5 GrpE (dimer).

Chaperone-mediated refolding was stopped by complexion of Mg^{2+} with EDTA or CDTA, which inhibits the chaperone ATPase activity. If, however, the subsequent enzymatic reaction for determination of the folding status was also inhibited by EDTA or CDTA, chaperone-mediated folding was stopped by quick hydrolysis of remaining ATP in the folding reaction with apyrase.

3.4.1.1 DAPA refolding

25 µM DAPA was denatured in 6 M GdnHCl in buffer A containing 10 mM DTT for 1 h at 25 °C and diluted 100-fold into buffer A containing 10 mM Na-pyruvate and 5 mM ATP in the absence or presence of chaperones as indicated. At specified time points, aliquots of the reactions were stopped with a final concentration of 12.5 mM CDTA. DAPA activity was determined colorimetrically as described (Vauterin *et al.*, 2000). The assay buffer contained 200 mM imidazole pH 7.4, 35 mM Na-Pyruvate, 4 mM *o*-aminobenzaldehyde and 2 mM L-aspartate-β-semialdehyde (ASA, a generous gift from Dr. B. Laber, Aventis, Frankfurt, Germany). The substrate ASA was stored in 4 M HCl at -20 °C and was neutralized with an equal volume of 4 M NaOH prior to usage.

3.4.1.2 DCEA refolding

DCEA was denatured with 6 M GdnHCl in buffer A containing 8 mM DTT for 1 h at 25 °C and diluted 100-fold (to 1 µM) into buffer A containing 15 µM pyridoxal 5-phosphate and 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, aliquots (25 µl) of the different refolding reactions were stopped with 1 U apyrase. DCEA activity was measured at 37 °C in a coupled enzymatic assay, by following the production of NADPH and corresponding increase in absorbance 340 nm as described (De Biase *et al.*, 1996).

3.4.1.3 ENO refolding

100 µM ENO was denatured in 6 M GdnHCl in buffer A containing 10 mM DTT for 1 h at 25 °C and diluted 100-fold into buffer A containing 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, aliquots of the refolding reactions were stopped by transferring them to enzyme assay solution containing 50 mM Tris-HCl pH 8.1, 100 mM KCl, 1 mM 2-phosphoglyceric acid, 1 mM $MgSO_4$ and 10 µM EDTA. ENO activity was measured essentially as described by Spring and Wold (1975); as a modification ENO activity measurements were stopped with 100 nM HCl to allow UV absorption at 230 nm.

3.4.1.4 GATD refolding

100 μM GATD was denatured in 6 M GdnHCl in buffer A containing 5 mM DTT for 1 h at 25 °C and diluted 100-fold into buffer A containing 50 μM $MnCl_2$ and 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, aliquots of the reactions were stopped with 0.1 U/μl apyrase. GATD activity was measured as described (Anderson and Markwell, 1982). The assay buffer contained 50 mM Tris, pH 8.2, 50 μM $MnCl_2$, 5 mM NAD^+ and 9 mM D-galactitol-6-phosphate. The substrate D-galactitol-6-phosphate was prepared by reduction of D-galactose-6-phosphate according to Wolff and Kaplan (1956).

3.4.1.5 METF refolding

METF concentrations were determined based on the absorption of bound FAD at 447 nm (ε=14300 $M^{-1}cm^{-1}$) (Sheppard *et al.*, 1999). 50 μM METF was denatured with 4.35 M GdnHCl in buffer A containing 10 mM DTT for 1 h at 25 °C and diluted 100-fold into buffer A containing 50 μM FAD, 1 g/L BSA and 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, aliquots of the reactions were stopped by 40 mM CDTA. METF activity was measured at 25 °C utilizing an NADH-menadione oxidoreductase assay, essentially as described (Sheppard *et al.*, 1999). The assay buffer was 50 mM Tris pH 7.2, 2 mM EDTA, 1 g/l BSA, 180 μM menadione and 200 μM NADH.

3.4.1.6 METK refolding

METK was denatured with 6 M GdnHCl in buffer A containing 8 mM DTT for 1 h at 25 °C and diluted 100-fold (to 500 nM) into buffer A containing 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, refolding reactions were stopped with a 26-fold molar excess of EL-D87K (GroEL-Trap) (Farr *et al.*, 1997), which binds to non-native protein but due to an inhibited ATPase is unable to release it. METK activity was measured at 25 °C essentially as described (Markham *et al.*, 1983) except that L-[^{35}S]-methionine (specific activity 50 Ci/mol) was used.

3.4.1.7 SYT refolding

50 μM SYT was denatured in 6 M GdnHCl in buffer A containing 10 mM DTT for 1 h at 25 °C and diluted 100-fold into buffer A containing 5 mM ATP in the absence or presence of indicated chaperones. At specified time points, 2 μl aliquots of the refolding reactions were transferred to 18 μl of an enzymatic assay reaction containing 20 mM MOPS pH 7.4, 100 mM KCl, 10 mM $MgCl_2$, 10 mM NaF, 2 mM threonine, 5 mM ATP and 2 mM [^{32}P]-PPi (0.5 MBq/μmol) at 37 °C (Bullard *et al.*, 2000). After 10 min, 2 μl aliquots were spotted onto PEI-cellulose plates and separated by thin layer chromatography using 4 M Urea, 0.75 M KH_2PO_4 as mobile phase. The formation of [^{32}P]-ATP was quantified on a FLA-2000 phosphoimager with Aida 2.31 imaging software.

3.4.2 *In vivo co-expression of chaperones with substrates*

BL21 (DE3) Gold cells, harboring either the pBAD33-ESL or pBAD33-EL plasmids

were transformed with individual pT7-substrate plasmids. Single colonies were picked and grown at 37 °C in LB medium with 0.1 g/l ampicillin (amp), 0.04 g/l chloramphenicol (cam), 0.2 % (w/v) glucose and 0.2 % (w/v) glycerol to OD_{600nm}=0.4. Cells were then divided and further grown in LB (amp, cam, glycerol), supplemented with 0.2 % arabinose, for 1 h to induce the chaperonins (the control reaction missing chaperone expression was grown in LB medium supplemented with 0.2 % glucose during this time). Following chaperone induction, the medium was changed back to glucose and supplemented with 1 mM IPTG for 1 h to induce the substrate protein. Equivalent numbers of cells were taken for preparation of total, soluble and insoluble protein fractions. Cells were pelleted and the material for total protein preparation was resuspended in SDS-PAGE sample buffer. The material for solu- ble/insoluble protein preparation was resuspended in lysis buffer (50 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM EDTA, 0.001 % (w/v) Tween 20 and 0.4 mg/ml lysozyme), incubated on ice for 3 h and subjected to multiple freeze-thaw cycles in the presence of benzonase (Roche). Insoluble from soluble material was separated by centrifugation (20000 g, 30 min) and resuspended in SDS-PAGE sample buffer. Total, soluble and insoluble extracts were prepared in identical volumes to facilitate comparison. The levels of proteins were compared following 12 % or 16 % SDS-PAGE and Coomassie Blue staining.

3.4.3   GroEL/GroES depletion

A GroEL depletion strain (SC3, a derivative of the *E. coli* TG1 strain), where the chromosomal groE promoter is replaced with the araC gene and the pBAD promoter, and a kanamycin resistance (Kan[R]) cassette is immediately upstream of groE (transcribed in the reverse orientation) was a generous gift from P. A. Lund (University of Birmingham, UK). A bacteriophage P1 lysate grown on the SC3 strain was used to transduce *E. coli* MC4100 to KanR (MC4100 SC3 Kan[R]) as previously described (Ang and Georgopoulos, 1989).

LB medium containing kanamycin (kan, 0.05 g/l) and 0.2 % arabinose was inocu- lated with MC4100 GroE PBAD. After growth for two hours, cells were washed with sugar free minimal medium and resuspended in pre-warmed minimal medium containing 0.2 % glucose and 0.05 g/ml kan to initiate GroE depletion. Growing cells were diluted into fresh pre-warmed depletion medium once their optical density reached OD_{600nm} >0.5. Samples were taken at indicated times and resuspended in lysis buffer (50 mM Tris-HCl, pH 8,100 mM NaCl, Complete protease inhibitor, 0.01 % Tween 20, 0.05 g/l lysozyme and 10 U/ml Benzonase). After incubation on ice for 2 h, samples were freeze-thawed repeatedly and an aliquot for total cell protein was taken. The remainder was centrifuged (20000 g) at 4 °C for 15 min and divided into soluble and insoluble fractions. The levels of endogenous substrates and depleted chaperonins were detected following 12 or 16 % SDS-PAGE and immunoblot- ting.

3.4.4   In vivo *capture of GroEL/GroES cis-cavity substrates*

E. coli MC4100 cells, transformed with the pBAD18-*Ec*ES-cHis, pBAD18-*Ec*ES, pBAD18-*Mm*ES-cHis or pBAD18-*Mm*ES plasmids, were grown at 30 °C or 37 °C in 4 l LB medium with 0.2 % glucose, 0.2 % glycerol, 0.1 g/l amp to exponential phase (OD_{600nm}~0.4).

Cells were further grown for 30 min in LB-amp with 0.2 % arabinose, 0.2 % glycerol for 30 min to induce GroES-C(His)$_6$ (or untagged GroES in control experiments). Spheroplasts were prepared from ~2 g cells at 4 °C as previously described (Ewalt *et al.*, 1997). The resulting ~0.8 g spheroplasts were resuspended in 50 ml M63 medium supplemented with 0.2 % glucose, 0.2 % glycerol, 0.5 mM L-amino acids, 1 mM MgSO$_4$ and 0.25 M sucrose and grown at 30 °C or 37 °C for 15 min. Lysis was carried out by rapid dilution of the spheroplasts into an equal volume of 25 °C hypo-osmotic lysis buffer (50 mM Tris-HCl (pH 8), 0.01 % (w/v) Tween 20, 10 mM MgCl$_2$, 25 U/ml benzonase, 2 mM Pefabloc, 10 mM glucose and 20 U/ml hexokinase). 10 s following lysis, ADP (pH 8) was added to a final concentration of 10 mM. All subsequent steps were carried out at 4 °C. The supernatant was cleared at 30000 x g for 10 min and incubated for 30 min with ~4 ml of Talon resin pre-equilibrated in buffer B (50 mM Tris-HCl (pH 8), 200 mM NaCl, 20 mM MgCl$_2$, 50 mM KCl and 10 mM ADP). The resin was washed twice for 10 min with 100 ml of buffer B then further washed in a chromatography column (at gravity flow) with 12 ml of 50 mM imidazole in buffer B. GroEL/GroES/substrate complexes were eluted with 12 ml of 200 mM imidazole in buffer B and 0.5 ml fractions were collected.

3.4.5   *Proteinase K digestion of GroEL/GroES/substrate complexes*

GroEL/GroES/substrate complexes were prepared and handled as before, but without the final elution from the Talon resin. An equal volume of buffer B supplemented with Proteinase K (33 µg/ml) was added at 25 °C. At indicated times, samples were removed and the digestion was stopped with 10 mM PMSF and 100 mM EDTA. Identical control reactions were performed with *in vitro* preformed GroEL/GroES complexes with purified native substrates. The samples were subjected to SDS-PAGE and immunoblotted against the indicated proteins.

3.4.6   *Sample preparation for protein identification by mass spectrometry*

Mass spectrometry and data analysis of the mass spectrometric output was performed by Dr. Yasushi Ishihama (Center for Experimental BioInformatics, Department of Biochemistry and Molecular Biology, University of Southern Denmark – Odense).

600 µL of the respective GroEL/GroES/substrate sample solution were separated by SDS-PAGE (16 %, 1.5 mm, 200 V for 2 h). The gel was Coomassie brilliant blue stained, the entire lanes were cut out and sliced into 5 pieces, in-gel reduced, alkylated, and digested with trypsin as described (Lasonder *et al.*, 2002). Following extraction of peptides from gel pieces using 3 % trifluoroacetic acid (TFA) and 30 % acetonitrile, the sample volume was partially reduced by vacuum evaporation and the residual solutions were applied to StageTip to desalt, filtrate and concentrate the peptide samples (Rappsilber *et al.*, 2003).

For measurement of a total *E. coli* lysate MC4100, wild type cells were cultured with SILAC medium containing arginine-^{13}C6 or leucine-D3 (Ong *et al.*, 2002; Ong *et al.*, 2003). The same procedure as the starting material for GroEL-substrate identification was performed to extract the soluble lysate with SILAC labeling. The labeled soluble lysate was mixed with the unlabeled proteins purified by IMAC with the ratio of 1:50, fractionated by

SDS-PAGE, reduced, alkylated, digested by trypsin, and purified by StageTip as described above.

3.4.7 Coupled liquid chromatography – mass spectrometry system (LC-MS/MS)

An LCMS system consisting of a QSTAR Pulsar quadrupole TOF tandem mass spectrometer (MDS-Sciex, Toronto, Canada) and an Agilent 1100 binary capillary pump (Palo Alto, CA, USA) was used throughout this study. Reprosil-Pur 120 C18-AQ materials (3 μm, Dr. Maisch-GmbH, Ammerbuch, Germany) or Vydac 218MSB3 (3 μm prototype C18 material, a generous gift from Grace Vydac, Hesperia, CA, USA) were packed into a needle (100 μm ID, 6 μm opening, 150 mm length) pulled by a Sutter P-2000 capillary puller (Novato, CA, USA). This needle worked as an ES needle as well as an analytical column where particles formed a self-assembled particle frit (SAP-frit) at the tapered end of the needle according to the principle of the stone arch bridge (Ishihama *et al.*, 2002). The packed needle was held on a nanoelectrospray ion source *via* a Valco titanium union (Houston, TX, USA). Peptides from each gel slice were divided into three fractions and were loaded onto the analytical column using an HTC-PAL autosampler (CTC analytics, Switzerland) with a Valco custom-made 10-port injection valve. Three different mobile phases containing (i) 0.5 % acetic acid, (ii) 0.5 % acetic acid plus 0.005 % TFA, and (iii) 0.5 % acetic acid plus 0.005 % heptafluorobutyric acid (HFBA) were employed to maximize the number of unique peptides by changing the elution times of peptides. Multi-step linear gradient elution programs from 4 % to 80 % acetonitrile in 80 min (protein identification) or 110 min (enrichment factor measurement) were applied for each mobile phase condition. A survey scan from m/z = 350 to 1400 for 1 s with subsequent 4 MSMS scans for 1.5 s each was performed and fragmented peptides were excluded from sequencing for 120 s, as described (Rappsilber *et al.*, 2002).

3.4.8 Mass spectrometry data analysis

Peak lists were created by scripts in Analyst QS (MDS-Sciex) on the basis of the recorded fragmentation spectra and were submitted to Mascot database searching engine (Matrix Sciences, London, UK) against the *E. coli* SwissProt database to identify proteins. The following search parameters were used in all Mascot searches: maximum of one missed trypsin cleavage, cysteine carbamidomethylation, methionine oxidation, and a maximum 0.25 Da error tolerance in both the MS and MSMS data. The output data from Mascot was submitted to in-house software in order to re-calibrate the obtained MS and MSMS spectra using identified peptide masses iteratively. The averaged parent ion mass deviation from the theoretical values resulted in approximately 10-15 ppm. All peptides with the scores <15 or the rank >1 were automatically discarded. Protein hits with score >50 were considered identified with no manual inspection. All other hits were manually verified by using accepted rules for peptide fragmentation. In addition, we used the parent ion mass accuracy (mass deviation <50 ppm), the predicted retention times (Meek, 1980) (difference <10 min), and protein molecular weight estimated from the gel slice as additional requirements for protein identification. For the measurement of enrichment of substrates on GroEL compared to the lysate, in-house

software was developed to calculate the peak areas of the pair of labeled and unlabeled peptides from each MS chromatogram. The peak area ratio of unlabeled peptide to the corresponding labeled peptide was described as enrichment factor without any normalization. This software allows searching the unidentified peptide pairs using external data with accurate parent ion masses and their retention times, which were measured from different LCMS runs.

Absolute protein concentrations were estimated by evaluation of mass spectrometric data with the exponentially modified Protein Abundance Index (emPAI) (Y. Ishihama, manuscript submitted to J. Proteome Res.). emPAI is defined as $m\pi = 10^{PAI} - 1$, with PAI being the number of observed peptides in mass spectrometry, divided by the number of theoretically observable peptides (Rappsilber et al., 2002). emPAI was shown to correlate with protein concentration linearly over a wide range with errors similar or better than by determination of protein staining

3.5   Bioinformatic methods

3.5.1   *Structural comparison of GroE substrates*

Structural comparison of GroEL substrates and *E. coli* lysate proteins with the SCOP database (Structural Classification Of Proteins, Lo Conte *et al*., 2002) was carried out by D. Frishman (Institute for Bioinformatics, German National Center for Health and Environment (GSF), Neuherberg). Pairwise all-on-all sequence comparisons of GroEL substrate protein sequences in each experimental dataset were carried out using PSI-BLAST (Altschul *et al.*, 1997). Sequences sharing significant similarity (BLAST score greater than 45) were joined into single-linkage clusters as described (Frishman, 2002). For homology-based fold assignments, sensitive similarity searches using the IMPALA software (Schäffer *et al.*, 1999) were carried out with each query protein sequence against the SCOP database of structural domains (Lo Conte *et al.*, 2002).

Assignment of CATH topologies (Class, Architecture, Topology, Homologous superfamily, Orengo *et al.*, 1997) for *E. coli* proteins was derived from the Gene3D database (version 2.0), which provides structural links to CATH for completely sequenced genomes. Gene3D was downloaded from ftp://ftp.biochem.ucl.ac.uk/pub/Gene3D/. Assignments in Gene3D are based on PSI-BLAST and subsequent verification (Buchan *et al.*, 2003).

3.5.2   *Protein sequence analyses*

Theoretical isoelectric points (pI) of protein sequences were calculated based on published pK values of amino acids (Bjellqvist *et al.*, 1993), utilizing the "Compute pI/Mw tool". This tool is available at the ExPASy proteomics server (http://www.expasy.org/tools/pi_tool.html).

The Grand Average of Hydropathy of protein sequences (GRAVY) (Kyte and Doolittle, 1982) was calculated using the software package CodonW (Peden, 1999). The program is available for download at http://www.molbiol.ox.ac.uk/cu/. A web interface can be accessed at http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html.

Functional assignment of *E. coli* proteins was derived from the COG database (Clusters of Orthologous Groups of proteins) (Tatusov *et al.*, 1997). COGs are based on phylogenetic classification of proteins encoded by multiple complete genomes. 17 distinct functional categories are assigned to COGs, which can be further summarized as subgroups of information storage and processing, cellular processes, metabolism, and poorly characterized proteins. Roughly 70 % of *E. coli* proteins are annotated in COGs. The database can be accessed *at* http://www.ncbi.nlm.nih.gov/COG/. Flat data files were downloaded from ftp://ncbi.nlm.nih.gov/pub/COG.

4   Results

This work was performed in close collaboration with Dr. Dean Naylor and Tobias Maier in the laboratory of Prof. Dr. F. Ulrich Hartl. Dean Naylor was involved in planning the experimental design as well as in the characterization of GroEL substrates identified by Houry *et al.* (1999). Tobias Maier purified the GroEL/GroES/substrate complexes, performed the GroEL depletion experiments and was involved in the second part of co-expression experiments. Mass spectrometry was performed by Dr. Yasushi Ishihama in the laboratory of Prof. Dr. Matthias Mann (University of Southern Denmark, Odense).

4.1   Classification of GroEL dependence by *in vitro* refolding assays

Since its discovery in the 1980s, the bacterial chaperonin system has been the subject of many detailed studies of its cellular function and molecular mechanism so that currently it is probably the best characterized molecular chaperone system. Despite this, studies on the chaperonin molecular mechanism have been mostly conducted with *E. coli* GroEL/GroES and heterologous substrates like *R. rubrum* rubisco, pig heart mitochondrial malate dehydrogenase or bovine liver mitochondrial rhodanese. Relatively little is known about the natural substrates of *E. coli* GroEL and how they fold in the living cell. These questions began to be addressed when 52 GroEL interacting proteins were initially identified by Houry *et al.* (1999). A number of these proteins (Table 1) were chosen for further analysis based on their readily assayable enzymatic activities, which can be used to monitor their *in vitro* refolding. An additional protein, dihydrodipicolinate synthase (DAPA), that had indirectly been reported to be GroEL dependent (McLennan and Masters, 1998) was also selected for further analysis. The behavior of these proteins in refolding reactions upon dilution from denaturant allowed their classification into three distinct classes of chaperone dependence.

In vitro refolding was followed by measurement of enzymatic activity of the respective protein at several time points following dilution from chemical denaturant into buffer solutions containing various combinations of chaperones and nucleotide. GroEL- and DnaK-mediated folding can be efficiently stopped by inhibition of their ATPase activity with EDTA or apyrase, when these compounds do not interfere with the subsequent enzymatic assay. Stopped folding reactions were then usually tested for enzymatic activity after completion of the time course. Spontaneous folding in buffer without chaperones was either followed by direct measurement of enzymatic activity during the refolding experiment, or folding was stopped by binding of unfolded polypeptide to GroEL in the presence of EDTA or to GroEL-Trap (GroEL-D87K, a mutant unable to hydrolyze ATP and thus unable to release bound polypeptide). Refolding yields are expressed as a ratio of regained enzymatic activity relative to activity of the native enzyme.

### 4.1.1   *Class I proteins are able to fold spontaneously*

Class I proteins were found to be able to fold spontaneously even without chaperone assistance. As shown in Figure 10A, enolase (ENO, a homodimeric protein of 45.5 kDa subunits) denatured in 6 M GdnHCl, reached roughly 55 % of its initial activity with a $t_{1/2}$ of

	SwissProt ID	Description	Molecular mass (monomer)	Quaternary structure
Class I	**ENO**	Enolase (2-phosphoglycerate dehydra-tase)	45.5 kDa	homodimer
	TDH	L-Threonine 3-dehydrogenase	37.2 kDa	homotetramer
	G3P1	Glyceraldehyde 3-phosphate dehydroge-nase (GAPDH-A)	35.4 kDa	homotetramer
Class II	**GATD**	Galactitol-1-phosphate 5-dehydrogenase	37.4 kDa	tetramer (probable)
	DCEA	Glutamate decarboxylase alpha	52.7 kDa	homohexamer
	SYT	Threonyl-tRNA synthetase (ThrRS)	74 kDa	homodimer
Class III	**METF**	5,10-Methylenetetrahydrofolate reduc-tase	33.1 kDa	homotetramer
	METK	S-Adenosylmethionine synthetase	41.8 kDa	homotetramer
	DAPA	Dihydrodipicolinate synthase (DHDPS)	31.3 kDa	homotetramer
	GATY	Tagatose-1,6-bisphosphate aldolase gatY (TBPA)	30.8 kDa	unknown

Table 1: E. coli GroEL interacting proteins, analyzed in vitro and in vivo in the course of this study.

~30 s upon 100 fold dilution into buffer A (see Materials and Methods, section 3.1.3.2) at 37 °C. Folding was monitored by measuring the enolase dehydration activity of 2-phosphoglycerate to phosphoenolpyruvate at the indicated times. The 55 % yield gained by spontaneous refolding could be increased to ~80 – 95 % through the addition of GroEL alone or GroEL combined with its cochaperone GroES. When ATP was omitted from a refolding reaction containing GroEL, folding did not occur, demonstrating efficient binding of un-folded enolase to GroEL. The DnaK/DnaJ/GrpE chaperone system was also able to increase the yield of ENO refolding to ~90 % at similar apparent rates as spontaneous and GroEL-mediated folding.

Proteins that fold spontaneously in dilute *in vitro* systems may, however, be depend-ent on chaperones inside the cell. This can be approached experimentally by the use of mac-romolecular crowding agents like Ficoll 70 in refolding experiments. These substances cre-ate a situation thought to mimic the environment in the living cell with high concentrations of macromolecules. Crowding agents decrease the available volume of solvent in a solution. This excluded volume effect leads to reduced diffusion coefficients and increased activity coefficients of the solutes (for review see Ellis, 2001, and Minton, 2001). Thus, macromo-lecular crowding often enhances protein aggregation, although in certain cases it can also
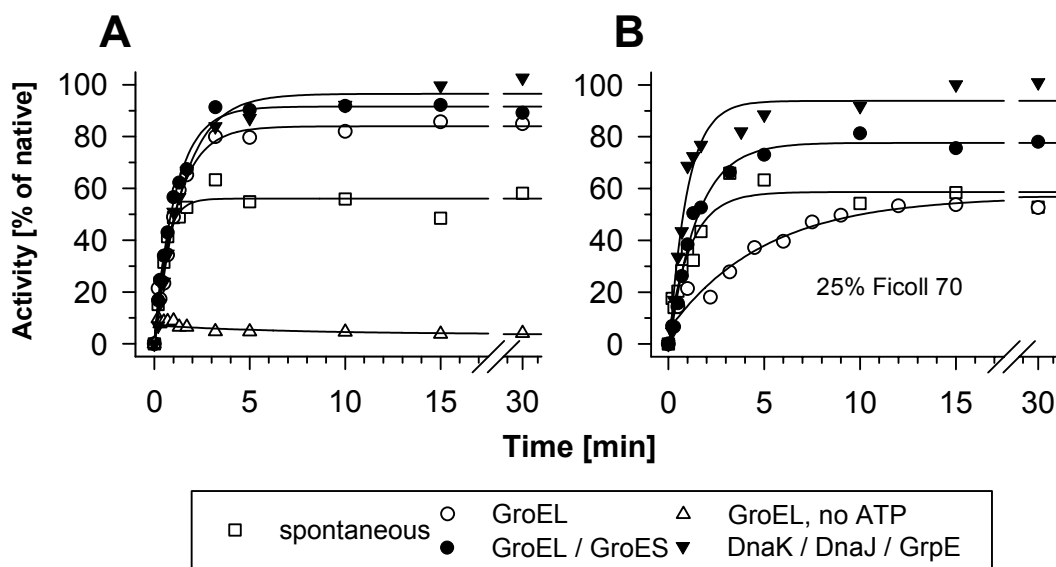
Figure 10: In vitro refolding of the class I protein ENO.

The class I protein ENO is able to fold spontaneously *in vitro* in diluted buffer and under conditions of macromolecular crowding. Chaperones can elevate the final yield of refolding. Denatured ENO was diluted 100-fold to 1 µM at 37 °C into (A) buffer A containing 5 mM ATP or (B) buffer A containing 5mM ATP and 25 % (w/v) of the crowding agent Ficoll 70 in presence of the indicated chaperone systems. The amount of refolding was assessed by measurement of enzymatic activity of ENO at indicated times.

accelerate protein folding (see introductory section 2.2.1).

As shown in Figure 10B, the refolding behavior of *E. coli* enolase both with and without assistance of chaperones did not change considerably in the presence of 25 % Ficoll 70 when compared to folding in diluted buffer. Spontaneous folding still occurred with a similar rate and yield, as did also GroEL/GroES and DnaK/DnaJ/GrpE assisted folding. The only notable difference was the slower apparent folding rate of the GroEL folding reaction without the cochaperonin GroES. This can be best explained by enhanced rebinding of unfolded enolase to the apical domains of GroEL. Lacking GroES to close the chaperonin cage, the unfolded substrate cannot undergo the normal chaperonin cycle to fold inside the GroEL/ GroES cavity, but is rather released directly into the bulk solution. The enhanced affinity of the GroEL apical domains to unfolded proteins, induced by macromolecular crowding (Martin and Hartl, 1997), results in quick rebinding of the still unfolded enolase and therefore slower refolding. This demonstrates the importance of the GroES cofactor for efficient chaperonin action in the cell.

The *S. cerevisiae* homolog of *E. coli* enolase was previously found to aggregate during spontaneous folding under conditions of macromolecular crowding (Martin, 2002). At the amino acid level, *E. coli* enolase is 46 % identical and 62 % similar to yeast enolase, so

the two proteins most probably share the same fold. In spite of this high similarity, the dependence on chaperone assistance for folding of these two proteins in a crowded environment appears to be different. The successful spontaneous refolding of *E. coli* enolase in contrast to yeast enolase in crowded solution shows the robustness of *E. coli* enolase folding. It therefore seems plausible that in the *E. coli* cytosol, the class I protein enolase is able to fold without or with little assistance of chaperones after leaving the ribosome-associated chaperones.

4.1.2   Class II proteins utilize DnaK or GroEL for folding

The second class of GroEL interacting proteins was highly aggregation sensitive and its members were dependent on chaperone interaction for efficient refolding from the denatured state. At 37 °C, glutamate decarboxylase alpha chain (DCEA; 52.7 kDa subunits, homohexameric, Figure 11A) and galactitol-1-phosphate 5-dehydrogenase (GATD; 37.4 kDa subunits, a putative homotetramer, Figure 11C) could not regain any detectable enzymatic activity upon dilution from denaturant into buffer without chaperones. The full GroEL/GroES system was capable of effectively refolding the two proteins. However, GroEL without GroES could not facilitate the folding of DCEA at 37 °C and showed only minimal folding activity for GATD at this temperature. At less stringent conditions, by decreasing the temperature from 37 °C to 25 °C, GroEL could fold the class II proteins DCEA and GATD without the GroES cofactor, as shown in Figures 11B and D, although with lower yields when compared to reactions with the full chaperonin system. GATD showed some spontaneous refolding without chaperones at 25 °C, albeit to a final yield of only ~10 %. This temperature-dependent folding behavior and the variation in GroES dependence suggested that these proteins may not constitute obligate GroEL/GroES substrates. Indeed, the DnaK system was similarly efficient in refolding DCEA and GATD (Figures 12A and B) and a combination of the GroEL and DnaK systems showed a noticeable additive effect on DCEA folding.

Threonyl-tRNA synthetase (SYT, a homodimer of 74 kDa subunits), with a molecular mass expected to exceed the size limit for encapsulation by the GroEL/GroES cavity, exhibited a slightly different refolding behavior. GroEL-mediated folding was only ~20 % efficient, irrespective of the presence of the cofactor GroES (Figure 13). However, the DnaK system was much more efficient in mediating the folding of SYT, leading to a final yield of ~70 % active protein.

A recent study has reported that yeast mitochondrial aconitase (a monomer of 82 kDa) can be efficiently folded by GroEL in a *trans* folding reaction that requires GroES binding to the opposite ring for release of nonnative aconitase without prior encapsulation (Chaudhuri *et al.*, 2001). A similar mechanism could also be shown (albeit with lower efficiency than *cis* folding) for smaller proteins that would be expected to be encapsulated inside GroEL/GroES (Farr *et al.*, 2003). In contrast to these studies, GroEL mediated refolding of SYT was found to be GroES-independent. To further examine the possibility of GroES-dependence, additional folding reactions were carried out in the presence of Ficoll 70 and 500 mM KCl. Both conditions are expected to impede the release of substrates from GroEL,
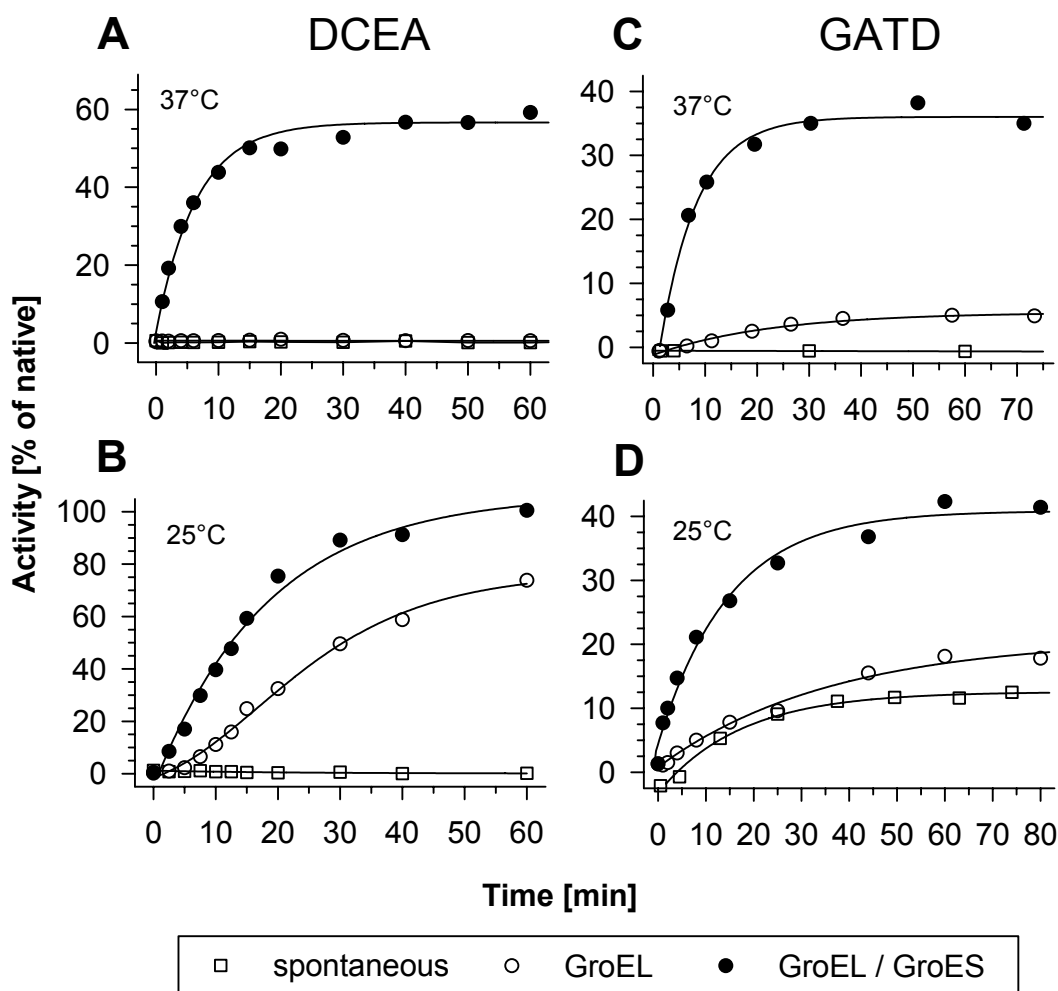
Figure 11: Spontaneous, GroEL- and GroEL/GroES-mediated in vitro refolding of the class II proteins DCEA and GATD.

Refolding of DCEA **(A, B)** and GATD **(C, D)** upon dilution from denaturant into buffer A containing the indicated chaperones and 5 mM ATP. Dependence on the cochaperonin GroES is temperature dependent; spontaneous refolding is observable for GATD at 25 °C only to a low extent. The complete chaperonin system is best suited to facilitate efficient refolding of DCEA and GATD. DCEA experiments were performed in cooperation with D. Naylor.

but GroES dependence of SYT refolding was still not observed (data not shown). There was also no observable difference between refolding temperatures of 25 °C and 37 °C.

Thus, while it appears likely that DnaK and GroEL can efficiently fold a number of common substrates in the preferred size range of GroEL (~20 to 60 kDa), proteins larger than 60 kDa are probably better suited for the DnaK system, consistent with the reported enrichment of DnaK substrates >60 kDa (Deuerling *et al.*, 1999; Mogk *et al.*, 1999).
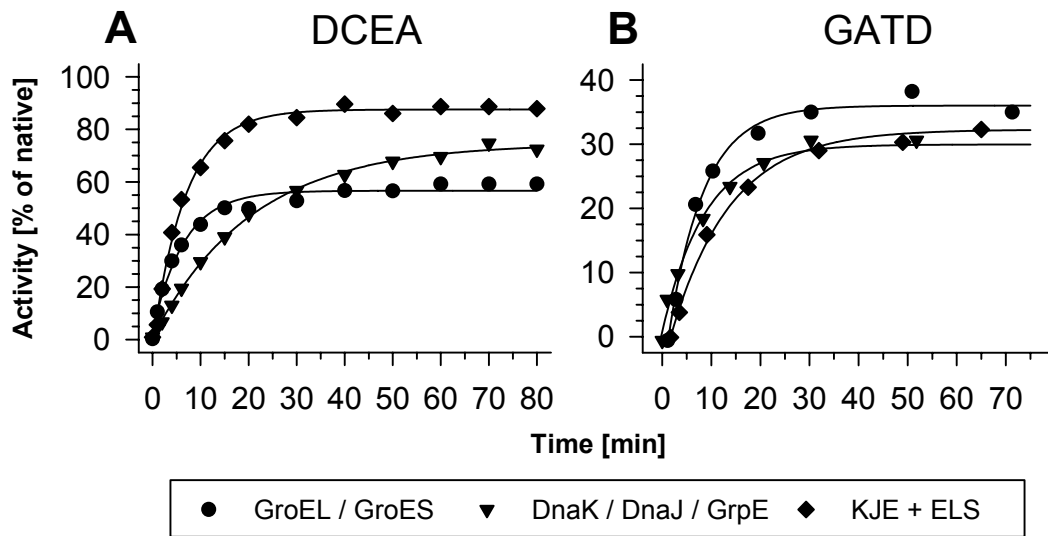
Figure 12: Influence of the DnaK system on refolding of the class II proteins DCEA and GATD.

GroEL/GroES- and DnaK/DnaJ/GrpE-mediated refolding of GdnHCl denatured class II proteins DCEA and GATD at 37 °C. Both chaperone systems can assist refolding of the two proteins to a similar degree, with a noticeable additive effect (KJE+ELS) in case of DCEA.

4.1.3   Class III proteins can be stabilized by DnaK but fold only upon transfer to GroEL

GroEL and GroES are essential gene products, so it seems reasonable that there exists at least one essential protein with absolute chaperonin dependence. A third class of GroEL interacting proteins comprises obligate chaperonin substrates. They include 5,10-methylenetetrahydrofolate reductase (METF, a homotetramer of 33.1 kDa subunits), S-adenosyl methionine synthetase (METK, a homotetramer of 42 kDa subunits) and dihydro-dipicolinate synthase (DAPA, a homotetramer of 31.3 kDa subunits). Notably, DAPA and METK are essential gene products (McLennan and Masters, 1998; Wei and Newman, 2002), and disruption of the *metF* gene leads to methionine auxotrophy (Blanco *et al.*, 1998). Figure 14 reveals that only the complete GroEL/GroES system is able to mediate the efficient refolding of METF ($t_{1/2}$ ~10 s, Figure 14A), METK ($t_{1/2}$ ~30 s, Figure 14C), and DAPA ($t_{1/2}$ ~4 min, Figure 14E) at 37 °C. Surprisingly, METF and METK were folded by GroEL/GroES at relatively fast rates compared to DAPA and the model substrates mitochondrial rhodanese (a monomer of 33 kDa) and bacterial rubisco (a homodimer of 50 kDa subunits, Brinker *et al.*, 2001). This suggests that the contribution of GroEL to protein folding in *E. coli* may be significantly greater than what has been previously estimated with heterologous substrates (Lorimer, 1996).
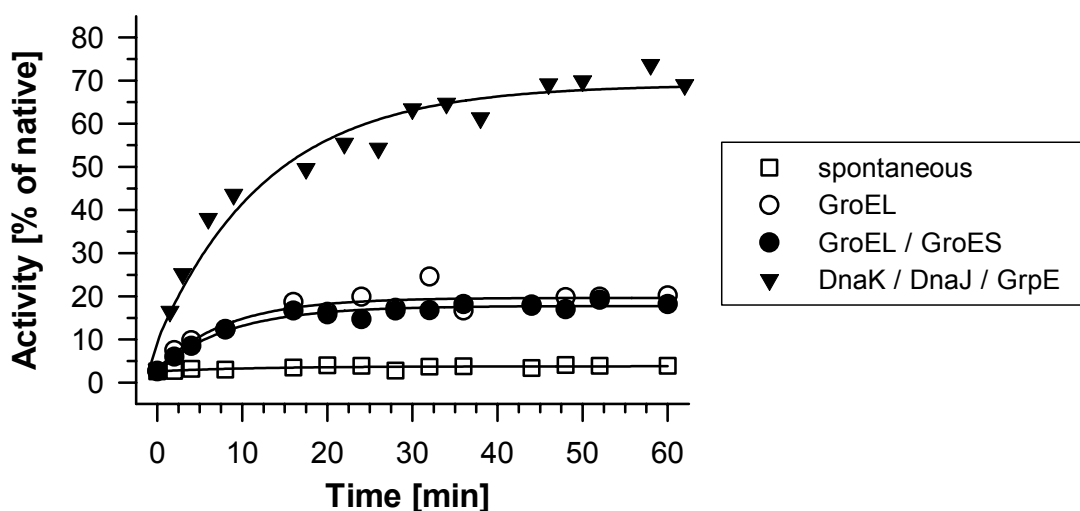
Figure 13: In vitro refolding of the 74 kDa class II protein SYT.

Denatured SYT was diluted at 37 °C into buffer A containing the indicated chaperones and 5 mM ATP. SYT exceeds the size limit of the GroEL/GroES cavity and thus cannot be encapsulated. Folding of SYT with the DnaK system is more efficient than chaperonin-mediated folding.

Unlike class I and II proteins, the DnaK system could not mediate the refolding of METF, METK and DAPA. However, DnaK could bind and thereby stabilize aggregation-prone, nonnative forms of these substrates and efficiently transfer them to GroEL for subsequent folding (Figures 14B, D, and F). The successive action of DnaK and GroEL in protein folding was first observed with mitochondrial rhodanese (Langer *et al.*, 1992a). This transfer can undoubtedly serve as a general pathway for the successful movement of aggregation-prone nascent-chains from the ribosome to GroEL and for efficient capture of aggregation-prone GroEL substrates denatured by stress. In agreement with this concept, METF and METK refolding yields were slightly higher when the unfolded substrates were first captured by DnaK (Figures 14B and D), consistent with the superior ability of DnaK, compared to GroEL, to capture aggregation-prone, non-native polypeptides (Mogk *et al.*, 1999). Furthermore, GroEL only binds a fraction of all newly synthesized polypeptides (~10 - 15 %) and the cellular levels of DnaK (~50 μM) are in molar excess over both ribosomes (~30 μM) and GroEL 14-mer (~3 μM, Ewalt *et al.*, 1997, Mogk *et al.*, 1999).

In summary, *E. coli* METF, METK and DAPA constitute the first obligate *E. coli* GroEL substrates identified. METK and DAPA are essential gene products, explaining the importance of GroEL/GroES for *E. coli* viability. DnaK can bind and stabilize METF, METK and DAPA but folding is only achieved upon transfer to GroEL. In this manner DnaK appears to function as a substrate reservoir for GroEL, facilitating the efficient capture of nascent chains and stress denatured proteins.
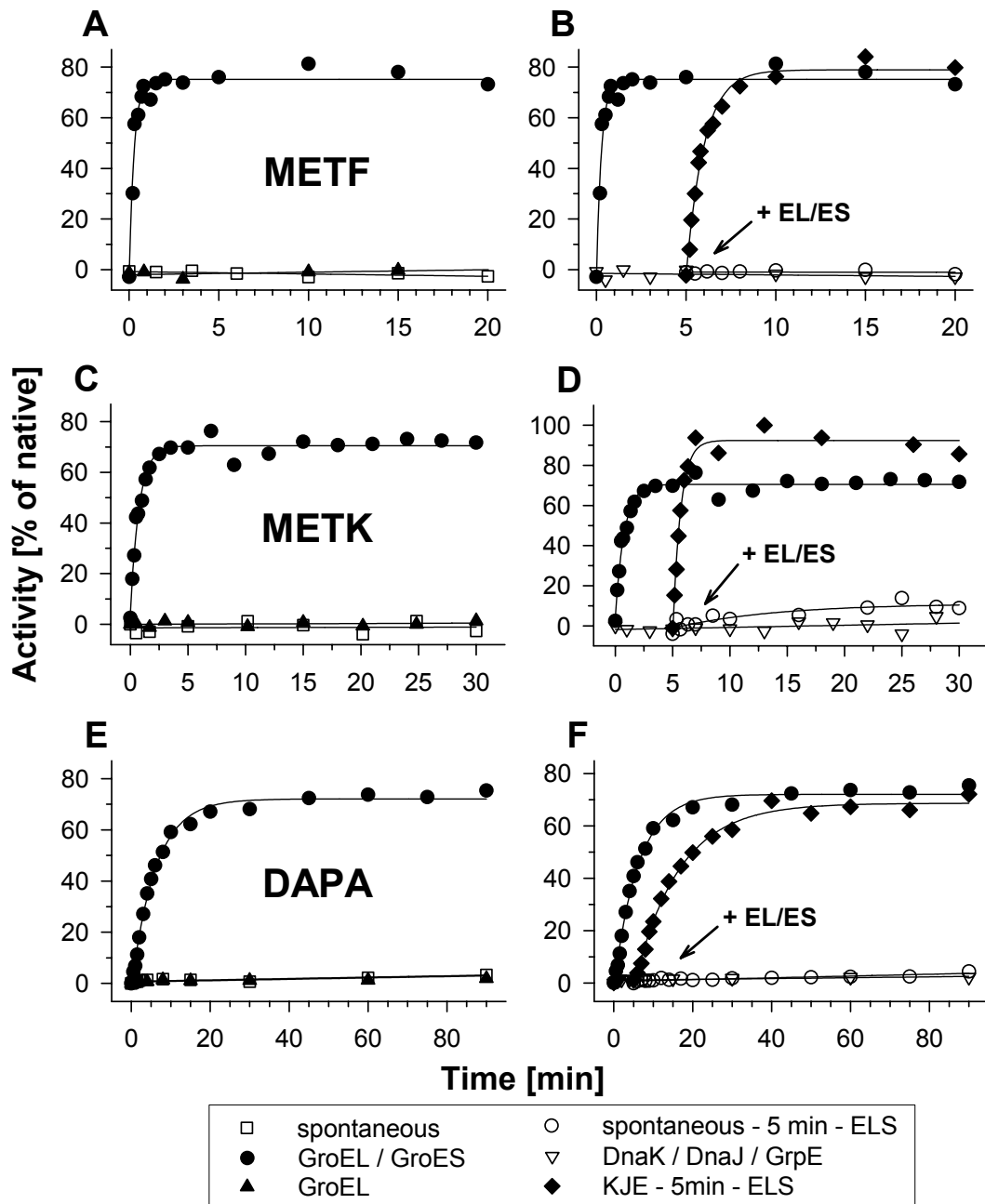
Figure 14: In vitro refolding of the class III substrates METF, METK and DAPA.

Efficient refolding of GdnHCl denatured METF **(A, B)**, METK **(C, D)** and DAPA **(E, F)** is de-
pendent on the full chaperonin system and is not detectable spontaneously or with GroEL in
absence of GroES **(A, C, E)**. DnaK/DnaJ/GrpE (KJE) cannot facilitate correct folding of class
III substrates **(A,C,E)** but are able to bind the unfolded polypeptides and transfer them to the
GroEL system (ELS), when this is added later in the reaction scheme (see arrow) **(B, D, F)**.
GroEL/GroES alone are not able to refold aggregated proteins upon attempted spontaneous
refolding. METK experiments were performed in cooperation with D. Naylor.

4.2 **The dependence of substrates on GroEL *in vivo***

#### *4.2.1 Overexpression of GroEL substrates in E. coli*

The variation in GroEL/GroES dependence of class I, II and III proteins was so far only examined *in vitro*. To investigate whether similar dependencies could also be observed in the cytosol of living *E. coli* cells, class I, II and III proteins were overexpressed in cells with either wild-type (~3μM) or elevated (~15 μM) levels of chaperonins. Different chaperonin combinations were employed for coexpression with the GroEL substrates, introducing either fully, or semi-productive chaperone machineries: (i) *E. coli* GroEL and GroES together, and (ii) GroEL without its cofactor GroES.

The overexpression of class I, II and III proteins, shown in Figure 15, was designed to saturate the available chaperones in wild type cells so that only a limited amount of protein became soluble. Comparison of these amounts permitted an assessment of the capacity of general chaperones to fold specific proteins. Elevating the levels of GroEL/GroES and GroEL ~5 fold before substrate overexpression allowed specific assessment of chaperonin contributions.

The class I proteins ENO and TDH were highly soluble in both cell lines, consistent with their ability to fold largely spontaneously without assistance from chaperones. In contrast, DCEA and GATD (class II) and METF, METK and DAPA (class III) were 60 - 70 % insoluble in wild type cells. Increasing the levels of the complete GroEL/GroES system, but not of GroEL alone, caused a 2 – 3 fold increase in solubility, indicating a limitation in endogenous chaperonin concentrations when overexpressing these proteins. Overexpression of GroEL alone, equivalent to a relative depletion of GroES, tended to reduce solubility of class III substrates, reflecting the requirement of these proteins for both GroEL and GroES.

The large class II protein SYT (74 kDa) was partially (~40 %) insoluble in wild type cells but was unaffected by chaperonin overexpression, both with and without the cofactor GroES. This supports the *in vitro* observation where the DnaK system was found to be superior to GroEL in assisting the folding of this large protein. GroEL would have been not expected to have a major effect on solubility of this protein.

A putative class III substrate is tagatose 1,6-bisphosphate aldolase (GATY, a probable homotetramer of 31.1 kDa subunits), which was almost completely insoluble upon overexpression in wild type cells and only slightly soluble upon additional expression of GroEL and GroES. Genetic analysis suggests that GATY is stabilized and its activity enhanced by GATZ (Brinkkötter *et al.*, 2002), whose function is otherwise unknown. Therefore, GATY and GATZ were coexpressed together in cells with wild type chaperone levels and in cells with elevated levels of GroEL/GroES. No improvement in the solubility of GATY was evident in either cellular background. Milligram amounts of soluble GATY could be purified successfully *via* an amino-terminal hexahistidine tag from cells with elevated levels of GroEL/GroES. However, the protein aggregated within ~30 min when stored in a variety of buffers. GATZ was not co-purified with GATY, even when GATZ (~40 % soluble) and GATY were overexpressed together (at similar levels). The unstable nature of GATY combined with its low activity level, prevented reliable *in vitro* refolding studies to
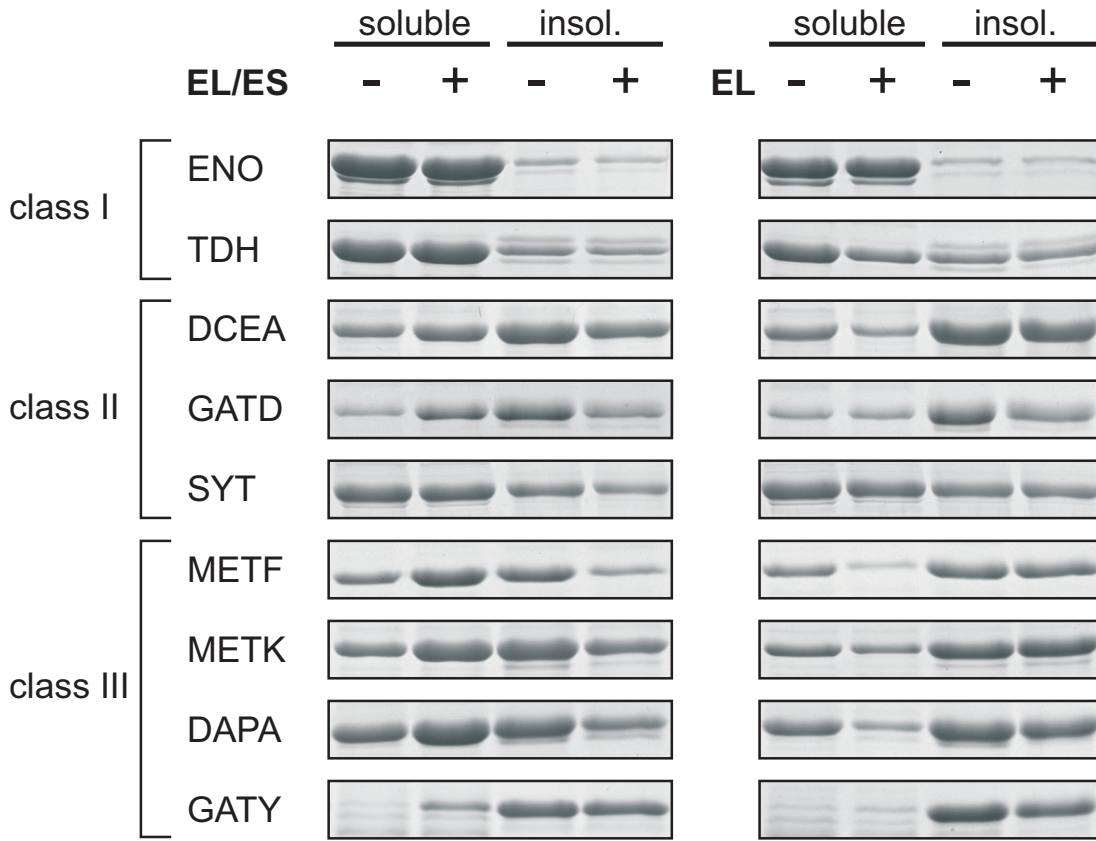
Figure 15: Solubility of GroEL substrates upon co-expression in E. coli with chaperonins.

E. coli cells were grown at 37 °C to exponential phase in LB-medium containing 0.4 % (w/v) glucose. Chaperones were induced (+) from a pBAD promoter with 0.4 % arabinose or further repressed (-) with glucose. Substrates were subsequently induced at 37 °C (or at 30 °C for GATY) from a T7 promoter with 1 mM IPTG. Samples were resolved on 12 or 16 % SDS-PAGE. Equivalent amounts of soluble and insoluble fractions were loaded. Prepared in co-operation with D. Naylor.

be undertaken.

4.2.2 Cellular depletion of GroEL/GroES results in misfolding of obligate substrates

The fate of GroEL interacting proteins was next examined in cells depleted of GroEL/GroES to establish chaperonin dependency at endogenous concentrations of substrates. An *E. coli* strain in which the GroE promoter was exchanged by the arabinose-controlled araBAD (P_{BAD}) promoter was prepared by Dr. Costa Georgopoulos (Université de Genève). GroEL levels decreased by more than 90 % within 3 hours upon shifting the cells from arabinose to glucose containing medium (Figure 16). Utilizing such a strain, it was previously observed that GroEL/GroES depletion is accompanied by the loss of the protein DAPA, indirectly suggesting that DAPA could be an obligate substrate of chaperonins

(McLennan and Masters, 1998). Alternatively, it could be argued that a positive regulator of DAPA synthesis is lacking in chaperonin depleted cells. The results of McLennan and Masters (1998) could be confirmed and the presented *in vitro* refolding and chaperone coexpression experiments provide the first clear evidence that DAPA is indeed an obligate GroEL/GroES substrate. As an extension of the study by McLennan and Masters, the expression pattern of total, soluble and insoluble material of DAPA and other known class I, II and III proteins over a time course of GroEL/GroES depletion were examined. Class I proteins (ENO, TDH) remained soluble throughout GroEL/GroES depletion (Figure 16). Similarly, the class II protein GATD was not affected in its solubility but showed a non-uniform expression behavior during this experiment, a phenomenon probably linked to the medium change from arabinose to glucose (Nobelmann and Lengeler, 1996). SYT, a large class two protein, was expressed constitutively throughout the time course of the experiment and was
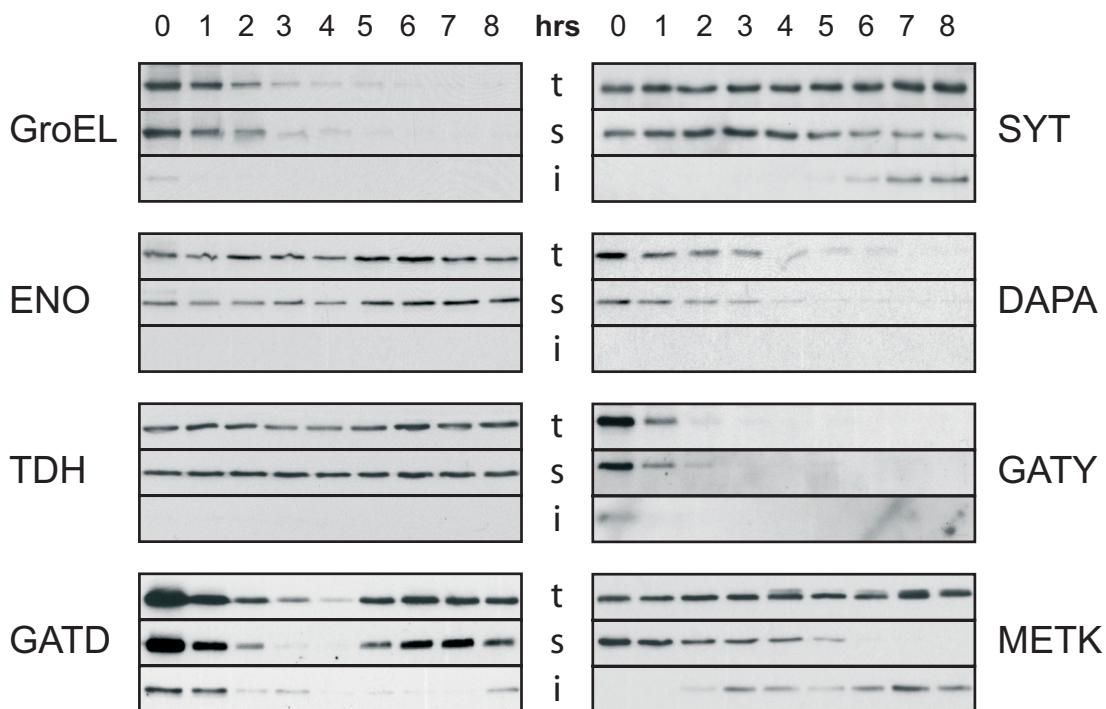
Figure 16: Solubility of GroEL substrates upon depletion of GroEL/GroES.

In the depletion strain, derived from *E. coli* MC4100, the native GroE promoter has been replaced with the araBAD promoter and the regulatory *araC* gene, generating a strain dependent on arabinose for continued GroEL/GroES production and viability. For GroEL/GroES depletion, mutant cells grown at 37 °C in LB-medium with 0.2 % (w/v) arabinose were shifted to LB-medium containing 0.2 % (w/v) glucose (T=0 hr). At indicated times, equivalent numbers of cells were taken for preparation of total (t), soluble (s) and insoluble (i) protein fractions. Samples were separated by SDS-PAGE and immunoblotted with antibodies against GroEL, ENO, TDH, GATD, SYT, DAPA, GATY and METK. Experiments performed in cooperation with T. Maier.

only partially insoluble upon prolonged chaperonin depletion. In contrast, the class III substrates showed an absolute requirement for GroEL/GroES. DAPA disappeared from the total and soluble fractions without accumulating in the insoluble fractions, suggesting that endogenous levels of this protein are efficiently degraded when unable to fold and aggregation only prevails upon overexpression. Rapid disappearance was observed also with GATY, suggesting a strong chaperonin dependence of this protein. Note that GATY is coexpressed with GATD as the two proteins are encoded in the same operon (Nobelmann and Lengeler, 1996). The rapid decrease in the levels of the substrates, following chaperonin depletion, could also suggest that GATY and DAPA are intrinsically unstable and may require returning to GroEL/GroES throughout their life times for conformational maintenance. It is noteworthy to mention that the depletion experiments were carried out in an MC4100 strain, while the substrate/chaperone coexpression studies were performed in BL21(DE3) Gold cells, that are deficient in the Lon protease, the major protease for degradation of unfolded proteins in *E. coli* (Goldberg *et al.*, 1994). Therefore, the ability to detect aggregated DAPA and GATY in the coexpression experiments is probably due in part to the absence of the Lon protease as well as to the use of substrate over-expression, causing saturation of both chaperones and available proteases.

Other proteins, such as METK, maintained stable cellular levels but disappeared from the soluble fraction and accumulated as aggregates upon GroEL/GroES depletion. These observations confirm the classification of GroEL-interacting proteins based on their *in vitro* refolding properties and allow a clear distinction between substrates grouped in classes II and III. Because the reduction in available GroEL capacity does not decrease the folding efficiency of class II substrates, these proteins appear to rely largely on the DnaK system or other chaperones for folding and rather use GroEL/GroES more extensively only when overexpressed.

Several studies (including this) have noted that during GroEL/GroES depletion or inactivation, there is a subsequent increase in the soluble levels of numerous proteins, including DnaK, ClpB and METE. Intriguingly, while DnaK and ClpB levels are raised ~2 - 4 fold, presumably to assist in dissaggregation and folding of chaperonin substrates, METE levels increase so substantially that it becomes the most abundant cellular protein. The synthesis of METE in *E. coli* is repressed in part by vitamin B12 and in part by the METJ repressor protein and its co-repressor S-adenosyl methionine (Cai *et al.*, 1992). Production of vitamin B12 requires both functional METF and METH while synthesis of S-adenosyl methionine requires functional METK. Based on the known regulative mechanisms of METE expression and the results described so far, the drastic induction in METE synthesis is most likely due to a loss in functional METF and METK. These observations again emphasize the strong correlation between the observed findings *in vitro* and *in vivo* for the dependencies of various proteins on GroEL/GroES for efficient folding.

4.3   Identification of the *in vivo* GroEL/GroES substrate proteome

In an attempt to identify the feature(s) that determines chaperonin dependency, the GroEL/GroES proteome was identified and exhaustively analyzed. A previous study of

GroEL interacting proteins by Houry *et al.* (1999) used coimmunoprecipitation with GroEL antibodies under nucleotide-free conditions followed by 2D-gel mass spectrometry. 52 GroEL-bound proteins were identified and a preference for multiple αβ domains was established. In the present study, further examination of a subset of these proteins revealed that they faithfully constitute GroEL interacting proteins. However, the occurrence of multiple αβ domains is not characteristic for a specific class of GroEL dependent proteins, suggesting that the original dataset of 52 proteins, with 24 amenable to structural analysis at the time, may have been too small to reveal a more detailed recognition motif.

In the present study, substrates were trapped within the *cis*-cavity of GroEL under a (His)$_6$-tagged GroES lid (Figure 17). The GroEL/GroES-(His)$_6$ complexes were fixed in the ADP-bound state upon lysis of live spheroplasts in the presence of glucose and hexokinase to rapidly (in < 3 sec) convert cellular ATP to ADP, followed by isolation of the complexes by immobilized metal affinity chromatography (IMAC). Special care was taken to exclude post lysis exchange of substrates associated with GroEL/GroES (see below). This method allowed the purification of sufficient amounts of GroEL/GroES/substrate complexes to perform a large-scale, high-accuracy mass spectrometric analysis that lead to the identification of the virtually complete set of GroEL interacting proteins in *E. coli* (Figure 18). The complexes were separated on one-dimensional SDS-PAGE, followed by in-gel trypsin digestion of sliced gel fragments and reversed phase liquid chromatography coupled to quadrupole time-of-flight mass spectrometry for peptide identification (LCMS) (Lasonder *et al.*, 2002).

GroEL/GroES complexes formed with *E. coli* GroES-(His)$_6$ proved to be of limited stability during isolation, raising the possibility of post-lysis loss or exchange of substrates. However, efficient recovery of GroEL complexes was achieved upon short-term expression
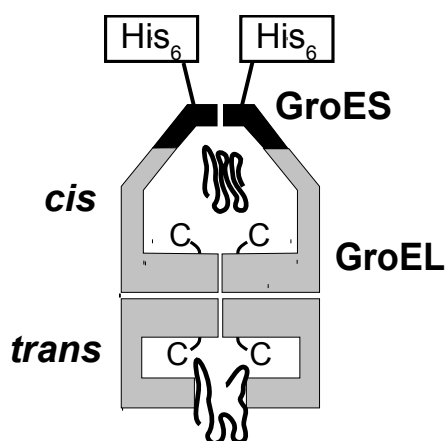

Figure 17: Illustration depicting the capture of a substrate protein within the central cis-cavity of GroEL/GroES in vivo.

Stable GroEL/GroES/substrate complexes can be isolated in the ADP state by immobilized metal affinity chromatography *via* seven hexahistidine tags on heptameric GroES.

of the highly similar GroES of *Methanosarcina mazei* (Mm). *Mm*GroES can functionally replace *E. coli* GroES *in vivo* (Figueiredo *et al.*, 2004; Klunker *et al.*, 2003), but was found to bind more stably to GroEL in the presence of ADP. It is important to note that GroEL but not GroES is responsible for substrate selection. This ensures that the captured proteins represent authentic substrates.

The seven carboxy-terminal hexahistidine tags on the GroES oligomer permitted very stringent washing conditions. Substrate complexes were not removed from a Talon IMAC resin with 50 mM imidazole but were efficiently released with 200 mM imidazole (Figure 18, lane 2). Utilizing the same conditions for cells where GroES lacking the hexahistidine tag was expressed (Figure 18, lane 3), only seven proteins were detected by mass spectrometry. These proteins were thus considered to be non-specifically bound to the IMAC resin and were excluded from further analysis. They include EFTU, FABZ, FUR, GLMS, RL32, RS15, and SLYD (see Supplementary Table S1). To test potential post-lysis exchange of GroEL bound substrates, intact cells with overexpressed (His)$_6$-tagged GroES were mixed with Arg-^{13}C6 labeled wild type cells and lysed together as described above. In addition to the non-specific binding proteins (with the exception of RL32), only a further 25 Arg-^{13}C6 labeled proteins could be identified as associated with GroEL in the resulting complexes
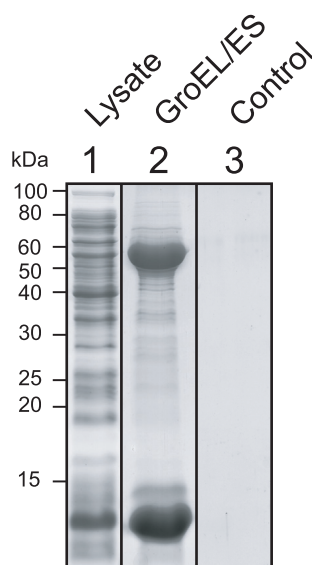
Figure 18: Purification of GroEL/GroES complexes with encapsulated substrate proteins.

Spheroplasts of *E. coli* MC4100 expressing (His)$_6$-tagged GroES were rapidly lysed in presence of glucose and hexokinase to convert all cellular ATP to ADP (lane 1). Stable GroEL/GroES/substrate complexes were bound to IMAC resin and eluted with 200 mM imidazole (lane 2). To identify unspecific binding of proteins to IMAC, GroEL/GroES/substrate complexes were prepared using a non tagged version of GroES and treated as before (lanes 3). Samples were subjected to 16 % SDS-PAGE, followed by Coomassie staining. Experiments performed in cooperation with T. Maier.

(Supplementary Table S1). This result demonstrates that post-lysis cycling of GroEL/GroES/ substrate complexes and re-binding of different proteins is very limited and does not significantly influence the results. There were 19 ribosomal proteins among the non-specific interactors. Therefore ribosomal proteins were excluded from further analysis.

Chaperonin-substrate complexes were isolated from MC4100 cells grown under a number of different cellular conditions, including rich and minimal medium at 37 °C, and rich medium at 30 °C and 23 °C. Isolated complexes were resolved by SDS-PAGE to separate the vast excess of GroEL and GroES from the captured substrates. Horizontal gel slices were subjected to trypsin digestion, the peptides extracted and analyzed by LCMS. This procedure identified repeatedly more than 200 GroEL interacting proteins for cells grown under the various conditions from a total of ~2 500 soluble cytosolic proteins. This number is expected to be close to the entire population of chaperonin substrates, which was previously estimated to be ~10 – 15 % of cytosolic proteins by mass (Houry *et al.*, 1999).

The GroEL substrate list (Supplementary Table S2) contained all of the above analyzed class I, II and III proteins, with the exception of the class I protein TDH (probably due to its very low cytosolic concentration, see below).

4.3.1   GroEL/GroES complexes contain encapsulated and trans ring-bound substrates

GroEL-GroES complexes arrested in the ADP state should contain substrate proteins encapsulated under the lid of GroES (the *cis* cavity) undergoing folding. In addition, some proteins may be bound to the GroEL *trans* ring (Farr *et al.*, 2003) (Figure 17). Indeed, several proteins predicted to be too large for encapsulation (> 60 kDa, Sigler *et al.*, 1998) were identified, including the class II substrate SYT (74 kDa).

Previous studies had shown that GroES can prevent the entry of proteinase K (PK) into the cavity of the GroEL ring to which it is bound, thereby effectively protecting the flexible carboxy-termini of GroEL and captured non-native substrates from degradation. On the other hand, unfolded substrates bound in the *trans*-ring of GroEL were easily degraded by PK, allowing the protease to cleave the carboxy-termini of GroEL in the open ring (Langer *et al.*, 1992a; Mayhew *et al.*, 1996; and see Figure 17). To test the above hypothesis, similar experiments were performed with isolated GroEL/GroES/substrate complexes in presence of ADP. When treated with PK, only about half of the carboxy-termini of GroEL were protected, giving rise to a characteristic GroEL double-band on SDS-PAGE (Figure 19). Western blot analysis revealed that the tested substrates <60 kDa in subunit size were protected from the protease. This suggests their efficient encapsulation under the GroES lid. The same substrates were either partially (ENO) or completely (METK) digested when tested with purified proteins in their native state, not bound to GroEL. The 74 kDa protein SYT, however, was quickly and completely degraded, consistent with its inability to be encapsulated (Figure 19).
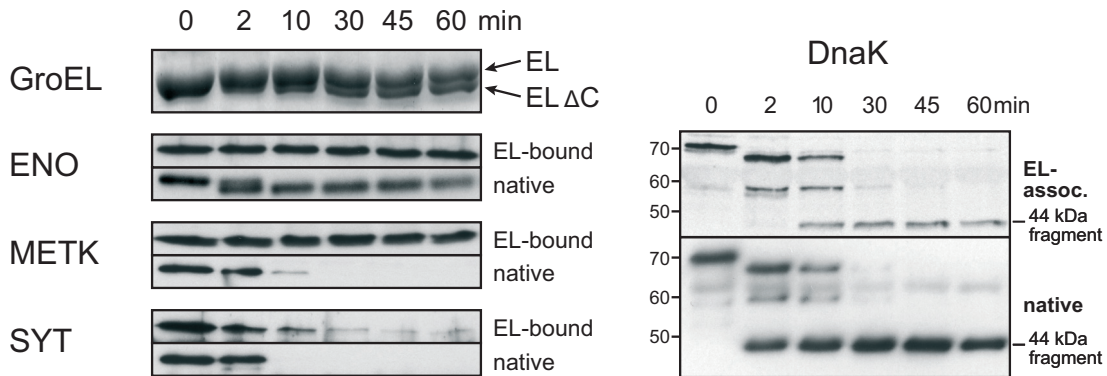
Figure 19: PK digestion of GroEL/GroES/substrate complexes.

Complexes and the native purified proteins, as controls, were subjected to PK digestion and the reactions were stopped at indicated times. Samples were subjected to SDS-PAGE and silver stained (GroEL) or immunoblotted for the proteins indicated (ENO, METK, SYT and DnaK).

A number of chaperones, including DnaK, DnaJ, GrpE and TF, were observed among the list of GroEL interacting proteins. Given the ability of the DnaK system to target aggregation-prone unfolded proteins to GroEL, the possibility that DnaK was interacting in its native state with large substrates (>60 kDa) bound to the *trans* ring of GroEL was tested. When PK-treated complexes were immunoblotted for DnaK, it could be shown that the chaperone was cleaved into its stable 44 kDa ATPase domain (Liberek *et al.*, 1991), resulting in a similar pattern on SDS-PAGE as native purified DnaK when subjected to protease treatment (Figure 19). This indicates that DnaK was indeed folded and not encapsulated, most likely as a result of being associated with unfolded substrates bound to GroEL in *trans*.

4.3.2   *GroEL interacting proteins are enriched in class III proteins*

As described above, only a subset of GroEL interacting proteins are obligate class III substrates that depend strictly on GroEL for successful folding. Other interactors are either largely chaperone independent (class I) or can use alternative chaperone systems for folding (class II). Based on this classification and the known members of each individual class, criteria were deduced to identify class III substrates among all GroEL interacting proteins. The capacity of GroEL in *E. coli* is limited: the cytosolic concentration of oligomeric GroEL is estimated to be ~3 µM (Ellis and Hartl, 1996) and about 10 – 15 % of all cytoplasmic proteins were shown to interact with GroEL under normal growth conditions (Ewalt *et al.*, 1997). Therefore, obligate class III proteins are not expected to be of high abundance but should be enriched among all GroEL interacting proteins, since they cannot be folded by other chaperone systems in the cell. This becomes evident when considering that the general pathway of chaperone-mediated folding upon protein synthesis at the ribosome includes the primary action of TF and the DnaK system, followed by transfer of not yet fully folded pro-

teins to the chaperonin system (Deuerling *et al.*, 1999; Langer *et al.*, 1992a; Teter *et al.*, 1999). Class I and II proteins are therefore expected to complete their folding along this pathway and not reach GroEL/GroES.

To test the hypothesis of class III protein enrichment on GroEL, a quantitative mass spectrometric method of relative concentration determination was employed. Lysate was prepared from Arg-[13]C6 labeled wild type cells and mixed in known amounts with unlabeled purified GroEL/GroES/substrate complexes (Figure 20). Using this 'stable isotope labeling by amino acids in cell culture' (SILAC, Ong *et al.*, 2002; Ong *et al.*, 2003), it was possible to accurately determine relative concentrations of GroEL substrate proteins in the complex preparations in comparison to their concentrations in the cell lysate. Peptides derived from tryptic digestion of a protein from Arg-[13]C6 labeled cell lysate and peptides of the same protein from isolated (unlabeled) GroEL complexes are detected as separate peaks in mass spectrometry, due to their differing mass. Areas of the different isotope peaks in the same spectrum can then be directly compared, thus allowing relative concentration measurement by mass spectrometry. Using SILAC, the GroEL bound fraction of substrate proteins was compared to their total cellular amount. GroEL is in large excess in the complex preparations, so the peak areas of GroEL derived peptides could not be directly used to internally normalize the two samples in regard to their GroEL content, as a result of overloading the mass spectrometer with GroEL peptides. However, GroEL levels were quantified by immuno-blotting, thus allowing determination of the total cellular fraction of substrate proteins bound to GroEL at any time.

A total of 252 different proteins were found to interact with GroEL in wild type MC4100 cells (Supplementary Table S2). A single pulldown experiment typically resulted in detection of 200 - 250 protein, confirmed by multiple mass spectrometry experiments. For comparison, mass spectrometry experiments performed with wild type MC4100 cell lysate resulted in detection of $\sim 800 - 1000$ individual *E. coli* lysate proteins in a single experiment. To enhance specificity of GroEL substrate assignment, a protein was only considered to be a real GroEL interactor if it could be detected under at least two different conditions of substrate preparation (37, 30 or 23 °C, in rich and minimal growth media) or if the concentrations of the protein could be determined by SILAC. SILAC measurements involved repeated manual inspection of the mass spectrometry data, thus offering higher reliability. However, not every protein was amenable to SILAC concentration measurement, since both peaks from lysate and complex preparations have to be above noise level in the same spectrum.
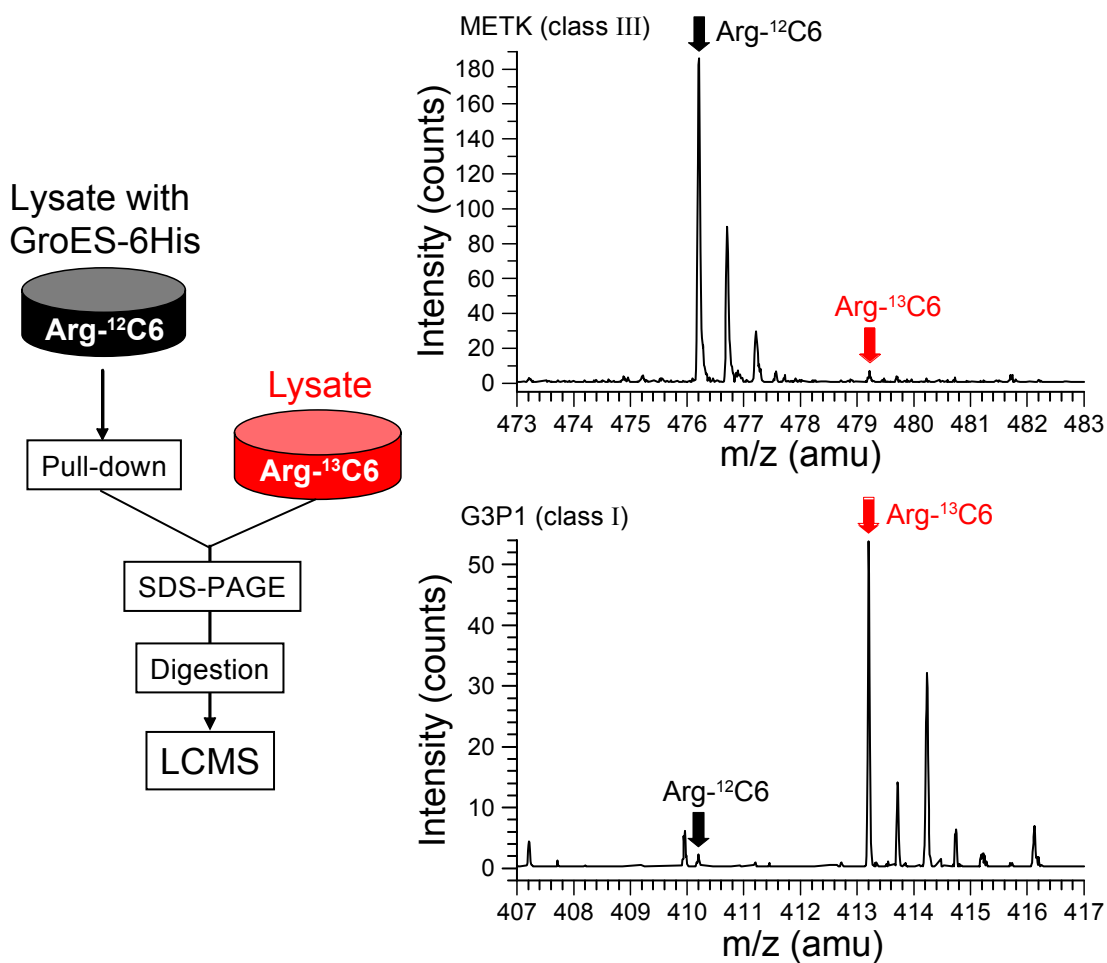
Figure 20: Measurement of the fraction of GroEL-associated substrate protein by SILAC.

The **left panel** illustrates the experimental set-up. GroEL/GroES/substrate complexes isolated from unlabeled cells containing GroES-(His)$_6$ were mixed at an appropriate ratio with soluble lysate proteins from cells labeled with Arg-^{13}C6. The mixture was subsequently separated by SDS-PAGE, followed by reduction/alkylation, in-gel digestion by trypsin, and analysis by LC-MS/MS. **Right panel**: Mass spectra of peptide FFINPTGR (doubly charged) from the class III protein METK (top) and of peptide VGINGFGR (doubly-charged) from the class I protein G3P1 (bottom). Ratios of unlabeled (Arg-^{12}C6) to Arg-^{13}C6-labeled peak intensities were measured by the software MSQuant and converted to calculate the fraction of GroEL-associated protein, taking the amounts of GroEL in the starting material and the mixing ratio into consideration.

The validated class III proteins METK, METF and GATY were among the most highly enriched substrates, with $3 - 6\%$ of their total cellular content being GroEL-associated. More than 3 % of a protein is expected to be GroEL-associated when all folding must proceed via GroEL (assuming a doubling time of E. coli of $30 - 40$ min and an average $t_{1/2}$ of GroEL-assisted folding of ~60 s; see Ewalt et al., 1997). An additional 81 proteins

accumulated in GroEL/GroES complexes to 4 % or more of total and thus were assigned to class III (see Supplementary Table S2). This group comprises proteins with a wide variety of cellular functions. It includes thirteen essential proteins, as defined by Gerdes *et al.* (2003) who used a genetic footprinting technique for genome-wide assessment of genes required for robust aerobic growth of *E. coli* MG1655. No preference to a particular class of enzymes or other functional group of proteins could be observed (see below).

In contrast, for validated class I proteins, less than 0.02 % of total was found to be GroEL-associated, indicating that their folding is essentially GroEL independent. The substrate set contained 42 other proteins sharing this property (Supplementary Table S2). They included eighteen essential proteins and represented highly abundant cytosolic proteins, as will be described below.

The remaining ~125 proteins were tentatively grouped in class II. For these proteins, including GATD and SYT, between 0.1 and 2.6 % of total was recovered on GroEL, indicative of partial GroEL dependence. This group included 36 essential proteins.

It is important to point out that the classification of the identified GroEL interaction proteome on the basis of relative concentrations measured by SILAC is to be utilized as a predictive tool rather than an absolute assignment. The boundaries between different classes of GroEL dependence might indeed be variable and dependent on different factors including cellular stress situations. However, as will be shown below, the trends of GroEL dependence, in particular for obligate class III substrates, are very well predicted by this approach.

4.3.3   *General properties of GroEL interacting proteins*

To explain the observation that GroEL is essential for cell viability under all tested conditions (Fayet *et al.*, 1989), it would be expected that GroEL is required to promote the efficient folding of at least one other essential protein. However, essential gene products are not enriched among obligate GroEL substrates (Figure 21). On the contrary, the average fraction of ~20 – 25 % essential gene products in the *E. coli* genome and in the experimentally determined *E. coli* lysate is increased to ~50 % in class I proteins, while obligate class III substrates include only ~17 % essential proteins. The total average of GroEL interacting proteins, as well as class II substrates, are slightly enriched in essential proteins (~30 %).

The molecular mass distribution of predicted class III substrates displays a median of 37.7 kDa, which is larger than that of total lysate proteins (median 33.3 kDa) and displays a relatively sharp cut-off for proteins >50 kDa (Figure 22A). About 80 % of class III proteins are between 20 and 50 kDa and only five are larger than 65 kDa, including one essential protein, subunit A of topoisomerase IV (84 kDa). This size distribution is consistent with a dependence on the encapsulation mechanism for the vast majority of class III substrates. Also, small proteins exhibiting molecular masses <20 kDa are of low abundance among class III proteins, in line with the idea that they are able to fold spontaneously. Interestingly, the other GroEL interactors (class I and II) do not show such a strong preference for a size that fits the central cavity of the GroEL/GroES complex. The deviation of their size distribution from lysate proteins is much smaller. Only small proteins <10 kDa are predominantly absent in the set of GroEL interactors.

	essential	dispensable
genome	620	2506
lysate	262	737
GroEL interacting	67	160
class I	18	19
class II	36	78
class III	13	63
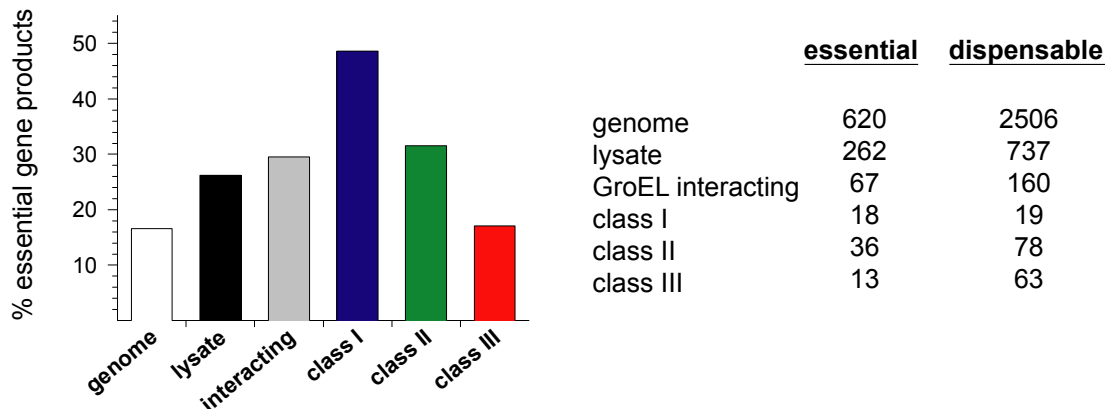
Figure 21: Essential gene products among E. coli proteins and different classes of GroEL interacting proteins.

Percentage of gene products, known to be essential for growth of *E. coli* (left) and absolute numbers of known essential and dispensable gene products for *E. coli* proteins and different classes of GroEL interacting proteins (right). Essentiality and dispensability of gene products as determined by Gerdes *et al.* (2003).

The distribution of theoretical isoelectric points (pI) of GroEL interactors deviated from that of total lysate proteins. A substantially greater fraction of the former exhibited pI values closer to 7 (Figure 22B). This indicates that at physiological pH, these proteins have a lower net charge than an average cytosolic protein (in both absolute terms and when expressed relative to size; data not shown), a property known to enhance the tendency of proteins to aggregate upon attempted refolding (Chiti *et al.*, 2002).

Hydrophobicity of GroEL interacting proteins was examined by calculating theoretical GRAVY (GRand AVerage of hydropathY) values. The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence (Kyte and Doolittle, 1982). More positive values indicate more hydrophobic sequences and more negative values more hydrophilic sequences. The GRAVY distribution of class III substrates tends to display higher values and class III proteins are thus predicted to be more hydrophobic than an average lysate protein (Figure 23). Class I and II proteins, on the other hand, do not exhibit such a shift of hydropathy. Although this effect is apparently quite small, in addition to the observed pI shift towards less charged proteins, it reveals that class III proteins are generally aggregation prone proteins, since both factors contribute to increase the tendency for nonspecific aggregation of unfolded polypeptides (Chiti *et al.*, 2002; Fink, 1998).
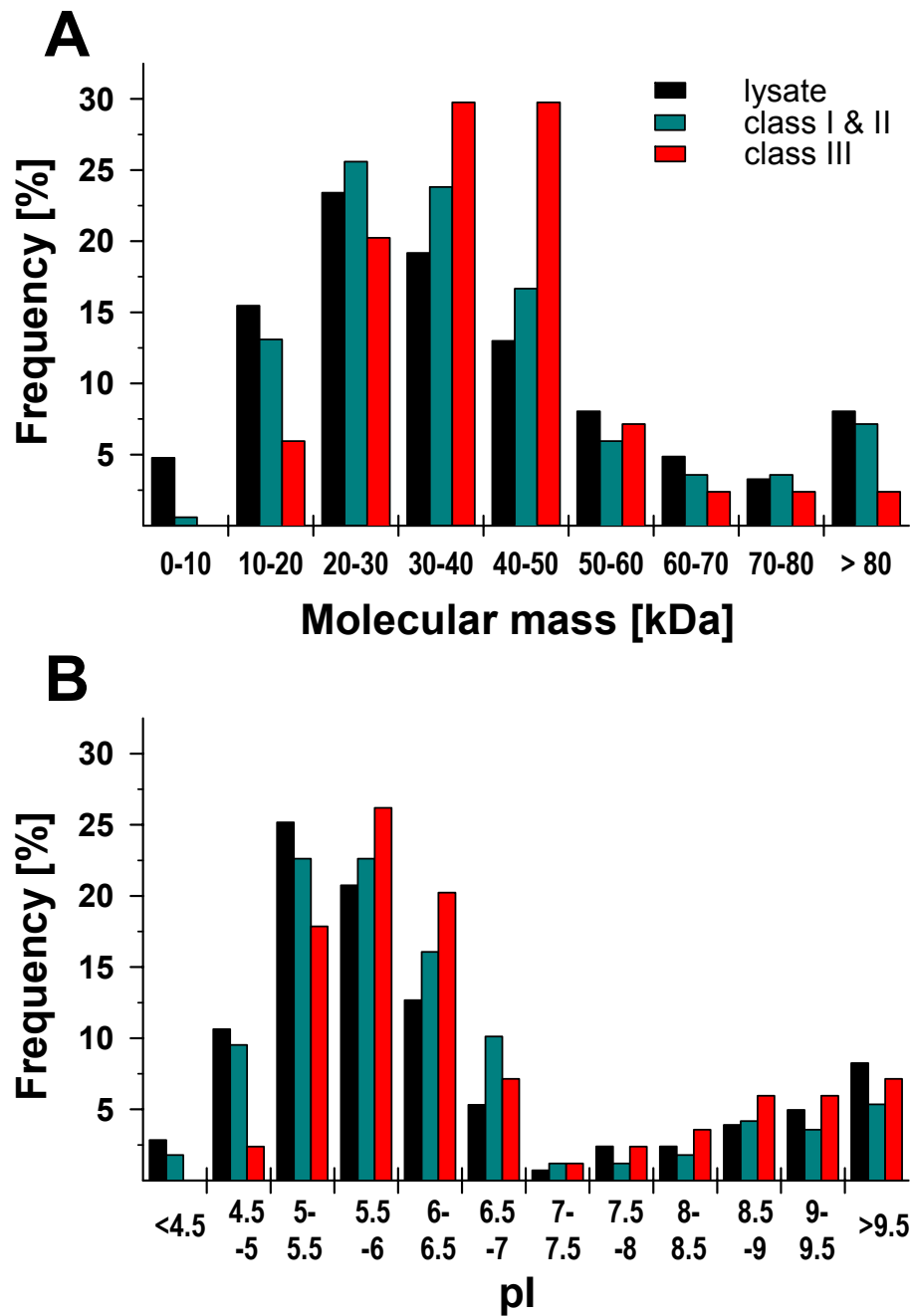
Figure 22: Properties of GroEL interacting proteins.

(A) Molecular mass distribution. Class III substrates preferentially exhibit molecular masses that can fit the GroEL/GroES cavity. **(B)** Distribution of isoelectric points. Class III proteins are shifted towards more neutral pI.
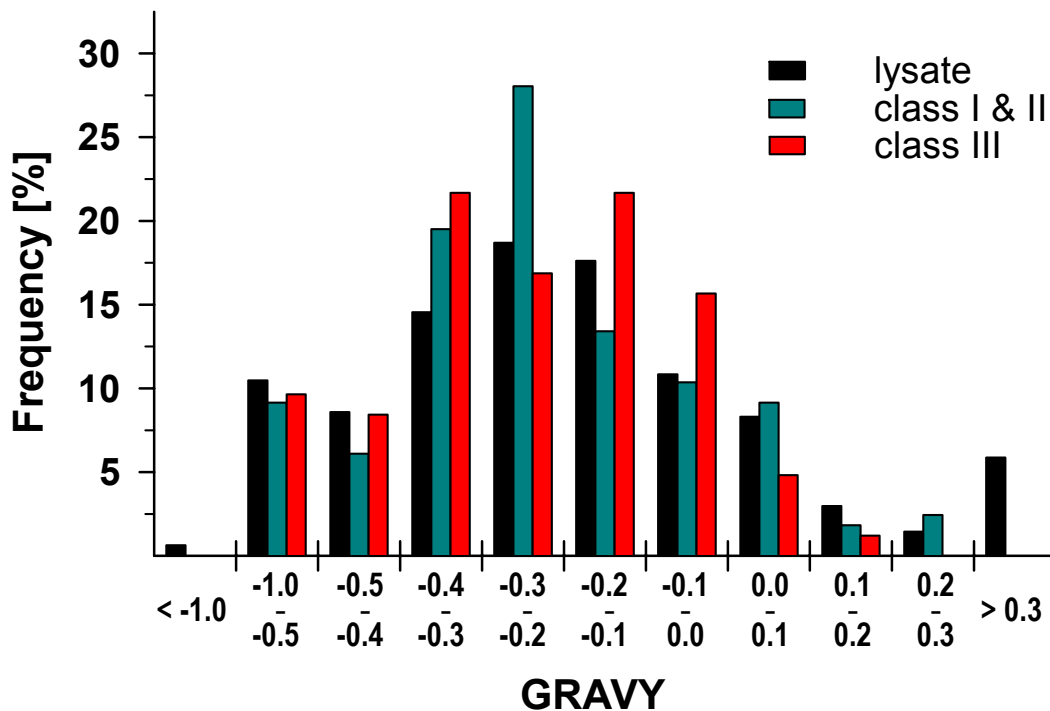
Figure 23: Hydropathy distribution of GroEL substrate classes.

A higher GRAVY (grand average of hydropathy) value indicates a more hydrophobic protein (Kyte and Doolittle, 1982). Class III substrates display a small tendency to be more hydrophobic than the average of *E. coli* lysate proteins.

4.3.4   *Class III GroEL substrates are enriched in the TIM (βα)₈ barrel fold*

To determine, whether GroEL dependence correlates with a specific type of fold or folding motif, a homology-based fold assignment search was performed with the set of GroEL interacting proteins, the experimentally identified lysate proteins and the complete *E. coli* proteome by querying the protein sequences against the SCOP database (Structural Classification Of Proteins; Lo Conte *et al.*, 2002; in collaboration with D. Frishman, Institute for Bioinformatics, German National Center for Health and Environment, GSF, Neuherberg). In a second approach, the sequences were matched to the CATH database (Class, Architecture, Topology, Homologous superfamily, Orengo *et al.*, 1997) by utilizing the CATH supplementary database Gene3D which provides structural assignment links to CATH for proteins within completely sequenced genomes (Buchan *et al.*, 2003).

SCOP and CATH both organize proteins in hierarchies: SCOP uses classes, folds, superfamilies and families while CATH uses classes, architectures, topologies and homologous superfamilies. SCOP classification is mainly performed manually, unlike CATH which incorporates some degree of automation in classifying structures. Differences between the two systems arise, amongst others, from the subjective task of separating proteins into do-

mains and differing class assignments. An outstanding example is the Rossmann fold family which represents the most highly populated fold in CATH and is classified into multiple families in SCOP: Flavodoxin-like, NAD(P)-Rossmann fold domains, P-loop-containing nucleotide triphosphate hydrolases, S-adenosyl-L-methionine-dependent methyltransferases and more. The reason for this discrepancy is due to the fact that CATH uses a broader, more geometrical sense, for classification of the general Rossmann fold motif, while SCOP uses rather more evolutionary directed criteria (Hadley and Jones, 1999).

The observed fold distribution for the *E. coli* proteome was nearly identical to the one found in the experimentally identified *E. coli* cell lysate for both SCOP and CATH, as shown for CATH in Figure 24. Further comparison could therefore be restricted to the lysate data. The advantage of utilizing this dataset is that these proteins were in fact detected by the mass spectrometry approach. In this manner only proteins that could be identified by this method were utilized for comparisons.

3.40.50	Rossmann fold
1.10.10	Arc Repressor Mutant, subunit A
3.30.70	Alpha-Beta Plaits
3.20.20	TIM Barrel
3.40.190	D-Maltodextrin-Binding Protein; domain 2
2.60.40	Immunoglobulin-like
2.40.50	OB fold (Dihydrolipoamide Acetyltransferase, E2P)
3.40.640	Aspartate Aminotransferase; domain 2
3.40.630	Aminopeptidase
1.10.287	Helix Hairpins
3.30.565	Heat Shock Protein 90
3.50.50	FAD/NAD(P)-binding domain
3.40.30	Glutaredoxin
1.10.260	434 Repressor (Amino-terminal Domain)
3.30.420	Nucleotidyltransferase; domain 5
1.25.40	Serine Threonine Protein Phosphatase 5, Tetratricopeptide repeat
3.90.550	Spore Coat Polysaccharide Biosynthesis Protein SpsA; Chain A
3.90.77	UDP-n-acetylmuramoyl-l-alanine
1.20.120	Four Helix Bundle (Hemerythrin (Met), subunit A)
3.10.20	Ubiquitin-like (UB roll)

Figure 24: Distribution of CATH topologies, assigned to the theoretical E. coli proteome and experimentally identified lysate proteins.

Gene3D (Buchan *et al.*, 2003) assigned folds for the complete *E. coli* proteome and for the experimentally derived lysate show only minor differences in the distribution of CATH topologies.

GroEL substrates were enriched in the TIM barrel fold with a high statistical significance (Figure 25) both in SCOP (fold c.1) and in CATH (b, fold 3.20.30). In SCOP, this fold is shared by 6.8 % of all proteins with an identified structural homology in the lysate (55 out of 814 lysate proteins). The complete set of GroEL/GroES interacting proteins, however, contains 17 % protein sequences with strong homology to the TIM barrel fold (35 out of 210), and stringent GroEL class III substrates are further enriched in TIM barrel proteins to 28 % (18 out of 65). These results strongly indicate an extended need of TIM $(\beta\alpha)_8$ barrel proteins on GroEL/GroES for effective folding, as this particular fold is fourfold more abundant among the subset of stringent GroEL proteins compared with the total cell lysate. Importantly, more than a quarter of all class III substrates share this fold. The CATH database fold distribution reveals a nearly identical profile for the TIM barrel proteins, since both databases classify this particular fold very consistently.

All of the TIM barrel representatives interacting with GroEL exhibit molecular masses between 23 and 54 kDa (30 – 54 kDa for class III substrates) and are thus likely to fold in a *cis* reaction inside the GroEL/GroES complex. When the fold distribution analysis was restricted to just proteins in the size range of 15 – 60 kDa, which is the predicted for encapsulation as well as the observed size range for most GroEL interactors (see above), no significant difference to the analysis including proteins of all molecular masses was observed. In particular, the ratio of TIM barrels remained constant (data not shown). Therefore, the enrichment in the TIM $(\beta\alpha)_8$ fold cannot simply be explained as a particular feature of polypeptides exhibiting the mentioned size range.

Among the class III TIM barrel proteins, the experimentally confirmed class III substrates METF, DAPA and GATY were found. To confirm the correlation that proteins with TIM barrel fold display a dependence on GroEL for efficient folding, the properties of eight representatives of newly identified class III TIM barrels were further analyzed [ADD (36.4 kDa), ALR2 (38.9 kDa), END4 (31.5 kDa), HEM2 (35.5 kDa), LLDD (42.7 kDa), NANA (32.5 kDa), XYLA (49.7 kDa), and YAJO (36.4 kDa)]. Their dependence on GroEL *in vivo* was tested by expression in the two cell lines harboring elevated levels of complete or incomplete chaperonin systems (GroEL/GroES or GroEL alone), as described above. Indeed, every one of the TIM barrel proteins tested displayed a behavior as would be predicted according to the previously validated class III proteins (Figure 26, compare with Figure 15). All proteins were soluble to some degree when overexpressed in a wild type background and solubility could be considerably increased with elevated amounts of GroEL/GroES. Like GATY, soluble ALR2 was only detectable when expressed in presence of elevated GroEL/GroES, indicating the absolute requirement of these proteins for GroEL/GroES. Coexpression of GroEL alone with the TIM barrel proteins, equivalent to a lack of GroES in the system, led to a decrease in the amount of soluble protein in every case, although to different degrees. The effect on ALR2, ADD, END4 and LLDD was mild. On the other hand, HEM2, NANA, XYLA and YAJO exhibited a more drastic decrease in soluble protein amounts upon coexpression with GroEL alone. The behavior of these proteins appears to be
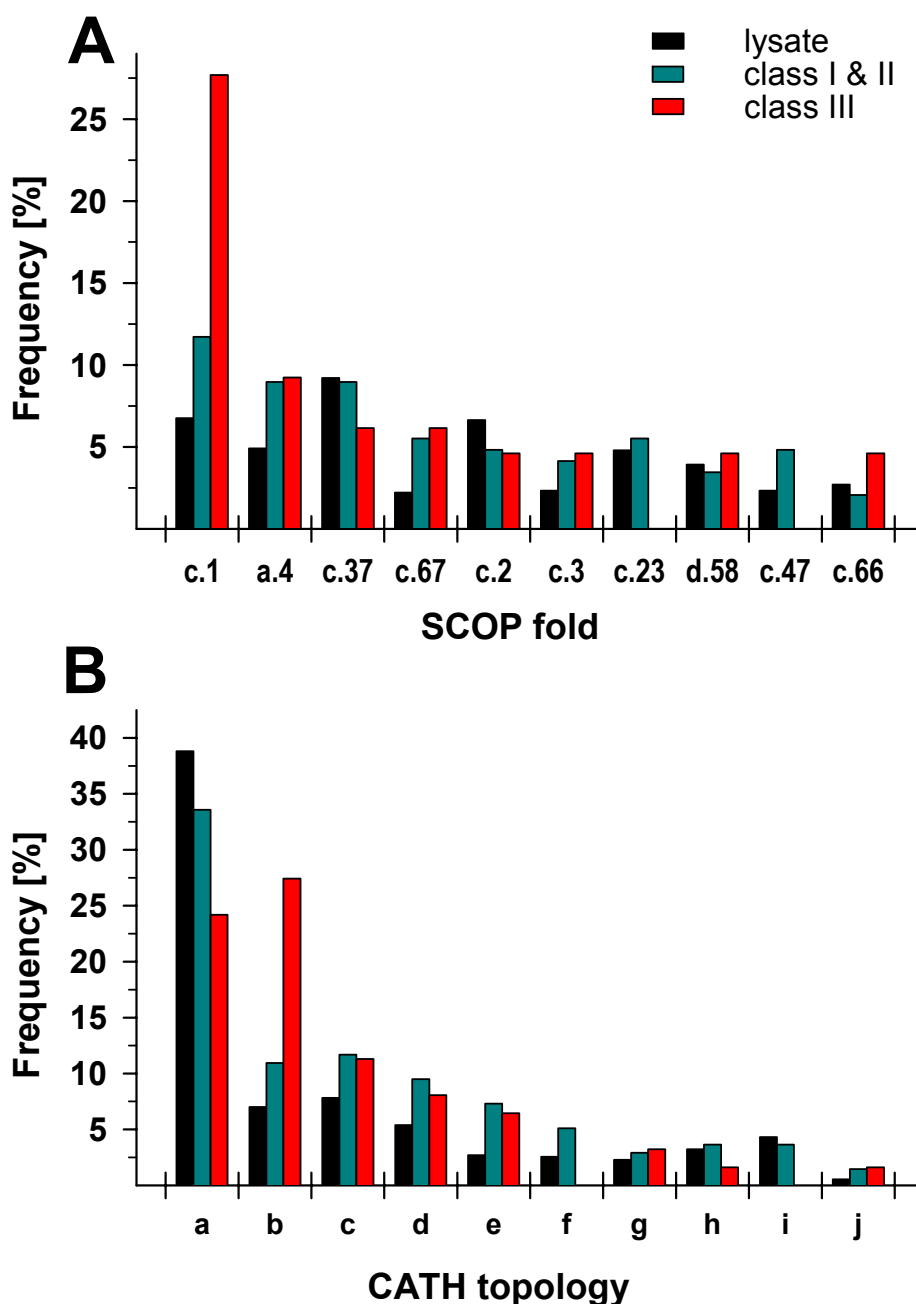
Figure 25: Fold distribution of GroEL substrate proteins according to the SCOP (A) and CATH (B) databases.
(A) c.1: TIM β/α-barrel; a.4: DNA/RNA-binding 3-helical bundle; c.37: P-loop containing nucleotide triphosphate hydrolases; c.67: PLP-dependent transferases; c.2: NAD(P)-binding Rossmann-fold domains; c.3: FAD/NAD(P)-binding domain; c.23: flavodoxin-like; d.58: ferredoxin-like; c.47: thioredoxin fold; c.66: S-adenosyl-L-methionine-dependent methyl-transferases. **(B)** a: Rossmann fold (3.40.50); b: TIM barrel (3.20.20); c: α-β plaits (3.30.70); d: Arc repressor mutant, subunit A (1.10.10); e: aspartate aminotransferase; domain 2 (3.40.640); f: glutaredoxin (3.40.30); g: FAD/NAD(P)-binding domain (3.50.50); h: D-maltodextrin-binding protein; domain 2 (3.40.190); i: OB fold (dihydrolipoamide acetyltrans-ferase, E2P) (2.40.50); j: enolase-like; domain 1 (3.30.390).
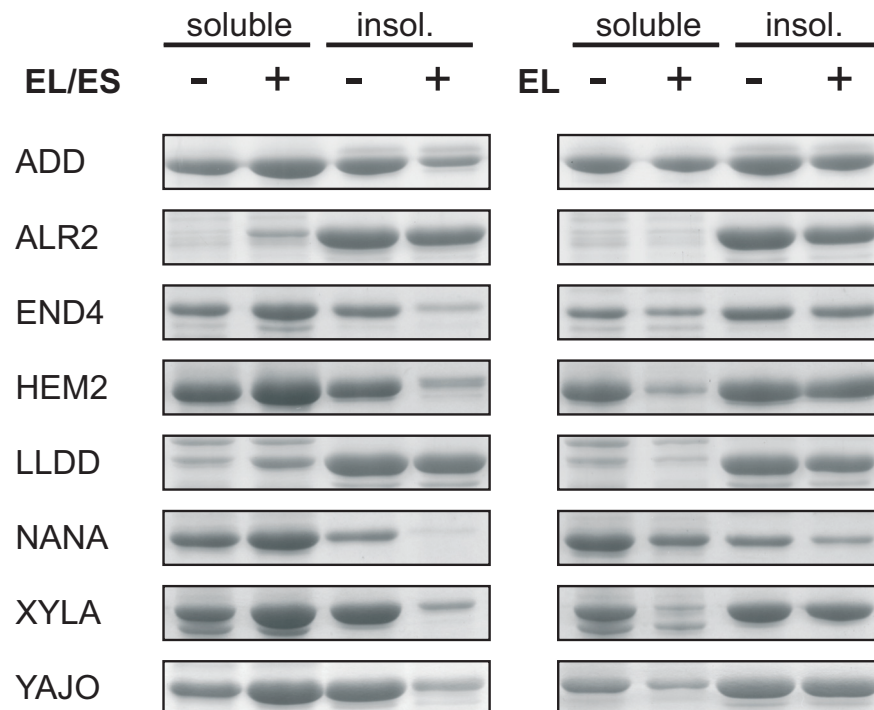
Figure 26: Solubility of newly identified class III TIM barrels upon co-expression in E. coli with chaperonins.

Experimental conditions were as described for Figure 15. The chaperones GroEL/GroES (EL/ES) and GroEL only (EL) were induced (+) or repressed (-), followed by class III TIM barrel substrate expression. Equivalent amounts of soluble and insoluble material was resolved on SDS-PAGE. For identities of tested proteins, please see main text and Supplementary Table S2.

correlated with the amount of soluble protein produced in the absence of chaperonin. If solubility is relatively high when the protein is overexpressed by itself, co-expression of GroEL (without GroES) leads to a very marked decrease in solubility.

These experiments confirm that the additionally tested TIM barrel proteins are stringent class III substrates, with the exception of ALR2. Due to its intrinsic insolubility under the employed experimental conditions it is not possible to assign it to either classes II or III. However, like GATY, the high enrichment on GroEL would strongly suggest that ALR2 is an obligate GroEL dependent protein.

Heterologous expression in the eukaryotic cytosol, which lacks a bacterial-type chaperonin, provides a stringent system to independently test the validity of the classification of newly-synthesized GroEL substrates. A set of class III proteins, including METK and 10 TIM barrel substrates, were moderately expressed in different wt *S. cerevisiae* strains from galactose-inducible promoters. Remarkably, all of these proteins accumulated in the insolu-

ble fraction, but were essentially soluble when both GroEL and GroES were expressed in addition (Results Hung-Chun Chang, Hartl laboratory, data not shown). Thus, the requirement of the class III proteins for GroEL/GroES is specific and independent of the bacterial machinery of protein synthesis. In contrast, ENO (class I) as well as 3 class II proteins tested were soluble upon expression in wt yeast. Substantial aggregation of the class II proteins was observed in the mutant strain *Δydj1* that lacks the yeast Hsp70 cofactor Ydj1p, supporting the conclusion that class II proteins are chaperone-dependent but can utilize either the Hsp70 system or GroEL/GroES for folding.

The TIM barrel fold is one of the most common domain structures present in proteins. Accordingly, it is very abundant in the *E. coli* cytosol: at ~7 % in the *E. coli* lysate, it is among the three most abundant SCOP folds. In CATH, it also represents a highly populated topology although not as common as the CATH Rossmann fold. Clearly, the TIM barrel fold *per se* cannot be the sole criterion that determines whether a protein displays absolute chaperonin dependence to reach its native structure. One example of this is the protein enolase (ENO), which adopts a TIM barrel fold in its native state but (as shown above) is a very robustly folding protein in the absence of any chaperones. Therefore, a more detailed search for a common structural feature of the obligate GroEL/GroES TIM barrel substrates was undertaken. A slight preference for the SCOP TIM barrel superfamily of metallo-dependent hydrolases was observed. All proteins of this superfamily which were identified as GroEL interacting proteins were predicted as class III. However, this result may not be statistically significant, since the stringent GroEL/GroES substrate set comprises only a total of 18 TIM barrel proteins. Further sub-classification of this small dataset renders any detailed analysis statistically insignificant.

The TIM barrel fold represents the only strongly enriched fold among class III substrates, but substrates belonging to class III also adapt other types of folds. Enriched among the total of GroEL interacting proteins, but not further enriched among class III substrates, is the DNA/RNA-binding 3-helical bundle fold (SCOP fold a.4). 5 % of the lysate proteins share this fold, compared to 10 % in total GroEL interacting as well as also 10 % of only the class III proteins. This fold is rather small. It consists of three bundled or partly opened α-helices and was mostly detected as an additional domain in substrate proteins. It is not very likely that this fold *per se* would render a protein dependent on GroEL for efficient folding. On the other hand, addition of a further domain indeed increases the complexity of protein folding and it is therefore not surprising to observe an enrichment of multi-domain proteins among GroEL substrates.

Other folds, *e.g.* Flavodoxin-like (SCOP fold a.23), are not found in the stringent substrate set but they are present in the lysate and total GroEL interacting proteins. This fold appears to be generally not dependent on GroEL for folding. The OB fold (CATH topology 2.40.50) is also not present in class III proteins, while observed in the lysate (~4 %) and in the total GroEL interactors (~3 %). It is especially enriched among class I substrates (8 %). The highly populated CATH Rossmann fold topology is less abundant among class III proteins. This behavior is not easily detectable when searching SCOP, since the Rossmann fold is divided into 13 different folds in this database.

When utilizing the distribution of CATH architectures, which reside one level above topologies in the CATH hierarchy, the results obtained are very similar. The *E. coli* lysate reveals a nearly identical distribution to the total proteome, and the architecture containing the TIM barrel topology, 3.20 ($\alpha\beta$ barrel topology), is the only statistically significantly enriched topology among the total GroEL interactors (data not shown). This enrichment is most pronounced for class III substrates (~27 %) while there is little enrichment of the barrel topology among class I and II proteins (11 % and 12 %, respectively, *versus* 8 % in the lysate). Two different architectures are noteworthy: class I proteins are about twofold enriched in proteins adopting the 2-layer sandwich architecture (3.30), with no significant change in the ratio of class II and III proteins compared to the lysate. The β-barrel architecture (2.40) is also enriched about twofold in class I proteins. Interestingly, class II and especially class III proteins with this architecture were much less abundant when compared to the lysate. The aforementioned OB fold, which is not populated in class III proteins, is a topology of the β-barrel architecture. The 2-layer sandwich and especially the β-barrel architecture thus appear to exhibit a generally low chaperone requirement.

4.3.5   Functional categories among GroEL interacting proteins

It was previously not known whether GroEL interacts preferably with groups of proteins belonging to the same cellular pathway or performing related tasks. In order to perform such a comparison for GroEL substrates *versus* the experimentally derived lysate proteins, a particular function had to be assigned to each of these proteins. The available classification of protein function provided by annotated comments in the SwissProt database, a curated protein sequence database available *via* ExPASy (Gasteiger *et al.*, 2003), is not very systematic and contains too many empty entries to be useful for this purpose. On the other hand, the EcoCyc database (Karp *et al.*, 2002), a pathway and genome database that describes the metabolic and signal-transduction pathways of *E. coli* as well as its enzymes, transport proteins and mechanisms of transcriptional control of gene expression, proved to be too detailed for such a comparison. The optimal source for functional annotation was the database of Clusters of Orthologous Groups of proteins (COGs) (Tatusov *et al.*, 1997). COGs are derived by comparing protein sequences from multiple complete genomes. Each COG consists of individual orthologous proteins or orthologous groups of paralogs from minimally three lineages and therefore corresponds to a conserved domain of ancient origin (Orthologs are defined as direct evolutionary counterparts related by vertical descent, whereas paralogs are homologous genes within the same organism, derived from gene duplication). Typically, proteins belonging to the same COG share a specific function. The *E. coli* genome is fully annotated in the COG database and thus the information therein proved to be very useful for comparative analysis of the function of the GroEL substrates.

When comparing the *E. coli* lysate to the total of GroEL interacting proteins, the observed distribution of functional categories was found to be similar (data not shown). The largest differences were detected in proteins involved in translation (4.3 % of lysate proteins, 7.1 % of GroEL interactors), amino acid transport and metabolism (5.1 % of the lysate, 8.7 % of GroEL interactors), carbohydrate transport and metabolism (13 % of the lysate,

10.7 % of GroEL interactors) and energy production and conversion (4.0 % of the lysate, 7.1 % of GroEL interactors). These differences, however, do not appear to be very significant. Thus, GroEL does not obviously interact primarily with certain kinds of functional groups of proteins. By further dividing this distribution into the three classes of GroEL substrates (Figure 27), proteins involved in energy production and conversion (lysate: 4.0 %, class III: 9.5 %) and secondary metabolites biosynthesis, transport and catabolism (lysate: 10.7 %, class III: 15.5 %) were found to be slightly enriched among class III substrates. For the different classes of GroEL substrates, there was also no clear preference for a specific functional category. The discovered significant enrichment of TIM barrel proteins in class III is not in conflict with this observation, since the TIM barrel fold provides an extremely versatile structure to carry out various enzymatic reactions. TIM barrel proteins perform a wide variety of different tasks in the cell and are not restricted to certain functional categories. GroEL dependence is therefore probably rather the result of other structural features that are not exclusively found in certain functional groups of proteins.

4.3.6   Concentration of GroEL substrate proteins in the E. coli lysate and in GroEL/GroES/substrate complexes

In order to assess the contribution of GroEL to protein folding in *E. coli* in general, and to measure the quantitative distribution of GroEL substrate classes within the total of GroEL interacting proteins, it was necessary to obtain information on protein concentrations of the individual GroEL substrates in the *E. coli* lysate as well as in GroEL/GroEL/substrate complexes. Therefore, a semi-quantitative method of mass spectrometry data interpretation was employed (Y. Ishihama, manuscript submitted to Molecular Cellular Proteomics, February 2005). Protein concentrations can be correlated to the number of observed peptides of a protein: the more peptides that are found for a protein in mass spectrometry, the higher its concentration in the sample.

emPAI (exponentially modified Protein Abundance Index) is a refinement of the protein abundance index (PAI) (Rappsilber *et al.*, 2002), which itself is defined as the number of observed peptides in mass spectrometry divided by the number of theoretically observable peptides. emPAI has been shown to be, within some error range, linearly proportional to protein concentration (Y. Isihama, manuscript submitted). It can be calculated from PAI as $emPAI = 10^{PAI} - 1$. The PAI is not only correlated to the abundance of a protein but is also dependent on its specific response to the mass spectrometry methodology. This varies according to digestion efficiency, peptide solubility, extraction, ionization and fragmentation. Thus, although emPAI has been shown to correlate more linearly with protein concentration than PAI, it is not an extremely accurate measure. However, it may be employed as a guide to distinguish high from low abundance proteins and its accuracy has been shown to be in the same range or better than protein staining techniques. For the average of large protein datasets, it is clearly sufficient to indicate their relative concentrations.

emPAI values were first calculated for proteins observed in the *E. coli* lysate. The distribution of lysate concentrations of all GroEL interacting proteins (class I+II+III) was very similar to the distribution observed for the entirety of identified *E. coli* lysate proteins.
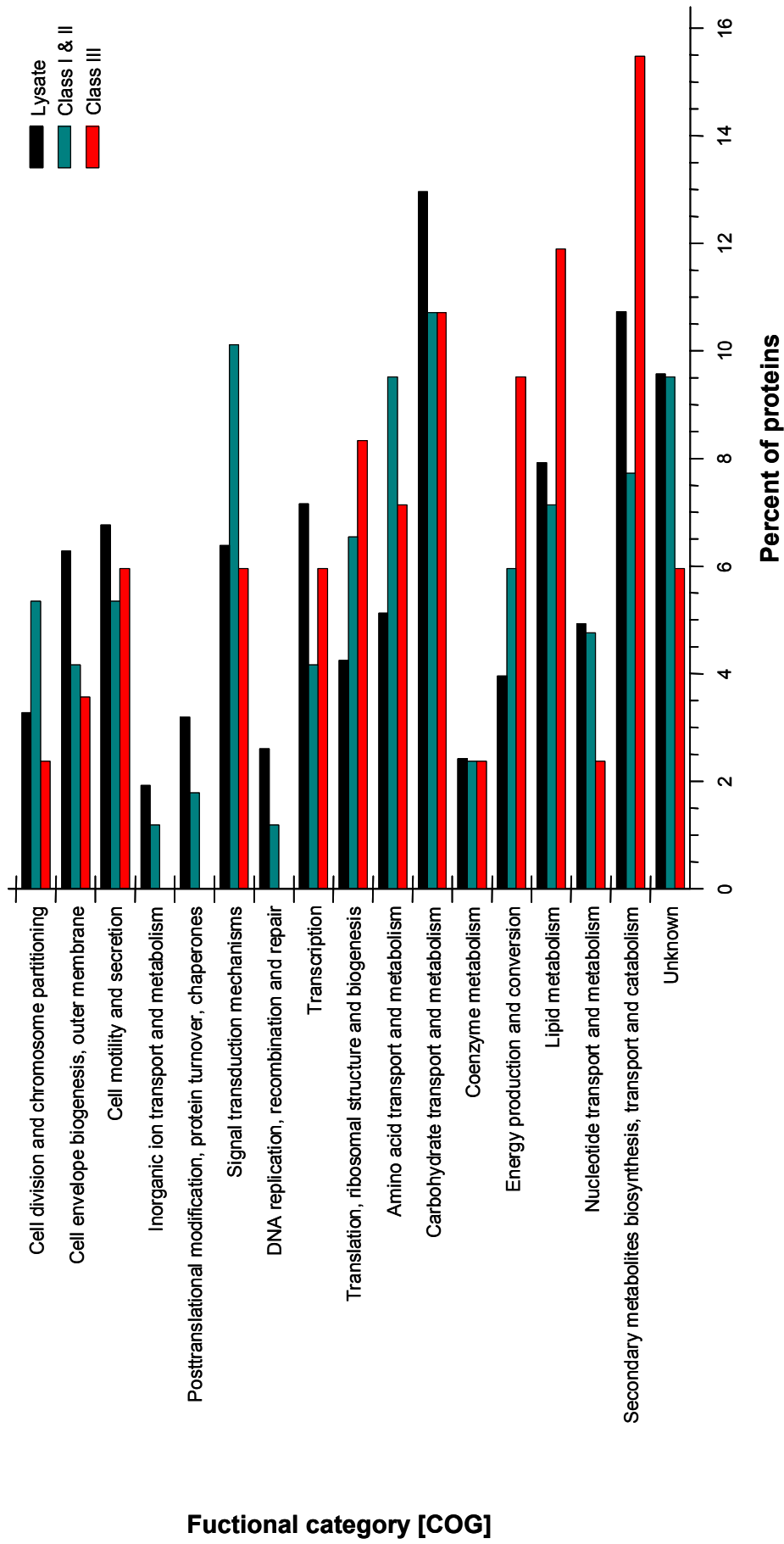
Figure 27: Distribution of functional categories among GroEL interacting proteins.

Functional assignment based on protein classification by the COG database (Clusters of Orthologous Groups of proteins) (Tatusov et al., 1997).

GroEL interacting proteins were, however, slightly enriched in highly abundant proteins (data not shown). When the distribution was divided into the three GroEL substrate classes, it was observed that the large majority of class I substrates are highly abundant proteins (Figure 28). The distribution of class I proteins deviated greatly from the lysate towards high cellular abundance. It can therefore be concluded that the identified largely chaperone independent class I proteins probably represent only a small fraction of all class I substrates present in the cell. The reason for the significant association of the identified class I proteins with GroEL is likely to be explained by their high abundance and the general ability of GroEL to bind all unfolded proteins exposing hydrophobic surfaces.

Class II proteins reveal a distribution of cellular concentration very similar to that of the whole lysate, while, importantly, stringent class III substrates are enriched in low abundant proteins below 50 ppm. Note that the data presented in Figure 28 even underestimates low concentration class III proteins, as 44 % of all class III proteins were not detectable in the lysate, most probably due to very low abundance (compared to 29 % of class II and only 3 % of class I). Their detection as GroEL substrates by LC-MS/MS was probably due to enrichment in GroEL/GroES/substrate complexes as well as the reduced complexity in this
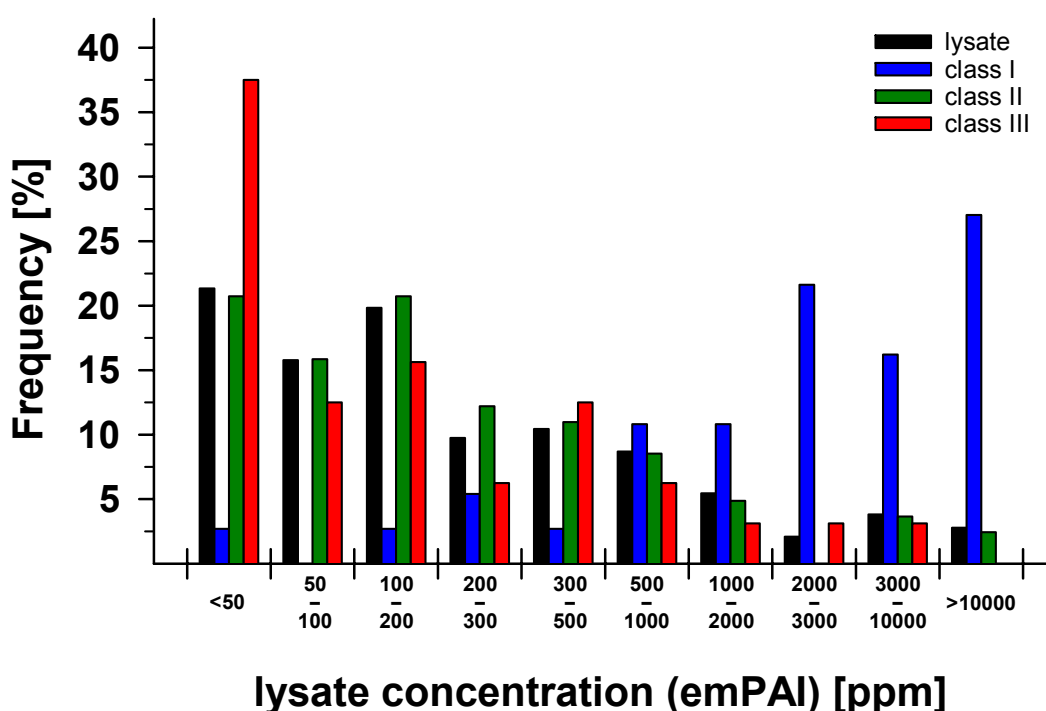
Figure 28: Distribution of cellular concentration of E. coli lysate proteins and GroEL substrates.

Cellular concentrations as estimated by their emPAI value (expressed in molar ppm, see main text) measured in the *E. coli* lysate.

less diverse mixture .

Concentration evaluation of proteins in GroEL/GroES/substrate complex prepara-tions using emPAI is much more difficult than in *E. coli* lysate samples. The large majority of peptides in the complex preparations are derived from GroEL and GroES, since these are in at least seven-fold molar excess over the sum of all their substrates (when considering both *cis*- as well as *trans*-bound peptides in one complex). Fairly reliable concentrations for single selected proteins when bound to GroEL are thus difficult to obtain. However, valuable information can be obtained from using accumulated emPAI concentrations for large groups of proteins, due to the statistically reduced error in such evaluations.

The TIM barrel fold, which is enriched in GroEL/GroES/substrate complexes when compared to the number of proteins sharing this fold (see section 4.3.4 and Figure 25), is also absolutely enriched in concentration on GroEL. The lysate exhibited ~6 % molar con-centration of TIM barrel proteins, while ~34 molar % of all specifically GroEL associated proteins (class I+II+III) were TIM barrel proteins. When considering class III proteins exclu-sively, the molar amount of proteins with TIM barrel structure increased to ~40 %. This un-derscores the importance of the TIM barrel-fold among GroEL substrates in total and among class III proteins in particular. No other fold was similarly increased in concentration among GroEL interactors.

When determining the total concentration of different GroEL substrate classes among GroEL interactors, it was found that class III proteins occupy nearly all of the capac-ity of GroEL (Figure 29). Approximately 78 % of GroEL substrates (in molar concentration) were classified as stringent class III substrates, while 18.5 % were class II and only ~1.5 % class I. Approximately 2 % of the total molar concentration of GroEL interactors were not
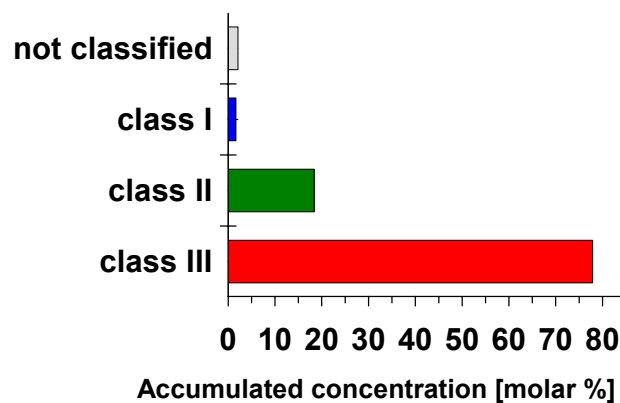

Figure 29: Concentration distribution of substrate classes in GroEL/GroES/substrate complexes.

Accumulated concentrations of GroEL substrate classes in GroEL/GroES/substrate com-plexes, estimated by their emPAI values.

assigned a substrate class and were not considered to be specific GroEL interactors.

The above findings suggest, that in the cell GroEL is mostly utilized for the folding of proteins that absolutely depend on it. Class II proteins occupy only a minor fraction of the GroEL capacity, while class I proteins are nearly negligible in terms of total mass associated with GroEL.

5   Discussion

The folding of many newly synthesized proteins in the cell requires molecular chaperones. The GroEL and DnaK chaperone systems of *E. coli* have been the subject of intensive study and the molecular mechanisms of these chaperones are now relatively well understood. However, little is known about the natural substrates of the distinct chaperones in *E. coli* or the details about their contribution to *in vivo* protein folding.

Previous work on entire substrate proteomes of *E. coli* chaperones involved in *de novo* protein folding has so far been limited mainly to their identification. Putative substrates of TF and DnaK have been identified by 2D-gel mass spectrometry (MS) of isolated aggregated proteins in *dnaK* deletion strains upon heat stress (Mogk *et al.*, 1999), in *tig* deleted cells during DnaK/DnaJ depletion (Deuerling *et al.*, 2003) and more recently in *E. coli* Δ*tig*Δ*dnaKJ* cells (Vorderwülbecke *et al.*, 2004). A subset of interacting proteins of the chaperonin GroEL has been identified under nucleotide free conditions by GroEL co-immunoprecipitation and subsequent 2D-gel MS (Houry *et al.*, 1999). However, these studies did not provide quantitative information on substrate interaction nor did they reveal direct insight into the degree of chaperone dependence of the identified substrate proteins.

In order to address these questions for the chaperonin GroEL, this study began with the analysis of the detailed *in vitro* and *in vivo* folding behavior of GroEL substrates from its natural host *E. coli*. Previous work on the GroEL/GroES mechanism had mainly been performed with a small number of heterologous substrates, such as bacterial rubisco, and the mitochondrial enzymes rhodanese, citrate synthase and malate dehydrogenase (Schneider, 2000). The characterization of the *in vitro* refolding behavior of seven endogenous GroEL substrates, undertaken as part of this work, provides a valuable tool for further characterization of the physiological GroEL/GroES mechanism in its own right. The identification and classification of the nearly complete *in vivo* GroEL interaction proteome, on the other hand, represents the most thorough study of this kind to date and the understanding of the chaperone network in living cells will undoubtedly benefit from the availability of this comprehensive analysis.

5.1   Classification of GroEL dependence

The body of work presented here firmly established that dependence of newly-synthesized and unfolded proteins on GroEL for efficient folding can be categorized in three classes (Figure 30), as had been proposed previously (Ewalt *et al.*, 1997). Class I proteins are largely chaperone independent *in vitro* and are able to fold spontaneously. Their refolding yield can, however, be optimized by chaperone interaction (Figure 10). *In vivo,* class I proteins were shown to be independent of the chaperonin system. Recombinant co-overexpression of class I proteins with complete or incomplete chaperonin systems of different functionality (GroEL/GroES and GroEL only) does not result in any detectable change in solubility (Figure 15). Furthermore, class I substrates remain completely soluble upon depletion of GroEL/GroES in *E. coli* (Figure 8). Identified proteins of class I represent highly abundant proteins of the *E. coli* cytosol (Figure 28). It may be safe to assume that additional
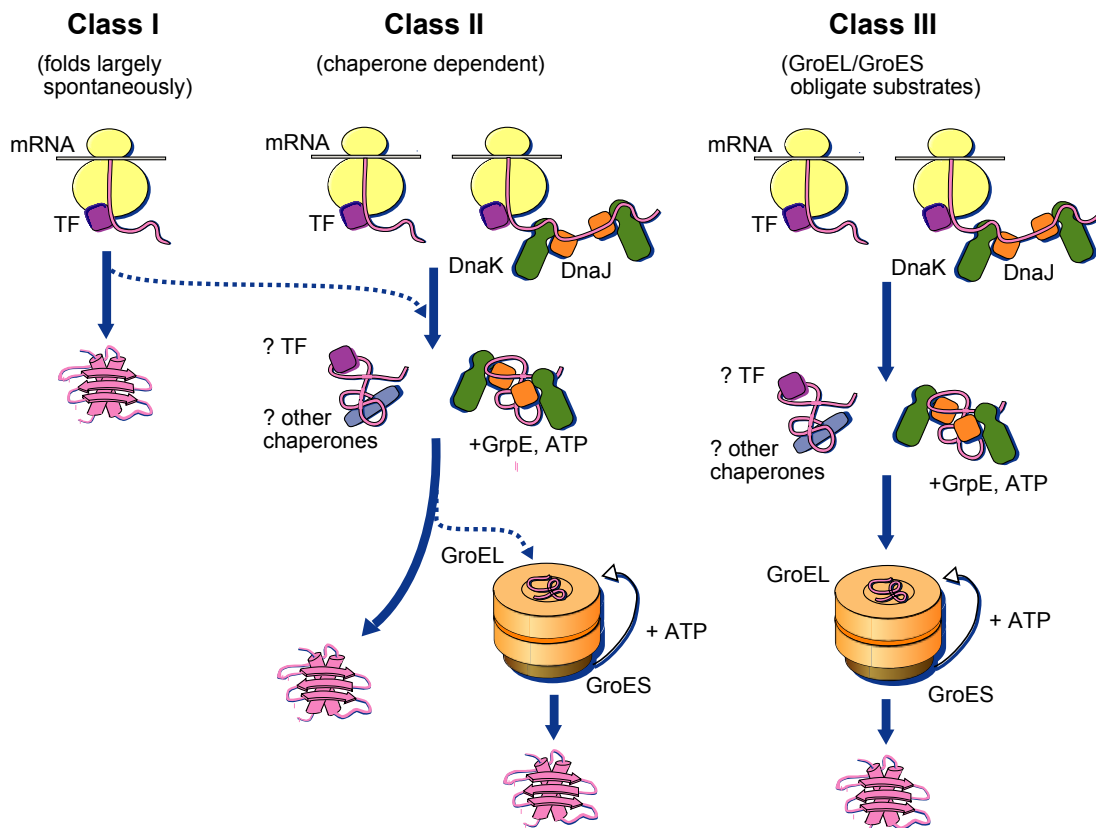
Figure 30: Chaperone usage of nascent polypeptides in E. coli upon synthesis on the ribosome.

This model distinguishes three classes of chaperonin dependence. Modified from Hartl and Hayer-Hartl (2002).

proteins in the cell behave as class I proteins in terms of their folding properties. GroEL has a general capacity in binding exposed hydrophobic surfaces (Coyle *et al.*, 1997) and can therefore bind nearly any protein along its folding pathway. However, the chaperones TF and DnaK/DnaJ/GrpE interact with nascent polypeptides upstream of the chaperonin system (Hartl and Hayer-Hartl, 2002) and most proteins with class I behavior will have had the opportunity to fold before reaching GroEL. Thus, only highly abundant proteins of class I are experimentally found to interact significantly with GroEL in the cell. The low relative amount of the total mass of a class I protein associated with GroEL (<0.05 %) does not allow detection of low abundance class I proteins bound to GroEL *in vivo*.

Class II proteins are unable to fold spontaneously under standard conditions *in vitro*. They depend on chaperone assistance for folding to their native structures. The DnaK system is as effective as the GroEL system in assisting folding of class II proteins *in vitro* (Figure 11) and *in vivo* (Figures 15 and 16). Class II proteins therefore do not strictly depend on encapsulation inside the chaperonin for folding, but rather represent highly aggregation-prone

proteins that need to be prevented from aggregating in their nonnative states. GroES dependence during *in vitro* GroEL-mediated folding of class II proteins is variable. In some cases (Figure 11) low temperature allows GroEL to mediate class II protein folding in the absence of GroES. Without GroES, GroEL acts like a general chaperone by binding and releasing non-native proteins without encapsulation. This mechanism is, however, less efficient when compared to that of DnaK. It can only promote successful folding at relatively mild conditions and for a limited number of substrate proteins. *In vivo*, the contribution of this particular mechanism of GroEL action without GroES is expected to be rather low. Affinity of the GroEL apical domains for GroES increases drastically upon ATP binding (Burston *et al.*, 1995), and the complete GroEL-GroES cycle is probably highly favored in the cell.

Class III GroEL substrates constitute a group of proteins, absolutely dependent on the assistance of the full chaperonin system for folding to their native state, both *in vitro* (Figure 14) and *in vivo* (Figures 15, 16, and 26). Unfolded class III substrates are highly aggregation prone and are unable fold spontaneously. The DnaK system is able to bind unfolded class III proteins effectively and thereby suppress their aggregation. However, DnaK cannot promote folding of class III substrates. Only upon transfer of the substrate to the chaperonin system does folding occur. GroEL-mediated folding of class III substrates is absolutely GroES dependent. Therefore, encapsulation inside GroEL/GroES is an essential feature of chaperonin-mediated class III substrate folding (for the limited set of class III substrates too large for encapsulation, please see below). *In vivo*, class III substrate folding also depends on GroEL and GroES. Co-overexpression of class III substrates with the complete chaperonin system enhances their solubility as expected. However, co-overexpression with GroEL only (which results in an incomplete and only partially functional chaperonin system) leads to aggregation of overexpressed class III proteins. Depletion of GroEL/GroES leads to an inability of endogenous class III substrates to fold, resulting in efficient degradation or adverse aggregation. Thus, no other chaperone system of *E. coli* is able to substitute for the function of GroEL in class III protein folding.

5.1.1   Folding of class III GroEL substrates

A number of reagents and conditions are know to promote spontaneous protein folding. However, none promoted the efficient folding of the class III proteins METF and DAPA. These conditions included very low protein concentrations (<20 nM), slow dilution out of the denaturant (6M GdnHCl), low temperature (4 °C), high salt (0.5 - 1M NaCl or $(NH_4)_3SO_4$), and addition of glycerol, proline or Ficoll 70 (data not shown). Only addition of 500 mM arginine to the refolding buffer allowed refolding of DAPA, but not METF, without addition of chaperones (data not shown). How the chaotropic agent arginine supports folding of proteins is not clearly understood. It is known to reduce aggregation (Arakawa and Tsumoto, 2003), and it might have further effects on the energetic properties of the folding pathway of a protein. However, efficiency of DAPA folding in arginine buffer is lower and the apparent rate much slower when compared to chaperonin-mediated folding. The effect of the chaperonin cavity on class III protein folding is therefore essential and not easily mimicked in free solution.

Prevention of aggregation is clearly an important aspect of chaperone-mediated protein folding, both for the DnaK and the GroEL systems. In addition to this, encapsulation of substrate proteins by the chaperonin offers particular conditions that favor promotion to the native state. This appears to be a necessity for successful folding of class III substrates. Confinement in the chaperonin cage is known to be able to enhance the folding rate of certain proteins (Brinker *et al.*, 2001) by smoothing the folding energy landscape and destabilization of folding intermediates (Baumketner *et al.*, 2003; Jewett *et al.*, 2004; Takagi *et al.*, 2003). Without encapsulation, class III proteins probably are trapped in such intermediate states, which subsequently aggregate. Importantly, GroEL-mediated *in vitro* refolding of class III proteins indeed occurs with higher apparent rates when compared to those of refolding of classes I, II or the heterologous proteins studied previously (Schneider, 2000).

An additional explanation of why DnaK is unable to promote folding of class III substrates might result from the fact that substrate release from DnaK is probably not a concerted process. When considering binding of multiple DnaK molecules to a single unfolded polypeptide chain, folding could be inhibited by DnaK molecules still binding to one region of the polypeptide chain while a different region is released. The concerted release of the polypeptide from the apical domains of GroEL into the cavity triggered by binding of GroES effectively circumvents such an adverse effect of chaperone binding for efficient folding.

5.1.2   *Substrates too large for encapsulation inside the GroEL/GroES cavity*

GroEL interacting proteins larger than ~60 kDa represent a unique kind of GroEL substrates. These proteins are too large to fit inside the chaperonin central cavity (Sigler *et al.*, 1998) and interact with GroEL in the *trans* ring opposite to bound GroES. Such a reaction has been recognized between GroEL and the heterologous yeast mitochondrial aconitase (Chaudhuri *et al.*, 2001), an 82 kDa monomeric Fe_4S_4 cluster-containing enzyme, known to aggregate in chaperonin-deficient mitochondria (Dubaquié *et al.*, 1998). Efficient GroEL-mediated folding of aconitase is dependent on GroES. Binding of GroES to GroEL is necessary for release of non-native aconitase from the GroEL *trans* apical domains. Importantly, none of the *E. coli* aconitase homologs (ACO1, ACO2, YBHJ, LEU2) were found to interact with GroEL *in vivo*. ACO1 and ACO2 were found in the *E. coli* lysate, which demonstrates the ability to detect these proteins by the utilized methodology.

Among 252 GroEL interacting proteins, 30 were larger than 60 kDa. Of these, six belong to class I, eighteen to class II and six to class III. SYT, an endogenous substrate too large for GroEL encapsulation was studied further, both *in vivo* and *in vitro* (Figures 13, 15, and 16). This homodimeric protein of 74 kDa subunits behaves like a typical class II substrate *in vivo*. GroEL/GroES co-overexpression slightly enhances its solubility, but GroEL/ GroES depletion does not lead to significant SYT aggregation. In *vitro*, DnaK mediates folding of SYT much more efficiently than the GroEL system. Furthermore, GroEL-assisted folding of SYT is apparently GroES-independent. Thus, SYT appears to follow a different folding mechanism from the one previously observed for yeast mitochondrial aconitase, although both proteins bind to GroEL at the *trans* side as a consequence of their size (Chaudhuri *et al.*, 2001) (Figure 19). GroEL is clearly less well suited than DnaK to assist

folding of large class II proteins like SYT. This is consistent with the accumulation of large proteins among putative DnaK substrates (Deuerling *et al.*, 2003; Mogk *et al.*, 1999) as well as with the size distribution of class I and II proteins (14 % of classes I + II are >60 kDa) compared to class III GroEL substrates (7 % >60 kDa) (Figure 22A). However, six large proteins have been identified among class III substrates. These proteins likely interact with GroEL in a *trans*-mediated mechanism similar to that of aconitase. This would involve a mechanism of GroEL-mediated *trans* folding that the DnaK system is unable to provide.

5.2   The GroEL interactome

5.2.1   *Completeness and quality of the dataset*

In the course of this study, the nearly complete set of *E. coli* GroEL interacting proteins was identified. Several observations strongly argue that the identified set, at least for class III substrates, is nearly complete. First, the amount of identified substrates did not increase in repeated experiments, both in multiple rounds of complex purification and in multiple LC-MS/MS experiments. Second, as shown in Figures 28 and 29, class III GroEL substrates are highly enriched on GroEL. They are therefore more easily detected by MS than class II and especially class I substrates, which occur in much lower concentration in the GroEL/GroES/substrate samples. Potentially missing proteins among GroEL interactors would thus most likely belong to class II and especially class I, a group which plays a minor role among GroEL interacting proteins by total mass. Third, experiments repeated with a new 'hybrid linear ion trap – Fourier transform ion cyclotron resonance mass spectrometer' (Thermo LTQ-FT), which offers an increased sensitivity of about one order of magnitude compared to the previously used QSTAR Pulsar mass spectrometer, did not result in additional identification of GroEL-interacting proteins.

5.2.2   *Overlap with previously identified GroEL interactors*

23 substrates identified in this study were found to overlap with 52 GroEL interacting proteins identified previously in the same laboratory (Houry *et al.*, 1999). Another four proteins (EFTU, RL9, RL7, RS2) of the earlier study were excluded from this analysis due to their non-specific binding to the purification matrix during GroEL/GroES/substrate isolation. The overlapping substrate fraction contains six class I, nine class II and eight class III proteins. 46 of the 52 GroEL substrates proteins from Houry *et al.* (1999) were either found in the lysate or in GroEL substrate complexes, which indicates that most of the previously identified substrates are detectable by the method applied here. Although there is clearly a significant overlap of GroEL substrates between the two studies, it is not complete. Approximately 50 % of GroEL interactors reported by Houry *et al.* (1999) are not found in this new study. A major difference of substrate detection between the two projects is the nucleotide state of GroEL during complex purification. Houry *et al.* (1999) used EDTA for stabilization of GroEL-substrate interactions, thereby precluding GroES binding and substrate encapsulation. The present study used a different approach: rapid conversion of cellular ATP to ADP with glucose plus hexokinase and purification of GroEL/GroES/substrate complexes in the presence of excess ADP. The ADP approach employed here probably closer resembles the

situation inside the cell. It can be thought as a snapshot taken at a late time in the GroEL reaction cycle by trapping the GroEL/GroES *cis* complex with encapsulated substrate, thus making the experiment less dependent on the characteristics of substrate binding to GroEL. In the experiment performed by Houry *et al.* (1999), pre-existing GroEL/GroES/substrate complexes were disrupted by EDTA. Subsequent permanent binding of substrates from solution then occurred in presence of EDTA. This situation is clearly different to the normal cellular environment, which provides a constant supply of ATP. Thus, differences in experimental conditions likely explain why the overlap in substrates between the two studies is only partial.

5.2.3   *Substrate classification of the GroEL interaction proteome*

The proposed substrate classification based on *in vitro* and *in vivo* experiments of selected GroEL interacting proteins was found to correlate well with their enrichment in GroEL/GroES/substrate complexes. Utilization of the SILAC methodology allowed the direct comparison of substrate protein concentrations in the *E. coli* lysate with those of the GroEL/GroES/substrate complexes. Previously characterized class I proteins were found to have only 0.05 % or less of their total cytosolic amount associated with GroEL. Class III proteins, on the other hand, are much more enriched on GroEL: 3 % to 5 % or more of the total cytosolic content of a class III protein is associated with GroEL at any time. Class II proteins exhibited relative concentrations on GroEL between those of class I and III (0.2 % - 2.5 %).

Based on the above observations, substrate classification was extended to the rest of the identified GroEL interaction proteome. This resulted in the categorization of a total of 252 GroEL interactors as follows: 42 class I, 126 class II and 84 class III. This extended classification was subsequently confirmed experimentally (Figures 26 and data not shown**Fehler! Verweisquelle konnte nicht gefunden werden.**, as well as in yeast co-expression experiments, H.-C. Chang, data not shown) and proved to be correct in all tested cases. Determination of protein enrichment on GroEL therefore appears to be a reliable way of predicting GroEL substrate classification.

5.2.4   *Class III substrate enrichment on GroEL*

A large fraction of GroEL was found to interact with class III substrates while a more limited amount of GroEL interacted with class II substrates under wild type conditions (Figure 29). Nearly 80 % of all polypeptide chains associated with GroEL are members of class III; less than 20 % belong to class II; and class I and unclassified substrates together are below 5 % of all GroEL interacting polypeptide chains. This high concentration of class III proteins on GroEL is in great contrast to their known cellular concentrations (Figure 28). Class III proteins are generally very low abundance proteins of the *E. coli* cytosol, whereas class I proteins represent the most abundant soluble proteins of *E. coli*.

High enrichment of class III substrates on GroEL in the cell is probably a consequence of mainly two factors. First, unfolded class III substrates have been shown in preliminary experiments to exhibit higher affinity to GroEL than class I or class II substrates

(Tobias Maier and Michael Kerner, data not shown). Class III substrates therefore preferentially bind to GroEL when competing with class I or II substrates. Second, the chaperone network in *E. coli* (Young *et al.*, 2004) has a 'filtering' effect for class I proteins and to a large extent also for class II substrates, so they are hardly expected to reach GroEL. Interaction with TF upon synthesis at the ribosome might already be sufficient for correct folding of most class I proteins. The DnaK system, which also acts upstream of GroEL in the chaperone network (Figure 30), promotes folding of almost all remaining class I and many class II substrates. Only upon DnaK interaction are substrates transferred to the chaperonin. Class III proteins, which do not experience folding assistance from DnaK, but can be prevented from aggregation, all eventually arrive at the GroEL system. Most class II and especially class I substrates have already completed their folding by this stage and do not need to interact further with chaperones.

This hypothesis is further supported by identification of GroEL substrates from *E. coli* cells lacking both DnaK and TF (Tobias Maier and Michael Kerner, data not shown). Cells missing only one of either of these chaperones do not exhibit a significantly different GroEL substrate spectrum from wild type *E. coli* cells. However, combined deletion of the genes encoding TF and DnaK considerably increases the number of identifiable GroEL substrates. 354 substrates of GroEL were found in a single experiment from Δ*tig*Δ*dnaKdnaJ* cells, while a typical single experiment from wild type cells resulted in identification of only ~200 proteins. 157 of the Δ*tig*Δ*dnaKdnaJ* GroEL substrates overlap with the total 254 wild type GroEL substrates, of which 25 are members of class I, 76 of class II and 56 of class III. The remaining newly identified 197 substrates from Δ*tig*Δ*dnaKdnaJ* had not been identified as specific GroEL substrates previously. These most likely represent substrates of TF and DnaK that need chaperone assistance for correct folding and thus require interaction with GroEL when these upstream chaperone systems are lacking.

5.2.5   *Interaction of trigger factor and DnaK with GroEL*

A general overlap of DnaK and GroEL substrates clearly exists, raising the question to which degree the two chaperone systems can replace each other in the cell. Proteins reported to be putative DnaK substrates (Deuerling *et al.*, 2003) were not abundantly detected among the total of GroEL interacting proteins in wild type *E. coli* or among stringent (class III) GroEL substrates. Of 94 putative DnaK/TF substrates, only three appear to be stringently GroEL dependent. 31 of the DnaK/TF substrates were found among all GroEL interactors and nearly all of them could be detected in the lysate (90 out of 94), demonstrating the capability of the applied method to detect these proteins.

Identification of DnaK/TF substrates has been achieved by analysis of aggregating proteins in an *E. coli* Δ*tig* strain during DnaK/DnaJ depletion (Deuerling *et al.*, 2003) as well as in Δ*tig*Δ*dnaKdnaJ* cells, which resulted in an overlap of >90 % of identified proteins (Vorderwülbecke *et al.*, 2004). Since GroEL can partially take over the function of TF and DnaK (Genevaux *et al.*, 2004), most proteins prevented from aggregation and folded by GroEL would not be expected to be found in the aggregating fraction and thereby miss detection as DnaK/TF substrates by this methodology. This is consistent with the observed

enrichment of identified DnaK substrates among the high molecular mass proteins. As shown above, the folding of these large proteins is less likely to benefit from GroEL interaction, allowing significant aggregation upon DnaK/TF depletion. Thus, the substrate spectrum of DnaK is probably larger than previously reported if proteins <50 – 60 kDa, effectively folded by either DnaK or GroEL, are taken into account.

GroEL substrates, identified from the Δ*tig*Δ*dnaKdnaJ E. coli* strain, reveal an overlap of only 35 proteins with the 94 aggregating proteins suggested to be DnaK substrates (Deuerling *et al.*, 2003). 17 of these proteins had not previously been identified as interacting with GroEL and can thus be considered to be mainly DnaK substrates. These probably associate, and possibly fold, with GroEL when TF and DnaK are absent. However, nearly two-thirds of the DnaK substrates found by Deuerling *et al.* could not be identified as GroEL substrates in the Δ*tig*Δ*dnaKdnaJ* strain, which is in support of the above hypothesis.

5.2.6   The essential role of GroEL

DnaK and TF, in addition to GroEL, play an important role in *de novo* protein folding in *E. coli*. However, these two chaperones are not essential for cell viability under laboratory conditions. Their loss can be compensated for (Genevaux *et al.*, 2004), whereas no other chaperone system in *E. coli* can substitute for loss of GroEL (Fayet *et al.*, 1989). This means that the GroEL/GroES system is the only essential chaperone in *E. coli* under all conditions tested. The existence of thirteen essential and stringently GroEL-dependent class III substrates (see Supplementary Table S2) now provides an explanation for the essential nature of GroEL. In the absence of the chaperonin system, these essential proteins fail to fold to their native state, are unable to fulfill their cellular functions and thus render GroEL itself essential. DnaK and TF, on the other hand, do not have any essential substrate that absolutely depends on either one of them for folding into the native state. DnaK and TF have overlapping substrate spectra (Deuerling *et al.*, 1999; Teter *et al.*, 1999) and each can compensate for the loss of the other. However, a combined deletion of both the genes encoding for TF and DnaK is only lethal at temperatures > 30 °C, although at temperatures < 30 °C high levels of aggregation occur (Genevaux *et al.*, 2004). Notably, GroEL is overexpressed in this strain to compensate for the loss of TF and DnaK. Thus, GroEL is very likely to be able to undertake the folding of at least all essential DnaK and TF substrates in their absence.

The number of essential proteins is decreased among class III proteins. 17 % of class III, 32 % of class II and ~50 % of class I GroEL substrates and are essential for cell growth, compared to 26 % of the *E. coli* lysate proteins (Gerdes *et al.*, 2003) (Figure 21). This possibly indicates an evolutionary tendency to favor less chaperone dependence for essential proteins. The high amount of essential proteins among class I GroEL substrates is very likely the result from the high cytosolic concentration of class I proteins. Highly concentrated proteins of the *E. coli* cytosol, as estimated by their emPA index (see section 4.3.6), were observed to be enriched in essential proteins (data not shown).

5.2.7   GroEL substrate solubility in S. cerevisiae

In a closely related study, it was found that the eukaryotic cytosol does not allow the

folding of recombinantly expressed newly synthesized stringent GroEL substrates (results Hung-Chun Chang, Hartl laboratory, manuscript in press). GroEL substrates expressed heterologously in *S. cerevisiae* were tested for solubility. It was shown that, while class I GroEL substrates were highly soluble, class II proteins were soluble but aggregated in the absence of Ydj1, a yeast DnaJ homolog. Class III proteins, on the other hand, were highly insoluble, and no degradation of the aggregated material was detectable. Only upon co-overexpression of *E. coli* GroEL/GroES with class GroEL substrates was solubility observed.

The general ability of the eukaryotic cytosol to fold multi-domain proteins and presence of the type II chaperonin TRiC are not sufficient to compensate for the lack of GroEL during the folding of class III substrates. The eukaryotic cytosol therefore does not exhibit a generally superior ability for the folding of this particular class of proteins compared to the bacterial cytosol. The two systems have rather co-evolved to meet their specific needs.

5.3 Properties of GroEL interacting proteins

5.3.1 Size dependence

Extension of the established GroEL substrate classification to the GroEL interactome (comprising ~250 proteins) revealed a strong preference for sizes between 30 and 50 kDa for the stringently GroEL dependent class III substrates. Only a very limited number were larger than the estimated size cutoff value of the GroEL cavity (Figure 22A). The GroEL/GroES cavity is ~85 000 Å^3, which would be able to accommodate a folded protein of ~70 kDa while an unfolded polypeptide of ~60 kDa (Sigler *et al.*, 1998). Although there have been reports of a single large fusion protein capable of being encapsulated (Chuang *et al.*, 1999; Song *et al.*, 2003; Song *et al.*, 2000), the size limit of ~60 kDa has been repeatedly confirmed both *in vivo* (Ewalt *et al.*, 1997; Houry *et al.*, 1999) and *in vitro* (Chaudhuri *et al.*, 2001; Huang and Chuang, 1999; Sakikawa *et al.*, 1999).

Unlike class III proteins, class I and II substrates did not reveal a particular size preference and exhibited a molecular mass distribution similar to the one observed for the *E. coli* lysate. Therefore, there is a need of class III GroEL substrates for folding in the environment provided by the GroEL *cis* cavity. The sharp cutoff values observed at ~50 – 55 kDa most likely reflect the optimal protein size for GroEL-assisted, and possibly accelerated, folding by confinement, as was previously shown for rubisco (Brinker *et al.*, 2001). This hypothesis has been further supported by theoretical considerations (Baumketner *et al.*, 2003; Jewett *et al.*, 2004; Takagi *et al.*, 2003). Confinement in the GroEL/GroES cavity is suggested to decrease the roughness of the folding energy landscape of an unfolded polypeptide and thus prevent the population of kinetically trapped intermediate folding states. This effect is expected to be size dependent and would have most impact on polypeptide sizes close to the volume of the surrounding cavity. The strong preference of class III GroEL substrates for molecular masses from ~30 – 50 kDa and the sharp cutoff values for masses >55 kDa suggests that this mechanism of confinement inside the chaperonin cage plays a major role for achieving effective folding of these proteins.

Class III proteins might represent polypeptides with particularly high energetic frus-

tration (Shea *et al.*, 1999) and thus a very rough folding landscape. Energetic frustration results from the formation of incorrect contacts between amino acids of the same polypeptide along the folding pathway from a random coil to the native structure. Many such adverse interactions during folding of a protein hinder quick folding or even render spontaneous folding impossible, and thus create chaperonin dependence. By smoothing the folding energy funnel of a protein inside the cage, chaperonins can facilitate folding of these proteins, that would otherwise be unable to fold into the native state.

5.3.2   Aggregation propensity

Class III GroEL substrates were observed to exhibit isoelectric points ~0.5 closer to 7 than average *E. coli* lysate proteins (Figure 22B). This results in a lower net charge of these proteins at physiological pH, a condition which is known to enhance the propensity of proteins to aggregate (Chiti *et al.*, 2002). This finding, together with the observation that class III proteins are generally more hydrophobic than an average lysate protein (although only to a small degree, see Figure 23), suggest that class III GroEL substrates are more aggregation prone than an average *E. coli* protein.

In an attempt to further advance this hypothesis, the software program TANGO was employed (Fernandez-Escamilla *et al.*, 2004; Linding *et al.*, 2004), which predicts aggregation propensities of polypeptides. The software was kindly made available before publication by Dr. L. Serrano, EMBL Heidelberg, Germany. The TANGO algorithm is based on the physicochemical principles of β-sheet formation and has been shown to correctly predict aggregation propensity of various peptides (Fernandez-Escamilla *et al.*, 2004). However, no significant differences between *E. coli* lysate proteins and all GroEL interacting proteins or the distinct classes of GroEL interactors could be identified (data not shown). This suggests that differences of aggregation propensities between the GroEL interactors and average lysate proteins are too small to be detected by this methodology. It is also possible that the TANGO algorithm (in its state of development status at the time of analysis) was not yet fully capable of predicting aggregation propensities of large proteins, since it was optimized for much shorter polypeptides than the proteins tested in the current study.

5.3.3   Structures of GroEL substrates: the TIM (βα) barrel

Analysis of protein folds of GroEL interactors revealed a broad spectrum of different structures recognized by GroEL. Notably, GroEL interacting proteins show a remarkable bias towards TIM barrel proteins, most pronounced for class III GroEL substrates. Class III proteins exhibit a fourfold enrichment of the TIM $(\beta\alpha)_8$ barrel fold when compared to lysate proteins, by both mass and number of different proteins (Figure 25). The TIM barrel is a highly ubiquitous topology and many representatives of this fold are not found among GroEL interacting proteins. Thus, a particular structural feature exhibited in the native fold or in intermediate states during the folding pathway must exist that distinguishes chaperonin independent, and even spontaneously folding TIM barrels, from GroEL dependent ones.

Folding of TIM $(\beta\alpha)_8$ barrel proteins has been previously studied in detail. For example, triosephosphate isomerase (Rietveld and Ferreira, 1998) and aldolase (Rudolph *et al.*,

1992), exhibit apparent two-state folding mechanisms, whereas alpha subunit of tryptophan synthase (TRPA, αTS; Wu and Matthews, 2002), indole-3-glycerol phosphate synthase (also referred to as N-(5'-phospho-ribosyl)anthranilate isomerase, TPRC; Sanchez del Pino and Fersht, 1997) and imidazole glycerol phosphate synthase subunit hisF (HIS6, HisF; Höcker *et al.*, 2001) feature apparent non two-state folding mechanisms. The latter are suggested to fold in partial fragments, which serve as autonomous folding units that provide an initial scaffold for the formation of the full native structure. Different folding fragments have been observed: 4+4 (βα) units for HisF, 6+2 for αTS. Such partial structure acquisition can accelerate folding, but might also result in partially stable folding intermediates with a significant energy barrier for final assembly of the native tertiary structure (Zitzewitz *et al.*, 1999). Interestingly, none of these TIM barrel proteins was found to interact with GroEL in this study, although TIM (TPIS) and αTS (TRPA) were detected in the *E. coli* lysate. However, in spite of the complicated folding pathway of many of these well studied TIM barrel proteins, all of them are able to fold spontaneously in free solution. This distinguishes them from the TIM barrel proteins among class III GroEL substrates, which have been shown to stringently depend on chaperonin assistance for correct folding.

The GroEL class III substrates with TIM barrel structures (for representative structures see Figure 31) thus probably constitute examples of proteins with particularly rough folding energy funnels. This would lead to frequent population of intermediate folding structures that may be highly aggregation prone, with a considerable energy barrier towards the native state. This barrier may be overcome by confinement through encapsulation inside the GroEL/GroES cage. Furthermore, concerted release from the apical domains of GroEL into the cavity triggered by GroES binding could be of specific importance for folding of these class III TIM barrel structures, as previously discussed.

Detailed inspection of class III TIM barrel structures and their evolutionary relationship (in collaboration with D. Frishman, GSF, Neuherberg, Germany and with A. Lupas and J. Soeding, MPI for Developmental Biology, Tübingen, Germany), however, did not reveal detectable differences among class I, II or lysate TIM barrel proteins. The difficulty in identification of GroEL dependence from a given native structure reflects the problem that folding routes of similar structures are not necessarily similar and are very challenging to predict. Furthermore, the population of GroEL interacting TIM barrels, although highly enriched, is probably too small to reveal subtle differences (35 TIM barrels were found to interact with GroEL, eighteen of these are predicted to be of class III). An extension of the current analysis of GroEL substrates to the chaperonin substrate proteomes of different organisms will provide valuable additional data that may allow a more in depth structural study of the chaperonin dependence of certain TIM barrel proteins.
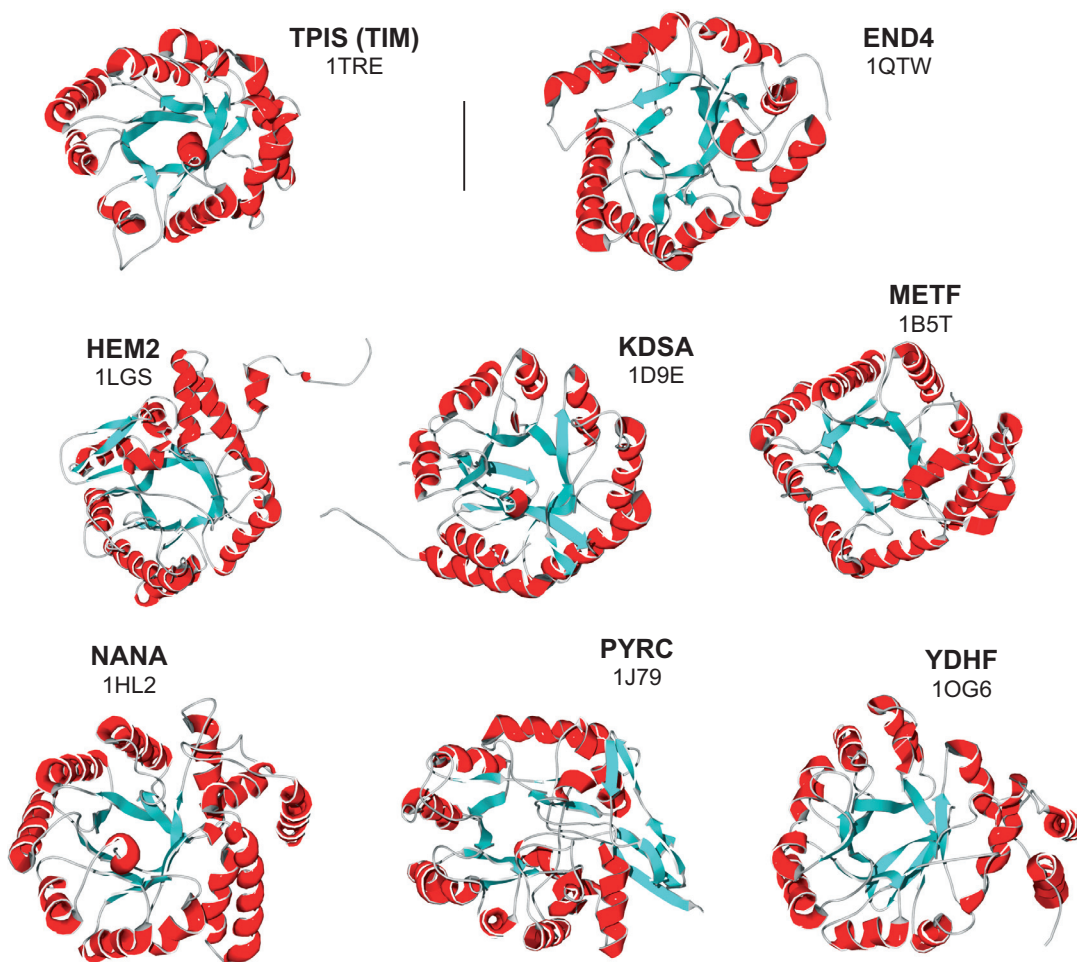
Figure 31: Structures of TIM (βα) barrel proteins.

Triosephosphate isomerase (TIM, SwissProt entry: TPIS_ECOLI, PDB code 1TRE) – the prototype TIM barrel fold, and GroEL class III substrates with known TIM barrel structures. Protein identities are indicated by their SwissProt entry name (please see Supplementary Table S2), followed by the pdb code used for generation of the image. Ribbon structure images were generated with SwissPDB Viewer (version 3.7) and rendered with PovRay (version 3.6).

5.3.4   Additional structures of GroEL substrates

The TIM barrel fold is clearly a very prominent structure among GroEL dependent proteins. However, proteins adopting this fold constitute only about a quarter to a third of all identified GroEL class III substrates (and about 35% of the GroEL substrates by mass). Thus, a very significant number of stringent GroEL dependent proteins display different topologies. Exemplary structures of these proteins are given in Figure 32. METK adopts the S-adenosylmethionine synthetase fold, consisting of three similar intertwined two-layer α+β domains. TYPH consists of three different domains: an all α domain with a methionine syn-

thase domain-like fold, an α/β domain of the nucleoside phosphorylase/phosphoribosyltransferase catalytic domain fold and an α+β domain of the αβ hammerhead fold. Correct classification of these two proteins into class III GroEL substrates was confirmed experimentally (Figures 14, 15, 16, and data not shown). In addition to the strong preference for the TIM barrel topology, no common structural or sequence-related feature could be identified among the structures of class III GroEL substrates. More in depth bioinformatic analyses, such as the use of artificial neural networks for identification of common sequence features, are currently underway to identify any common features.

As already described in section 4.3.4, no other topology in addition to the TIM barrel fold was similarly enriched among all GroEL interacting proteins. The DNA/RNA binding 3-helical bundle (for a representative image of this fold see Figure 33) was found to be mildly enriched among all GroEL interactors, but not particularly enriched among class III. This fold was mostly observed in one of multiple domains in substrate proteins, which possibly reflects an increased folding complexity in multi-domain proteins that renders them chaperone – but not necessarily chaperonin – dependent.

Structures which probably need less chaperonin assistance and are thus less abundant among class III GroEL substrates are the 2-layer sandwich architecture and the β-barrel architecture (Figure 33). The OB and the flavodoxin like folds, do not have a single representative among class III GroEL substrates, while multiple examples were found in the lysate. However, due to the limited number of stringent GroEL substrates, the statistical validity of this finding will have to be tested through future folding experiments or by chaperonin substrate identification and classification from additional organisms.
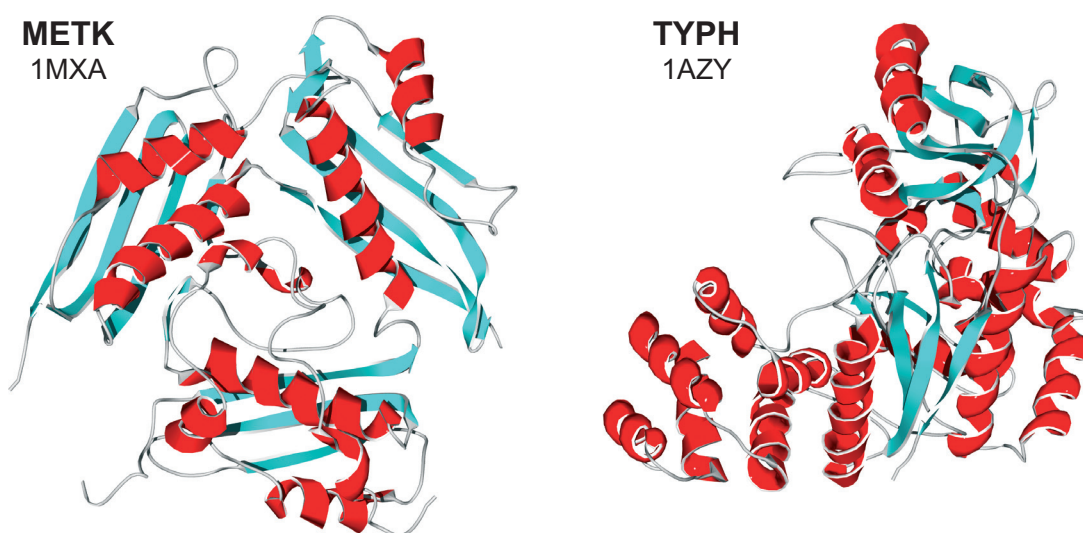
Figure 32: Ribbon structure images of the class III GroEL substrates METK and TYPH.
METK (PDB code 1MXA) consists of three similar αβ domains, TYPH (PDB code 1AZY) contains an all α and two αβ domains. Images were generated as in Figure 31.

DNA/RNA
3-helical bundle
(2HDD)

2-layer
sandwich
(1AA8)

β **barrel**
(1HAV)

flavodoxin like
(1AG9)
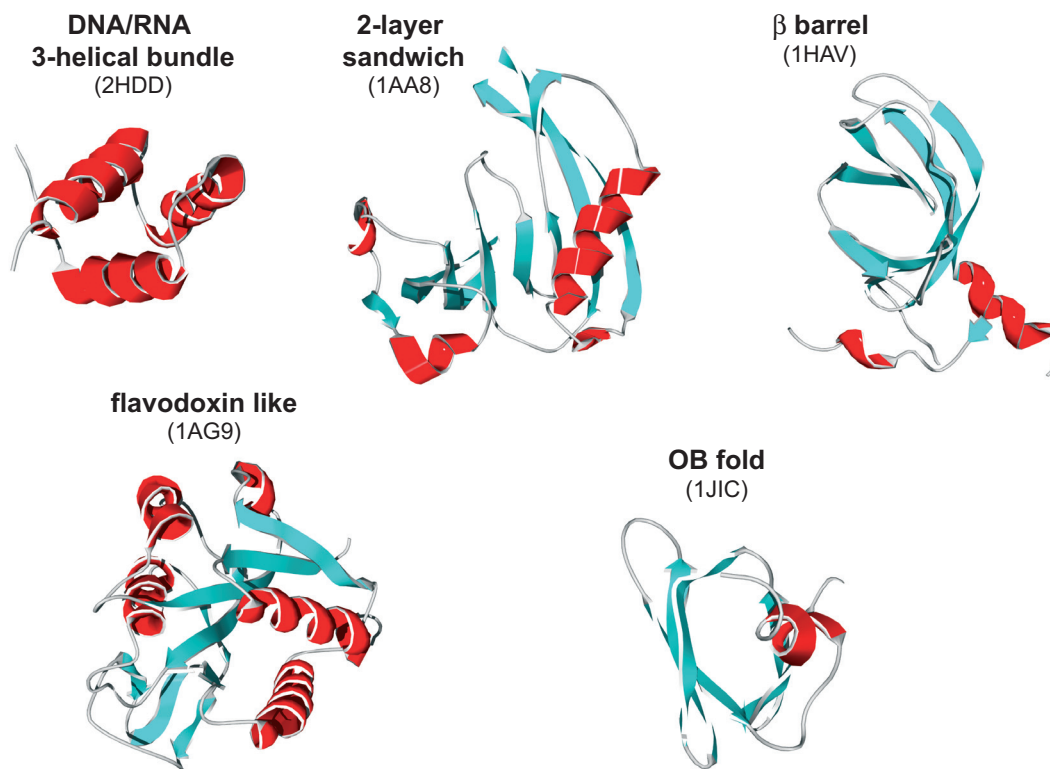
OB fold
(1JIC)

Figure 33: Representative structure images of the SCOP DNA/RNA 3-helical bundle,
the flavodoxin like fold, the CATH 2-layer sandwich and β barrel architectures and the
OB fold.

PDB codes of the utilized structure files are indicated in brackets. Images were generated as
in Figure 31.

Class III proteins are enriched in SCOP folds which contain more distinct SCOP su-
perfamilies than an average lysate protein (D. Frishman, data not shown). They might there-
fore offer structural diversity based on a common evolutionary ancestor. It can therefore be
speculated that GroEL was instrumental to assist the increased number of different, but
closely related, superfamilies of its substrates during evolution. Chaperonin encapsulation
could have helped to overcome folding difficulties which were introduced by mutations dur-
ing evolution. Identification of GroEL substrates from additional organisms will allow
evaluation of this hypothesis.

5.4   GroEL substrate homologs in GroEL-deficient organisms

Orthologs of class III substrates in GroEL deficient organisms (Wong and Houry,
2004) are less abundant than expected based on homology of the genomes. These organisms
contain 25 – 40 % orthologous proteins from *E. coli*, but only harbor 15 – 20 % of class III
orthologs (data not shown). These GroEL-deficient organisms thus appear to have dimin-

ished stringent GroEL dependent substrates in their genome. However, a number of orthologs of class III GroEL substrates, such as the TIM barrel proteins YCFH, GATY, and END4 are still present. These proteins must therefore have evolved to fold in a GroEL independent fashion. Detailed comparison of their folding properties by *in vitro* and *in vivo* experiments, as well as bioinformatic analysis, might reveal further insight into structural or folding properties that enforce strict chaperonin dependence.

6   Perspectives

The present characterization of *in vitro* refolding properties of natural *E. coli* GroEL substrates offers new possibilities for detailed analysis of the chaperonin reaction mechanism and its underlying principles. As has been discussed above, encapsulation by GroEL/GroES can accelerate the folding of substrate proteins (Brinker *et al.*, 2001; Lin and Rye, 2004). This finding is largely based on experiments utilizing the single heterologous substrate *R. rubrum* rubisco. Structural features in substrate proteins that are responsible for accelerated folding inside GroEL/GroES were thus hard to identify. A polypeptide size close to the volume of the surrounding cavity has been suggested to favor accelerated folding (Takagi *et al.*, 2003), but additional factors likely exist that influence such an effect. The availability of *in vitro* refolding assays of seven natural GroEL substrates now allows a more detailed exploration of accelerated folding inside the chaperonin cavity. In particular, DAPA might be a promising candidate for further investigation, since its slow refolding in presence of high concentrations of arginine (data not shown) is in great contrast to the quick refolding, when assisted by GroEL/GroES (Figure 14). On the other hand, it is also possible that certain class III substrates may benefit from the proposed unfolding action of GroEL.

The TIM barrel topology was identified in this study as being highly abundant among stringent GroEL/GroES substrates. However, the absolute number of different TIM barrel proteins among class III was too low for a further characterization of the structural features that determine chaperonin dependence in the TIM barrel architecture, despite intensive effort. Identification and characterization of GroEL interacting proteins from additional organisms, for instance from *M. mazei* and *S. cerevisiae*, are currently underway and may eventually allow a more detailed structural characterization of proteins that are dependent on chaperonin encapsulation for efficient folding. Furthermore, the analysis of additional chaperonin interaction proteomes will allow further examination of the observed chaperonin independence of certain topologies, like the OB and the flavodoxin like folds (see section 5.3.4). The limited number of stringent GroEL substrates in *E. coli* alone does not yet allow a statistically significant confirmation of such independence. Knowledge of chaperonin dependent and independent structures will add valuable parameters for the theoretical inspection and prediction of protein folding pathways.

Orthologs of stringent GroEL substrates have been detected to exist in GroEL deficient organisms (see section 5.4), although in reduced numbers. These orthologous proteins must therefore have evolved to fold independently of the chaperonin system. An example is METK, a stringent GroEL substrate in *E. coli* with an orthologous protein present in the GroEL deficient organism *Ureaplasma urealyticum*. Preliminary *in vivo* chaperonin co-expression experiments suggest that folding of *U. urealyticum* METK is GroEL independent, when heterologously expressed in *E. coli* (data not shown). Preparation of chimera proteins, by replacing *E. coli* METK domains with those from *U. urealyticum* METK, might reveal which part of the *E. coli* METK structure determines chaperonin dependence.

An experimental approach to analyze the development of chaperonin independence, as is observed in GroEL deficient organisms, may be performed in the cytosol of *S. cere-*

visiae, which is not dependent on GroEL/GroES for viability. A promising candidate for such a study would be the class III protein METF, which exhibits an orthologous cytosolic protein in *S. cerevisiae*, MET13. *metF* deletion in *E. coli*, as also deletion of *met13* in *S. cerevisiae*, result in methionine auxotrophy (Blanco *et al.*, 1998, and non-referenced information in the Yeast Genome database, http://db.yeastgenome.org/cgi-bin/phenotype/ phenotype.pl?dbid=S000003093). If METF is able to replace MET13 in *S. cerevisiae* (which remains to be tested), multiple rounds of selection by expression of randomly mutagenized METF from a plasmid in Δ*met13 S. cerevisiae* (in methionine free medium), in combination with heterologous expression of GroEL/GroES might eventually result in chaperonin independent METF with conserved function. This can then be easily screened by omitting GroEL/GroES expression and testing for *S. cerevisiae* viability in absence of methionine.

By using a similar approach as described above for the development of chaperonin independence, it may also be examined experimentally whether during evolution GroEL could have assisted to allow mutation of proteins which led to development of different function, but concurrently inhibited successful spontaneous folding (Fares *et al.*, 2002). Consistent with this would be the enrichment of SCOP folds with a high number of different super-families among class III GroEL substrates, which suggests a broad structural divergence based on a common ancestor (see section 5.3.4).

7   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.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST - a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.

Anderson, R.L. and Markwell, J.P. (1982). D-Galactitol-6-phosphate dehydrogenase. *Methods Enzymol.*, **89 Pt D**, 275-277.

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

Ang, D. and Georgopoulos, C. (1989). The heat-shock-regulated grpE gene of Escherichia coli is required for bacterial growth at all temperatures but is dispensable in certain mutant backgrounds. *J. Bacteriol.*, **171**, 2748-2755.

Arakawa, T. and Tsumoto, K. (2003). The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochem. Biophys. Res. Commun.*, **304**, 148-152.

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

Baldwin, R.L. (1996). On-pathway versus off-pathway folding intermediates. *Fold. Des.*, **1**, R1-8.

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

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.

Ben-Zvi, A.P. and Goloubinoff, P. (2001). Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J. Struct. Biol.*, **135**, 84-93.

Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023-1031.

Blanco, J., Coque, J.J. and Martin, J.F. (1998). The folate branch of the methionine biosynthesis pathway in Streptomyces lividans: disruption of the 5,10-methylenetetrahydrofolate reductase gene leads to methionine auxotrophy. *J. Bacteriol.*, **180**, 1586-1591.

Boisvert, D.C., Wang, J., Otwinowski, Z., Horwich, A.L. and Sigler, P.B. (1996). The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat. Struct. Biol.*, **3**, 170-177.

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.

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.

Brinkkötter, A., Shakeri-Garakani, A. and Lengeler, J.W. (2002). Two class II D-tagatose-bisphosphate aldolases from enteric bacteria. *Arch. Microbiol.*, **177**, 410-419.

Brunel, C., Romby, P., Moine, H., Caillet, J., Grunberg-Manago, M., Springer, M., Ehresmann, B. and Ehresmann, C. (1993). Translational regulation of the Escherichia coli threonyl-tRNA synthetase gene: structural and functional importance of the thrS operator domains. *Biochimie*, **75**, 1167-1179.

Buchan, D.W.A., Rison, S.C.G., Bray, J.E., Lee, D., Pearl, F., Thornton, J.M. and Orengo, C.A. (2003). Gene3D: structural assignments for the biologist and bioinformaticist alike. *Nucl. Acids Res.*, **31**, 469-473.

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

Bukau, B. and Walker, G.C. (1990). Mutations altering heat shock specific subunit of RNA polymerase suppress major cellular defects of E. coli mutants lacking the DnaK chaperone. *EMBO J.*, **9**, 4027-4036.

Bullard, J.M., Cai, Y.C. and Spremulli, L.L. (2000). Expression and characterization of the human mitochondrial leucyl-tRNA synthetase. *Biochim. Biophys. Acta*, **1490**, 245-258.

Burston, S.G., Ranson, N.A. and Clarke, A.R. (1995). The origins and consequences of asymmetry in the chaperonin reaction cycle. *J. Mol. Biol.*, **249**, 138-152.

Cai, X.Y., Jakubowski, H., Redfield, B., Zaleski, B., Brot, N. and Weissbach, H. (1992). Role of the metF and metJ genes on the vitamin B12 regulation of methionine gene expression: involvement of N5-methyltetrahydrofolic acid. *Biochem. Biophys. Res. Commun.*, **182**, 651-658.

Chaudhuri, T.K., Farr, G.W., Fenton, W.A., Rospert, S. and Horwich, A.L. (2001). GroEL/GroES-mediated folding of a protein too large to be encapsulated. *Cell*, **107**, 235-246.

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.

Chen, L. and Sigler, P.B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell*, **99**, 757-768.

Chen, S., Roseman, A.M., Hunter, A.S., Wood, S.P., Burston, S.G., Ranson, N.A., Clarke, A.R. and Saibil, H.R. (1994). Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature*, **371**, 261-264.

Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. and Dobson, C.M. (2002). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci. USA*, **99**, 16419-16426.

Chuang, J.L., Wynn, R.M., Song, J.L. and Chuang, D.T. (1999). GroEL/GroES-dependent reconstitution of alpha2 beta2 tetramers of humanmitochondrial branched chain alpha-ketoacid decarboxylase. Obligatory interaction of chaperonins with an alpha beta dimeric intermediate. *J. Biol. Chem.*, **274**, 10395-10404.

Coyle, J.E., Jaeger, J., Gross, M., Robinson, C.V. and Radford, S.E. (1997). Structural and mechanistic consequences of polypeptide binding by GroEL. *Fold. Des.*, **2**, R93-104.

Daggett, V. and Fersht, A.R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.*, **28**, 18-25.

De Biase, D., Tramonti, A., John, R.A. and Bossa, F. (1996). Isolation, overexpression, and biochemical characterization of the two isoforms of glutamic acid decarboxylase from Escherichia coli. *Protein Expr. Purif.*, **8**, 430-438.

Deuerling, E., Patzelt, H., Vorderwülbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H. and Bukau, B. (2003). Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.*, **47**, 1317-1328.

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

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.

Dobson, C.M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.*, **24**, 329-332.

Dubaquié, Y., Looser, R., Fünfschilling, U., Jenö, P. and Rospert, S. (1998). Identification of in vivo substrates of the yeast mitochondrial chaperonins reveals overlapping but non-identical requirement for hsp60 and hsp10. *EMBO J.*, **17**, 5868-5876.

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

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

Ellis, R.J. (2003). Protein folding: importance of the Anfinsen cage. *Curr. Biol.*, **13**, R881-R883.

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

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.

Fares, M.A., Ruiz-Gonzalez, M.X., Moya, A., Elena, S.F. and Barrio, E. (2002). Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature*, **417**, 398.

Farr, G.W., Fenton, W.A., Chaudhuri, T.K., Clare, D.K., Saibil, H.R. and Horwich, A.L. (2003). Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes. *EMBO J.*, **22**, 3220-3230.

Farr, G.W., Furtak, K., Rowland, M.B., Ranson, N.A., Saibil, H.R., Kirchhausen, T. and Horwich, A.L. (2000). Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell*, **100**, 561-573.

Farr, G.W., Scharl, E.C., Schumacher, R.J., Sondek, S. and Horwich, A.L. (1997). Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms. *Cell*, **89**, 927-937.

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. and Horwich, A.L. (2003). Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q. Rev. Biophys.*, **36**, 229-256.

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.

Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C. and Radford, S.E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.*, **286**, 1597-1608.

Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, **22**, 1302-1306.

Fersht, A.R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.*, **7**, 3-9.

Fersht, A.R. and Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573-582.

Figueiredo, L., Klunker, D., Ang, D., Naylor, D.J., Kerner, M.J., Georgopoulos, C., Hartl, F.U. and Hayer-Hartl, M. (2004). Functional characterization of an archaeal GroEL/GroES chaperonin system. *J. Biol. Chem.*, **279**, 1090-1099.

Fink, A.L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold. Des.*, **3**, R9-23.

Freund, J. and McDermot, K. (1942). Sensitization to horse serum by means of adjuvants. *Proc. Soc. Exp. Biol. Med.*, **49**, 548.

Frishman, D. (2002). Knowledge-based selection of targets for structural genomics. *Protein Eng.*, **15**, 169-183.

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.

Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M. and Brinkman, F.S.L. (2004). PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, bti057.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res.*, **31**, 3784-3788.

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. USA*, **98**, 3762-3767.

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.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M.V., Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Barabasi, A.-L., *et al.* (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J. Bacteriol.*, **185**, 5673-5684.

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

Goldberg, A.L., Moerschell, R.P., Chung, C.H. and Maurizi, M.R. (1994). ATP-dependent protease La (lon) from Escherichia coli. *Methods Enzymol.*, **244**, 350-375.

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.

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.

Hadley, C. and Jones, D.T. (1999). A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. *Structure*, **7**, 1099-1112.

Harlow, E. and Lane, D. (1988) *Antibodies. A laboratory manual.* Cold Spring Harbor Laboratory Press, New York, USA.

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.

Hayer-Hartl, M.K., Ewbank, J.J., Creighton, T.E. and Hartl, F.U. (1994). Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. *EMBO J.*, **13**, 3192-3202.

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. USA*, **93**, 4437-4441.

Hlodan, R., Tempst, P. and Hartl, F.U. (1995). Binding of defined regions of a polypeptide to GroEL and its implications for chaperonin-mediated protein folding. *Nat. Struct. Biol.*, **2**, 587-595.

Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. and Sterner, R. (2001). Dissection of a (beta alpha)(8)-barrel enzyme into two folded halves. *Nat. Struct. Biol.*, **8**, 32-36.

Horwich, A.L., Low, K.B., Fenton, W.A., Hirshfield, I.N. and Furtak, K. (1993). Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. *Cell*, **74**, 909-917.

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.

Huang, Y.S. and Chuang, D.T. (1999). Mechanisms for GroEL/GroES-mediated folding of a large 86-kDa fusion polypeptide in vitro. *J. Biol. Chem.*, **274**, 10405-10412.

Ishihama, Y., Rappsilber, J., Andersen, J.S. and Mann, M. (2002). Microcolumns with self-assembled particle frits for proteomics. *Journal of Chromatography. A.*, **979**, 233-239.

Jackson, S.E. and Fersht, A.R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428-10435.

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. USA*, **101**, 13192-13197.

Jordan, R. and McMacken, R. (1995). Modulation of the ATPase activity of the molecular chaperone DnaK by peptides and the DnaJ and GrpE heat shock proteins. *J. Biol. Chem.*, **270**, 4563-4569.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C. and Gama-Castro, S. (2002). The EcoCyc Database. *Nucl. Acids Res.*, **30**, 56-58.

Karplus, M. and Weaver, D.L. (1976). Protein-folding dynamics. *Nature*, **260**, 404-406.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, **14**, 1-63.

Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA*, **92**, 9029-9033.

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.

Kim, P.S. and Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.*, **59**, 631-660.

Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D.J., Pfeifer, G., Müller, V., Deppenmeier, U., Gottschalk, G., Hartl, F.U. and Hayer-Hartl, M. (2003). Coexistence of group I and group II chaperonins in the archaeon Methanosarcina mazei. *J. Biol. Chem.*, **278**, 33256-33267.

Koradi, R., Billeter, M. and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51-55, 29-32.

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.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. *Nature*, **419**, 171-174.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105-132.

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., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. and Hartl, F.U. (1992a). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, **356**, 683-689.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W. and Hartl, F.U. (1992b). 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.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G. and Mann, M. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature*, **419**, 537-542.

Levinthal, C. (1969) How to Fold Graciously. In DeBrunner, J.T.P. and Munck, E. (eds.), *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.* University of Illinois Press, pp. 22-24.

Liberek, K., Skowyra, D., Zylicz, M., Johnson, C. and Georgopoulos, C. (1991). The Escherichia coli DnaK chaperone, the 70-kDa heat shock protein eukaryotic equivalent, changes conformation upon ATP hydrolysis, thus triggering its dissociation from a bound target protein. *J. Biol. Chem.*, **266**, 14491-14496.

Lin, Z. and Rye, H.S. (2004). Expansion and Compression of a Protein Folding Intermediate by GroEL. *Mol. Cell*, **16**, 23-34.

Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. and Serrano, L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.*, **342**, 345-353.

Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.*, **30**, 264-267.

Lorimer, G.H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J.*, **10**, 5-9.

Markham, G.D., Hafner, E.W., Tabor, C.W. and Tabor, H. (1983). S-adenosylmethionine synthetase (methionine adenosyltransferase) (Escherichia coli). *Methods Enzymol.*, **94**, 219-222.

Martin, J. (2002). Requirement for GroEL/GroES-dependent protein folding under nonpermissive conditions of macromolecular crowding. *Biochemistry*, **41**, 5050-5055.

Martin, J. and Hartl, F.U. (1997). The effect of macromolecular crowding on chaperonin-mediated protein folding. *Proc. Natl. Acad. Sci. USA*, **94**, 1107-1112.

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.

Matagne, A., Radford, S.E. and Dobson, C.M. (1997). Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.*, **267**, 1068-1074.

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.

Mayor, U., Guydosh, N.R., Johnson, C.M., Grossmann, J.G., Sato, S., Jas, G.S., Freund, S.M., Alonso, D.O., Daggett, V. and Fersht, A.R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, **421**, 863-867.

McLennan, N. and Masters, M. (1998). Groe is vital for cell-wall synthesis. *Nature*, **392**, 139.

Meek, J.L. (1980). Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. *Proc. Natl. Acad. Sci. USA*, **77**, 1632-1636.

Minton, A.P. (2001). The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.*, **276**, 10577-10580.

Mogk, A. and Bukau, B. (2004). Molecular chaperones: structure of a protein disaggregase. *Curr. Biol.*, **14**, R78-R80.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rüdiger, S., Roder, D., Langen, H. and Bukau, B. (1999). Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.*, **18**, 6934-6949.

Motojima, F., Chaudhry, C., Fenton, W.A., Farr, G.W. and Horwich, A.L. (2004). Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl. Acad. Sci. USA*, 0406132101.

Naylor, D.J. and Hartl, F.U. (2001). Contribution of molecular chaperones to protein folding in the cytoplasm of prokaryotic and eukaryotic cells. *Biochem. Soc. Symp.*, **68**, 45-68.

Nobelmann, B. and Lengeler, J.W. (1996). Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. *J. Bacteriol.*, **178**, 6790-6795.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics*, **1**, 376-386.

Ong, S.E., Kratchmarova, I. and Mann, M. (2003). Properties of C-13-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.*, **2**, 173-181.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997). CATH--a hierarchic classification of protein domain structures. *Structure*, **5**, 1093-1108.

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.

Otzen, D.E., Itzhaki, L.S., elMasry, N.F., Jackson, S.E. and Fersht, A.R. (1994). Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. USA*, **91**, 10422-10425.

Peden, J.F. (1999). Analysis of codon usage. *Department of Genetics*. University of Nottingham, Nottingham, UK.

Privalov, P.L. (1996). Intermediate states in protein folding. *J. Mol. Biol.*, **258**, 707-725.

Ptitsyn, O.B. and Rashin, A.A. (1975). A model of myoglobin self-organization. *Biophys. Chem.*, **3**, 1-20.

Radford, S.E., Dobson, C.M. and Evans, P.A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, **358**, 302-307.

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

Ranson, N.A., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L. and Saibil, H.R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, **107**, 869-879.

Rappsilber, J., Ishihama, Y. and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.*, **75**, 663-670.

Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002). Large-scale proteomic analysis of the human splicesome. *Genome Res.*, **12**, 1231-1245.

Richardson, A., Schwager, F., Landry, S.J. and Georgopoulos, C. (2001). The importance of a mobile loop in regulating chaperonin/ co-chaperonin interaction: humans versus Escherichia coli. *J. Biol. Chem.*, **276**, 4981-4987.

Rietveld, A.W. and Ferreira, S.T. (1998). Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins. *Biochemistry*, **37**, 933-937.

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.

Rudolph, R., Siebendritt, R. and Kiefhaber, T. (1992). Reversible unfolding and refolding behavior of a monomeric aldolase from Staphylococcus aureus. *Protein Sci.*, **1**, 654-666.

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

Sakikawa, C., Taguchi, H., Makino, Y. and Yoshida, M. (1999). On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J. Biol. Chem.*, **274**, 21251-21256.

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

Sanchez del Pino, M.M. and Fersht, A.R. (1997). Nonsequential unfolding of the alpha/beta barrel protein indole-3-glycerol-phosphate synthase. *Biochemistry*, **36**, 5560-5565.

Schäffer, A.A., Wolf, Y.I., Ponting, C.P., Koonin, E.V., Aravind, L. and Altschul, S.F. (1999). IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. *Bioinformatics*, **15**, 1000-1011.

Schaffner, J., Winter, J., Rudolph, R. and Schwarz, E. (2001). Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins. *Appl. Environ. Microbiol.*, **67**, 3994-4000.

Schechter, A.N., Chen, R.F. and Anfinsen, C.B. (1970). Kinetics of folding of staphylococcal nuclease. *Science*, **167**, 886-887.

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

Schiene, C. and Fischer, G. (2000). Enzymes that catalyse the restructuring of proteins. *Curr. Opin. Struct. Biol.*, **10**, 40-45.

Schiene-Fischer, C., Habazettl, J., Schmid, F.X. and Fischer, G. (2002). The hsp70 chaperone DnaK is a secondary amide peptide bond cis-trans isomerase. *Nat. Struct. Biol.*, **9**, 419-424.

Schneider, C. (ed.). (2000) *Chaperonin protocols* (Methods Mol. Biol., vol. 140). Humana Press Inc., Totowa, New Jersey.

Shea, J.-E., Onuchic, J.N. and Brooks, C.L., III. (1999). Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA*, **96**, 12512-12517.

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.

Shewmaker, F., Kerner, M.J., Hayer-Hartl, M., Klein, G., Georgopoulos, C. and Landry, S.J. (2004). A mobile loop order-disorder transition modulates the speed of chaperonin cycling. *Protein Sci.*, **13**, 2139-2148.

Shinde, U.P., Liu, J.J. and Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. *Nature*, **389**, 520-522.

Shomura, Y., Dragovic, Z., Chang, H.-C., Tzvetkov, N., Young, J.C., Brodsky, J.L., Guerriero, V., Hartl, F.U. and Bracher, A. (2005). Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. *Mol. Cell*, **17**, 367-379.

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

Siegers, K., Bölter, B., Schwarz, J.P., Böttcher, U.M.K., Guha, S. and Hartl, F.U. (2003). TRiC/CCT cooperates with different upstream chaperones in the folding of distinct protein classes. *EMBO J.*, **22**, 5230-5240.

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.

Song, J.-L., Li, J., Huang, Y.-S. and Chuang, D.T. (2003). Encapsulation of an 86-kDa assembly intermediate inside the cavities of GroEL and its single-ring variant SR1 by GroES. *J. Biol. Chem.*, **278**, 2515-2521.

Song, J.L., Wynn, R.M. and Chuang, D.T. (2000). Interactions of GroEL/GroES with a heterodimeric intermediate during alpha2 beta2 assembly of mitochondrial branched-chain alpha-ketoacid dehydrogenase. cis capping of the native-like 86-kDa intermediate by GroES. *J. Biol. Chem.*, **275**, 22305-22312.

Spring, T.G. and Wold, F. (1975). Enolase from Escherichia coli. *Methods Enzymol.*, **42**, 323-329.

Sternlicht, H., Farr, G.W., Sternlicht, M.L., Driscoll, J.K., Willison, K. and Yaffe, M.B. (1993). The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin in vivo. *Proc. Natl. Acad. Sci. USA*, **90**, 9422-9426.

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. USA*, **100**, 11367-11372.

Tanford, C. (1962). Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.*, **84**, 4240-4247.

Tatusov, R.L., Koonin, E.V. and Lipman, D.J. (1997). A genomic perspective on protein families. *Science*, **278**, 631-637.

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.

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. USA*, **93**, 4030-4035.

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. USA*, **76**, 4350-4354.

Vauterin, M., Frankard, V. and Jacobs, M. (2000). Functional rescue of a bacterial dapA auxotroph with a plant cDNA library selects for mutant clones encoding a feedback-insensitive dihydrodipicolinate synthase. *Plant J.*, **21**, 239-248.

Vetsch, M., Puorger, C., Spirig, T., Grauschopf, U., Weber-Ban, E.U. and Glockshuber, R. (2004). Pilus chaperones represent a new type of protein-folding catalyst. *Nature*, **431**, 329-333.

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.

Vorderwülbecke, S., Kramer, G., Merz, F., Kurz, T.A., Rauch, T., Zachmann-Brand, B., Bukau, B. and Deuerling, E. (2004). Low temperature or GroEL/ES overproduction permits growth of Escherichia coli cells lacking trigger factor and DnaK. *FEBS Lett.*, **559**, 181-187.

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

Wei, Y. and Newman, E.B. (2002). Studies on the role of the metK gene product of Escherichia coli K-12. *Mol. Microbiol.*, **43**, 1651-1656.

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.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*, **70**, 697-701.

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.

Wolff, J.B. and Kaplan, N.O. (1956). D-Mannitol 1-phosphate dehydrogenase from Escherichia coli. *J. Biol. Chem.*, **218**, 849-869.

Wong, P. and Houry, W.A. (2004). Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. *J. Struct. Biol.*, **146**, 79-89.

Wu, Y. and Matthews, C.R. (2002). Parallel channels and rate-limiting steps in complex protein folding reactions: Prolyl isomerization and the alpha subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol.*, **323**, 309-325.

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.

Yaffe, M.B., Farr, G.W., Miklos, D., Horwich, A.L., Sternlicht, M.L. and Sternlicht, H. (1992). TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature*, **358**, 245-248.

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.

Young, J.C., Barral, J.M. and Ulrich Hartl, F. (2003). More than folding: localized functions of cytosolic chaperones. *Trends Biochem. Sci.*, **28**, 541-547.

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.

Zitzewitz, J.A., Gualfetti, P.J., Perkons, I.A., Wasta, S.A. and Matthews, C.R. (1999). Identifying the structural boundaries of independent folding domains in the alpha subunit of tryptophan synthase, a beta/alpha barrel protein. *Protein Sci.*, **8**, 1200-1209.

Zylicz, M., Ang, D. and Georgopoulos, C. (1987). The grpE protein of Escherichia coli. Purification and properties. *J. Biol. Chem.*, **262**, 17437-17442.

Zylicz, M., Yamamoto, T., McKittrick, N., Sell, S. and Georgopoulos, C. (1985). Purification and properties of the dnaJ replication protein of Escherichia coli. *J. Biol. Chem.*, **260**, 7591-7598.

8 Appendices

8.1 Supplementary Tables

Supplementary Table S1: Proteins identified non-specifically among GroEL interactors. Identification method: (a) Proteins identified to non-specifically interact with the IMAC resin in GroEL/GroES/substrate complex purification. (b) Arg-[13]C6 labeled proteins associated with GroEL/GroES/substrate complexes upon lysis of unlabeled *E. coli* cells expressing GroES-(His)$_6$ in presence of Arg-[13]C6 labeled lysate.

SwissID	Description	Molecular mass [kDa]	identification method*
atda_ecoli	Spermidine N(1)-acetyltransferase (EC 2.3.1.57) (Diamine-acetyltransferase) (SAT).	21.8	b
eftu_ecoli	Elongation factor Tu (EF-Tu) (P-43).	43.2	a, b
fabz_ecoli	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (EC 4.2.1.-); ((3R)-hydroxymyristoyl ACP dehydrase) (17 kDa actomyosin component).	17.0	a, b
fur_ecoli	Ferric uptake regulation protein (Ferric uptake regulator).	16.8	a, b
glms_ecoli	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]; (EC 2.6.1.16); (Hexosephosphate aminotransferase); (D-fructose-6-phosphate amidotransferase); (GFAT); (L-glutamine-D-fructose-6-phosphate- amidotransferase); (Glucosamine-6-phosphate synthase).	66.8	a, b
hdfr_ecoli	HTH-type transcriptional regulator hdfR (H-NS-depdendent flhDC regulator).	31.7	b
hfq_ecoli	Hfq protein (Host factor-I protein) (HF-I) (HF-1).	11.0	b
mpl_ecoli	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (EC 6.3.2.-) (Murein peptide ligase).	49.9	b
puru_ecoli	Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4) hydrolase).	31.9	b
rl1_ecoli	50S ribosomal protein L1.	24.6	b
rl11_ecoli	50S ribosomal protein L11.	14.7	b
rl14_ecoli	50S ribosomal protein L14.	13.5	b
rl15_ecoli	50S ribosomal protein L15.	15.0	b
rl17_ecoli	50S ribosomal protein L17.	14.4	b

SwissID	Description	Molecular mass [kDa]	identifi-cation method*
rl2_ecoli	50S ribosomal protein L2.	29.7	b
rl28_ecoli	50S ribosomal protein L28.	8.9	b
rl32_ecoli	50S ribosomal protein L32.	63.2	a
rl9_ecoli	50S ribosomal protein L9.	15.8	b
rs1_ecoli	30S ribosomal protein S1.	61.2	b
rs12_ecoli	30S ribosomal protein S12.	13.6	b
rs15_ecoli	30S ribosomal protein S15.	10.1	a, b
rs19_ecoli	30S ribosomal protein S19.	10.3	b
rs2_ecoli	30S ribosomal protein S2.	26.6	b
rs20_ecoli	30S ribosomal protein S20.	9.6	b
rs3_ecoli	30S ribosomal protein S3.	25.9	b
rs4_ecoli	30S ribosomal protein S4.	23.3	b
rs5_ecoli	30S ribosomal protein S5.	17.5	b
rs7_ecoli	30S ribosomal protein S7.	19.9	b
rs9_ecoli	30S ribosomal protein S9.	14.7	b
rsua_ecoli	Ribosomal small subunit pseudouridine synthase A (EC 4.2.1.70) (16S pseudouridylate 516 synthase) (16S pseudouridine 516 synthase) (Uracil hydrolyase).	25.9	b
slyd_ecoli	FKBP-type peptidyl-prolyl cis-trans isomerase slyD (EC 5.2.1.8); (PPIase) (Rotamase) (Histidine-rich protein) (WHP).	20.9	a, b
ybha_ecoli	Hypothetical protein ybhA.	30.2	b

Supplementary Table S1

Supplementary Table S2: Substrates of GroEL, identified by LC-MS/MS analysis of GroEL/*Mm*GroES-(His)$_6$/substrate complexes. For predicted GroEL substrate class and enrichment factor, please see main text. Protein names, identifiers and oligomeric state are according to the SwissProt database (Gasteiger *et al.*, 2003, http://www.expasy.org/). For SCOP fold abbreviations please see http://scop.mrc-lmb.cam.ac.uk/scop/parse/dir.des.scop.txt_1.65 (Lo Conte *et al.*, 2002). Essentiality of *E. coli* proteins was determined by Gerdes *et al.* (2003) (1: essential, 0: not essential for *E. coli* MG1655 cell growth). Prediction of subcellular localization with PSORTb version 2.0 (Gardy *et al.*, 2004) was derived from the pre-computed genome file of *E. coli* K12, available at http://www.psort.org/genomes/NC_000913.out. Functional categorization is according to the COG database (Tatusov *et al.*, 1997, http://www.ncbi.nlm.nih.gov/COG/). For COG categories please see Supplementary Table S3.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
thi2_ecoli	P33636	Thioredoxin 2 (EC 1.8.1.8) (Protein-disulfide reductase) (Disulfide reductase) (Trx2).	1	15.6	c.47	1	5		Cytoplasmic	Cytoplasmic	O
tpx_ecoli	P37901	Thiol peroxidase (EC 1.11.1.-) (Scavengase P20).	1	17.7	c.47	0	4.75		Periplasmic	Unknown	O
ptga_ecoli	P08837	PTS system, glucose-specific IIA component (EIIA-GLC) (Glucose-permease IIA component) (Phosphotransferase enzyme II, A component) (EC 2.7.1.69) (EIII-GLC).	1	18.1	b.84	0	4.73		Cytoplasmic	Cytoplasmic	G
faba_ecoli	P18391	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.60) (Beta-hydroxydecanoyl thioester dehydrase).	1	18.8	d.38	1	6.16	Homodimer	Cytoplasmic	Cytoplasmic	I
ipyr_ecoli	P17288	Inorganic pyrophosphatase (EC 3.6.1.1) (Pyrophosphate phospho-hydrolase) (PPase).	1	19.6	b.40	1	5.03	Homohexamer	Cytoplasmic	Cytoplasmic	C
ahpc_ecoli	P26427	Alkyl hydroperoxide reductase subunit C (EC 1.6.4.-) (Alkyl hydroperoxide reductase protein C22) (SCRP-23) (Sulfate starvation-induced protein 8) (SSI8).	1	20.6	c.47	0	5.03	Homodimer (By similarity)	Cytoplasmic	Cytoplasmic	O
rrf_ecoli	P16174	Ribosome recycling factor (Ribosome releasing factor) (RRF).	1	20.6	d.67	1	6.44		Cytoplasmic	Cytoplasmic	J
grpe_ecoli	P09372	GrpE protein (HSP-70 cofactor) (Heat shock protein B25.3) (HSP24).	1	21.8	b.73	1	4.68		Cytoplasmic	Cytoplasmic	O

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
deod_ecoli	P09743	Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase) (PNP).	1	25.8	c.56; c.48	0	5.42	Homohexamer		Cytoplasmic	F
thig_ecoli	P30139	Thiazole biosynthesis protein thiG.	1	26.9	c.1	0	5.36		Cytoplasmic	Cytoplasmic	F
deoc_ecoli	P00882	Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase) (DERA).	1	27.7	c.1	0	5.5	Monomer and homodimer	Cytoplasmic	Cytoplasmic	F
panb_ecoli	P31057	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11) (Ketopantoate hydroxymethyltransferase).	1	28.2	c.78; c.1	0	5.15	Hexamer (Potential)		Unknown	H
gpma_ecoli	P31217	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (EC 5.4.2.1) (Phosphoglyceromutase) (PGAM) (BPG-dependent PGAM) (dPGM).	1	28.4	c.60	0	5.86	Homodimer		Unknown	G
efts_ecoli	P02997	Elongation factor Ts (EF-Ts).	1	30.3	d.43; a.5	1	5.22	Heterotetramer composed of two EF-Ts.EF-Tu dimer complex.	Cytoplasmic	Cytoplasmic	J
rbsb_ecoli	P02925	D-ribose-binding periplasmic protein precursor.	1	31.0	c.93	0	5.99		Periplasmic	Periplasmic	G
blat_ecoli	P00810	Beta-lactamase TEM precursor (EC 3.5.2.6) (TEM-1) (TEM-2) (TEM-3) (TEM-4) (TEM-5) (TEM-6) (TEM-8/CAZ-2) (TEM-16/CAZ-7) (TEM-24/CAZ-6) (IRT-4) (Penicillinase).	1	31.5	e.3	0	5.46		Periplasmic	Periplasmic	M
g3p1_ecoli	P06977	Glyceraldehyde 3-phosphate dehydrogenase A (EC 1.2.1.12) (GAPDH-A).	1	35.4	c.2; d.81	1	6.58	Homotetramer	Cytoplasmic	Cytoplasmic	G

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
rpoa_ecoli	P00574	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).	1	36.5	a.60; d.74; d.181		4.98	Homodimer. The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta' and 1 omega subunit. When a sigma factor is associated with the core the holoenzyme is formed, which can initiate transcription.		Cytoplasmic	K
ompa_ecoli	P02934	Outer membrane protein A precursor (Outer membrane protein II*).	1	37.2	f.4	0	5.6	Monomer (Probable)	Integral membrane protein. Outer membrane.	OuterMembrane	M
ynce_ecoli	P76116	Hypothetical protein yncE precursor.	1	38.6	b.69; b.70; b.68	1	8.8			Unknown	S
ompc_ecoli	P06996	Outer membrane protein C precursor (Porin ompC) (Outer membrane protein 1B).	1	40.4	f.4	0	4.48	Homotrimer	Integral membrane protein. Outer membrane.	OuterMembrane	M
pgk_ecoli	P11665	Phosphoglycerate kinase (EC 2.7.2.3).	1	41.0	c.86	1	5.08	Monomer	Cytoplasmic	Cytoplasmic	G
fabb_ecoli	P14926	3-oxoacyl-[acyl-carrier-protein] synthase I (EC 2.3.1.41) (Beta-ketoacyl-ACP synthase I) (KAS I).	1	42.6	c.95	1	5.35	Homodimer	Cytoplasmic	Cytoplasmic	I
acka_ecoli	P15046	Acetate kinase (EC 2.7.2.1) (Acetokinase).	1	43.3	c.55	1	5.85	Homodimer	Cytoplasmic	Cytoplasmic	C

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
sera_ecoli	P08328	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (PGDH).	1	44.0	c.2; d.58; c.23	0	5.93	Homotetramer		Cytoplasmic	E
glya_ecoli	P00477	Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase) (SHMT).	1	45.3	c.67	1	6.03	Homotetramer	Cytoplasmic	Cytoplasmic	E
eno_ecoli	P08324	Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase).	1	45.5	d.54; c.1	0	5.32	Homodimer	Cytoplasmic	Cytoplasmic	G
pura_ecoli	P12283	Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase) (AdSS) (AMPSase).	1	47.2	c.37	0	5.32	Homodimer	Cytoplasmic	Cytoplasmic	F
tig_ecoli	P22257	Trigger factor (TF).	1	48.2	d.26	0	4.83	Homodimer and monomer		Cytoplasmic	O
kpy1_ecoli	P14178	Pyruvate kinase I (EC 2.7.1.40) (PK-1).	1	50.7	b.58; c.49; c.1	0	5.77	Homotetramer		Unknown	G
6pgd_ecoli	P00350	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).	1	51.5	c.2; a.100	0	5.05			Unknown	G
syn_ecoli	P17242	Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase) (AsnRS).	1	52.4	b.40; d.104	1	5.17	Homodimer	Cytoplasmic	Cytoplasmic	J
oppa_ecoli	P23843	Periplasmic oligopeptide-binding protein precursor.	1	60.9	c.94	0	5.85		Periplasmic	Periplasmic	E
odp2_ecoli	P06959	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12) (E2).	1	66.0	b.84; a.9; c.43	1	5.09	Forms a 24-polypeptide structural core with octahedral symmetry.		CytoplasmicMembrane	C
cira_ecoli	P17315	Colicin I receptor precursor.	1	73.9	f.4	0	5.03		Outer membrane	OuterMembrane	P

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
pnp_ecoli	P05055	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide phosphorylase) (PNPase).	1	77.1	d.51; d.14; d.101; d.52; b.40; a.4	1	5.11	Homotrimer	Cytoplasmic	Cytoplasmic	J
efg_ecoli	P02996	Elongation factor G (EF-G).	1	77.5	d.14; c.37; d.58; b.43	1	5.24		Cytoplasmic	Cytoplasmic	J
odp1_ecoli	P06958	Pyruvate dehydrogenase E1 component (EC 1.2.4.1).	1	99.5	c.48; c.36	0	5.46	Homodimer		Unknown	C
muli_ecoli	P02937	Major outer membrane lipoprotein precursor (Murein-lipoprotein).	1 or 2	8.3	-		8.12		Attached to the outer membrane by a lipid anchor	Unknown	N
if3_ecoli	P02999	Translation initiation factor IF-3.	1 or 2	20.6	d.15; d.68	1	9.54	Monomer	Cytoplasmic	Cytoplasmic	J
yhgi_ecoli	P46847	Protein yhgI.	1 or 2	21.0	-	0	4.52			Cytoplasmic	O
dldh_ecoli	P00391	Dihydrolipoamide dehydrogenase (EC 1.8.1.4) (E3 component of pyruvate and 2-oxoglutarate dehydrogenases complexes) (Glycine cleavage system L protein).	1 or 2	50.6	c.3; d.87; c.4	1	5.79	Homodimer	Cytoplasmic	Cytoplasmic	C
ppic_ecoli	P39159	Peptidyl-prolyl cis-trans isomerase C (EC 5.2.1.8) (PPIase C) (Rotamase C) (Parvulin).	2	10.1	d.26	0	9.23		Cytoplasmic	OuterMembrane	O
yohl_ecoli	P76424	Hypothetical protein yohL.	2	10.1	-	1	8.79			Cytoplasmic	S
ydhd_ecoli	P37010	Protein ydhD.	2	12.9	c.47	1	4.75			Unknown	O
nikr_ecoli	P28910	Nickel responsive regulator.	2	15.1	-	0	5.77	Homotetramer		Cytoplasmic	K
yjbq_ecoli	P32698	Hypothetical protein yjbQ.	2	15.7	-	0	6.49			Cytoplasmic	S

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
uspg_ecoli	P39177	Universal stress protein G.	2	15.9	c.29	0	6.03	Interacts with groEL.		Cytoplasmic	T
moac_ecoli	P30747	Molybdenum cofactor biosynthesis protein C.	2	17.3	d.58	0	6.58	Homohexamer		Unknown	H
yfhp_ecoli	P77484	Hypothetical protein yfhP.	2	17.3	-	0	6.82			Unknown	K
grea_ecoli	P21346	Transcription elongation factor greA (Transcript cleavage factor greA).	2	17.6	a.2; d.26	0	4.71			Cytoplasmic	K
ppib_ecoli	P23869	Peptidyl-prolyl cis-trans isomerase B (EC 5.2.1.8) (PPIase B) (Rotamase B).	2	18.2	b.62	1	5.51		Cytoplasmic	Cytoplasmic	O
moab_ecoli	P30746	Molybdenum cofactor biosynthesis protein B.	2	18.5	c.57	0	5.73			Cytoplasmic	H
dps_ecoli	P27430	DNA protection during starvation protein.	2	18.6	a.25	0	5.72	Associates into a complex of 12 subunits forming two stacked hexameric rings.		Unknown	L
nuoe_ecoli	P33601	NADH-quinone oxidoreductase chain E (EC 1.6.99.5) (NADH dehydrogenase I, chain E) (NDH-1, chain E) (NUO5).	2	18.6	c.47	0	5.4	Composed of 13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		Cytoplasmic	C
mug_ecoli	P43342	G/U mismatch-specific DNA glycosylase (EC 3.2.2.-) (Mismatch-specific uracil DNA-glycosylase) (UDG).	2	18.7	c.18	0	9.17		Cytoplasmic (Potential)	Cytoplasmic	L
luxs_ecoli	P45578	S-ribosylhomocysteinase (EC 3.13.1.-) (Autoinducer-2 production protein LuxS) (AI-2 synthesis protein).	2	19.3	d.185	0	5.18	Homodimer (By similarity)		Cytoplasmic	T
arok_ecoli	P24167	Shikimate kinase I (EC 2.7.1.71) (SKI).	2	19.4	c.37	1	5.26		Cytoplasmic (Probable)	Cytoplasmic	E

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
yfbu_ecoli	P76492	Protein yfbU.	2	19.5	-	0	6.07			Cytoplasmic	S
seqa_ecoli	P36658	SeqA protein.	2	20.3	-	1	8.82			Cytoplasmic	L
nusg_ecoli	P16921	Transcription antitermination protein nusG.	2	20.4	-	0	6.33			Cytoplasmic	K
rimm_ecoli	P21504	16S rRNA processing protein rimM (21K).	2	20.6	-	1	4.61		Cytoplasmic (Potential)	Cytoplasmic	J
riml_ecoli	P13857	Ribosomal-protein-serine acetyltransferase (EC 2.3.1.-) (Acetylating enzyme for N-terminal of ribosomal protein L7/L12).	2	20.7	d.108	1	5.86		Cytoplasmic	Cytoplasmic	J
nudh_ecoli	Q46930	(Di)nucleoside polyphosphate hydrolase (EC 3.6.1.-) (Ap5A pyrophosphatase).	2	20.8	d.113	0	10.05	Monomer		Unknown	L
pth_ecoli	P23932	Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH).	2	21.1	c.56	1	9.04	Monomer	Cytoplasmic	CytoplasmicMembrane	J
hemg_ecoli	P27863	Protoporphyrinogen oxidase (EC 1.3.3.4) (PPO).	2	21.2	c.23	1	9.68	Belongs to a multi-protein complex.		Unknown	C
tehb_ecoli	P25397	Tellurite resistance protein tehB.	2	22.5	c.66	0	6.84		Cytoplasmic (Potential)	Cytoplasmic	Q
yihx_ecoli	P32145	Hypothetical protein yihX.	2	22.7	c.108	0	5.18			Cytoplasmic	R
thie_ecoli	P30137	Thiamine-phosphate pyrophosphorylase (EC 2.5.1.3) (TMP pyrophosphorylase) (TMP-PPase) (Thiamine-phosphate synthase).	2	23.0	c.1	0	5.54			Unknown	H
ycio_ecoli	P45847	Protein yciO.	2	23.2	d.115	0	5.97			Cytoplasmic	J
engb_ecoli	P24253	Probable GTP-binding protein engB.	2	23.6	c.37	1	6.86			Unknown	D
rcsb_ecoli	P14374	Capsular synthesis regulator component B.	2	23.7	c.23; a.4	0	6.85			Cytoplasmic	T
ycbl_ecoli	P75849	Hypothetical protein ycbL.	2	23.8	d.157	0	4.95			Unknown	R

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
uvry_ecoli	P07027	Response regulator uvrY.	2	23.9	c.23; a.4	0	6.53		Cytoplasmic (Probable)	Cytoplasmic	T
glr2_ecoli	P39811	Glutaredoxin 2 (Grx2).	2	24.4	c.47; a.45	0	7.72			Unknown	O
yoda_ecoli	P76344	Protein yodA.	2	24.8	-	0	5.66			Unknown	R
yadf_ecoli	P36857	Protein yadF.	2	25.1	c.53	1	6.16			Cytoplasmic	P
ygea_ecoli	P03813	Hypothetical protein ygeA.	2	25.2	c.78	0	5.09			Cytoplasmic	M
inaa_ecoli	P27294	Protein inaA.	2	25.3	-	0	9.12			Cytoplasmic	S
trmh_ecoli	P19396	tRNA (Guanosine-2'-O-)-methyltransferase (EC 2.1.1.34) (tRNA [GM18] methyltransferase).	2	25.3	c.116	0	6.71		Cytoplasmic (Potential)	Cytoplasmic	J
arad_ecoli	P08203	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose isomerase).	2	25.5	c.74	0	5.73			Unknown	G
phop_ecoli	P23836	Transcriptional regulatory protein phoP.	2	25.5	c.23; a.4	0	5.1		Cytoplasmic (Probable)	Cytoplasmic	T
fabg_ecoli	P25716	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl- acyl carrier protein reductase).	2	25.6	c.2	1	6.76			Cytoplasmic	Q
proq_ecoli	P45577	ProP effector.	2	25.9	a.136	0	9.66		Cytoplasmic (Potential)	Unknown	T
cpxr_ecoli	P16244	Transcriptional regulatory protein cpxR.	2	26.3	c.23; a.4	0	5.39		Cytoplasmic (Probable)	Cytoplasmic	T
pyrf_ecoli	P08244	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase) (OMPDCase) (OMPdecase).	2	26.4	c.1	1	5.81	Homodimer		Unknown	F
ompr_ecoli	P03025	Transcriptional regulatory protein ompR.	2	27.4	c.23; a.4	0	6.04	Monomer and multimer	Cytoplasmic	Cytoplasmic	T

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
lpxa_ecoli	P10440	Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase (EC 2.3.1.129) (UDP-N-acetylglucosamine acyltransferase).	2	28.1	b.81	1	6.63	Homotrimer	Cytoplasmic	Cytoplasmic	M
ybff_ecoli	P75736	Putative esterase/lipase ybfF (EC 3.1.-.-).	2	28.4	c.69	0	5.86			Unknown	R
pstb_ecoli	P07655	Phosphate import ATP-binding protein pstB (EC 3.6.3.27) (Phosphate- transporting ATPase) (ABC phosphate transporter).	2	28.9	c.37	0	6.12	The complex is composed of two ATP-binding proteins (pstB), two transmembrane proteins (pstC and pstA) and a solute- binding protein (pstS) (Probable).	Inner membrane-associated	CytoplasmicMembrane	P
yjjv_ecoli	P39408	Putative deoxyribonuclease yjjV (EC 3.1.21.-).	2	28.9	c.1	1	6.14			Cytoplasmic	L
cyse_ecoli	P05796	Serine acetyltransferase (EC 2.3.1.30) (SAT).	2	29.3	b.81	1	6.05	Homohexamer. Dimer of a homotrimer.	Cytoplasmic	Cytoplasmic	E
kdsa_ecoli	P17579	2-dehydro-3-deoxyphosphooctonate aldolase (EC 2.5.1.55) (Phospho-2- dehydro-3-deoxyoctonate aldolase) (3-deoxy-D-manno-octulosonic acid 8-phosphate synthetase) (KDO-8-phosphate synthetase) (KDO 8-P synthase) (KDOPS).	2	30.8	c.1	0	6.32	Homotrimer	Cytoplasmic	Cytoplasmic	M
yffs_ecoli	P76550	Hypothetical protein yffS.	2	31.0	-	1	5.45			Cytoplasmic	S
ypt2_ecoli	Q99390	Hypothetical 31.7 kDa protein in TRAX-FINO intergenic region (ORFC).	2	31.7	c.69		5.37			Unknown	R
ybbn_ecoli	P77395	Protein ybbN.	2	31.8	a.118; c.47	0	4.5			Cytoplasmic	O

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
yijo_ecoli	P32677	Hypothetical transcriptional regulator yijO.	2	32.1	a.4	0	8.34			Cytoplasmic	K
ymia_ecoli	P77739	Hypothetical protein yniA.	2	32.5	d.144		4.98			Unknown	S
hslo_ecoli	P45803	33 kDa chaperonin (Heat shock protein 33) (HSP33).	2	32.5	d.193		4.35		Cytoplasmic	Cytoplasmic	O
rlub_ecoli	P37765	Ribosomal large subunit pseudouridine synthase B (EC 4.2.1.70) (Pseudouridylate synthase) (Uracil hydrolyase).	2	32.7	d.58; d.66	1	10.03			Cytoplasmic	J
rob_ecoli	P27292	Right origin-binding protein.	2	33.1	d.60; a.4	0	6.66			Cytoplasmic	K
ycjz_ecoli	P77333	Putative HTH-type transcriptional regulator ycjZ.	2	33.5	c.94; a.4	0	9.07			Cytoplasmic	K
yadb_ecoli	P27305	Hypothetical protein yadB.	2	33.6	c.26	0	6.42			Cytoplasmic	J
ydhf_ecoli	P76187	Hypothetical oxidoreductase ydhF (EC 1.-.-.-).	2	33.6	c.1	0	5.77			Cytoplasmic	R
kprs_ecoli	P08330	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1) (RPPK) (Phosphoribosyl pyrophosphate synthetase) (P-Rib-PP synthetase) (PRPP synthetase).	2	34.1	c.61	1	5.23		Cytoplasmic	Cytoplasmic	F
oxyr_ecoli	P11721	Hydrogen peroxide-inducible genes activator (Morphology and auto- aggregation control protein).	2	34.3	c.94; a.4	0	5.96	Homodimer and homotetramer		Cytoplasmic	K
yeat_ecoli	P76250	Putative HTH-type transcriptional regulator yeaT.	2	34.6	c.94; a.4		6.21			Unknown (This protein may have multiple localization sites.)	K
ybib_ecoli	P30177	Hypothetical protein ybiB.	2	35.0	a.46	0	6.38			Cytoplasmic	E

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
acca_ecoli	P30867	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (EC 6.4.1.2).	2	35.1	-	1	5.76	Acetyl-CoA carboxylase is a heterohexamer of biotin carboxyl carrier protein, biotin carboxylase and the two subunits of carboxyl transferase in a 2:2 complex.		Cytoplasmic	I
cysb_ecoli	P06613	HTH-type transcriptional regulator cysB (Cys regulon transcriptional activator).	2	36.2	c.94; a.4	0	6.87	Homotetramer (By similarity)	Cytoplasmic	Cytoplasmic	K
dusa_ecoli	P32695	tRNA-dihydrouridine synthase A (EC 1.-.-.).	2	36.8	c.1	0	6.06			Cytoplasmic	J
mreb_ecoli	P13519	Rod shape-determining protein mreB.	2	37.0	c.55	1	5.19			Cytoplasmic	D
moaa_ecoli	P30745	Molybdenum cofactor biosynthesis protein A.	2	37.3	-	0	8.15			Cytoplasmic	H
gatd_ecoli	P37190	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251).	2	37.4	c.2; c.66; b.35	1	5.94			Cytoplasmic	E
rsmc_ecoli	P39406	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52) (rRNA (guanine-N(2)-)-methyltransferase) (16S rRNA m2G1207 methyltransferase).	2	37.5	c.66	0	6			Unknown	J
inh5_ecoli	P76071	Transposase insH for insertion sequence element IS5Y.	2	37.8	-	1	9.56			Cytoplasmic	L
yihe_ecoli	P32127	Hypothetical protein yihE.	2	38.1	d.144	0	4.99			Cytoplasmic	R
tas_ecoli	Q46933	Tas protein.	2	38.5	c.1	0	6.27			Unknown	C
pyrc_ecoli	P05020	Dihydroorotase (EC 3.5.2.3) (DHOase).	2	38.7	c.1	0	5.77	Homodimer		Cytoplasmic	F

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
yghz_ecoli	Q46851	Hypothetical protein yghZ.	2	38.8	c.1	0	6.72			Unknown	C
alf_ecoli	P11604	Fructose-bisphosphate aldolase class II (EC 4.1.2.13) (FBP aldolase).	2	39.0	c.1	1	5.52	Homodimer		Unknown	G
dcup_ecoli	P29680	Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).	2	39.2	c.1	1	5.88		Cytoplasmic (Probable)	Unknown	H
insh_ecoli	P03837	Transposase insH for insertion sequence element IS5.	2	39.3	-	0	9.58			Cytoplasmic	L
ompf_ecoli	P02931	Outer membrane protein F precursor (Porin ompF) (Outer membrane protein 1A) (Outer membrane protein B).	2	39.3	f.4	0	4.64	Homotrimer	Integral membrane protein. Outer membrane.	OuterMembrane	M
serc_ecoli	P23721	Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).	2	39.7	c.67	0	5.37	Homodimer	Cytoplasmic	Cytoplasmic	H
ybbb_ecoli	P33667	Hypothetical protein ybbB.	2	41.1	c.46	0	5.93			Unknown	R
entc_ecoli	P10377	Isochorismate synthase entC (EC 5.4.99.6) (Isochorismate mutase).	2	42.9	d.161	0	5.49	Monomer		Unknown	H
argm_ecoli	P77581	Succinylornithine transaminase (EC 2.6.1.-) (Succinylornithine aminotransferase) (Carbon starvation protein C).	2	43.7	c.67	0	5.91			Unknown	E
odo2_ecoli	P07016	Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (E2).	2	43.9	b.84; a.9; c.43	1	5.58	Forms a 24-polypeptide structural core with octahedral symmetry.		Cytoplasmic	C
iscs_ecoli	P39171	Cysteine desulfurase (EC 4.4.1.-) (ThiI transpersulfidase) (NifS protein homolog).	2	45.1	c.67	1	5.94			Cytoplasmic	E

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
glpb_ecoli	P13033	Anaerobic glycerol-3-phosphate dehydrogenase subunit B (EC 1.1.99.5) (G-3-P dehydrogenase).	2	45.4	c.3	0	5.75	Composed of a catalytic glpA/B dimer and of membrane bound glpC.	Loosely bound to the cytoplasmic membrane often occurring in vesicles associated with fumarate reductase	Cytoplasmic	E
gsa_ecoli	P23893	Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA) (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).	2	45.4	c.67	1	4.73	Homodimer	Cytoplasmic (Potential)	Cytoplasmic	H
yade_ecoli	P31666	Hypothetical protein yadE precursor.	2	46.3	-	0	9.81			Unknown	G
avta_ecoli	P09053	Valine--pyruvate aminotransferase (EC 2.6.1.66) (Transaminase C) (Alanine--valine transaminase).	2	46.7	c.67	0	5.65	Homodimer (By similarity)	Cytoplasmic (By similarity).	Cytoplasmic	E
rhlb_ecoli	P24229	ATP-dependent RNA helicase rhlB (EC 3.6.1.-).	2	47.0	c.37	0	7.28	Component of the degradosome complex. Binds to RNase E and PNPase. Forms multimers.		Unknown	L
rho_ecoli	P03002	Transcription termination factor rho.	2	47.0	c.37; a.140; b.40	0	6.75	Homohexamer		Cytoplasmic	K
dhna_ecoli	P00393	NADH dehydrogenase (EC 1.6.99.3).	2	47.2	c.4; c.3	0	8.96		Membrane	Unknown	C
cisy_ecoli	P00891	Citrate synthase (EC 2.3.3.1).	2	48.0	a.103	0	6.21	Homohexamer		Cytoplasmic	C
ygaf_ecoli	P37339	Hypothetical protein ygaF.	2	48.6	c.3		9.16			Unknown	R
paak_ecoli	P76085	Phenylacetate-coenzyme A ligase (EC 6.2.1.30) (Phenylacetyl-CoA ligase) (PA-CoA ligase).	2	49.0	e.23	0	6.18			Cytoplasmic	H

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
nuof_ecoli	P31979	NADH-quinone oxidoreductase chain F (EC 1.6.99.5) (NADH dehydrogenase I, chain F) (NDH-1, chain F) (NUO6).	2	49.3	-	0	6.44	Composed of 13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		Cytoplasmic	C
accc_ecoli	P24182	Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA carboxylase) (EC 6.4.1.2) (ACC).	2	49.3	d.142; c.1; b.84; c.30	0	6.65	Acetyl-CoA carboxylase is a heterohexamer of biotin carboxyl carrier protein, biotin carboxylase and the two subunits of carboxyl transferase in a 2:2 complex.		Cytoplasmic	I
yegd_ecoli	P36928	Hypothetical chaperone protein yegD.	2	49.4	c.55	0	5.11			Unknown	O
ycaj_ecoli	P45526	Hypothetical protein ycaJ.	2	49.6	c.37	0	6			Cytoplasmic	L
stha_ecoli	P27306	Soluble pyridine nucleotide transhydrogenase (EC 1.6.1.1) (STH) (NAD(P)(+) transhydrogenase [B-specific]).	2	51.4	d.87; c.3; c.4; c.2	0	6.09	Homooligomer; probable homooctamer	Cytoplasmic	Cytoplasmic	C
dcea_ecoli	P80063	Glutamate decarboxylase alpha (EC 4.1.1.15) (GAD-alpha).	2	52.7	c.67	0	5.22	Homohexamer			E
tnaa_ecoli	P00913	Tryptophanase (EC 4.1.99.1) (L-tryptophan indole-lyase) (TNase).	2	52.8	c.67	0	5.88	Homotetramer		Cytoplasmic	E
ydcr_ecoli	P77730	Hypothetical protein ydcR.	2	52.8	a.4; c.67	0	8.85			Cytoplasmic	K

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
pcnb_ecoli	P13685	Poly(A) polymerase (EC 2.7.7.19) (PAP) (Plasmid copy number protein).	2	54.7	-	0	9.68	Monomer. Interacts with csdA, rhlE and srmB.		Cytoplasmic	J
mgla_ecoli	P23199	Galactoside transport ATP-binding protein mglA.	2	56.4	c.37	0	7.21		Inner membrane-associated (Potential)	CytoplasmicMembrane	G
typa_ecoli	P32132	GTP-binding protein typA/BipA (Tyrosine phosphorylated protein A).	2	65.4	c.37; d.58; b.43	0	5.1			Unknown (This protein may have multiple localization sites.)	N
nucd_ecoli	P33599	NADH-quinone oxidoreductase chain C/D (EC 1.6.99.5) (NADH dehydrogenase I, chain C/D) (NDH-1, chain C/D) (NUO3/NUO4).	2	68.7	e.18	0	5.98	Composed of 13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		Cytoplasmic	C
dnak_ecoli	P04475	Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).	2	69.0	e.20; c.55	1	4.83			Cytoplasmic	O
gida_ecoli	P17112	Glucose inhibited division protein A.	2	69.5	c.2; c.4; c.3	0	6.2			Cytoplasmic	D
ydcp_ecoli	P76104	Putative protease ydcP precursor (EC 3.4.-.-).	2	72.7	-	0	6.69			Unknown	O
syt_ecoli	P00955	Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--tRNA ligase) (ThrRS).	2	74.0	d.66; c.51; d.67; d.104	1	5.8	Homodimer	Cytoplasmic	Cytoplasmic	J

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
spot_ecoli	P17580	Guanosine-3',5'-bis(Diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2) ((ppGpp)ase) (Penta-phosphate guanosine-3'-pyrophosphohydrolase).	2	79.3	d.66		8.89			Cytoplasmic	T
lon_ecoli	P08177	ATP-dependent protease La (EC 3.4.21.53).	2	87.4	c.37	0	6.01	Homotetramer	Cytoplasmic	Cytoplasmic	O
rnr_ecoli	P21499	Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein).	2	92.1	b.40	0	8.78	Monomer		Cytoplasmic	K
adhe_ecoli	P17547	Aldehyde-alcohol dehydrogenase [Includes: Alcohol dehydrogenase (EC 1.1.1.1) (ADH) Acetaldehyde dehydrogenase [acetylating] (EC 1.2.1.10) (ACDH) Pyruvate-formate-lyase deactivase (PFL deactivase)].	2	96.0	c.82; e.22	0	6.33	Seems to form a rod shaped polymer composed of about 40 identical subunits.		Cytoplasmic	C
gyra_ecoli	P09097	DNA gyrase subunit A (EC 5.99.1.3).	2	97.0	e.11	1	5.09	Made up of two chains. The A chain is responsible for DNA breakage and rejoining; the B chain catalyzes ATP hydrolysis. The enzyme forms an A2B2 tetramer.		Cytoplasmic	L
if2_ecoli	P02995	Translation initiation factor IF-2.	2	97.4	c.37; a.114; c.20; b.43	1	5.8		Cytoplasmic	Cytoplasmic	J
glnd_ecoli	P27249	[Protein-PII] uridylyltransferase (EC 2.7.7.59) (PII uridylyl-transferase) (Uridylyl removing enzyme) (UTase).	2	102.4	-	1	6.22			CytoplasmicMembrane	O

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
odo1_ecoli	P07015	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha- ketoglutarate dehydrogenase).	2	105.1	c.36	1	6.04	Homodimer		Cytoplasmic	C
pur4_ecoli	P15254	Phosphoribosylformylglycinamidine synthase (EC 6.3.5.3) (FGAM synthase) (FGAMS) (Formylglycinamide ribotide amidotransferase) (FGARAT) (Formylglycinamide ribotide synthetase).	2	141.4	c.23	0	5.23	Monomer	Cytoplasmic	Unknown	F
hrpa_ecoli	P43329	ATP-dependent helicase hrpA.	2	149.0	c.37	0	7.89			CytoplasmicM embrane	L
rpob_ecoli	P00575	DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase beta chain) (RNA polymerase beta subunit).	2	150.6	e.29	1	5.15	The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta' and 1 omega subunit. When a sigma factor is associated with the core the holoenzyme is formed, which can initiate transcription.		Cytoplasmic	K

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
rpoc_ecoli	P00577	DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase beta' chain) (RNA polymerase beta' subunit).	2	155.2	e.29	1	6.67	The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta' and 1 omega subunit. When a sigma factor is associated with the core the holoenzyme is formed, which can initiate transcription (By similarity).		Cytoplasmic	K
ybak_ecoli	P37175	Protein ybaK.	3	17.1	d.116	0	9.02			Cytoplasmic	S
hlpa_ecoli	P11457	Histone-like protein HLP-1 precursor (DNA-binding 17 kDa protein).	3	17.7	-	0	9.52	Homotetramer	Either in the nucleoid (chromatin) or in the outer membrane	OuterMembrane	M
ssrp_ecoli	P32052	SsrA-binding protein (Small protein B).	3	18.1	b.111	0	9.9		Cytoplasmic (Potential)	Cytoplasmic	O
rsd_ecoli	P31690	Regulator of sigma D.	3	18.2	-	1	5.65			Unknown	K
ubic_ecoli	P26602	Chorismate--pyruvate lyase (EC 4.-.-.-).	3	18.6	d.190	0	7.73	Monomer	Cytoplasmic	Cytoplasmic	H
yqab_ecoli	P77475	Hypothetical protein yqaB.	3	20.8	c.108	0	5.51			Cytoplasmic	R
ycfp_ecoli	P75950	Hypothetical protein ycfP.	3	21.2	-	1	6.13			Cytoplasmic	R
rfbc_ecoli	P37745	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) (dTDP-4-keto-6- deoxyglucose 3,5-epimerase) (dTDP-L-rhamnose synthetase).	3	21.3	b.82	0	5.48	Homodimer (By similarity)		Cytoplasmic	M

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
rimj_ecoli	P09454	Ribosomal-protein-alanine acetyltransferase (EC 2.3.1.128) (Acetylating enzyme for N-terminal of ribosomal protein S5).	3	22.7	d.108	0	9.2		Cytoplasmic	Cytoplasmic	J
yajb_ecoli	P21515	Hypothetical protein yajB.	3	23.0	-	0	5.88			Cytoplasmic	S
yqji_ecoli	Q46872	Hypothetical protein yqjI.	3	23.4	-	0	6.27			Cytoplasmic	K
crp_ecoli	P03020	Catabolite gene activator (cAMP receptor protein) (cAMP-regulatory protein).	3	23.6	b.82; a.4		8.38	Binds DNA as a dimer		Cytoplasmic	T
ftse_ecoli	P10115	Cell division ATP-binding protein ftsE.	3	24.4	c.37	1	9.37			CytoplasmicMembrane	D
gch1_ecoli	P27511	GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).	3	24.7	d.96	1	6.93	Homodecamer, composed of a dimer of pentamers.		Unknown	H
trmb_ecoli	P32049	tRNA (guanine-N(7)-)-methyltransferase (EC 2.1.1.33) (tRNA(m7G46)- methyltransferase).	3	27.3	c.66	0	6.42	Monomer		Cytoplasmic	J
fucr_ecoli	P11554	L-fucose operon activator.	3	27.4	a.4; c.35	0	7.75			Cytoplasmic	K
pfla_ecoli	P09374	Pyruvate formate-lyase 1 activating enzyme (EC 1.97.1.4) (PFL- activating enzyme).	3	28.1	-	0	6		Cytoplasmic	Cytoplasmic	O
trmd_ecoli	P07020	tRNA (Guanine-N(1)-)-methyltransferase (EC 2.1.1.31) (M1G- methyltransferase) (tRNA [GM37] methyltransferase).	3	28.4	-	1	5.5	Monomer	Cytoplasmic (Potential)	Cytoplasmic	J
glcc_ecoli	P52072	Glc operon transcriptional activator.	3	28.8	a.4	0	9.16			Cytoplasmic	K
suhb_ecoli	P22783	Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1- phosphatase) (I-1-Pase).	3	29.2	e.7	1	6.45	Monomer		Cytoplasmic	G
ycfh_ecoli	P37346	Putative deoxyribonuclease ycfH (EC 3.1.21.-).	3	29.8	c.1	0	5.19			Cytoplasmic	L
yafd_ecoli	P30865	Hypothetical protein yafD.	3	30.0	d.151	0	9.62		Cytoplasmic (Potential)	Unknown	S

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
gaty_ecoli	P37192	Tagatose-1,6-bisphosphate aldolase gatY (EC 4.1.2.-) (TBPA).	3	30.8	c.1	0	5.87			Cytoplasmic	G
dapa_ecoli	P05640	Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDPS).	3	31.3	c.1	1	5.98	Homotetramer	Cytoplasmic	Unknown	E
amia_ecoli	P36548	Probable N-acetylmuramoyl-L-alanine amidase amiA precursor (EC 3.5.1.28).	3	31.4	-	0	9.9			Unknown	M
end4_ecoli	P12638	Endonuclease IV (EC 3.1.21.2) (Endodeoxyribonuclease IV).	3	31.5	c.1	0	5.43	Monomer		Cytoplasmic	L
ypt1_ecoli	P29368	Hypothetical 31.7 kDa protein in TRAX-FINO intergenic region.	3	31.8	c.69	0	5.49			Unknown	R
yneb_ecoli	P76143	Putative aldolase yneB (EC 4.2.1.-).	3	31.9	c.1	0	6.09			Unknown	G
nana_ecoli	P06995	N-acetylneuraminate lyase (EC 4.1.3.3) (N-acetylneuraminic acid aldolase) (N-acetylneuraminate pyruvate-lyase) (Sialic acid lyase) (Sialate lyase) (Sialic acid aldolase) (NALase).	3	32.5	c.1	0	5.62	Homotetramer	Cytoplasmic	Unknown	E
yhbj_ecoli	P33995	Hypothetical UPF0042 protein yhbJ.	3	32.5	c.37	0	6.72			Cytoplasmic	R
metf_ecoli	P00394	5,10-methylenetetrahydrofolate reductase (EC 1.7.99.5).	3	33.1	c.1	0	6	Homotetramer		Cytoplasmic	E
arac_ecoli	P03021	Arabinose operon regulatory protein.	3	33.4	b.82; a.4	0	6.46	Homodimer	Cytoplasmic	Cytoplasmic	K
icia_ecoli	P24194	Chromosome initiation inhibitor (OriC replication inhibitor).	3	33.5	c.94; a.4	0	6.4	Behaves as an homodimer in solution.		Cytoplasmic	K
dusc_ecoli	P33371	tRNA-dihydrouridine synthase C (EC 1.-.-.-).	3	35.2	c.1	0	6.12			Cytoplasmic	J
hem2_ecoli	P15002	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen synthase) (ALAD) (ALADH).	3	35.5	c.1	1	5.25	Homooctamer		Cytoplasmic	H

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
dusb_ecoli	P25717	tRNA-dihydrouridine synthase B (EC 1.-.-.).	3	35.9	c.1	0	6.27			Cytoplasmic	J
rluc_ecoli	P23851	Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70) (Pseudouridylate synthase) (Uracil hydrolyase).	3	36.0	d.66; d.58	0	9.85			Cytoplasmic	J
lipa_ecoli	P25845	Lipoic acid synthetase (Lip-syn) (Lipoate synthase).	3	36.1	-	0	8.08	Monomer or homodimer	Cytoplasmic	Cytoplasmic	H
add_ecoli	P22333	Adenosine deaminase (EC 3.5.4.4) (Adenosine aminohydrolase).	3	36.4	c.1	0	5.36			Cytoplasmic	F
yajo_ecoli	P77735	Hypothetical oxidoreductase yajO (EC 1.-.-.).	3	36.4	c.1	0	5.19			Cytoplasmic	C
ltae_ecoli	P75823	Low-specificity L-threonine aldolase (EC 4.1.2.5) (Low-specificity L- TA).	3	36.5	c.67	0	5.81	Homotetramer (Probable)		Unknown	E
nagz_ecoli	P75949	Beta-hexosaminidase (EC 3.2.1.52) (N-acetyl-beta-glucosaminidase) (Beta-N-acetylhexosaminidase).	3	37.6	c.1	0	5.86	Monomer (Potential)	Cytoplasmic	Cytoplasmic	G
yfif_ecoli	P33635	Hypothetical tRNA/rRNA methyltransferase yfiF (EC 2.1.1.-).	3	37.8	c.116	0	8.94			Unknown	J
alf1_ecoli	P71295	Fructose-bisphosphate aldolase class I (EC 4.1.2.13) (FBP aldolase).	3	38.0	-	0	6.24	Homooctamer or homodecamer	Cytoplasmic (Probable)	Unknown	G
ybjs_ecoli	P75821	Hypothetical protein ybjS.	3	38.1	c.2	1	8.81			Unknown	M
alr2_ecoli	P29012	Alanine racemase, catabolic (EC 5.1.1.1).	3	38.8	c.1; b.49	0	6.56			Cytoplasmic	M
yjiu_ecoli	P39407	Hypothetical protein yjiU.	3	39.8	-	0	8.71			Unknown	R
dhas_ecoli	P00353	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA dehydrogenase) (ASADH).	3	40.0	c.2; d.81	1	5.37	Homodimer		Unknown	E
his7_ecoli	P06987	Histidine biosynthesis bifunctional protein hisB [Includes: Histidinol-phosphatase (EC 3.1.3.15) Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) (IGPD)].	3	40.3	c.108	0	5.76		Cytoplasmic	Cytoplasmic	E

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
phol_ecoli	P77349	PhoH-like protein.	3	40.7	-	0	6.24		Cytoplasmic (Potential)	Cytoplasmic	T
thik_ecoli	P21151	3-ketoacyl-CoA thiolase (EC 2.3.1.16) (Fatty oxidation complex beta subunit) (Beta-ketothiolase) (Acetyl-CoA acyltransferase).	3	40.9	c.95	0	6.31	Tetramer of two alpha chains and two beta chains	Cytoplasmic	Cytoplasmic	I
dnaj_ecoli	P08622	Chaperone protein dnaJ (Heat shock protein J) (HSP40).	3	41.0	a.2; b.4; a.4; a.138; g.54	0	8.03	Homodimer	Cytoplasmic	Cytoplasmic	O
biof_ecoli	P12998	8-amino-7-oxononanoate synthase (EC 2.3.1.47) (AONS) (8-amino-7- ketopelargonate synthase) (7-keto-8-amino-pelargonic acid synthetase) (7-KAP synthetase) (L-alanine--pimelyl CoA ligase).	3	41.6	c.67	0	6.64	Homodimer		Unknown	H
metk_ecoli	P04384	S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine adenosyltransferase) (AdoMet synthetase) (MAT).	3	41.8	d.130	1	5.1	Homotetramer	Cytoplasmic	Cytoplasmic	H
trma_ecoli	P23003	tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)- methyltransferase) (RUMT).	3	42.0		0	5.71			Unknown	J
rsmd_ecoli	P42596	Putative ribosomal RNA small subunit methyltransferase D (EC 2.1.1.52) (rRNA (guanine-N(2)-)-methyltransferase) (16S rRNA m2G966 methyltransferase).	3	42.3	c.66	0	6.32			Cytoplasmic	J
arge_ecoli	P23908	Acetylornithine deacetylase (EC 3.5.1.16) (Acetylornithinase) (AO) (N-acetylornithinase) (NAO).	3	42.3	d.58; c.56	0	5.54	Homodimer	Cytoplasmic (Probable)	Cytoplasmic	E
lldd_ecoli	P33232	L-lactate dehydrogenase (Cytochrome) (EC 1.1.2.3).	3	42.7	c.1	0	6.33			Unknown	C
fabf_ecoli	P39435	3-oxoacyl-[acyl-carrier-protein] synthase II (EC 2.3.1.41) (Beta- ketoacyl-ACP synthase II) (KAS II).	3	42.9	c.95	0	5.71	Homodimer		Cytoplasmic	I

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
phea_ecoli	P07022	P-protein [Includes: Chorismate mutase (EC 5.4.99.5) (CM) Prephenate dehydratase (EC 4.2.1.51) (PDT)].	3	43.1	a.130; d.58	0	6.21		Cytoplasmic	Cytoplasmic	E
thih_ecoli	P30140	Thiazole biosynthesis protein thiH.	3	43.2	-		6.61			Unknown	H
ints_ecoli	P37326	Putative prophage CPS-53 integrase.	3	44.1	d.163		9.4			Cytoplasmic	L
csdb_ecoli	P77444	Selenocysteine lyase (EC 4.4.1.16) (Selenocysteine reductase) (Selenocysteine beta-lyase) (SCL).	3	44.4	c.67	0	5.89	Homodimer		Unknown	E
yfbq_ecoli	P77727	Probable aminotransferase yfbQ (EC 2.6.1.-).	3	45.5	c.67	0	5.85	Homodimer (By similarity)	Cytoplasmic (By similarity)	Cytoplasmic	E
rspa_ecoli	P38104	Starvation sensing protein rspA.	3	46.0	d.54; c.1	0	5.68			Cytoplasmic	H
gatz_ecoli	P37191	Putative tagatose 6-phosphate kinase gatZ (EC 2.7.1.144).	3	47.1	-	0	5.5			Cytoplasmic	G
typh_ecoli	P07650	Thymidine phosphorylase (EC 2.4.2.4) (TdRPase).	3	47.2	d.41; a.46; c.27	0	5.21	Homodimer		Unknown	F
dada_ecoli	P29011	D-amino acid dehydrogenase small subunit (EC 1.4.99.1).	3	47.6	c.5; d.16; c.3; c.4; c.2	1	6.17	Heterodimer of a small and a large subunit	Inner membrane-bound	Cytoplasmic	E
pmba_ecoli	P24231	PmbA protein (TldE protein).	3	48.4	-	0	5.4		Cytoplasmic	Cytoplasmic	O
eutb_ecoli	P19635	Ethanolamine ammonia-lyase heavy chain (EC 4.3.1.7) (Ethanolamine ammonia-lyase large subunit).	3	49.4	-	0	4.79	Heterodimer of two nonidentical chains		Unknown	C
xyla_ecoli	P00944	Xylose isomerase (EC 5.3.1.5) (D-xylulose keto-isomerase).	3	49.7	c.1	0	5.75	Homotetramer	Cytoplasmic	Cytoplasmic	G
rhle_ecoli	P25888	Putative ATP-dependent RNA helicase rhlE.	3	50.0	c.37	0	10.06	Interacts with pcnB.		Cytoplasmic	L

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
pepq_ecoli	P21165	Xaa-Pro dipeptidase (EC 3.4.13.9) (X-Pro dipeptidase) (Proline dipeptidase) (Prolidase) (Imidodipeptidase).	3	50.2	d.127	0	5.6			Cytoplasmic	E
tldd_ecoli	P46473	TldD protein.	3	51.4	-	0	4.93			Cytoplasmic	O
uxac_ecoli	P42607	Uronate isomerase (EC 5.3.1.12) (Glucuronate isomerase) (Uronic isomerase).	3	54.0	c.1		5.44			Cytoplasmic	G
ampa_ecoli	P11648	Cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP) (Leucyl aminopeptidase) (Aminopeptidase A/I).	3	54.9	c.50; c.56	0	6.82	Homohexamer		Cytoplasmic	E
araa_ecoli	P08202	L-arabinose isomerase (EC 5.3.1.4).	3	56.1	-	0	5.95			Unknown	G
aldb_ecoli	P37685	Aldehyde dehydrogenase B (EC 1.2.1.22) (Lactaldehyde dehydrogenase).	3	56.3	c.82		5.44			Cytoplasmic	C
dhsa_ecoli	P10444	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1).	3	64.4	a.7; c.3; d.168	0	5.85	Part of an enzyme complex containing four subunits: a flavoprotein, an iron-sulfur, cytochrome b-556, and an hydrophobic anchor protein.		Unknown (This protein may have multiple localization sites.)	C
frda_ecoli	P00363	Fumarate reductase flavoprotein subunit (EC 1.3.99.1).	3	65.8	a.7; c.3; d.168	0	5.87	Fumarate dehydrogenase forms part of an enzyme complex containing four subunits: a flavoprotein, an iron-sulfur, and two hydrophobic anchor proteins.		Periplasmic	C

Supplementary Table S2: GroEL substrates.

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essentiality	pI	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	Subcellular Localization (PSORTb v2.0)	COG Functional Category
dead_ecoli	P23304	Cold-shock DEAD-box protein A (ATP-dependent RNA helicase deaD).	3	70.4	c.37	0	8.76		Cytoplasmic (Probable)	Cytoplasmic	L
ycby_ecoli	P75864	Hypothetical protein ycbY.	3	78.9	c.66	0	8.96			Unknown	L
parc_ecoli	P20082	Topoisomerase IV subunit A (EC 5.99.1.-).	3	83.8	e.11	1	6.24	Composed of two subunits: parC and parE.	Membrane-associated	Cytoplasmic	L
phsm_ecoli	P00490	Maltodextrin phosphorylase (EC 2.4.1.1).	3	90.3	c.87	0	7.11	Homodimer		Unknown	G

Supplementary Table S2: GroEL substrates.

Supplementary Table S3: Functional categorization according to the COG database (Clusters of Orthologous Groups of proteins, http://www.ncbi.nlm.nih.gov/COG/) (Tatusov *et al.*, 1997).

Information storage and processing	**J**	Translation, ribosomal structure and biogenesis
	K	Transcription
	L	DNA replication, recombination and repair
Cellular processes	**D**	Cell division and chromosome partitioning
	M	Cell envelope biogenesis, outer membrane
	N	Cell motility and secretion
	O	Posttranslational modification, protein turnover, chaperones
	P	Inorganic ion transport and metabolism
	T	Signal transduction mechanisms
Metabolism	**C**	Energy production and conversion
	E	Amino acid transport and metabolism
	F	Nucleotide transport and metabolism
	G	Carbohydrate transport and metabolism
	H	Coenzyme metabolism
	I	Lipid metabolism
	Q	Secondary metabolites biosynthesis, transport and catabolism
Poorly characterized	**R**	General function prediction only
	S	Function unknown

Supplementary Table S3

8.2 Abbreviations

Units are expressed according to the international system of units (SI), including outside units accepted for use with the SI.
Amino acids are abbreviated with their one or three letter symbols.
Protein names are abbreviated according to their SWISSPROT database entries.

ADD	adenosine deaminase
ADP	adenosine 5'-diphosphate
ALR2	alanine racemase, catabolic
Amp	ampicillin
APS	ammonium peroxodisulfate
ASA	L-aspartate-β-semialdehyde
ATP	adenosine 5'-triphosphate
BIA(core)	Biomolecular Interaction Analysis
BLAST	Basic Local Alignment Search Tool
β-NADH	β-nicotinamide adenine dinucleotide
BSA	albumin bovine serum
Cam	chloramphenicol
CATH	database of protein structures (Class, Architecture, Topology, Homologous superfamily)
CDTA	*trans*-1,2-diaminocyclohexane-*N,N,N′,N′*-tetraacetic acid
ClpB	chaperone clpB
COG	Clusters of Orthologous Groups of proteins
CRP	catabolite gene activator
DAPA	dihydrodipicolinate synthase (DHDPS)
DCEA	glutamate decarboxylase alpha
DNA	deoxyribonucleic acid
DnaJ	bacterial Hsp40 chaperone
DnaK	bacterial Hsp70 chaperone
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
emPAI	exponentially modified (→) PAI
END4	endonuclease IV
ENO	enolase (2-phosphoglycerate dehydratase)
FAD	flavine adenine dinucleotide
FPLC	Fast Protein Liquid Chromatography
g	acceleration of gravity, 9.81 m/s^2
G3P1	glyceraldehyde 3-phosphate dehydrogenase (GAPDH-A)
GATD	galactitol-1-phosphate 5-dehydrogenase
GATY	tagatose-1,6-bisphosphate aldolase gatY (TBPA)
GdnHCl	guanidinium hydrochloride

Gene3D	supplementary to the CATH database, contains links to CATH for proteins in complete proteomes
GRAVY	GRand AVerage of HydropathY
GroEL	bacterial Hsp60 chaperonin
GroES	bacterial Hsp10 cochaperonin
GrpE	bacterial nucleotide exchange factor of DnaK
GTP	guanosin 5'-triphosphate
HEM2	delta-aminolevulinic acid dehydratase
HFBA	heptafluorobutyric acid
hr	hour
IMAC	Immobilized Metal Affinity Chromatography
IMPALA	Integrating Matrix Profiles And Local Alignments
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	Luria Bertani
LCMS	coupled Liquid Chromatography - Mass Spectrometry
Leu-D3	L-leucine-5,5,5-D3
LLDD	L-lactate dehydrogenase
LTAE	low-specificity L-threonine aldolase
M. mazei	*Methanosarcina mazei*
MES	2-morpholinoethanesulfonic acid
METE	5-methyltetrahydropteroyltriglutamate-homocysteine methyl-transferase
METF	5,10-methylenetetrahydrofolate reductase
METH	methionine synthase
METJ	Met repressor
METK	S-adenosylmethionine synthetase
*Mm*GroEL	($\rightarrow$) *M. mazei* ($\rightarrow$) GroEL
*Mm*GroES	($\rightarrow$) *M. mazei* ($\rightarrow$) GroES
MOPS	3-(N-morpholino)propanesulfonic acid
MS	mass spectrometry
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate
NANA	N-acetylneuraminate lyase
NTA	nitrilo-triacetic acid
OAc	acetate
OD	optical density
PAGE	PolyAcrylamide Gel Electrophoresis
PAI	Protein Abundance Index
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/
PEDANT	Protein Extraction, Description and ANalysis Tool
Pefabloc	4-(2-aminoethyl)benzenesulfonylfluoride HCl
PEI	polyethyleneimine

PSI-BLAST	Position Specific Iterative (→) BLAST
S. cerevisiae	*Saccharomyces cerevisiae*
SCOP	database of protein structures (Structural Classification of Proteins)
SDS	sodiumdodecylsulfate
SILAC	Stable Isotope Labeling with Amino acids in Cell culture
Sulfo-GMBS	*N*-(4-maleimidobutyryloxy)sulfosuccinimide
SYT	threonyl-tRNA synthetase (ThrRS)
TCA	trichloroacetic acid
TDH	L-threonine 3-dehydrogenase
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TF	trigger factor
TFA	trifluoroacetic acid
TLC	Thin Layer Chromatography
TOF	Time Of Flight. Mass spectrometry ion detector.
Tris HCl	tris(hydroxymethyl)aminomethane hydrochloride
TYPH	thymidine phosphorylase
U. urealyticum	*Ureaplasma urealyticum*
XYLA	xylose isomerase
YAJO	hypothetical oxidoreductase yajO
YHBJ	hypothetical UPF0042 protein yhbJ

8.3   Publications

8.3.1   Journal articles

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.

Figueiredo, L., Klunker, D., Ang, D., Naylor, D.J., **Kerner, M.J.**, Georgopoulos, C., Hartl, F.U. and Hayer-Hartl, M. (2004). Functional characterization of an archaeal GroEL/GroES chaperonin system. ***J. Biol. Chem.***, 279, 1090-1099.

Shewmaker, F., **Kerner, M.J.**, Hayer-Hartl, M., Klein, G., Georgopoulos, C. and Landry, S.J. (2004). A mobile loop order-disorder transition modulates the speed of chaperonin cycling. ***Protein Sci.***, 13, 2139-2148.

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. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. ***Cell, 2005, in press***.

8.3.2   Oral presentations

"The *E. coli* GroEL interaction proteome: identification and classification." Structural Genomics & Proteomics, European Union Projects Meeting, December 2004, Barcelona, Spain.

"The chaperonin interaction proteome of *E. coli*: understanding the essential role of GroEL-assisted protein folding." General meeting of the SFB 594 (German Federal Funding Organization), December 2004, Munich, Germany.

8.3.3   Posters

Backes, A.C., **Kerner, M.J.**, Siehl H.U. Experimental and quantum chemical NMR investigation of reactive intermediates - new silyl substituted cyclopropylmethyl carbocations. 19. Nuclear Magnetic Resonance Section Meeting of GDCh (German Chemical Society), October 1997, Friedrichroda, Germany.

Hayer-Hartl, M., Brinker, A., Pfeifer, G., **Kerner, M.J.**, Naylor, D.J. and Hartl, F.U. Dual function of protein confinement in chaperonin-assisted protein folding. Gordon Research Conference - Protein Folding Dynamics, January 2002, Ventura, CA, USA.

Naylor, D.J., **Kerner, M.J.**, Stines, A.P., Maier, T., Ishihama, Y., Mann, M., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., and Hartl, F.U. Obligate substrates of the GroEL/GroES chaperonin system. FASEB (Federation of American Societies for Experimental Biology) Summer Research Conferences – Protein Folding in the Cell, July 2002, Saxtons River, Vermont, USA.

Kerner, M.J., Maier, T., Tang, C., Hirtreiter, A., Haas, B., Figueiredo, L., Klunker, D., Naylor, D.J., Stines, A.P., Hayer-Hartl, M. and Hartl, F.U. Mechanism of chaperonin assisted protein folding, Symposium of the SFB 594 (German Federal Funding Organization), November 2003, Munich, Germany.

Kerner, M.J., Maier, T., Naylor, D.J., Chang, H.-C., Stines, A.P., Georgopoulos, C., Hayer-Hartl, M., Frishman, D., Mann, M., Ishihama, Y. and Hartl, F.U. The *E. coli* GroEL interaction proteome: identification and classification. Structural Genomics & Proteomics, European Union Projects Meeting, December 2004, Barcelona, Spain.

8.4   Curriculum vitae

Michael Johannes Kerner

Geboren am 13. April 1974 in München

Wissenschaftlicher Werdegang

Dissertation

Dez 1999 – heute	Experimentelle Doktorarbeit in der Abteilung Zelluläre Biochemie von Prof. Dr. F. Ulrich Hartl am Max-Planck-Institut für Biochemie, Martinsried: *The Escherichia coli GroEL interaction proteome: identification and classification.*

Hochschulstudium

Sept. 1999	Diplom (Univ.) in Chemie
März - Sept. 1999	Diplomarbeit in der Sektion Polymere von Prof. Dr. Hartmut Seliger an der Universität Ulm: *Subklonierung und Expression von Mutanten des spannungsabhängigen Na$^+$-Ionenkanals der menschlichen Skelettmuskelzelle auf Basis des hybriden MVA/T7pol Expressionssystems.*
Feb. 1999:	Diplomhauptprüfung Fächer: Anorganische Chemie, Organische Chemie, Physikalische Chemie, Makromolekulare Chemie.
Praktika:	Juni-Sept. 1997, Prof. H.-U. Siehl, Universität Ulm: *Synthese von Vorstufen für die Darstellung von 1-Cyclopropyl-2-trialkylsilyl-ethyl Kationen.* März-April 1997, Boehringer Ingelheim Pharma KG, Biberach: *Organische Synthese von pharmakologischen Testkandidaten für die Tumorsuppression.*
Okt. 1996:	Diplomvorprüfung Fächer: Anorganische und Analytische Chemie, Organische Chemie, Physikalische Chemie, Physik.
Okt. 1994 – Sept. 1999	Studium der Chemie an der Universität Ulm.

Hochschulreife

1993:	Allgemeine Hochschulreife, Schubart Gymnasium Ulm.