Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Proteome-Wide Analysis of Chaperonin-Dependent Protein Folding in *Escherichia coli*

Tobias Maier

aus

Heilbronn

2006

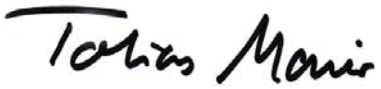
Erklärung

Diese Dissertation wurde im Sinne von § 13 Absatz 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Professor Dr. F. Ulrich Hartl betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig und ohne unerlaubte Hilfe erarbeitet.

München, 26. Oktober 2005



Dissertation eingereicht am 28. Oktober 2005

1. Gutachter: Professor Dr. F. Ulrich Hartl

2. Gutachter: PD Dr. Konstanze Winklhofer

Mündliche Prüfung am 01. Februar 2006

1. Summary 1
2. Introduction 2
2.1. From DNA to protein 2
2.1.1. Synthesis of proteins 2
2.1.2. The protein folding problem 3
2.1.3. Protein folding mechanisms 5
2.1.4. Protein folding *in vitro* and *in vivo* 7
2.1.5. Diseases related to protein folding 8
2.2. Molecular chaperones 9
2.2.1. Ribosome-associated chaperones 10
2.2.2. Hsp70 chaperones 11
2.2.3. The reaction cycle of DnaK, DnaJ and GrpE 13
2.3. Hsp60 and Hsp10: The chaperonins 13
2.3.1. *E. coli* chaperonins GroEL and GroES 14
2.3.2. The structure of GroEL and GroES 15
2.3.3. The mechanism of GroEL and GroES mediated protein folding 18
2.3.4. The substrates of GroEL and GroES 21
2.4. Chaperone networks in *E. coli* 22
2.5. Introduction to proteomics 24
2.5.1. Principles of mass spectrometry 24
2.5.2. Technical possibilities and applications 24
3. Materials and Methods 26
3.1. Growth media and buffers 26
3.1.1. Growth media 26
3.1.2. Buffers and stock solutions 26
3.2. Bacterial strains and plasmids 27
3.2.1. *E. coli* strains 27
3.2.2. Plasmids 27
3.3. DNA analytical methods 28
3.3.1. PCR amplification 28
3.3.2. DNA restriction, ligation and plasmid isolation 29
3.4. Competent cells and transformation 30
3.5. Protein purification 31
3.6. Protein analytical methods 32
3.6.1. Determination of protein concentration 32
3.6.2. SDS - PAGE 32
3.6.3. Silver staining 33
3.6.4. Western Blotting 33
3.6.5. Generation of antibodies 34
3.6.6. Size exclusion chromatography 34

3.7. Protein refolding 35
3.7.1. DAPA refolding 35
3.7.2. DCEA refolding 35
3.7.3. ENO refolding 36
3.7.4. GATD refolding 36
3.7.5. METF refolding 36
3.7.6. METK refolding 37
3.7.7. SYT refolding 37
3.8. *In vivo* co-expressions 37
3.8.1. Co-expressions of chaperones and substrates in *E. coli* 37
3.8.2. Co-expressions of chaperones and substrates in S. cerevisiae 38
3.9. GroEL/GroES depletion 39
3.10. GroEL/GroES-substrate complexes 39
3.10.1. Cell growth 39
3.10.2. Cell lysis and purification of complexes 40
3.10.3. Alternative purification method 40
3.10.4. Proteinase K digestion of GroEL/GroES/substrate complexes 41
3.11. Mass spectrometric methods 41
3.11.1. Sample preparation for protein identification by mass spectrometry 41
3.11.2. Coupled liquid chromatography – mass spectrometry (LC-MS/MS) 42
3.11.3. Analysis of mass spectrometric data 43
3.12. Bioinformatic methods 44
3.12.1. Structural comparison of GroE substrates 44
3.12.2. Protein sequence analyses 45
4. Results 46
4.1. Identification of GroEL substrates 46
4.1.1. Experimental approach 46
4.1.2. Stability of GroEL/GroES complexes 47
4.1.3. Processing of GroEL/GroES/substrate complexes 48
4.1.4. Experimental controls 49
4.1.5. Influence of other chaperone systems on GroEL substrate diversity 51
4.1.6. Quantification of GroEL interacting proteins 51
4.2. Properties of GroEL substrates 53
4.2.1. Mass distribution of proteins associated with GroEL 53
4.2.2. Distinction between cis and *trans* bound polypeptides to GroEL 54
4.2.3. Essentiality of GroEL substrates 56
4.2.4. Functional categories among GroEL interacting proteins 57
4.2.5. Structural categories among GroEL interacting proteins 60
4.2.6. Analyzed fold types 61
4.2.7. The TIM barrel fold 63
4.2.8. Proteins enriched on GroEL 63

4.2.9. Quantification of proteins on GroEL 65
4.3. *In vitro* refolding of GroEL substrates 71
4.3.1. Class I: Chaperone-independent refolding 71
4.3.2. Class II: Chaperone-dependent refolding 73
4.3.3. Class III: GroEL-dependent refolding 76
4.3.4. Substrate selection by GroEL 78
4.4. Chaperone-substrate co-expression 79
4.4.1. GroEL-independent folding 80
4.4.2. GroEL dependent folding 80
4.4.3. Correlation with proteomic and refolding data 82
4.5. GroEL/GroES depletion 82
4.5.1. *E. coli* GroEL depletion strain 83
4.5.2. Proteins not differentially affected by GroEL depletion 84
4.5.3. GroEL-dependent proteins 85
4.5.4. Other effects of GroEL depletion on *E. coli* cells 86
4.5.5. Co-expression of GroEL/GroES and substrates in *S. cerevisiae* 87
5. Discussion 89
5.1. Classes of GroEL substrates 89
5.1.1. Class I proteins 90
5.1.2. Class II proteins 90
5.1.3. Class III proteins 91
5.1.4. GroEL substrates expressed in *S. cerevisiae* 92
5.2. The GroEL interactome 93
5.2.1. Quality of the dataset 93
5.2.2. Methodological constraints 94
5.3. Properties of GroEL interactors 95
5.3.1. Size distribution of proteins associated with GroEL 95
5.3.2. Substrates too large to fit inside the GroEL/GroES cavity 95
5.4. Structures of GroEL substrates 96
5.4.1. The TIM barrel fold 96
5.4.2. Other folds and substrate orthologs in other organisms 98
5.5. Classification of GroEL interactors 99
5.5.1. Extension of the classification to all GroEL interacting proteins 99
5.5.2. Calculations on GroEL transit of substrate proteins 101
5.5.3. Chaperone networks in *E. coli* 103
5.5.4. The essentiality of GroEL, GroES and other chaperone systems 104
5.6. Evolutionary considerations 105
6. References 107
7. Supplementary Material 118

Figure 1: Amino acids connected by peptide bonds 4
Figure 2: The folding funnel 7
Figure 3: Folding states of polypeptide chains 9
Figure 4: DnaK reaction mechanism 12
Figure 5: Structural depiction of GroEL and GroES 15
Figure 6: Structural rearrangements in GroEL upon GroES binding 16
Figure 7: Structural depiction of GroES 18
Figure 8: The GroEL/GroES reaction cycle 19
Figure 9: Model of chaperone folding pathways in *E. coli* 23
Figure 10: Basic set up of MS based experiments 25
Figure 11: Model of GroEL and GroES with bound substrate polypeptide 47
Figure 12: Purification of GroEL/GroES/substrate complexes 50
Figure 13: Quantification of GroEL bound proteins by SILAC 52
Figure 14: Figure: Mass distribution of *E. coli* proteins 54
Figure 15: Proteinase K digests of GroEL and bound substrates 55
Figure 16: Proteinase K digests of GroEL and DnaK 56
Figure 17: Distribution of functional categories among GroEL interacting proteins. 59
Figure 18: Fold types enriched on GroEL 61
Figure 19: Fold types on GroEL and in the *E. coli* cytosol 62
Figure 20: Enriched fold types in the GroEL substrate set 65
Figure 21: Distribution of cellular concentrations of *E. coli* lysate proteins and GroEL substrates 66
Figure 22: Mass distributions of GroEL interacting proteins 68
Figure 23: *In vitro* refolding of enolase (ENO) 72
Figure 24: *In vitro* refolding of DCEA and GATD I 73
Figure 25: *In vitro* refolding of DCEA and GATD II 74

Figure 26: *In vitro* refolding of SYT 75

Figure 27: *In vitro* refolding of METF, METK and DAPA 76

Figure 28: Competition of class I and class II proteins with class III proteins... 78

Figure 29: Solubility of GroEL substrates upon co-expression in *E. coli* with chaperonins 80

Figure 30: Solubility of GroEL substrates upon co-expression in *E. coli* with chaperonins 81

Figure 31: Experimental setup for GroEL/GroES depletion experiments 83

Figure 32: Solubility of GroEL substrates upon depletion of GroEL and GroES I 84

Figure 33: Solubility of GroEL substrates upon depletion of GroEL and GroES II 85

Figure 34: Coexpressions of *E. coli* GroEL, GroES and substrate proteins in yeast 88

Figure 35: Exemplary structures of GroEL interacting proteins. 97

Figure 36: Classification of all GroEL substrates 101

Figure 37: Essentiality of proteins by classes 105

Table 1: Typical PCR reaction 29

Table 2: Typical PCR cycling conditions 29

Table 3: Sample buffer preparation for SDS PAGE 32

Table 4: Gel preparation for SDS PAGE 33

Table 5: Essential GroEL substrate proteins with enrichment factors >100 ... 57

Table 6: GroEL interacting proteins analysed individually in this study 70

Table 7: GroEL interacting proteins sorted into substrate classes 100

1. Summary

In *Escherichia coli*, the cylindrical chaperonin GroEL and its cofactor GroES promote the folding of a fraction of newly synthesized polypeptide chains by acting as an Anfinsen cage. GroEL recognizes substrate proteins with its apical domains of the tetradecameric structure. Exposed hydrophobic side chains in non-native proteins interact with GroEL and bound substrates are subsequently encapsulated under the GroES lid, where they can fold in a protected environment. Despite the detailed knowledge about structural and mechanistic features of GroEL and GroES, little is known about its genuine *in vivo* substrate proteins.

Here, the nearly complete set of GroEL interacting proteins *in vivo* was identified and quantified by an approach using affinity chromatography for the isolation of GroEL/GroES/substrate complexes and subsequent analysis by mass spectrometric methods. GroEL substrate proteins were analyzed with respect to their fold types and functional classes, revealing a preference for proteins which fold into the versatile TIM barrel fold to interact with GroEL.

Further *in vivo* and *in vitro* experiments with individual proteins identified as GroEL substrates verified the data obtained by the proteomic approach and allowed conclusions on the usage of the other main chaperone system in E. coli: DnaK/DnaJ/GrpE. Taken together, the results culminated in the classification of GroEL interacting proteins according to their dependence on chaperones for folding.

Class I proteins are largely independent of chaperones but their folding yield can be increased by chaperone interaction. Class II proteins do not refold efficiently in the absence of chaperones *in vitro*, but can utilize either the DnaK or the GroEL/GroES systems for folding. Class III substrates are fully dependent on GroEL. DnaK can bind class III proteins and thus prevent their aggregation, but folding is achieved only upon transfer to GroEL.

2. Introduction

Proteins are very diverse biomolecules. Their building blocks are 20 different amino acids with different physicochemical properties. The amino acids are linked to each other as chains of various lengths by covalent bonds, the so called peptide bonds. Proteins constitute the majority of the dry mass of a cell and are involved in many diverse cellular functions. Proteins provide structural building blocks, catalyze anabolic and catabolic reactions of the metabolism and as integral parts of membranes they form channels allowing selective transport of substances. They are responsible for infections, as well as for the immune response. Proteins are involved in the synthesis of ATP, the cellular energy currency. They are central components of DNA replication, DNA damage repair, regulated gene expression and stress response as well as in their own degradation. Even the ribosome, the molecular machine in the cell where proteins are synthesized from an mRNA template, is to a good part composed of proteins itself, although the catalytic reactions that form peptide bond are performed by RNA.

2.1. From DNA to protein

2.1.1. Synthesis of proteins

Segments of DNA, the genes which code for individual proteins, are transcribed to mRNA molecules and subsequently translated into amino acid sequences on the ribosome. These chains of covalently bonded amino acids (Figure 1) exit the ribosome in a sequential manner and generally adopt unique three-dimensional structures during and upon release from the ribosome. The acquisition of a three-dimensional structure from a linear sequence of amino acids is called protein folding.

Amino acids have a carboxyl group, an amino group, an H atom and variable side chains attached to their central Cα atom (Figure 1). The side chains determine the physicochemical properties of individual amino acid residues.

Twenty different amino acids are commonly used in naturally occurring proteins. They are connected by so called peptide bonds, forming a stable backbone structure (Figure 1).

The basic chain-like alignment of amino acids is referred to as the primary sequence. Hydrophobic interactions, van der Waals forces such as dipole interactions, salt bridges and H-bonds between amino acid residues lead to the formation of secondary structure elements. The secondary structure thus describes local three-dimensional structure, usually restricted to only parts of a polypeptide chain. Common secondary structures are α-helices and β-strands. The tertiary structure of a protein describes the arrangement of secondary structure elements within the entire protein chain, determining the final shape of the protein subunit. Covalent disulfide bonds between cysteine residues grant additional stability and correct assembly. Many proteins are only functional in homo- or hetero-oligomeric complexes. This final assembly resulting in an active and functional protein is referred to as quaternary structure.

2.1.2. The protein folding problem

In 1972 Christian Anfinsen was awarded the Nobel Prize in chemistry for his work on ribonuclease concerning the connection between the amino acid sequence and the biologically active conformation. Anfinsen showed that correct refolding of unfolded Ribonuclease A into its native and enzymatically active structure occurs spontaneously in free solution (Taniuchi and Anfinsen, 1969). All information determining the native structure is fully contained in the amino acid sequence of a protein (Anfinsen, 1973).

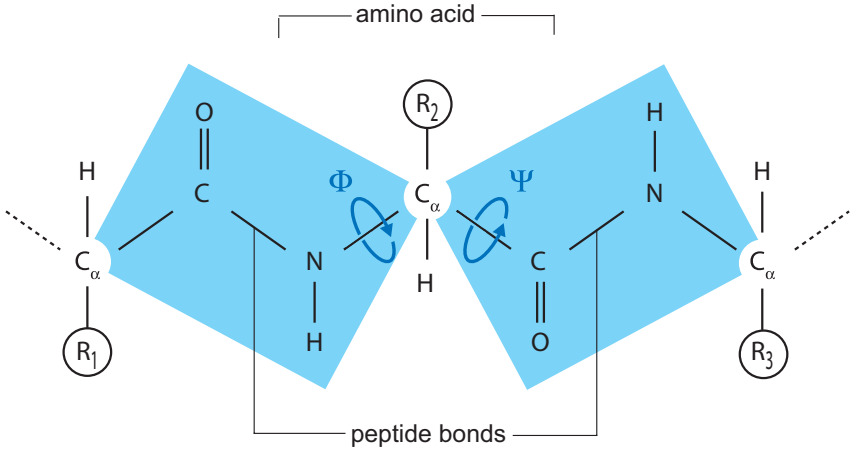


Figure 1: Amino acids connected by peptide bonds

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

Amino acid chains are, with limits, freely rotatable around the Cα-carboxyl group and the Cα-amino group bonds in their peptide backbone. The angles of these rotations are termed Ψ and Φ, respectively. In proteins, many of the potentially infinite conformers are excluded because of sterical clashes of amino acid side chains. 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.

Theoretical calculations on the number of possible conformations of a polypeptide of 100 amino acids results in 2^{100} or about 10^{30} possible conformers, when considering only the two lowest energetic states of Ψ and Φ in the polypeptide backbone (Levinthal, 1969). The physical speed limit of interconversions of Ψ and Φ is about 10^{11} every second. It would therefore take

$$\frac{10^{30}}{10^{11}} = 10^{19}\, s \approx 317 \times 10^{9}\, years \qquad (1)$$

for a single protein molecule to fold by a random search of the most stable conformer (Dinner et al., 2000). This so called 'Levinthal paradox' leads to the conclusion that a protein cannot sample all possible conformations during the process of folding rather, 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 stabilized structural elements, mainly of secondary structural origin, in combination with unstructured regions. A pathway mechanism of folding drastically reduces the amount of possible conformations during the folding process, thus allowing effective protein folding during biologically relevant timescales.

Two main models for folding pathways are currently being discussed (Daggett and Fersht, 2003). One model predicts that initially formed stable secondary structural elements collapse into tertiary structures by diffusion and collision with other secondary structures. This model is referred to as framework model (Kim and Baldwin, 1982; Kim and Baldwin, 1990) or diffusion-collision model (Karplus and Weaver, 1976). The second model, the hydrophobic collapse model (Baldwin, 1989; Schellman, 1955; Tanford, 1962), is based on a rapid collapse of the hydrophobic polypeptide chain, upon which folding can proceed with significantly less possibilities for the formation of trapped intermediate folding states.

Recent observations show that proteins can actually fold without forming detectable intermediate structures (Jackson and Fersht, 1991) or that they form

secondary and tertiary structure in parallel during hydrophobic collapse (Otzen et al., 1994). These observations lead to the proposal of the nucleation-condensation mechanism (Fersht, 1997). This mechanism combines features of both the framework mechanism and the hydrophobic collapse.

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 oligopeptides and very small proteins and are not yet applicable to the large majority of proteins in the cell.

The probable existence of multiple folding pathways for different proteins led to the proposal of an energy surface model for protein folding (Figure 2). Rather than following a defined pathway, the folding process is described by an energy landscape or folding funnel with a vast array of down-hill routes to the native state in a more or less rugged surface (Baldwin, 1995; Dobson et al., 1998; Onuchic and Wolynes, 2004).

Typically the native state of a protein can be described thermodynamically as the free energy minimum of all possible structures (Radford, 2000; Schultz, 2000). Whether a denatured protein is prone to intramolecular aggregation or reaches the native state efficiently depends on the rate of the folding process. That is, how fast a globular structure is reached in which hydrophobic surfaces are minimally exposed. How a given amino acid sequence encodes a defined three-dimensional structure is however not yet fully understood.

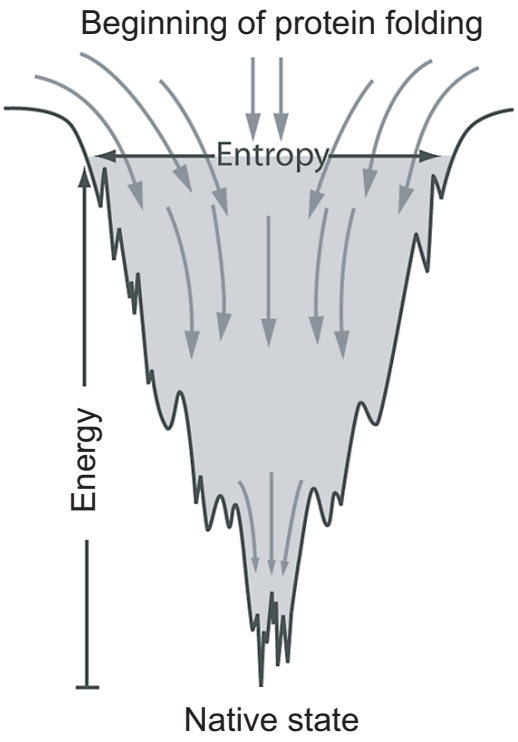


Figure 2: The folding funnel

The rugged folding landscape of a protein is funnel-like, with a preferred direction of flow toward a unique native state with the lowest energy level. Dents in the funnel wall indicate local energy minima in which proteins may get trapped in unfavorable intermediate states during folding. Figure adapted from (Onuchic and Wolynes, 2004).

2.1.4. Protein folding *in vitro* and *in vivo*

Protein folding can be experimentally followed by first unfolding and denaturing 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 folding, the proteins are transferred from the denaturing condition into an environment allowing the native state to be formed. The folding reaction itself can be analysed by a variety of methods such as circular dichroism, where secondary structure is monitored, fluorescence or absorption spectroscopy and light scattering measurements, as well as assays of enzymatic activity of refolded enzymes, to name a few.

In vitro, entire polypeptide chains are exposed to refolding upon dilution from denaturant, whereas *in vivo* an intimate coupling of biosynthesis and folding exists. Proteins sequentially emerging from the ribosome either fold co-translationally or have to be protected from aggregation and misfolding until translation is complete (Netzer and Hartl, 1998). Although it is firmly established from refolding experiments *in vitro* that the native fold of a protein is encoded in its amino acid sequence, protein folding inside cells is not generally a spontaneous process. The high concentration of macromolecules (300 g/l) in the cell provide a crowded, complex environment resulting in stronger competition of unproductive side reactions and aggregation with the productive folding pathway (Ellis, 1997; Hartl, 1996).

2.1.5. Diseases related to protein folding

In some cases, unfolded proteins are not cleared from cellular compartments by either refolding (Ben-Zvi and Goloubinoff, 2001) or degradation by proteases, but rather form stable aggregates, for example amyloid fibrils (Dobson, 1999). Amyloid formation can lead to protein folding diseases such as Alzheimer's, Huntington's or Parkinson's diseases (Dobson, 1999; Wanker, 2000). Also, spongiform encephalopathies with prions as causative disease agents are related to protein misfolding. Known diseases are Creutzfeldt-Jacob's disease and Kuru in humans or Scrapie and Bovine Spongiform Encephalopathy (BSE) in animals (Cohen, 2000).

Amyloid are highly ordered, fibrillar protein aggregates with a characteristic pattern and a typical structure as shown by X-ray diffraction and electron microscopy. Amyloid is thought to be a generic structural form that all proteins can adopt but usually do so only under extreme conditions of pH or temperature. The proteins involved in the above diseases, however, assemble into amyloid structures under physiological conditions more easily. Certain mutations in the causative proteins increase their probability to aggregate, leading to early onset cases of the respective illness.

Several other diseases probably related with misfolded proteins, such as hereditary spastic paraplegia SPG13, spastic ataxia of Charlevoix–Saguenay

(SACS), McKusick–Kaufman Syndrome (MKKS), Bardet–Biedel type 6 Syndrome (BBS6) and desmin-related myopathies are caused indirectly by mutations in genes coding for heat shock proteins or proteins with similarities to this class of proteins helping other proteins to fold, the so called molecular chaperones (Barral et al., 2004).

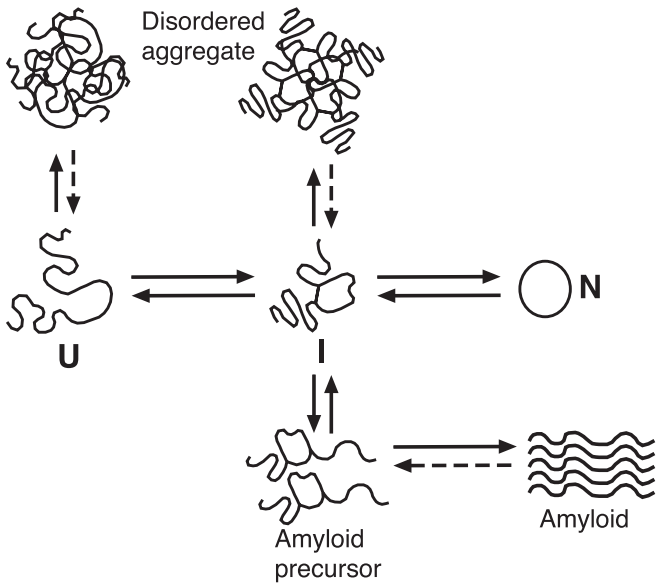


Figure 3: Folding states of polypeptide chains

Aggregation of unfolded polypeptide chains is a side-reaction of protein folding. Some chaperones can resolubilize aggregated protein species, indicated by dashed arrows. U: Unfolded polypeptide chain, I: partially structured folding intermediate, N: natively structured protein. Figure adapted from (Hartl and Hayer-Hartl, 2002).

2.2. Molecular chaperones

Evidence has accumulated over the last fifteen years that many newly synthesized proteins require a complex cellular machinery of molecular chaperones and the input of metabolic energy to reach their native states efficiently (Ellis and Hemmingsen, 1989; Gething and Sambrook, 1992; Hartl, 1996; Hartl and Hayer-Hartl, 2002). Unfolded and partially folded polypeptides often expose hydrophobic regions, which are energetically unfavorable in the

hydrophilic environment of the cytosol. Native proteins bury these hydrophobic regions inside their globular structure during folding. Chaperones capture nonnative polypeptide chains and assist their proper folding. They shield unproductive interactions of exposed hydrophobic surfaces and amino acid side chains of nascent or misfolded polypeptides and prevent subsequent malfunction and aggregation (Figure 3).

In general, chaperones do not actively fold their substrate proteins; they rather create a local environment favoring productive protein folding over functionally non-productive side reactions. Binding and release of substrate polypeptides by chaperones is often achieved by ATP-driven conformational changes, allowing multiple rounds of binding and rebinding between substrate and chaperone machinery, until a native structure is achieved. Typically then all hydrophobic areas are buried inside the core of the protein. One class of chaperones, the chaperonins can speed up folding of some proteins (Brinker et al., 2001), however, chaperones do not provide additional input of structural information in the folding process and hence chaperone action is in agreement with the dogma of protein folding described by Anfinsen: The final structure of a given protein is determined by the amino acid sequence of its polypeptide chain (Anfinsen, 1973).

Molecular chaperones are conserved throughout all kingdoms of life and act in the cell at all temperatures, but the levels of many are greatly upregulated under stress conditions. Therefore, molecular chaperones are also known as heat shock proteins (Hsps). Their respective molecular weight determines their names, *e.g.* Hsp 104, Hsp70, Hsp40, Hsp60, Hsp10. For reasons of simplicity and relevance for this study, the introduction to chaperone classes and the explanation of their function is limited to the most important chaperone systems and to the respective homologues in *Escherichia coli*.

2.2.1. Ribosome-associated chaperones

Polypeptides are generated and released into the cytosol sequentially from the ribosome and therefore expose large unstructured and hydrophobic regions during their synthesis. In order to prevent aggregation of 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).

The first chaperone that interacts with a nascent chain during their synthesis at the ribosome is trigger factor (TF). It is associated with the ribosome itself. The 48 kDa *E. coli* protein binds to a docking site at protein L23 of the large ribosomal subunit (Kramer et al., 2004). 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 are encountered. The TF reaction is not ATPase driven (Hesterkamp et al., 1996). TF also 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).

2.2.2. Hsp70 chaperones

The Hsp70 system constitutes a central part of the molecular chaperone arsenal of the cell. The common mode of DnaK action, the *E. coli* homologue of Hsp70, appears to be binding to short, extended hydrophobic peptide sequences in the substrate proteins with an ATP-regulated and ligand induced change in affinity for binding and release (Liberek et al., 1991). By shielding exposed hydrophobic surfaces, Hsp70 chaperones prevent further folding and aggregation of bound substrate proteins for the time they are bound. Native proteins do not usually expose such hydrophobic fragments and are thus not recognized by DnaK.

DnaK is active as a monomer of ~70 kDa and is comprised of two functional domains: a ~45 kDa amino-terminal ATPase domain and a ~25 kDa carboxy-terminal polypeptide binding domain whose structures have both been solved by X-ray crystallography independently (Harrison et al., 1997; Zhu et al., 1996) but not in an intact DnaK molecule. Communication between the two domains in the functional cycle results in efficient binding and release of substrate polypeptides.

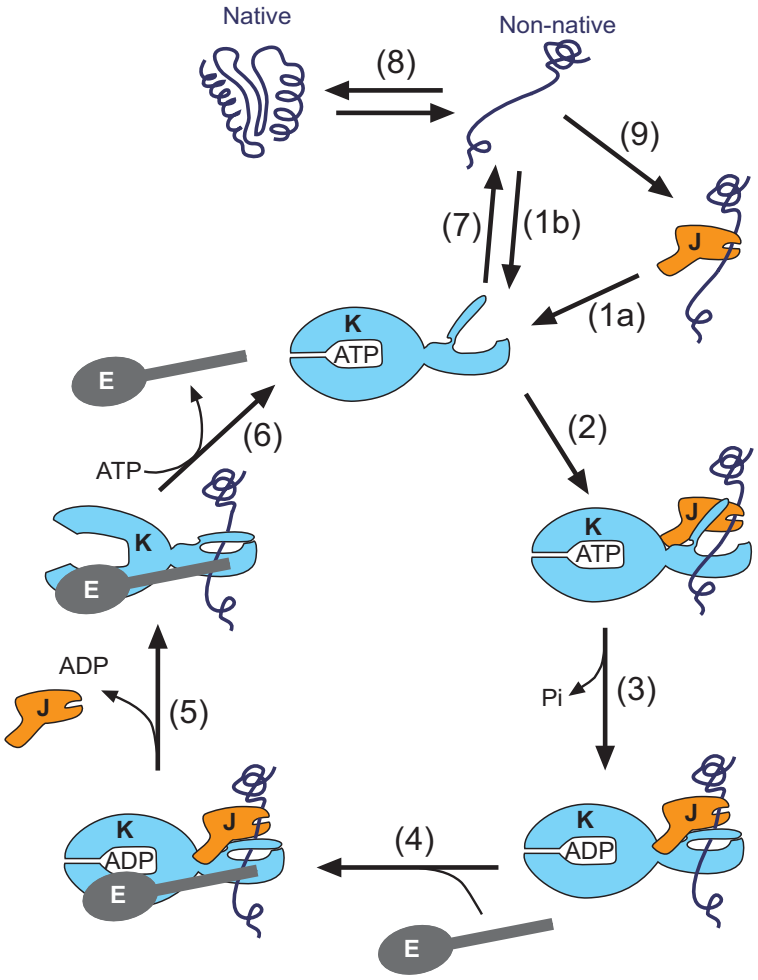


Figure 4: DnaK reaction mechanism

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 rebind to DnaJ (9) or DnaK (1b). Adapted from (Naylor and Hartl, 2001).

2.2.3. The reaction cycle of DnaK, DnaJ and GrpE

DnaK cooperates with partner proteins during their functional cycle called DnaJ (Hsp40) and the nucleotide exchange factor GrpE (Figure 4) (Bukau and Horwich, 1998; Naylor and Hartl, 2001). DnaK exists in two structural states. When ATP is bound and 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 and dissociation rates are low. Substrate binding thus occurs in the ATP bound state.

Upon substrate binding of DnaK in the ATP bound state, interaction with DnaJ triggers ATP hydrolysis with drastic structural rearrangements: The substrate binding pocket of DnaK traps the substrate inside. DnaJ is also capable of binding unfolded polypeptides itself and deliver them to DnaK. Substrate dissociation is induced by release of ADP and binding of new ATP. This requires interaction with the nucleotide exchange factor GrpE, which greatly accelerates release of ADP even if present only in small amounts, since it acts catalytically on DnaK. ATP induced opening of the substrate binding pocket of DnaK then allows release and exchange of substrate polypeptides.

2.3. Hsp60 and Hsp10: The chaperonins

The chaperonins constitute a conserved class of essential gene products encoded in the genome of almost every organism sequenced to date, distributed among eukaryotes, archaea and prokaryotic organisms (Fayet et al., 1989; Knapp et al., 1994; Ostermann et al., 1989). Chaperonins are large, multimeric, nearly 1 MDa complexes with a double-ring structure, forming two central cavities. They are divided into two 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). They have a seven-fold symmetry. 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 and the eukaryotic cytosol. The archaeal chaperonin is called thermosome, the eukaryotic homolog is called either TRiC (TCP1 Ring Complex) or CCT (Chaperonin Containing T-complex protein 1). Group II chaperonins do not interact with Hsp10-like cofactors, but the function provided by this factor is thought to be directly embedded into the structure of group II chaperonins themselves. TRiC has an eight-fold symmetry.

The following detailed introduction to structure and function of chaperonins is limited to the class I chaperonin homologues of *E. coli*, GroEL and GroES.

2.3.1. *E. coli* chaperonins GroEL and GroES

Early genetic studies identified the *Escherichia coli groES* and *groEL* genes because mutations in them blocked the growth of bacteriophages λ and T4 (Ang et al., 2000). Subsequent analyses and the finding that GroEL and GroES are overexpressed upon heat stress have shown that GroES and GroEL are part of the Hsps and constitute a chaperonin machine, essential for *E. coli* growth under all conditions tested, because it is needed for the correct folding of many of its proteins.

GroEL and GroES constitute the most intensively studied chaperone system to date (Bukau and Horwich, 1998; Fenton and Horwich, 1997; Hartl, 1996; Hartl and Hayer-Hartl, 2002; Sigler et al., 1998) and its structural properties as well as mechanistic features are well understood. Crystallographic (Boisvert et al., 1996; Braig et al., 1994; Xu et al., 1997) and electron microscopic studies (Langer et al., 1992a; Ranson et al., 2001; Saibil et al., 1991) showed that GroEL is a homo-tetradecamer of nearly 800 kDa. It is composed of two heptameric rings stacked back to back (Figures 5, 6). The rings of GroEL form two separated cavities.

The co-chaperone GroES is a dome-shaped homo heptameric structrure of 10 kDa subunits. GroES binds to either side of GroEL, forming cavities large enough to accommodate proteins up to 60 kDa inside the GroEL structure (Sigler et al., 1998). GroEL has an ATPase function and substrate and GroES binding and release as well as structural movements within GroEL are ATP/ ADP induced.

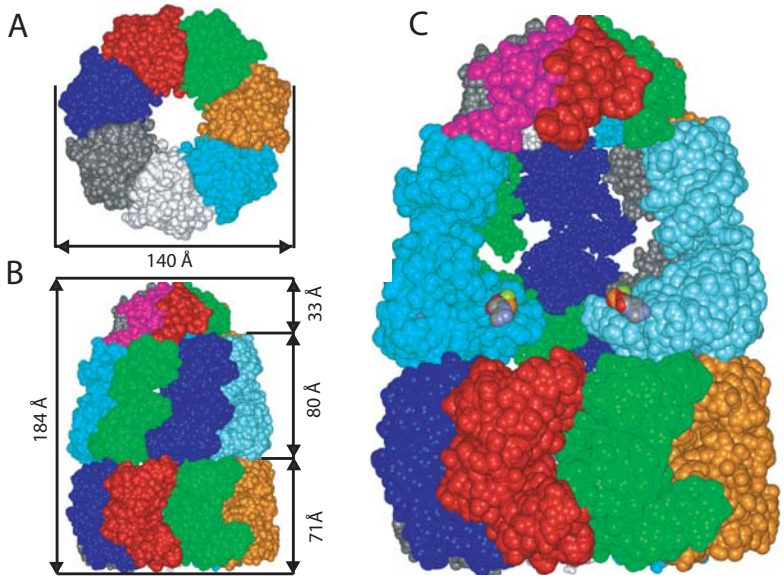


Figure 5: Structural depiction of GroEL and GroES

Space-filling models of GroEL/GroES (PDB 1AON, Xu et al., 1997). Subunits are colored individually. A: bottom view of a GroEL heptameric oligomer. B: Side view of the GroEL/GroES asymmetric complex. C: same view as in B but with two subunits removed from the upper GroEL heptrameric ring. The central cavity accommodating substrate proteins can be seen. Structures edited with ViewerPro software.

2.3.2. The structure of GroEL and GroES

Each GroEL subunit contains three domains (Figure 6). The equatorial domain is responsible for nucleotide binding as well as for stable contacts between the two heptameric ring structures. The apical domain exposes hydrophobic surfaces at the opening of the GroEL cavity. The exposed residues are responsible for substrate binding as well as for interaction with the co-chaperonin GroES. Only three consecutive apical domains in each GroEL heptamer are required for efficient substrate binding and cell viability (Farr et al.,

2000). The intermediate domain connects the equatorial and the apical domain. It is flexible and undergoes large structural rearrangements upon cooperative binding of seven ATP molecules to each ring and subsequent GroES binding.

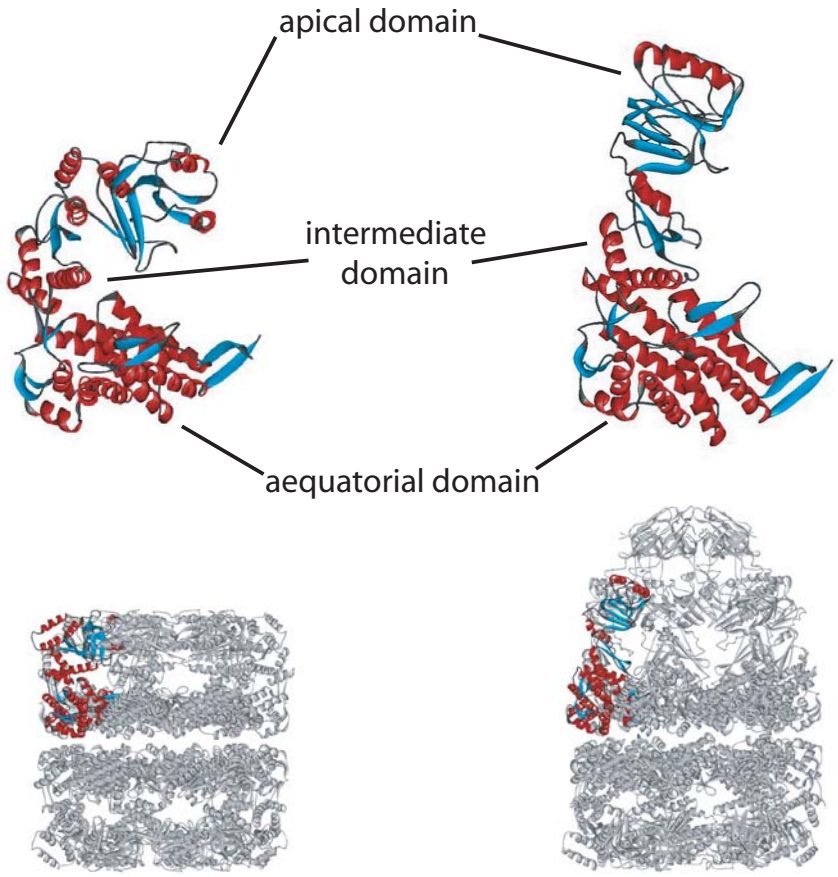


Figure 6: Structural rearrangements in GroEL upon GroES binding

Top panels show ribbon diagrams of single GroEL subunits, oriented as indicated in the bottom panels of GroEL (PDB code 1AON Boisvert et al., 1996) and GroEL/GroES (PDB code 1AON; Xu et al., 1997). GroEL monomers consist of three domains: the equatorial, intermediate and apical domains, as indicated in the top panels. Structures edited with ViewerPro software.

The rearrangements include burying of hydrophobic residues, allowing a switch from a hydrophobic to a hydrophilic lining inside the GroEL cavity (Xu et al., 1997) as well as an approximate two-fold increase of the cavity volume under the GroES lid in conjunction with an opening up and an outward and upward twisting movement of the hydrophobic apical domains (Figure 6) (Chen et al., 1994; Hayer-Hartl et al., 1996; Mayhew et al., 1996; Roseman et al., 1996; Weissman et al., 1994).

GroES binds to the apical domains of either one of the two GroEL rings depending on the ATP or ADP bound state of GroEL. GroES binding is mediated by a mobile loop at the base of the GroES dome (Figure 7) (Landry et al., 1993; Richardson et al., 2001). This loop contains 16 amino acids which fold into a β-hairpin structure upon association with its GroEL docking site. GroES is not involved in substrate recognition. Under physiological conditions, a single GroES heptamer binds to each GroEL tetradecamer, thereby forming an asymmetric GroEL/GroES complex with a cavity underneath the GroES lid (Figure 5). The GroES bound ring of GroEL is called *cis*, the unliganted GroEL ring opposite to bound GroES is called *trans*. Both N- and C-termini of the GroES heptamer are oriented towards the outside of the cavity (Figure 7).

GroES and substrate binding sites partly overlap (Chen and Sigler, 1999; Fenton et al., 1994; Xu et al., 1997). Therefore, it is assumed that GroES binding to GroEL together with the associated substrate protein, in combination with the structural rearrangements of the apical domains, virtually pushes the substrate protein inside the central cavity, where it is then allowed to fold.

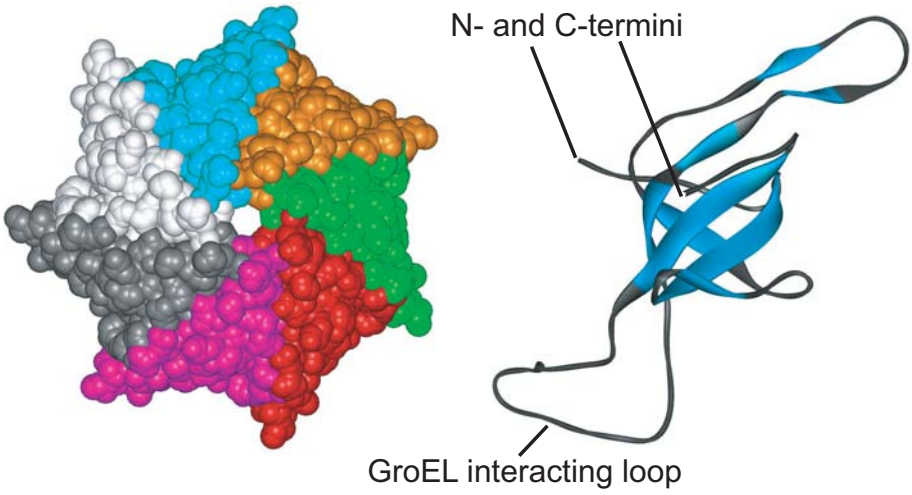


Figure 7: Structural depiction of GroES

The left panel shows a space filling model of a top view of a GroES heptamer with subunits colored individually. The right panel shows a side view of a single GroES subunit in a ribbon display. Exposed termini as well as the GroEL interacting loop are indicated. Structures modified from (Xu et al., 1997) pdb: 1AON with ViewerPro software.

2.3.3. The mechanism of GroEL and GroES mediated protein folding

GroEL- and GroES-mediated protein folding involves encapsulation of unfolded substrate proteins under the GroES lid in a concerted, alternating fashion involving both rings of the GroEL tetradecamer. Under physiological conditions, only the *cis* ring of GroEL is occupied by substrate and GroES, the *trans* side is available for substrate binding. ATP hydrolysis in the *cis*- and subsequent binding of ATP to the *trans* ring leads to a release of GroES and substrate from one side and binding GroES to the other side, which now becomes the *cis* folding chamber. Released substrate can be rebound several times until productive folding has occurred. This cyclical mode of action is called the GroEL/GroES reaction cycle (Figure 8).

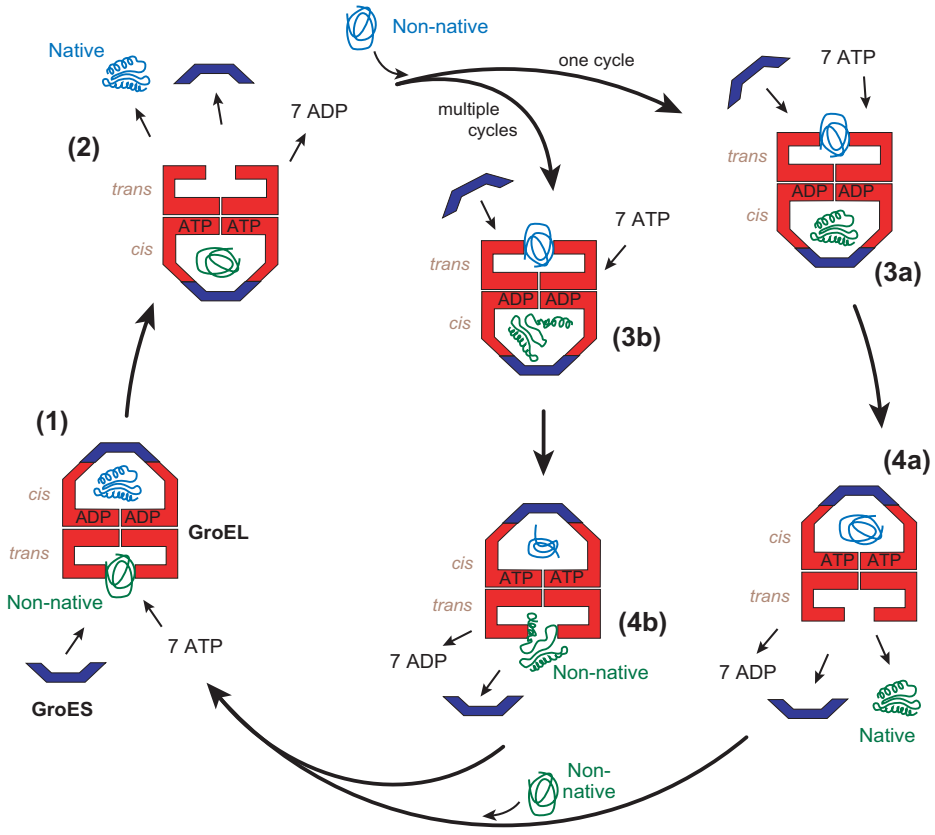


Figure 8: The GroEL/GroES reaction cycle

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. 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).

The mechanism of protein folding mediated by GroEL and GroES is fundamentally distinct from the mode of action of the Hsp70 chaperones with its principle of preventing aggregation by binding exposed hydrophobic stretches in substrate proteins. Initial recognition and binding of substrates by GroEL is mediated by hydrophobic interactions as well, but subsequent encapsulation into a hydrophilic folding chamber allows folding to proceed unimpaired by interaction with other macromolecules in the cellular environment.

Two models have been proposed to explain protein folding by GroEL. The iterative annealing model suggests an active role of GroEL in unfolding of misfolded polypeptides upon binding (Todd et al., 1996). The Anfinsen cage model describes GroEL as a passive box, in which a substrate protein can fold, unimpaired by interaction with other molecules, in "infinite dilution" inside the cavity (Ellis, 1996; Ellis and Hartl, 1996).

In the iterative annealing model, GroEL is believed to partially unfold or rearrange substrate polypeptides, before their release into the GroEL/GroES cage or back into solution. Forceful unfolding would imply the usage of ATP not only for domain movements in GroEL but also to lift bound substrates to an energetically higher level. In multiple rounds of binding and release, GroEL would actively help proteins out of energetically trapped intermediate forms and polypeptides would have repeated chances to reach their native conformation. Although this model seems appealing, supporting data is scarce (Shtilerman et al., 1999) and contradicting results have been published, showing no evidence for forceful unfolding by deuterium exchange experiments (Chen et al., 2001; Lin and Rye, 2004).

The Anfinsen cage model is in accordance with the dogma that the final structure of a protein is fully determined by its sequence of amino acids and no additional factors, like forceful unfolding of polypeptides is required. Encapsulation of substrate proteins would therefore merely protect them from unproductive interactions with other polypeptides and provide an ideal environment for folding. The single polypeptide chain inside the hydrophilic cavity of GroEL can be described as protein in infinite dilution (Ellis, 1996). The

finding that some *in vivo* chaperone dependent proteins can to a certain degree fold spontaneously *in vitro* at low concentration supports this hypothesis, as well as a large body of literature (Betancourt and Thirumalai, 1999; Brinker et al., 2001; Coyle et al., 1997; Ellis, 1994; Wang and Weissman, 1999). It is believed that the confined space of the GroEL cavity excludes certain unfavorable folding states by its size limitations, thereby promoting folding to a more compact state and increasing folding speed (Baumketner et al., 2003; Brinker et al., 2001; Jewett et al., 2004; Takagi et al., 2003). In this new view of the cage model the chaperonin modulates the way in which folding proceeds as compared to folding in bulk solution in the absence of aggregation.

2.3.4. The substrates of GroEL and GroES

In vitro, GroEL interacts with almost all unfolded proteins tested (Coyle et al., 1997; Viitanen et al., 1992), and many studies with heterologous substrate proteins such as malate dehydrogenase, DHFR, citrate synthase, *R. rubrum* RuBisCo and rhodanese have been published. The promiscuity of GroEL with respect to substrate recognition has been attributed to the plasticity of the hydrophobic binding sites in the apical domains (Chen and Sigler, 1999), allowing interaction with almost any polypeptide chain. Complementarily, substrate proteins have various conformational possibilities to interact with GroEL apical domains. Therefore, despite major efforts, no crystal structure of GroEL with bound substrate polypeptide could be determined until now.

Few attempts have been made to identify genuine GroEL substrates *in vivo*. Co-immunoprecipitation of *E. coli* GroEL with bound substrates and subsequent 2D-gel separation and analysis by mass spectrometry allowed the identification of a subset of *in vivo* GroEL substrates (Houry et al., 1999). In a separate study with an *E. coli* strain capable of down-regulating GroEL, DAPA, a protein involved in cell wall synthesis, was identified as an obligate GroEL substrate (McLennan and Masters, 1998).

In vivo, GroEL is only involved in the folding of about 10%-15% of cytosolic proteins (Ewalt et al., 1997; Houry et al., 1999). This finding from quantitative immunoprecipitation experiments contradicts the observed

promiscuity for GroEL in substrate recognition, but can be explained by the vectorial manner of protein synthesis in the cell, as well as the presence of different chaperones interacting with newly synthesized polypeptide chains. Research on the sequential interaction of different chaperones with non-native proteins resulted in hypotheses of chaperone pathways or chaperone networks (Langer et al., 1992a; Young et al., 2004).

2.4. Chaperone networks in *E. coli*

Most small proteins probably fold spontaneously upon release from the ribosome, even in the hazardous folding environment of the cytosol. Larger proteins with more complex folding pathways interact with chaperones to reach their native structure. Both DnaK and TF function in stabilizing nascent polypeptide chains, maintaining them in a folding competent state. They posses overlapping substrate spectra (Deuerling et al., 1999; Teter et al., 1999) but TF is believed to generally be the first chaperone to interact with newly synthesized proteins, since the chaperone is found associated with ribosomes. For some proteins, interaction with TF is probably sufficient to reach the native state. Many polypeptide chains reach the native state only upon interaction with the DnaK system. Neither the deletion of DnaK nor of TF is lethal, however a combined deletion leads to a severe growth defect at temperatures above 30°C (Deuerling et al., 1999; Genevaux et al., 2004; Teter et al., 1999). Strikingly, overproduction of GroEL and GroES can complement this growth defect to some extent (Genevaux et al., 2004; Vorderwülbecke et al., 2004). DnaK also cooperates with the Hsp104/Clp family of chaperones in resolubilization and degradation of protein aggregates (Mogk and Bukau, 2004).

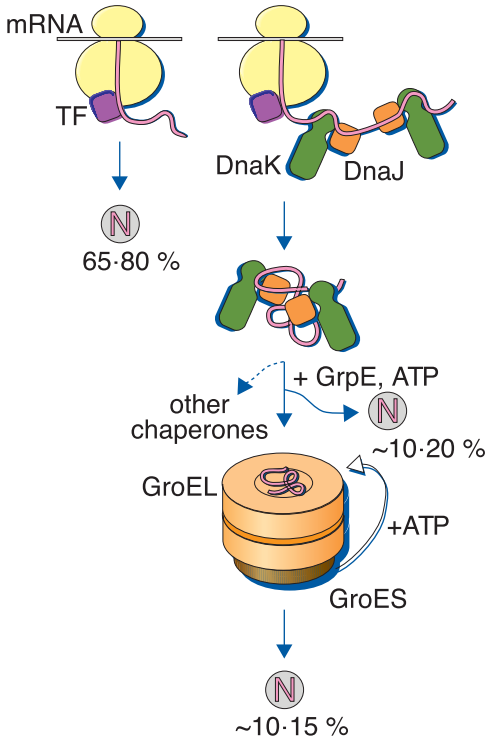


Figure 9: Model of chaperone folding pathways in *E. coli*

N: Natively folded protein, TF: trigger factor. Many proteins in the bacterial cytosol fold without further assistance upon release from the ribosome and TF. DnaK assists the remainder of proteins in folding, and can transfer substrates to the chaperonin system (GroEL/GroES). Adapted from (Hartl and Hayer-Hartl, 2002).

DnaK can deliver certain substrates to the chaperonin system which is located further downstream of the chaperone pathway (Ewalt et al., 1997; Horwich et al., 1993; Teter et al., 1999). These proteins are kept in a folding competent state and only upon transfer to GroEL, folding proceeds (Langer et al., 1992a). The chaperonin itself is unlikely to interact directly with emerging polypeptide chains from the ribosome, as only free polypeptide chains can be encapsulated. Indeed, GroEL has not yet been found associated to ribosome-attached nascent chains when TF or the DnaK is present. GroEL thus predominantly binds to the 10%-15% remaining proteins which failed to complete their folding with the assistance of upstream chaperones (Figure 9).

2.5. Introduction to proteomics

Proteomics is a young and increasingly powerful technology in molecular cell biology. It generally deals with large scale determinations of cellular function directly at the protein level. Mass spectrometry (MS) has increasingly become the method of choice for analysis of complex protein samples. MS based proteomics has only become possible by the availability of genome sequence databases and the discovery and development of protein ionization methods, as recognized by the 2002 Nobel Prize in chemistry for Koichi Tanaka (Aebersold and Mann, 2003).

2.5.1. Principles of mass spectrometry

A mass spectrometer consists of an ion source, a mass analyzer and a detector. A complex protein sample of choice is pre-treated by chromatographic methods to reduce complexity and by protease digestion with specific proteases such as trypsin to obtain defined peptides. Further reduction in complexity is achieved by subsequent high pressure liquid chromatography (HPLC). The eluted peptides are most commonly volatilized and ionized by either electrospray ionization (ESI) or matrix-assisted laser desorption/ ionization (MALDI). The mass to charge ratio is measured from protonated peptides in a mass spectrometer and mass spectra of detected peptides are recorded by the mass analyzer and the detector. Resulting spectra are commonly matched with databases to recognize specific peptides and to identify specific proteins (Figure 10).

2.5.2. Technical possibilities and applications

Recent developments allow analysis of very complex protein samples and large protein assemblies, like organelles or the ribosome. Refinements in sample preparation and labeling techniques permit the quantitative analysis of protein samples (Ong et al., 2002; Ong et al., 2003), and also comparison of proteomes at different time points or from different growth conditions. The increasing speed of analysis and refinement of methods makes this high throughput technique applicable to many problems encountered in cellular biochemistry (Aebersold and Mann, 2003; Pandey and Mann, 2000)

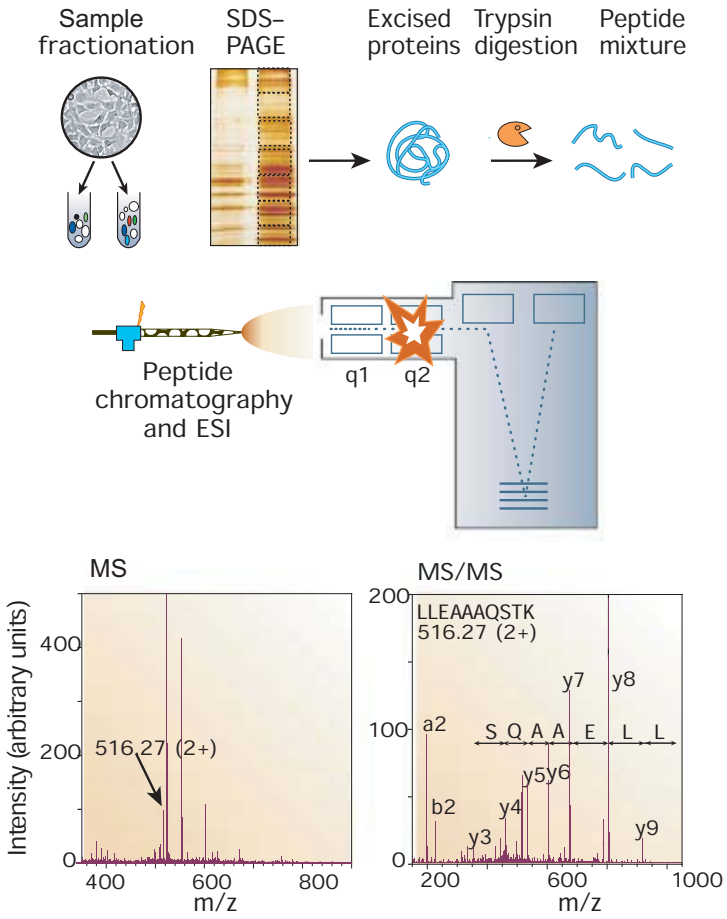


Figure 10: Basic set up of MS based experiments

The proteins to be analysed are isolated from cell lysate or tissues by biochemical fractionation. Proteins are subsequently degraded enzymatically to peptides, usually by trypsin. The peptides are separated by one or more steps of high-pressure liquid chromatography and eluted into an electrospray ion source. Multiply protonated peptides enter subsequently the mass spectrometer and a mass spectrum of the peptides eluting at this time point is recorded. A series of tandem MS (MS/MS) experiments with selected peptides follows. These consist of isolation of a given peptide ion, fragmentation by energetic collision with gas, and recording of the spectrum. Figure adapted from (Aebersold and Mann, 2003).

3. Materials and Methods

3.1. Growth media and buffers

3.1.1. Growth media

Growth media for *E. coli* were prepared with demineralized H_2O and autoclaved after preparation. LB medium: 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, (+ 15 g/l agar for solid medium). M63 minimal 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. pH was adjusted to 7.0 with KOH. Prior to usage, 1 ml/l 1M $MgSO_4$ x 7 H_2O, 10ml/l 20% carbon source (values in % are w/v unless otherwise stated) and L-amino acids to 0.5 mM final concentration were added (Sambrook et al., 1989). SILAC medium: as M63 medium, except Leucine was exchanged for Leu-D3 or Arg for Arg $^{13}C6$, respectively, in the amino acid mix. M63 sucrose medium for growth of spheroplasts: prepared as M63 medium but with addition of 250 mM sucrose, 1 mM $MgCl_2$ and 0.2% glycerol.

3.1.2. Buffers and stock solutions

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_2PO_4/K_2HPO_4, pH 7.4. TAE: 40 mM Tris-Acetate, 1 mM EDTA, pH 8.3. TBST: 25 mM Tris-HCl, 140 mM NaCl, 3 mM KCl, 0.05% (v/v) Tween 20 pH 8.0.

Antibiotic additives to growth media were prepared as 1000 x stock solutions and filter sterilized before usage: ampicilin: 100 g/l, kanamycin: 50 g/l, chloramphenicol: 20 g/l, spectinomycin: 50 g/l, tetracyclin: 5 g/l. Glucose and arabinose were prepared as 20% stock solutions, filter sterilized and diluted 100 fold before usage. Other buffers and solutions were prepared as convenient stock solutions and either autoclaved or filter sterilized before usage, if applicable.

3.2. Bacterial strains and plasmids

3.2.1. *E. coli* 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 ΔdnaK ΔdnaJ (Teter et al., 1999), MC4100 Δtig, MC4100 ΔdnaK ΔdnaJ Δtig (laboratory strain collection), MG1655 (American Type Culture Collection – ATCC 47076), SC3 (P. A. Lund, University of Birmingham, UK), MC4100 GroE PBAD, MC4100 ΔdnaK ΔdnaJ ΔgrpE (C. Georgopoulos, this study).

3.2.2. Plasmids

The following plasmids were generated for recombinant protein expression and *in vivo* experiments: pET11a amp^R *E. coli* GroES inserted at NdeI and BamHI restriction endonuclease sites. pET11a amp^R *E. coli* GroEL inserted at the NdeI and BamHI restriction endonuclease sites (pT7-ES, Brinker et al., 2001). pET22b amp^R *Methanosarcina mazei* GroES inserted at the NdeI and EcoRI restriction endonuclease sites (pT7-MmES, (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 (Novagen) for wild type proteins and into pET28b kan^R (Novagen) for amino-terminally hexahistidine tagged proteins at NdeI and HindIII, BamHI or EcoRI restriction endonuclease sites.

Construction of pT7-ES-C(His)$_6$ and pT7-MmES-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-MmES and inserted into pET22b amp^R (Novagen) at the NdeI and HindIII or EcoRI restriction endonuclease sites.

Construction of pBAD18-ES, pBAD18-ES-C(His)$_6$, pBAD18-MmES and pBAD18-MmES-C(His)$_6$: 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. Construction of pBAD33-EL and pBAD33-MmES: the ribosomal binding site and coding region of the pT7-EL and pT7-MmES plasmids was excised with XbaI and HindIII restriction endonucleases. Each fragment was inserted into the same sites of pBAD33 cam^R (Guzman et al., 1995). All constructs were verified by DNA sequencing.

3.3. DNA analytical methods

DNA concentrations were measured by UV absorption spectroscopy at λ = 260 nm. A solution of 50 µg/ml of double stranded DNA in H_2O 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 60-100 V. Primer synthesis was done by Metabion (Martinsried, Germany), DNA sequencing was performed by Medigenomix GmbH (Martinsried, Germany) or Sequiserve (Vaterstetten, Germany).

3.3.1. PCR amplification

PCR (polymerase chain reaction) mediated amplification of DNA was performed according to a standard protocol with minor modifications, when necessary. PCR running conditions also followed a standard protocol, annealing temperature and extention time varied according to primer composition and template length. For site directed mutagenesis of misincorporated bases in constructs, the Stratagene quick mutagenesis kit was used and the standard protocol was followed.

DNA template	25 ng – 250 ng
Primer	20 pmol each
dNTPs	1 mM
Polymerase buffer	1 x
Polymerase	2.5 U
Final volume	25µl, 50 µl or 100 µl

Table 1: Typical PCR reaction

Cycle count	30
Initial strand separation	95 °C, 5 min
Annealing	52 – 58 °C, 30 – 60 seconds
Cycle strand separation	95 °C, 30 – 60 seconds
Extension	72 °C, 1 min per kbp of DNA
Final exension	72 °C 10 – 20 min
Storage	4 °C or -80 °C

Table 2: Typical PCR cycling conditions

3.3.2. DNA restriction, ligation and plasmid isolation

DNA restriction was performed according to product instructions of the respective enzymes. Typically, a 20 µl reaction contained 1 µl of each restriction enzyme and 17 µl purified PCR product or 8 µl mini-prepped plasmid DNA in the appropriate reaction buffer. For ligation, 50-100 ng (~1-2 µl) 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 2 h or, for increased efficiency, at 16 °C overnight and transformed into chemically competent *E. coli* DH5α cells. Cultures with transformed cells were grown overnight in LB medium at 37°C and plasmids

were isolated using either the QIAprep Spin Miniprep kit or QIAGEN Plasmid Midi Kit or the Wizard Plus kit from Promega.

3.4. Competent cells and transformation

Transformation of bacterial cells with plasmid DNA was carried out following two different protocols. To obtain larger amounts of competent cells, 1 liter cultures of logarithmically growing *E. coli* cells were centrifuged, chilled, and carefully resuspended in 20 ml cold, filter sterilized competence buffer I (100 mM KCl, 30 mM KOAc, 60 mM $CaCl_2$, 15% glycerol pH 5.8, adjusted with acetic acid). Cells were cooled on ice for 60 min, washed with cold buffer I and resuspended in 5 ml of cold competence buffer II (10 mM MOPS, 10 mM KCl, 75 mM $CaCl_2$, 15% glycerol; pH 6.8, adjusted with NaOH, filter sterilized). 50 µl aliquots were frozen in liquid nitrogen and stored at -80 °C.

For transformation, competent cells were mixed with 1 µl plasmid DNA or 5 µl ligation reaction and incubated on ice for 20 min. Cells were heat-shocked at 42 °C for 45 s and subsequently placed on ice for 2 min. 450 µl of LB medium was added and upon phaenotypical expression for 60 min, the transformation reaction was plated on selective agar plates and incubated at the desired temperature, until colonies had developed.

For instant transformation of cells, the TSS transformation method was used. Recipient strains were grown until early logarithmic growth phase. Cells were centrifuged, 10 x concentrated in fresh LB medium and mixed with an equal volume of 2 x TSS (20% PEG-6000, 10% DMSO, 100 mM $MgSO_4$, dissolved in LB and autoclaved. 500 µl aliquots were stored at -20 °C). 1 µl of DNA was added to the transformation mixture and kept on ice for 20 min. After a 60 min phaenotypic expression time, cells were plated on selective media and incubated until colonies became visible.

3.5. Protein purification

All protein purifications steps were performed at 4 – 8 °C. The following proteins were obtained from the laboratory collection: DnaJ (Zylicz et al., 1985), GrpE (Zylicz et al., 1987), GrpE-$(His)_6$, GroES (Hayer-Hartl et al., 1996), GroEL (Hayer-Hartl et al., 1994) with modifications, GroEL-D87K (GroEL-Trap) (Fenton et al., 1994), (Weissman et al., 1994) with modifications, DnaK (Jordan and McMacken, 1995) with modifications, MetK (Markham et al., 1980), SYT (Brunel et al., 1993) with modifications.

All other GroEL substrates (ADD, ALR2, DAPA, DCEA, END4, ENO, G3P1, GATD, GATY, HEM2, LLDD, METF, NANA, TDH, XYLA, YAJO) were purified following a general protocol: *E. coli* BL21 (DE3) Gold cells harboring pET28 plasmids, which add an N-terminal hexahistidine tag to the protein of interest, were grown at 37 °C or 30 °C for aggregation prone proteins in 6 l LB medium containing 100 mg/l ampicillin. Plasmids were induced with 1 mM IPTG at an OD600 0.5 for 5 – 6 h and harvested by centrifugation for 30 min at 2500 g. Cells were resuspended in 50 mM Tris-HCl pH 7.3, 300 mM NaCl, complete protease inhibitor without EDTA (1 tablet/ 50 ml). The suspension was frozen in liquid nitrogen and thawed before addition of lysozyme (0.2 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 pore size). Lysate fractions were applied to ~10 – 15 ml Talon resin columns 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 nitrogen and stored at -80 °C.

3.6. Protein analytical methods

3.6.1. Determination of protein concentration

Protein concentrations were determined spectrophotometrically, based on the theoretical extinction coefficient of the respective protein at λ = 280 nm (Gill and von Hippel, 1989), as calculated by the ProtParam tool at the ExPASy proteomics server (http://www.expasy.org), unless otherwise stated. Molar concentrations of chaperones are expressed for the native state oligomers. GroEL substrates are expressed as monomers, since monomeric binding of substrates to chaperones is assumed.

3.6.2. SDS - PAGE

SDS - PAGE (sodiumdodecylsulfate polyacrylamide gel electrophoresis) was performed using a discontinuous buffer system (Laemmli, 1970) in BioRad electrophoresis chambers. Running buffer was 50 mM Tris-Base, 380 mM glycine, 0.1% SDS (pH 8.3). A constant current between 30 mA and 70 mA was applied per gel. Mini gels were prepared with an in-house gel casting system. SDS loading buffer was added to protein samples to 1 x concentration. Samples were heated at 96 °C for five minutes and centrifuged prior to loading.

Sample buffer	**2 x**	**5 x**
1 M Tris-HCl pH6.8 (MW 121.1)	2.4 ml	6 ml
SDS (MW 288.38)	0.8 g	2 g
Glycerol (MW 92.09)	3.2 ml	8 ml
DTT (MW 154.3)	0.82 g	1.54 g
H_2O up to	20 ml	20 ml

Table 3: Sample buffer preparation for SDS PAGE

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 background staining.

Chemicals	**Separating gel**				**Stacking gel**
	7.5%	12.5%	14%	16%	
Acrylamide stock 30.8%	9.8 ml	16.2 ml	18.1 ml	20.6 ml	3.3 ml
1.5 M Tris-HCl pH8.8	20 ml	20 ml	20 ml	19 ml	-
1 M Tris-HCl pH6.8	-	-	-	-	2.5 ml
10% SDS	400 ml	400 µl	400 µl	400 µl	200 µl
TEMED	30 ml	30 µl	30 µl	30 µl	20 µl
10% APS	200 ml	200 µl	200 µl	200 µl	200 µl
H_2O up to	40 ml	40 ml	40 ml	40 ml	20 ml

Table 4: Gel preparation for SDS PAGE

3.6.3. Silver staining

Polyacrylamide gels were placed in fixing solution (40% ethanol, 10% acetic acid) for 30 minutes. A subsequent oxidation in incubation solution (30% ethanol, 250 mM sodium acetate, 8 mM sodium thiosulfate x 5 H_2O) for 30 min was followed by three 5 min washing steps in water. The gels were incubated in silver solution (5 mM $AgNO_3$, 10 µl formaldehyde/ 100 ml) for 40 minutes and then placed in developing solution (250 mM Sodium carbonate, 10 µl formaldehyde/ 100 ml). After the gels were stained to sufficient intensity, they were placed in stop solution (40 mM EDTA, solubilised NaOH pellets) to prevent overincubation.

3.6.4. 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 a constant current of 150 mA (Towbin et

al., 1979). Western blots with one or two gels were run for 1.5 h, four and more gels for 2 h.

Nitrocellulose membranes were blocked in 5% skimmed milk powder in TBST for 1 h or overnight. The membranes were then incubated with a 1:1000 – 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). After extensive washing, protein bands were detected by incubating the membranes with ECL chemiluminescence solution and exposure to X-ray film or on a Fuji LAS 3000 machine.

3.6.5. Generation of antibodies

Rabbit polyclonal antibodies were generated at the animal facilities of the MPI for Biochemistry. 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 for test bleeds and the final bleed was taken ~10 days after injection (Harlow and Lane, 1988).

3.6.6. Size exclusion chromatography

For competitive binding experiments of GroEL substrates, proteins were mixed, incububated with GroEL and subjected to size exclusion chromatography on a SMART system. Substrate proteins ENO, GATD, DCEA, METK and DAPA were denatured in buffer A (20 mM MOPS pH 7.4, 100 mM KCl, 10 mM $MgCl_2$) with 4-6 M GdHCl and 10 mM DTT for 30 min at 25°C, mixed in defined ratios (Figure 35) and diluted 100-fold at 37°C into buffer A containing 0.25 µM GroEL. Samples were subjected to size exclusion chromatography on a SMART system with a Superose 6 column at 37°C. Fractions containing GroEL were pooled and further analyzed by immunoblotting and quantification software AIDA.

3.7. Protein refolding

Refolding reactions were done in collaboration with Michael Kerner and Dean Naylor. 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 ATPase activity of the chaperone. 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.

3.7.1. 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 gift from R. E. Viola and R. Moore, University of Toledo, Ohio, USA). 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.

3.7.2. 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 at 340 nm as described (De Biase et al., 1996).

3.7.3. ENO refolding

100 µM enolase 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 MgSO4 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.

3.7.4. 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 MnCl2, 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).

3.7.5. 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.

3.7.6. METK refolding

METK was denatured for 1 h at 25 °C with 6 M GdnHCl in buffer A containing 8 mM DTT 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., 1980) except that L-[^{35}S]-methionine (specific activity 50 Ci/mol) was used.

3.7.7. 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 MgCl2, 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.

3.8. *In vivo* co-expressions

3.8.1. Co-expressions of chaperones and substrates in *E. coli*

BL21 (DE3) Gold cells, harboring either the pBAD33-ESL, pBAD33-EL, or pBAD33-MmES 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% glucose and 0.2% glycerol to OD_{600nm} = 0.4. Chaperones were induced for 1 h by shifting cells from

glucose to arabinose (0.2%) containing medium. Control cells were grown continuously in LB with glucose. Following chaperone induction, the medium was changed back to glucose and supplemented with 1 mM IPTG for 1 h to induce expression of substrate proteins. Equivalent numbers of cells were taken for preparation of total, soluble and insoluble protein fractions. Cells were centrifuged and the material for total protein preparation was resuspended in SDS-PAGE sample buffer. The material for soluble/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. 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.

3.8.2. Co-expressions of chaperones and substrates in S. cerevisiae

Bacterial proteins were expressed in S. cerevisiae YPH499 cells transformed with expression plasmids p415Gal under galactose promoter control grown in SC–Leu medium at 30°C. Protein expression was induced at OD_{600} = 0.5 with 2% galactose for 4 hrs. Spheroplasts were prepared by Zymolyase treatment and lysed in PBS containing 0.1% Triton X-100 and EDTA-free protease inhibitors (Roche). Samples were fractionated into soluble and pellet by centrifugation (20,000 x g for 15 min). Protein amounts were analyzed by immunoblotting. For GroEL/GroES co-expression with substrates, the above strain was co-transformed with substrate plasmid and GroEL (pSI215) and GroES (p426ADH) plasmids under copper and ADH promoter control, respectively, in SC–Leu–Trp–Ura medium. GroEL was induced with 0.5 mM $CuSO_4$ for 3 hr before induction of bacterial GroEL substrates. To examine bacterial substrate solubility in yeast ydj1-deficient background, the strain wy1 and its isogenic wild-type strain DS10 (Becker et al., 1996) were analyzed as above.

3.9. 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 inserted 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 kan^R (MC4100 SC3 Kan^R) as previously described (Ang and Georgopoulos, 1989).

LB medium containing kanamycin (0.05 g/l) and 0.2% arabinose was inoculated with MC4100 GroE P_{BAD}. 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 mg/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 immunoblotting.

3.10. GroEL/GroES-substrate complexes

3.10.1. Cell growth

E. coli MC4100 cells were transformed with plasmid pBAD18 expressing either *E. coli* GroES or *M. mazei* GroES, with or without a C-terminal histidine-tag and grown on selective medium containing 0.2% glucose to prevent expression of plasmid genes. Single colonies were picked and grown at 30 °C or 37 °C in a pre-culture overnight. Cells were then diluted into pre-warmed LB medium with 0.2% glucose and grown to mid-exponential phase (OD_{600} ~ 0.4). After

centrifugation, cells were resuspended in pre-warmed LB medium, containing 0.2% arabinose to induce expression of plasmid genes for 30 min.

Cells were centrifuged and resuspended in buffer 1 (100 mM Tris-HCl, 500 mM sucrose, 5 mM EDTA, pH8.0). The resuspension was mixed with an equal volume of 0.3 mg/ml lysozyme in water and incubated on ice for 15 minutes. After careful addition of 1 M $MgSO_4$ to a final concentration of 20 mM, the cells were centrifuged for 15 min with 1000 g. The supernatant was carefully removed, the spheroplasts were resuspended in buffer 2 (50 mM Tris-HCl, 250 mM sucrose, 10 mM $MgSO_4$, pH8.0) and again centrifuged for 15 min at 1000 g.

3.10.2. Cell lysis and purification of complexes

Resulting spheroplasts were resuspended in 50 ml prewarmed M63 minimal 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 minutes. 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% 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.

After clearance of the supernatant by centrifugation at 30000 x g for 10 min and pH adjustment to ~8.0, soluble proteins of four liters of *E. coli* culture were 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 buffer B then further washed in a chromatography column with 12 ml of 50 mM imidazole in buffer B at gravity flow. GroEL/GroES/substrate complexes were eluted with 12 ml of 200 mM imidazole in buffer B and 0.5 ml fractions were collected. Protein content was analyzed by SDS-PAGE.

3.10.3. Alternative purification method

To isolate the less stable substrate complexes with *E. coli* GroES instead of MmES, the method described above was modified to increase preparation speed.

As a consequence, preparations of GroEL substrate complexes contained more unspecifically bound proteins.

After growth in LB medium, induction of GroES and centrifugation, cells were resupended in lysis buffer (50 mM Tris-HCl pH8.0, 0.2 mg/ml lysozyme, 0.2% glucose, 10U/ml Hexokinase EDTA free protease inhibitor, 10U/ml benzonase, 15 mM $MgCl_2$ 15 mM KCl 10 mM ADP), incubated at 4 °C for 5 min, and lysed by sonification. Remaining cell debris was spun down at 20 000 g for 10 minutes and the supernatant was adjusted to pH7.5. Equilibrated Talon beads (1 ml/l of cells, equilibrated in wash buffer) were incubated with the lysate for 30 min at 4°C. Beads were subsequently washed repeatedly in small volumes of wash buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 15 mM $MgCl_2$, 50 mM KCl, 10 mM ADP), also with low concentrations (20 mM) of imidazole. GroES and interacting proteins were finally eluted with 200 mM imidazole in wash buffer. Protein content of fractions was analysed by SDS-PAGE.

3.10.4. 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.

3.11. Mass spectrometric methods

3.11.1. Sample preparation for protein identification by mass spectrometry

Mass spectrometry and data analysis of the mass spectrometric output was performed in the laboratory of Professor Mathias Mann by Dr. Yasushi Ishihama and Morten Kirkegaard (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 per-formed 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.

3.11.2. Coupled liquid chromatography – mass spectrometry (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 electrospray 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 0.5% acetic acid,

0.5% acetic acid plus 0.005% TFA, and 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 (enrich-ment 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).

3.11.3. Analysis of mass spectrometric data

Peak lists were created by scripts in Analyst QS (MDS-Sciex) on the basis of the recorded fragmentation spectra and were submitted to the 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 $mπ = 10^{PAI} - 1$, 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

3.12. Bioinformatic methods

3.12.1. 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).

3.12.2. Protein sequence analyses

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 the laboratory of Prof. Dr. F. Ulrich Hartl. Michael Kerner was involved in data analysis and interpretation of GroEL substrate complexes, *in vitro* refolding experiments, cloning and purification of substrate proteins, as well as in co-expressions of GroE chaperones with substrate proteins. Dr. Dean Naylor was involved in cloning and purification of GroEL substrate proteins. He performed parts of the *in vitro* refolding experiments and was involved in experimental design. Hung-Chun Chang carried out the solubility studies of GroEL substrates in yeast.

Mass spectrometry was performed by Dr. Yasushi Ishihama and Morten Kirkegaard in the laboratory of Prof. Dr. Matthias Mann (University of Southern Denmark, Odense).

4.1. Identification of GroEL substrates

4.1.1. Experimental approach

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 chaperone system. However, until now, studies on the chaperonin molecular mechanism have been mostly conducted with *E. coli* GroEL/GroES and heterologous substrates such as *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.

In a previous study conducted in our laboratory, a co-immuno-precipitation approach was used to identify GroEL interacting proteins *in vivo* (Houry et al., 1999). In this research project, polyclonal antibodies were used to capture GroEL from cell lysate in its nucleotide free state, allowing stable binding of substrate proteins. Analysis by 2D SDS-PAGE and mass spectrometry allowed identification of 52 GroEL interacting proteins. Structural analysis revealed a

preference for multiple α-β domains among these. However, the dataset was too small to permit more detailed conclusions on what determines a protein to be a GroEL substrate.

To extend this study and to comprehensively identify all GroEL interacting proteins *in vivo*, a more sophisticated methodological approach was introduced. We isolated GroEL/GroES/substrate complexes, with substrate proteins being trapped within the *cis*-cavity of GroEL under a histidine-tagged GroES lid (Figure 11). These complexes are stable under ADP conditions and allowed rapid isolation by immobilized metal affinity chromatography (IMAC). With this approach, the purification of sufficient amounts of GroEL/GroES/substrate complexes was possible, which allowed us to perform large-scale, high-accuracy mass spectrometric analysis leading to the identification of the virtually complete set of GroEL interacting proteins in *E. coli* cells.

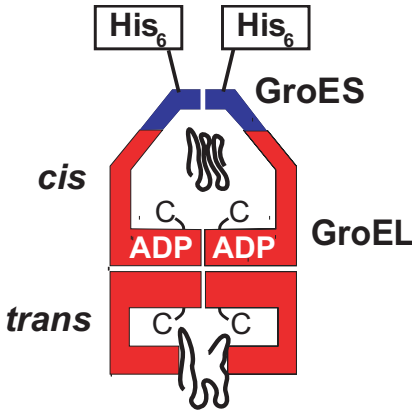


Figure 11: Model of GroEL and GroES with bound substrate polypeptide

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

4.1.2. Stability of GroEL/GroES complexes

Stability of GroEL/ES/substrate complexes is crucial to obtain a maximal yield during purification, while thorough washing conditions can be applied to remove proteins binding non-specifically to the purification matrix. Wild type *E.*

coli GroES did not allow purification of complexes with a reproducibly high yield (data not shown). We therefore used a GroES homolog of an archaeal organism, *Methanosarcina mazei* (MmES). Surface plasmon resonance experiments had shown that dissociation of either EcES or MmES from GroEL is prevented in the presence of ADP (Figueiredo et al., 2004; Klunker et al., 2003), verifying our experimental approach. It was further shown that GroEL formed significantly more stable complexes with MmES than with EcES, allowing isolation of GroEL/GroES/substrate complexes in reproducibly high quantities. MmES could replace the essential function of *E. coli* GroES both *in vivo* and *in vitro* refolding studies with *E. coli* GroEL (Figueiredo et al., 2004; Klunker et al., 2003). It is important to note however, that GroEL but not GroES is responsible for substrate selection, thereby ensuring that the captured proteins represent authentic substrates.

4.1.3. Processing of GroEL/GroES/substrate complexes

By expressing low levels of carboxy-terminally hexahistidine tagged *M. mazei* GroES in E. coli MC4100 cells and lysing spheroplasts in the presence of sufficient glucose-hexokinase to rapidly convert all cellular ATP to ADP, it was possible to isolate captured substrate complexes by IMAC. The seven hexahistidine tags on the GroES oligomer did not pose sterical constraints, since the carboxy-terminus of GroES is oriented towards the outside when bound to GroEL (Figure 7). Furthermore, binding of GroES with seven histidine tags to the purification matrix permitted stringent washing conditions. Substrate complexes were not removed while washing with 50 mM imidazole but were efficiently released with 200 mM imidazole (Figure 12).

Substrate complexes were separated by SDS-PAGE to isolate the vast excess of GroEL and GroES from the captured substrates. Gel slices perpendicular to the gel migration were subjected to trypsin digestion and the peptides extracted and analyzed by LCMS/MS. Identified proteins were verified by manual annotation (see materials and methods). A total of 402 different proteins were found to interact with GroEL in wild type MC4100 cells under all tested conditions (23°C, 30°C, 37 °C, rich and minimal media), while a single

pull-down experiment typically resulted in detection of 200 - 250 proteins. The number of 200-250 GroEL interacting proteins is close to the entire expected population of chaperonin substrates, which was previously estimated to be ~10% – 15% of cytosolic proteins by mass (Ewalt et al., 1997; Houry et al., 1999) by quantitative immuno precipitation. To enhance specificity of GroEL substrate assignment, a protein was only considered to be GroEL interacting if it was identified in at least two independent experiments or if the relative enrichment factor (REF, see below) of the protein could be determined. These resulted in a data set of 252 GroEL interacting proteins used throughout this study.

4.1.4. Experimental controls

Utilizing the same experimental parameters with cells where MmES lacking the hexahistidine tag was expressed, only seven proteins were detected by mass spectrometry (Figure 12). These proteins were thus considered to be non-specifically bound to the IMAC resin and were excluded from further analysis. They are EFTU, FABZ, FUR, GLMS, RL32, RS15, and SLYD.

To test potential post-lysis exchange of GroEL-bound substrates, intact cells with overexpressed histidine-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, 25 Arg-^{13}C6 labeled proteins could be identified as associated with GroEL in the resulting complexes. Post-lysis cycling of GroEL/GroES/substrate complexes and re-binding of different proteins is therefore very limited and does not significantly influence the results. The identified proteins comprised six polypeptides already identified as non-specifically interacting. Also 19 ribosomal proteins were identified among this set of non-specific interacting proteins. Consequently, ribosomal proteins were excluded from further analysis.

To additionally ensure that complexes were not formed after cell lysis, wild type spheroplasts, not containing hexahistidine tagged GroES, were lysed in the presence of glucose-hexokinase and an excess of purified carboxy-terminal hexahistidine tagged GroES. Following IMAC isolation, only a very small

amount of *E. coli* GroEL was captured, confirming that complex formation largely occurs *in vivo* (Figure 12).

For this study, we concentrated on a data set of 252 GroEL interacting proteins. It was estimated to be essentially complete, based on three criteria: Firstly, the number of GroEL interacting proteins did not increase in repeated analyses. Secondly, analysis by more sensitive FT-MS did not significantly increase the number of proteins found associated with GroEL and thirdly, the fact that about 1200 different proteins of a possible 2400 soluble proteins (Frishman et al., 2003) from an *E. coli* cell lysate were identified by the same technique indicated high enough sensitivity of the MS-MS approach used in this study.

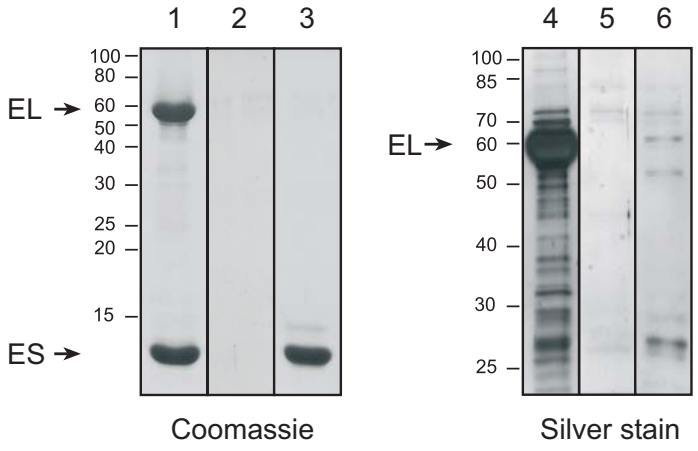


Figure 12: Purification of GroEL/GroES/substrate complexes

Spheroplasts of *E. coli* MC4100 expressing $(His)_6$-tagged GroES were rapidly lysed in presence of glucose and hexokinase to convert cellular ATP to ADP. Stable GroEL/GroES/substrate complexes were eluted with 200 mM imidazole (lanes 1, 4). To identify non-specific binding of proteins to IMAC, GroEL/GroES/substrate complexes were prepared using a non-tagged version of GroES (lanes 2, 5). Post lysis cycling of GroEL in complexes could be excluded by lysing cells expressing non-tagged GroES in presence of purified $(His)_6$-GroES (lanes 3, 6). Samples were subjected to 16% SDS-PAGE, followed by Coomassie (1, 2, 3) or silver staining (4, 5, 6).

4.1.5. Influence of other chaperone systems on GroEL substrate diversity

Isolation of GroEL/GroES/substrate complexes was repeated in cells with combinations of chaperone deletion genotypes. GroEL substrates isolated from cells with the genes for either TF or DnaK/DnaJ/GrpE deleted (Genevaux et al., 2004) showed no significant difference in number or composition to GroEL substrates isolated from wild type cells (data not shown). However, a combinded deletion of both chaperone systems at once increased the number of GroEL associated proteins by ~60%.

TF and the DnaK chaperone system are known to possess overlapping substrate pools (Deuerling et al., 2003; Genevaux et al., 2004; Teter et al., 1999), explaining the similarities between GroEL substrates isolated under either wild type conditions or deletion conditions of single chaperone systems. The chaperone systems can functionally compensate for the loss of either of them. A complete lack of upstream chaperones, as given in the combined deletion of TF and DnaK/DnaJ/GrpE, leads to a concurrent increase in GroEL associated proteins. Proteins, which would normally fold by either TF or DnaK/DnaJ/GrpE now remain in an unfolded state, exposing hydrophobic residues and stretches, recognized by GroEL.

4.1.6. Quantification of GroEL interacting proteins

The concentration of GroEL tetradecamer in the cell is assumed to be 3 µM (Ellis and Hartl, 1996). However, molar concentrations of newly folding proteins, identified as GroEL substrates, significantly exceed the GroEL capacity. GroEL, seen as a general folding machine, could therefore only fold a small fraction of all newly synthesized interacting proteins. Alternatively, GroEL could discriminate against certain proteins, and preferentially fold others. To address this crucial question with respect to substrate distribution on GroEL *in vivo*, as compared to an *E. coli* cell lysate, we utilized a novel technique for quantification by mass spectrometry called SILAC (stable isotope labeling by amino acids in cell culture) (Ong et al., 2002; Ong et al., 2003).

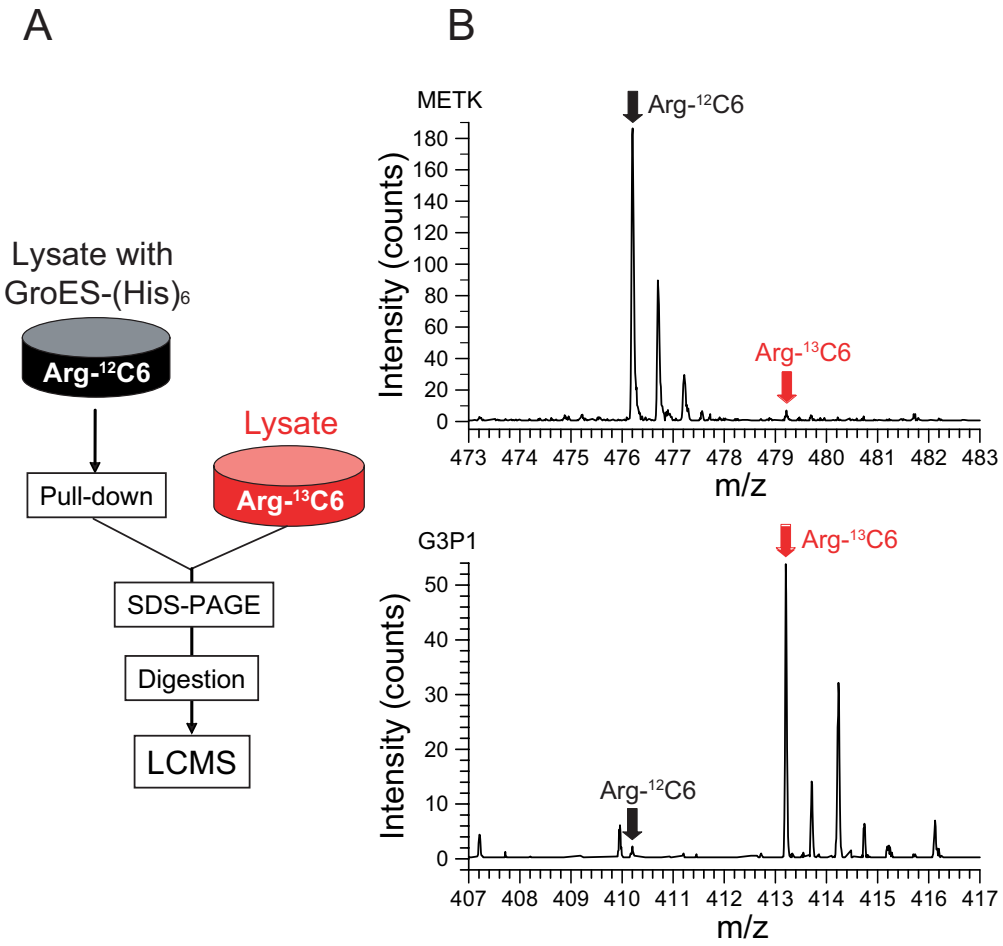


Figure 13: Quantification of GroEL bound proteins by SILAC

(A) Experimental setup and (B) mass spectra of peptide FFINPTGR (doubly charged) from enriched *E. coli* protein METK (top) and of peptide VGINGFGR (doubly-charged) from the not enriched protein G3P1 (bottom).

An arginine-^{13}C6 labeled cell lysate of wild type *E. coli* cells was mixed in a known ratio with unlabeled purified GroEL/GroES/substrate complexes and treated as described above. Peptides derived from tryptic digestion of a protein from Arg-^{13}C6 labeled cell lysate and peptides of the same protein bound to GroEL were detected as separate peaks in mass spectrometry, due to their differing mass (Figure 13). Areas of the different isotope peaks were then directly compared in the same spectrum, allowing their relative concentration to be determined. Ratios of unlabeled (Arg-^{12}C6) to Arg-^{13}C6-labeled peak intensities were measured by MSQuant software and, taking the amounts of GroEL in the

starting material and the mixing ratio into consideration, converted to a characteristic enrichment factor. These factors, with 1 being neither enriched nor depleted, ranged from <0.01 to >100 (Supplementary Table), indicating a clear preference of GroEL for some substrate proteins as compared to others. A detailed interpretation of enrichment factors is discussed below.

4.2. Properties of GroEL substrates

4.2.1. Mass distribution of proteins associated with GroEL

GroEL-GroES complexes arrested in the ADP state contain substrate proteins encapsulated under the lid of GroES (the *cis* cavity), where they undergo folding. GroEL can accommodate proteins up to a molecular weight of 60 kDa in its *cis* cavity (Sigler et al., 1998). In addition, some proteins may be bound to the GroEL *trans* ring (Farr et al., 2003), most likely those candidates of a size too large to be encapsulated (Figure 11). A size distribution of GroEL bound proteins and *E. coli* lysate proteins indeed revealed a preference of GroEL interacting proteins to fit inside the cavity. 77% of these polypeptides have a size between 20 kDa and 60 kDa, whereas only 62% of remaining polypeptides found in the lysate fit in the corresponding size range (Figure 14). About 12% of GroEL interacting polypeptides are larger than 60 kDa, being potential candidates for proteins folding with GroEL via an alternative *trans* mechanism (Farr et al., 2003). Small proteins <10 kDa are basically not found among GroEL interactors.

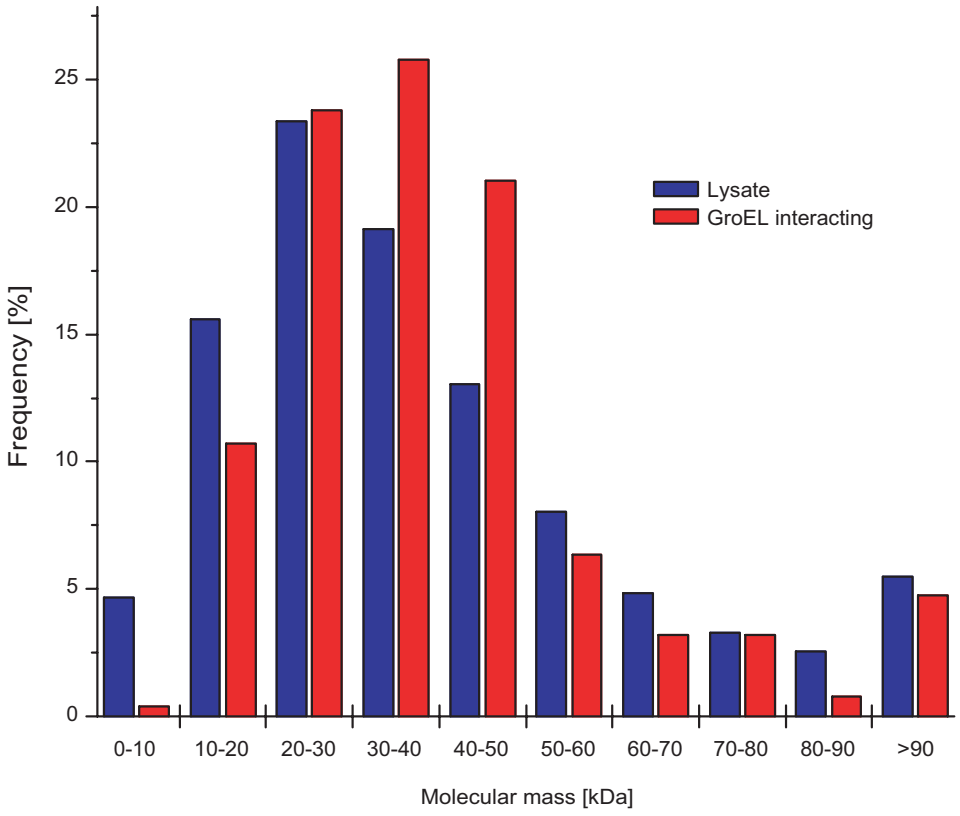


Figure 14: Figure: Mass distribution of *E. coli* proteins

Bar graph of the mass distribution of proteins from an *E. coli* lysate (blue bars) and GroEL interacting proteins (red bars). GroEL interacting proteins show a size preference between 20 kDa and 50 kDa. Some proteins on GroEL are larger than 60 kDa, the theoretical size limit of the GroEL/GroES cis cavity (Sigler et al., 1998).

4.2.2. Distinction between cis and *trans* bound polypeptides to GroEL

Previous studies have 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 capturing non-native substrates from degradation. In contrast, unfolded substrates bound to the apical domains of the *trans*-ring of GroEL are easily degraded by PK, allowing the protease also to cleave the 14 carboxy-terminal amino acids of GroEL in the open ring. This results in a characteristic double band, visible upon SDS-PAGE (Figure 15) and serves as indicator for proteinase K activity (Langer et al., 1992b;

Mayhew et al., 1996). It was reported that yeast mitochondrial aconitase, a monomer of 82 kDa, can fold productively by GroEL in *trans*; it requires GroES binding to the opposite ring for release of aconitase without encapsulation (Chaudhuri et al., 2001). To confirm *trans*-binding of substrates larger than the maximal cavity size, isolated GroEL/GroES/substrate complexes in presence of ADP were treated with proteinase K. 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. However, the 74 kDa protein SYT, a threonyl t-RNA synthase of *E. coli*, was quickly and completely degraded, in GroEL bound as well as in unbound form, consistent with its inability to become encapsulated (Figure 15).

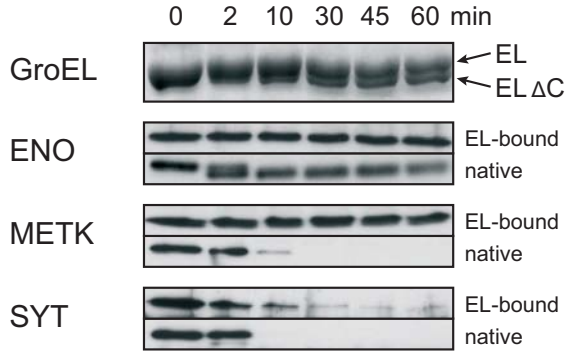


Figure 15: Proteinase K digests of GroEL and bound substrates

Complexes and the native purified proteins, as controls, were subjected to PK digestion and the reactions were stopped at indicated times. Samples were analyzed by SDS-PAGE and silver stained (GroEL) or immunoblotted for the proteins indicated (ENO, METK, SYT). EL: GroEL; ELΔC: C-terminally truncated GroEL by protease treatment.

DnaK, the *E. coli* Hsp70 homolog was also identified as a GroEL interacting protein. Since DnaK has a molecular weight of 69 kDa, folding inside the GroEL cavity could be excluded. PK digests did not result in complete DnaK degradation, as it would be expected for unfolded proteins, but resulted in a characteristic band pattern indicative of native DnaK, including its stable 44 kDa

ATPase domain (Liberek et al., 1991), Figure 16). It has been suggested that DnaK has the ability to target aggregation-prone unfolded proteins to GroEL *in vivo* (Langer et al., 1992a), and finding DnaK in its native form bound to GroEL, or more likely to a large GroEL bound substrate protein, further strengthens this concept of chaperone interplay during protein folding (see chapter 1.4).

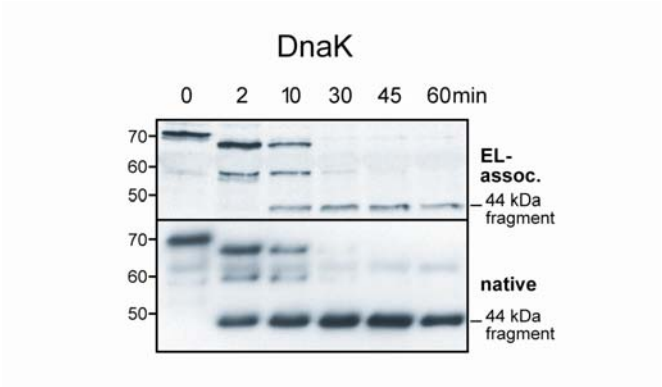


Figure 16: Proteinase K digests of GroEL and DnaK

Complexes and the native purified DnaK protein were subjected to PK digestion and the reactions were stopped at indicated times. Samples were analyzed by SDS-PAGE and immunoblotted with a DnaK antibody.

4.2.3. Essentiality of GroEL substrates

GroEL and GroES are two essential gene products (Fayet et al., 1989). This feature was logically attributed to the existence of at least one essential *E. coli* protein being exclusively folded by GroEL. Among GroEL interacting proteins, 26% are identified as essential (67 out of 252, essentiality determined after (Gerdes et al., 2003). Interaction of a substrate protein with GroEL however, does not necessarily mean that this protein makes use exclusively of GroEL for productive folding. Enrichment factors, as determined above, are a measure of how much of an individual protein is bound to GroEL at any given time. High enrichment factors therefore correlate with a high degree of GroEL usage for folding. Ten proteins with an enrichment factor >100 were identified as essential proteins, thereby explaining essentiality of GroEL for cell survival (Table 5).

Name	MW	Function
RSD	18.2	Regulator of sigma factor D
YCFP	21.2	Hypothetical protein
FTSE	24.4	Cell division ATP-binding protein
GCH1	24.7	GTP cyclohydrolase 1
TRMD	28.4	tRNA guanine methyltransferase
HEM2	35.5	Delta-aminolevulinic acid dehytratase
YBJS	38.1	Hypothetical protein
DHAS	40.0	Aspartate semialdehyde dehydrogenase
DADA	47.6	D-amino acid dehydrogenase small subunit
PARC	83.8	Topoisomerase IV subunit A

Table 5: Essential GroEL substrate proteins with enrichment factors >100

4.2.4. Functional categories among GroEL interacting proteins

It was previously not known whether GroEL interacts preferably with certain functional groups of proteins, belonging to the same cellular pathway or performing related tasks. To identify functional categories of substrates bound to GroEL and to compare them to *E. coli* lysate proteins, particular functions had to be assigned to all proteins. A comparison of functional classification data bases, namely SwissProt, EcoCyc and COGs, revealed that the latter, Clusters of Orthologous Groups of Proteins (COGs), served our purposes best (Tatusov et al., 1997). COGs are derived by comparing protein sequences from multiple complete genomes. Typically, proteins belonging to the same COG share a specific function. This data base offers a detailed, comprehensive and simple classification of fold types.

When comparing the *E. coli* lysate to GroEL interacting proteins, the observed distribution of functional categories was found to be similar (Figure 17). Differences in proteins related with cell division, secondary metabolism, cell motility, lipid metabolism and ion transport have to be considered insignificant, since values for GroEL interacting proteins merely relate to six proteins or less. Proteins involved in coenzyme metabolism, DNA replication, recombination and repair, transcription and protein turnover were slightly enriched among GroEL interactors, whereas cell envelope biogenesis proteins and proteins involved in translation, in ribosomal structure and biogenesis are found to a higher percentage in an *E. coli* lysate than among GroEL interacting proteins. The latter difference is probably due to the fact that ribosomal proteins were excluded from this study, since they were repeatedly identified as unspecific binders. 13.5% of the identified proteins in the lysate and 9.5% of GroEL interacting proteins were of unknown function. GroEL dependence is not strictly related to distinct functional characteristics of proteins, since the observed differences between lysate and GroEL interacting proteins do not appear to be very significant. From these functional classes it can also not be concluded whether evolutionary old proteins or relatively young proteins predominantly interact with GroEL, since too little is known about evolutionary descent of proteins and protein motives. Rather than functional characteristics, structural features of a protein might determine its GroEL dependence.

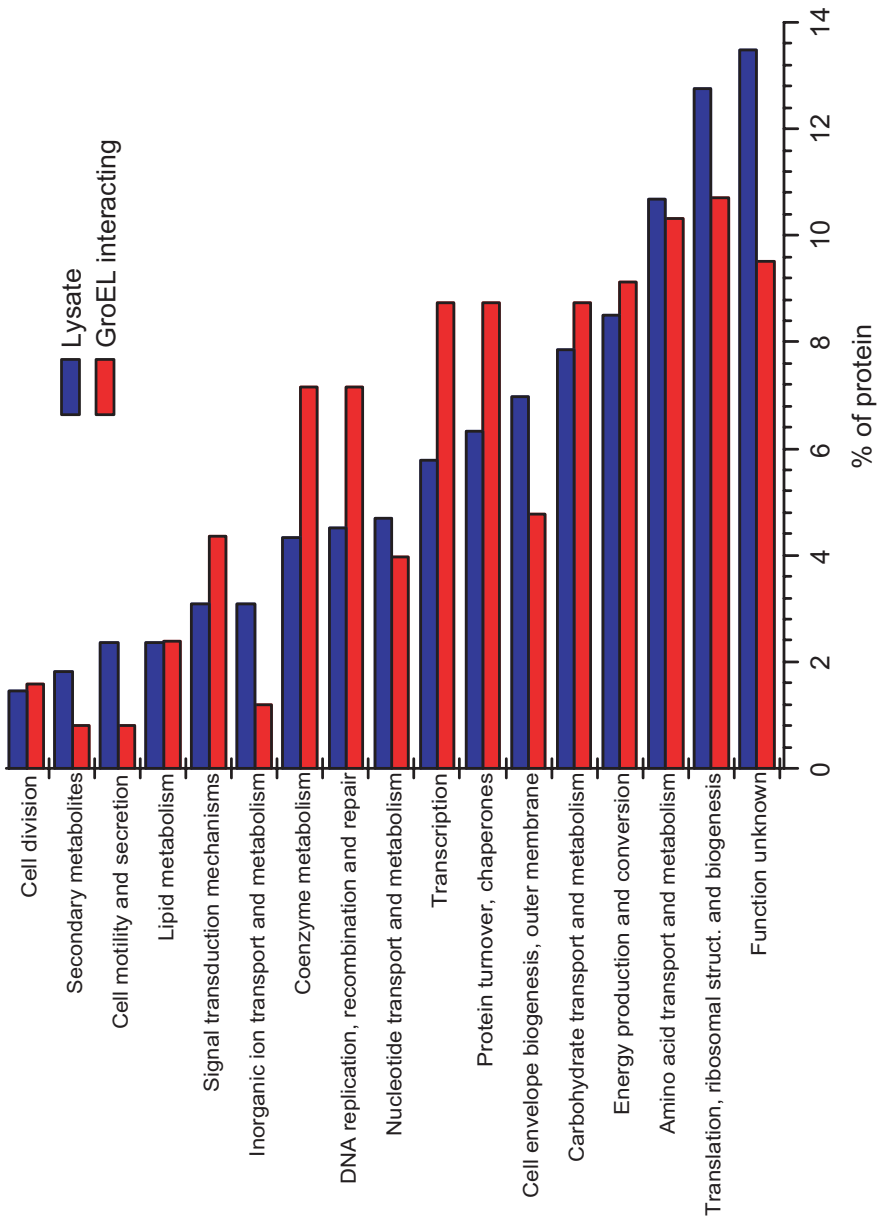


Figure 17:Distribution of functional categories among GroEL interacting proteins.

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

4.2.5. Structural categories among GroEL interacting proteins

For *E. coli* Hsp70 (DnaK) and trigger factor, exposed hydrophobic stretches have been identified as determinants in making a protein a substrate of the respective chaperones (Liberek et al., 1991). GroEL presumably acts further downstream on the protein folding pathway (Deuerling et al., 1999; Langer et al., 1992a; Teter et al., 1999), possibly recognizing more organized structural features exposed in substrate proteins. Depending on how far along a protein is on the folding pathway when it interacts with GroEL, fold motives present in its final fold might also be recognized by GroEL.

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. Proteins were queried against two databases: SCOP database (Structural Classification Of Proteins) (Lo Conte et al., 2002) and CATH database (Class, Architecture, Topology, Homologous superfamily) (Orengo et al., 1997). SCOP and CATH both organize proteins in hierarchies, its major difference lies in the fact that CATH uses an evolutionary approach for classification, while SCOP uses more structurally based 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 (data not shown). For reasons of simplicity, the analysis here is therefore limited to the SCOP data base. Since both, experimental lysate and the whole *E. coli* proteome delivered virtually identical fold distributions; further analysis was limited to the lysate data only. Restriction to this data set omits methodological biases in protein identification, since both GroEL interacting proteins and the experimentally determined lysate proteins were determined by the same mass spectrometric methods in this study. It was possible to assign known fold types to 211 of 252 GroEL interacting proteins and to 815 of 1134 proteins identified in the E.coli lysate. Single polypeptide chains can fold into proteins with multiple folds. All identified folds of a protein are considered individually.

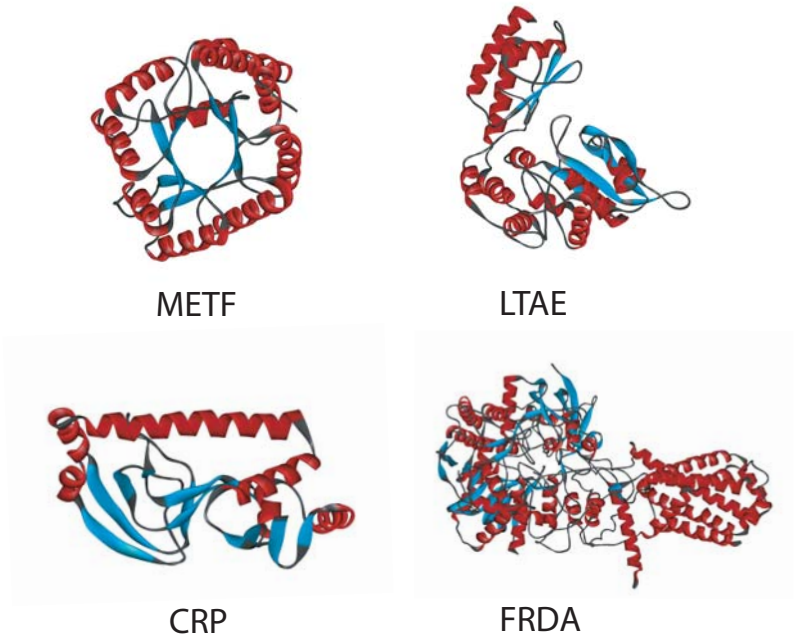


Figure 18: Fold types enriched on GroEL

Ribbon diagrams of example proteins for enriched fold types on GroEL. METF (PDB: 1B5T) is an example of the TIM barrel fold (c.1), CRP (PDB: 1CGP) is an example of DNA/RNA-binding 3-helical bundle (a.4), LTAE (PDB: 1M6S) is an example of PLP-dependent transferases (c.67), and FRDA (PDB: 1L0V) is an example of FAD/NAD(P)-binding domain (c.3). Structures were edited with ViewerPro software.

4.2.6. Analyzed fold types

Structural analysis was limited to the ten most abundant SCOP fold types identified in this study for reasons of statistical relevance. These are: TIM β/α-barrel (c.1); DNA/RNA-binding 3-helical bundle (a.4); P-loop containing nucleotide triphosphate hydrolases (c.37); PLP-dependent transferases (c.67); NAD(P)-binding Rossmann-fold domains (c.2); FAD/NAD(P)-binding domain(c.3); flavodoxin-like (c.23); ferredoxin-like (d.58); thioredoxin fold (c.47) and S-adenosyl-L-methionine-dependent methyl transferases (c.66). In general, more complex folds are more likely to be chaperone dependent for folding than simple, small folds. Complex folds are therefore also more likely to be found on GroEL.

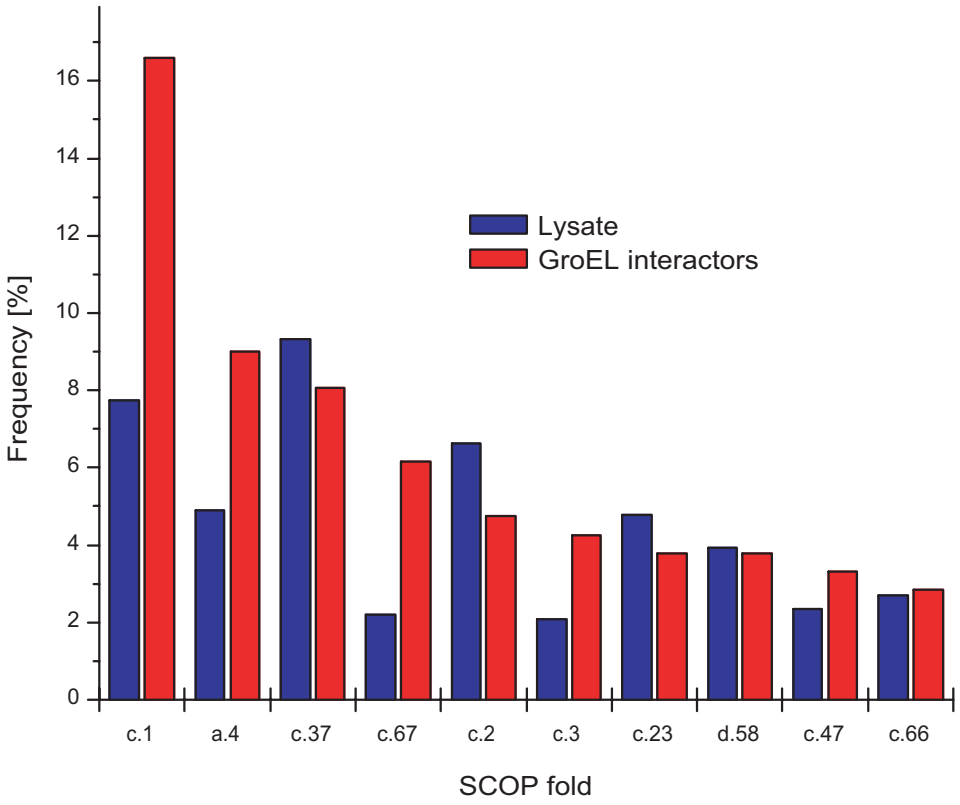


Figure 19: Fold types on GroEL and in the *E. coli* cytosol

Bar graph of the occurrence of distinct fold types in the *E. coli* lysate (blue bars) and among GroEL interacting proteins (red bars). Analyzed fold types are: TIM β/α-barrel (c.1); DNA/RNA-binding 3-helical bundle (a.4); P-loop containing nucleotide triphosphate hydrolases (c.37); PLP-dependent transferases (c.67); NAD(P)-binding Rossmann-fold domains (c.2); FAD/NAD(P)-binding domain(c.3); flavodoxin-like (c.23); ferredoxin-like (d.58); thioredoxin fold (c.47) and S-adenosyl-L-methionine-dependent methyl transferases (c.66).

Four fold types appear to be enriched among GroEL interacting proteins compared with the lysate: the TIM barrel fold (c.1), DNA/RNA-binding 3-helical bundle (a.4), PLP-dependent transferases (c.67), and FAD/NAD(P)-binding domain (c.3) (Figure 18). Fold type a.4 is rather small. It consists of three bundled or partly opened α-helices (SCOP) and was mostly detected as an additional domain in substrate proteins. It is not very likely that this fold *per se* makes a protein dependent on GroEL for efficient folding, but rather serves as a factor which increases the structural complexity of individual proteins, making it more

likely for these candidates to interact with GroEL. PLP-dependent transferase superfamily members (c.67) have a basic structure of three layers as a mixed beta-sheet of seven strands where strand seven is antiparallel to the rest (SCOP). This fold is rather complex, making it a good candidate for GroEL interaction. Fold c.3, the FAD/NAD(P)-binding domain also has a complex setup. It consists of three layers, with a central parallel beta-sheet of five strands and an antiparallel beta-sheet of three strands on top (SCOP) (Figure 18).

4.2.7. The TIM barrel fold

Most significantly, GroEL interacting proteins were enriched in the TIM barrel fold (c.1) (Figure 19). It contains a closed parallel beta-sheet barrel with eight-fold symmetry. This fold is one of the most common domain structures in *E. coli* proteins; accordingly, it is shared by 7.7% of all proteins with an identified structural homology in the lysate (63 out of 815 proteins with known fold type). The set of GroEL interacting proteins, however, contains 16.6% proteins with strong homology to the TIM barrel fold (35 out of 211) (Figure 19). All of the TIM barrel representatives interacting with GroEL exhibit molecular masses between 23 and 54 kDa and are thus likely to fold in a *cis* reaction inside the GroEL/GroES cavity. Restriction of the fold analysis to just proteins in the size range of 15 – 60 kDa in the *E. coli* lysate did not change the ratio of TIM barrels in the lysate to the ones identified in GroEL interacting proteins (data not shown). Hence, enrichment in the TIM β/α fold is not merely a consequence of size limitations among GroEL interacting proteins.

4.2.8. Proteins enriched on GroEL

Given an average folding time for GroEL interacting proteins of ~60 seconds (Ewalt et al., 1997) and a doubling time for *E. coli* cells of 30-35 min under the experimental conditions used in this study, all protein molecules of a kind would fold via GroEL, when ~3% of the specific protein is associated with GroEL at any given time.

Relative enrichment factors (REFs) determined by SILAC (Figure 13) provided values for the average affinity of individual proteins to GroEL during folding. REF detection involved manual inspection of mass spectrometry data.

However, not every protein was accessible to REF measurement, since peaks from lysate and complex preparations both had to be quantifiably detectable in the same spectrum.

REFs are in first instance simply arbitrary values equivalent to the direct ratio of observed peak areas of individual peptides from lysate and GroEL complex samples. In order to obtain optimal peak ratios, lysate samples were mixed with GroEL complex preparations 1:50 prior to mass spectrometry, yielding REF values from <0.01 to >100. GroEL is in large excess in the complex preparations, so the GroEL concentration could not be directly used to internally normalize the two samples as a result of overloading the mass spectrometer with GroEL peptides. However, based on quantification of GroEL levels by immuno-blotting (data not shown), it was found that GroEL is ~30-fold more concentrated in GroEL/GroES/substrate samples than in cell lysate samples. Considering the applied 1:50 dilution, an effective 1500 fold dilution of the cell lysate sample was obtained, when normalized to GroEL levels. A REF of 50, for a certain protein for example, thus indicates that ~3% (= 50/1500) of the total cytosolic amount of this protein is associated with GroEL. Hence, a REF of >50 is consistent with the calculations above, that virtually all of a specific protein is folded in a GroEL mediated manner.

Repetition of the fold type analysis with a set enrichment factor cut-off revealed preferences of certain SCOP classes to fold with GroEL (Figure 20). Strikingly, the frequency of TIM barrel proteins among GroEL interacting proteins increased to more than 25% of all identified proteins (15 of 58 proteins) when an artificial cutoff was set to a REF of 50. Further lowering the REF cut off to 10 did not decrease TIM barrel dominance among GroEL binding proteins (23 of 92 proteins, 25%).

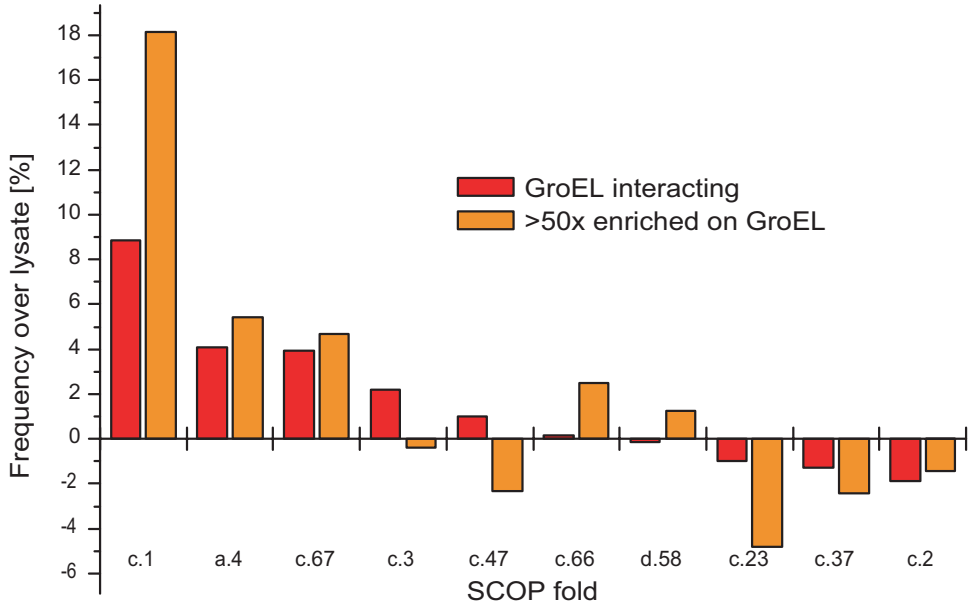


Figure 20: Enriched fold types in the GroEL substrate set

Bar graph with a gain- and- loss display of enriched and depleted fold types of GroEL interacting proteins (red bars) and GroEL interacting proteins with an enrichment factor >50 (orange bars), in relation to an *E. coli* lysate. Analyzed fold types are as in Figure 19.

4.2.9. Quantification of proteins on GroEL

In order to assess the contribution of GroEL to protein folding in *E. coli* in general and to measure the quantitative distribution of SCOP fold classes on GroEL within the total of GroEL interacting proteins, it was necessary to obtain information on protein concentrations in the *E. coli* lysate as well as in GroEL/GroES/substrate complexes. Mass spectrometry classically does not provide quantitative information; however, recent approaches allow generation of quantitative data, with certain limitations.

The protein abundance index (PAI) is defined as the number of observed peptides divided by the number of observable peptides *per* protein (Rappsilber et al., 2002).

$$PAI = \frac{N_{observed\ peptides}}{N_{observable\ peptides}} \qquad (2)$$

PAI shows a linear relationship with the logarithm of protein concentration. It was converted to exponentially modified PAI (EMPAI) for absolute quantification of proteins in a given sample (Y. Ishihama et al., in press).

$$EMPAI = 10^{PAI} - 1 \qquad (3)$$

PAI values are not only correlated to the abundance of a protein but are also dependent on its specific response to the mass spectrometry methodology. This varies according to digestion efficiency, peptide solubility, extraction, ionization and fragmentation and thereby is prone to error. Values for single proteins were hence not used for comparative quantification but rather quantitative data for groups of identified proteins, like GroEL interacting proteins, fold types among them or proteins with high or low enrichment factor.

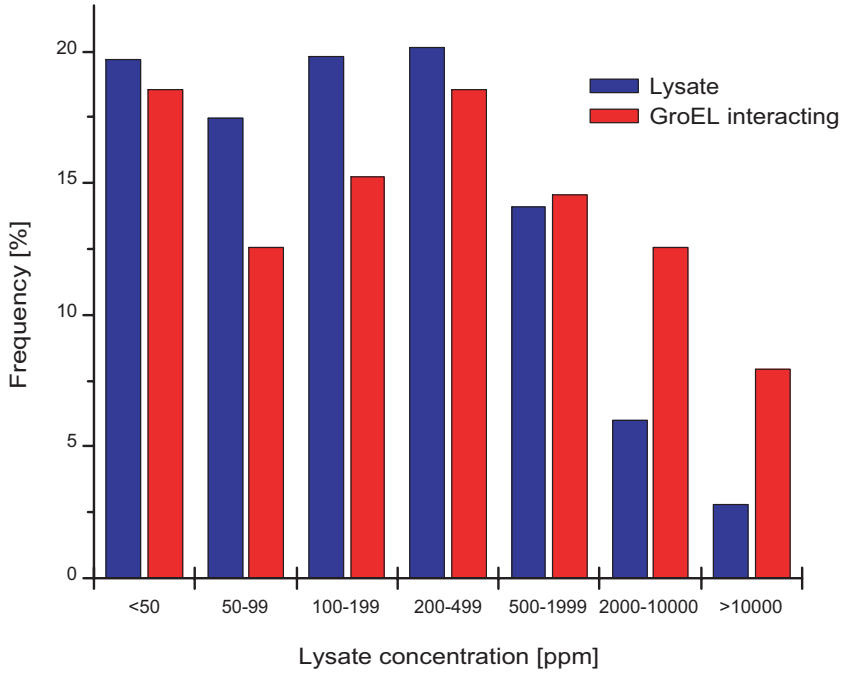


Figure 21: Distribution of cellular concentrations of *E. coli* lysate proteins and GroEL substrates

Bar graph showing the concentration distribution of *E. coli* lysate protiens (blue bars) and GroEL interacting proteins (red bars). Cellular concentrations as estimated by emPAI value expressed in molar ppm.

Protein concentrations in lysate proteins ranged from below 10 parts per million (ppm) to more than 50,000 ppm. The concentration distribution of lysate proteins is shown in Figure 21. When the lysate distribution for all GroEL interacting proteins was analyzed, no significant differences could be observed. The slight bias towards proteins of high abundance among GroEL interacting proteins is likely to be due to methodological constraints. GroEL and GroES were present in GroEL substrate samples with at least seven fold excess over the sum of all identified substrate proteins, given all GroES heptramer bound a GroEL tetradecamer with one substrate molecule in *cis* and one in *trans*. This excess in peptides derived from the chaperones possibly shields detection of rare peptides, thereby shifting the identified substrate protein spectrum to higher abundant proteins. Nevertheless, the similarity of concentration distributions in those two samples confirms identification by mass spectrometry to a satisfying degree.

Fold analysis with respect to EMPAI values determined for GroEL interacting proteins did not lead to conclusive results. TIM barrel proteins occupy 29% of the GroEL capacity according to this analysis (data not shown), more than any of the other SCOP fold classes.

However, 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. This can be understood by making the conceptual consideration that only folding intermediates serve as GroEL substrates, but not the folded protein. Proteins sharing a similar fold can have highly divergent folding pathways (Ferguson et al., 1999). One example for a GroEL independent TIM barrel is the protein enolase (ENO). It is a very robustly folding protein in the absence of any chaperones (Figure 29). Although many other *E. coli* TIM barrels proteins fold in a GroEL independent manner, they are not found among GroEL interacting polypeptides at all.

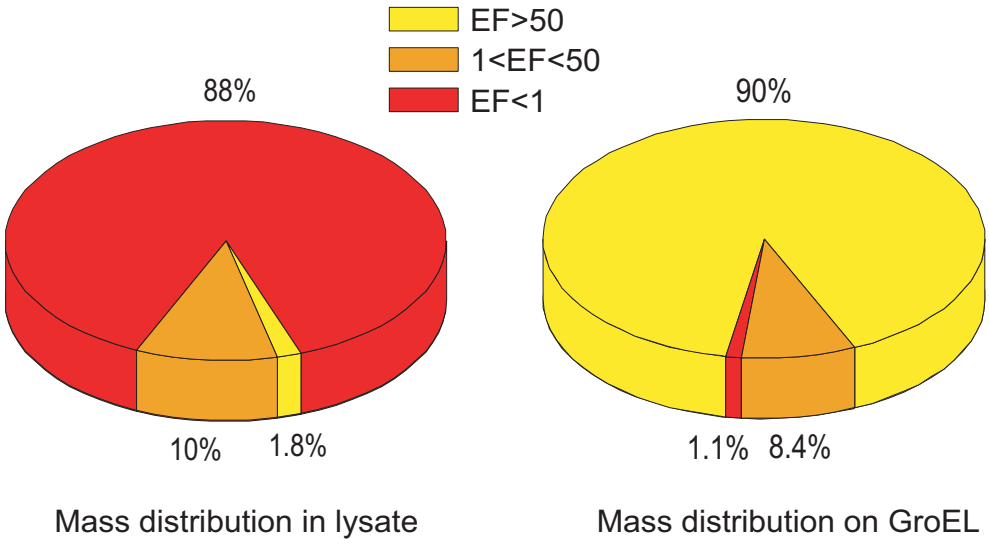


Figure 22: Mass distributions of GroEL interacting proteins

Proteins with enrichement factors >50 (yellow) contribute to 90% of the protein mass on GroEL. When only considering GroEL interacting proteins in an *E. coli* lysate, only 1.8% of those proteins by mass are proteins highly enriched on GroEL. Proteins with low enrichment (red) make up only 1.1% of protein mass interacting with GroEL, whereas their relative amount (only GroEL interacting proteins considered) in an *E. coli* lysate reaches 88%.

Specificity of substrate interaction with GroEL and discrimination for substrate proteins is already indicated by the finding that substrate proteins have diverse enrichment factors on GroEL. This finding is confirmed by analysis of the mass distribution of proteins on GroEL. EMPAI values of GroEL interacting proteins revealed that more than 90% of the GroEL capacity is occupied by proteins with an enrichment factor larger than 50 whereas only 1% of GroEL interacting proteins by mass have an enrichment factor below 1 (Figure 22). These findings are based on 171 out of 252 GroEL interacting proteins. Actual polypeptide distribution on GroEL is probably even more biased towards low abundant proteins, since enrichment factors could not be determined for the remaining 81 proteins. They were not identified in the *E. coli* lysate, due to their low cellular concentration. Identification on GroEL, however already indicates high enrichment for those proteins.

Clearly, GroEL discriminates among substrate proteins. It is a striking coincidence that proteins of low abundance are among the most frequently found GroEL interactors, raising interesting questions about the role of GroEL capacity in protein evolution. This aspect will be discussed later in this work. The mode of discrimination was subject of further analysis. Experimental goals were to investigate in what respect different affinities of proteins to GroEL govern their interaction with the chaperone and to what extent the other major chaperone machinery, the DnaK/DnaJ/GrpE system, is involved in substrate selection and folding by GroEL substrates. Further, the effects *in vivo* depletion of GroEL and GroES has on global cellular metabolism and on individual GroEL interacting proteins was investigated. *In vitro* refolding experiments helped to understand whether the enrichment factors observed *in vivo* are reflected by the necessity of a protein to fold via GroEL to obtain its native state *in vitro*. Co-expression of chaperones and substrate proteins in *E. coli* cells revealed different levels of chaperone dependence for individual substrates *in vivo*.

In an effort to address these questions, a number of genes coding for GroEL substrates were cloned, expressed and purified. For many of these proteins, antibodies were generated and many are available in various expression vectors (Table 6).

Name	Function	MW	RF	CE	DP
ADD	Adenosine deaminase	36397		x	
ALR2	Alanine racemase, catabolic	38844		x	
CRP	Catabolite gene activator	23640		x	
DAPA	Dihydrodipicolinate synthase	31270	x	x	x
DCEA	Glutamate decarboxylase alpha	52685	x	x	
END4	Endonuclease IV	31479		x	
ENO	Enolase (EC 4.2.1.11)	45523	x	x	x
GATD	Galactitol-1-phosphate	37390	x	x	x
GATY	Tagatose-1,6-bisphosphate aldolase	30812		x	x
HEM2	Delta-aminolevulinic acid dehydratase	35493		x	
LLDD	L-lactate dehydrogenase (Cytochrome)	42728		x	
LTAE	Low-specificity L-threonine aldolase	36494		x	
METF	5,10-methylenetetrahydrofolate red.ase	33102	x	x	
METK	S-adenosylmethionine synthetase	41820	x	x	x
NANA	N-acetylneuraminate lyase	32462		x	
SYT	Threonyl-tRNA synthetase	74014	x	x	x
TDH	L-threonine 3-dehydrogenase	37239		x	x
TYPH	Thymidine phosphorylase	47207		x	
XYLA	Xylose isomerase	49742		x	x
YAJO	Hypothetical oxidoreductase yajO	36420		x	
YHBJ	Hypothetical UPF0042 protein yhbJ.	32492		x	

Table 6: GroEL interacting proteins analysed individually in this study

Names are SwissProt entries, MW in [Da], RF: Refolding experiments, CE: Chaperone-substrate co-expression experiments, DP: GroE depletion experiments.

4.3. *In vitro* refolding of GroEL substrates

Observed enrichement factors by quantitative mass spectrometric analysis revealed a gradation in GroEL dependence of identified GroEL substrates. To verify these findings, and to elucidate the role the other major ATP driven chaperone system in the cell plays, consisting of proteins DnaK, DnaJ and GrpE, the GroEL dependence of substrates was studied with *in vitro* refolding experiments.

Several identified substrate proteins were cloned, expressed and purified (Table 6 and Materials and Methods section). *In vitro* refolding was followed by measurement of enzymatic activity over time, following dilution from chemical denaturant into buffer solutions containing various combinations of chaperones and nucleotides. 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 refolding reactions were subsequently assayed for enzymatic activity after completion of refolding series (up to 80 min). 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 unfolded polypeptide). Refolding yields are expressed as ratio of regained enzymatic activity relative to activity of the native enzyme.

Results from refolding experiments allowed classification of GroEL interacting proteins into three classes according to their chaperone requirements.

4.3.1. Class I: Chaperone-independent refolding

Enolase (ENO, a homodimeric protein of 45.5 kDa subunits) was found to fold spontaneously without addition of chaperones (Figure 23). The protein was denatured in 6 M GdnHCl and diluted 100-fold into buffer A (materials and methods). Upon refolding, ENO reached roughly 55% of its initial activity with a $t_{½}$ of about 30 s at 37 °C. Folding was monitored by measuring the enolase dehydration activity of 2-phosphoglycerate to phosphoenolpyruvate at the

indicated times. The 55% yield obtained by spontaneous refolding could be increased to ~80 – 95% by addition of GroEL alone or GroEL with GroES. When ATP was omitted from a refolding reaction containing GroEL, folding did not occur, demonstrating efficient binding of unfolded enolase to GroEL, supporting the finding that ENO interacts with GroEL *in vivo*, even though the chaperonins are not needed for efficient refolding. The DnaK/DnaJ/GrpE chaperone system was also able to increase the yield of ENO refolding to ~90% of native control at similar apparent rates to spontaneous and GroEL-mediated folding.

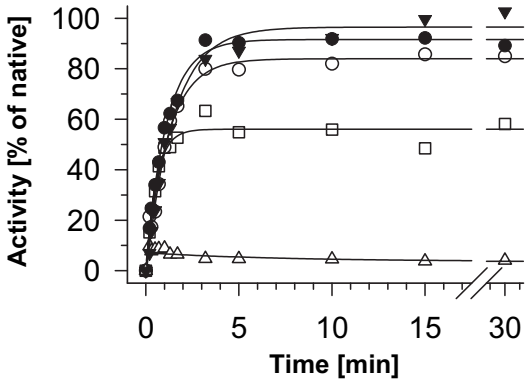


Figure 23:*In vitro* refolding of enolase (ENO)

Chaperones can help increase the yield of ENO during refolding. Denatured ENO was diluted 100-fold into buffer. Refolding reactions were stopped at indicated time points and enzymatic activity was determined. Protein concentrations used are given in chapter 1.7. □: Spontaneous refolding; ○: GroEL only; ●: GroEL and GroES; △: GroEL, no ATP; ▼: DnaK/DnaJ/GrpE.

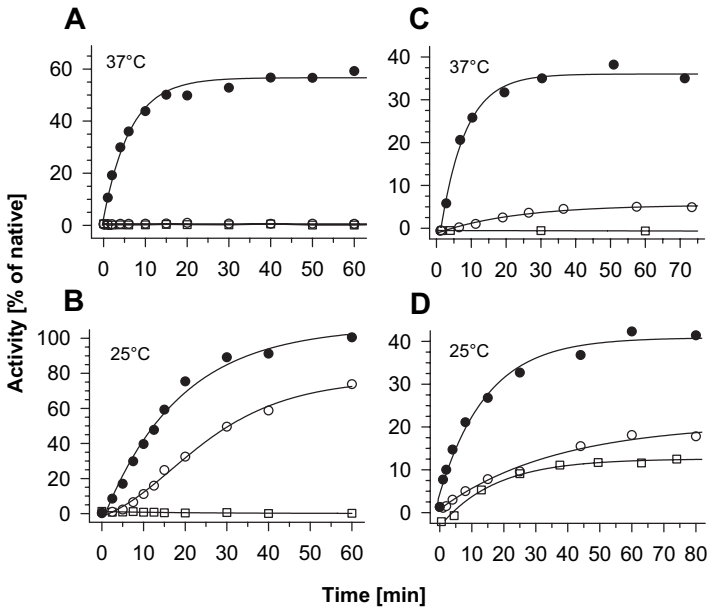


Figure 24:*In vitro* refolding of DCEA and GATD I

Temperature dependent increase in GroEL mediated refolding yield. Denatured DCEA (A, B) or GATD (C, D) was diluted 100-fold into buffer. Refolding reactions were stopped at indicated time points and enzymatic activity was determined. Protein concentrations used are given in chapter 1.7. □: Spontaneous refolding; ○: GroEL only; ●: GroEL and GroES

4.3.2. Class II: Chaperone-dependent refolding

Some tested GroEL interacting proteins were 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) and galactitol-1-phosphate 5-dehydrogenase (GATD; 37.4 kDa subunits, a putative homotetramer) 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 (Figure 24). Switching to less stringent conditions by decreasing the temperature from 37 °C to 25 °C allowed GroEL to fold the proteins DCEA and GATD without the GroES cofactor, although to lower yields when compared to reactions with the full chaperonin system (Figure 24). GATD even showed some

spontaneous refolding without chaperones at 25 °C, albeit to a final yield of only about 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 (Figure 25) and a combination of the GroEL and DnaK systems showed a noticeable additive effect on DCEA folding.

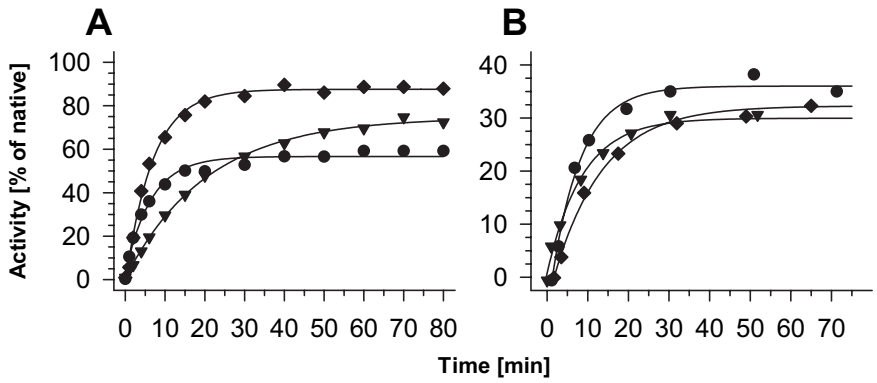


Figure 25:*In vitro* refolding of DCEA and GATD II

Both chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE can mediate refolding of DCEA (A) and GATD (B) with comparable rates and to comparable yields. Denatured DCEA or GATD was diluted 100-fold into buffer. Refolding reactions were stopped at indicated time points and enzymatic activity was determined. Protein concentrations used are given in chapter 1.7. ●: GroEL/GroES; ▼: DnaK/DnaJ/GrpE; ◆: DnaK/DnaJ/GrpE and GroEL/GroES

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 26). However, the DnaK system was much more efficient in mediating the folding of SYT, leading to a final yield of ~70% active protein.

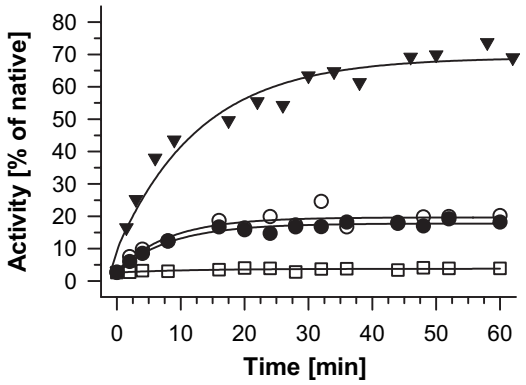


Figure 26:*In vitro* refolding of SYT

GroEL alone and with GroES can mediate SYT refolding. High yields are only obtained when refolded with chaperones DnaK/DnaJ/GrpE present. Denatured SYT was diluted 100-fold into buffer. Refolding reactions were stopped at indicated time points and enzymatic activity was determined. Protein concentrations used are given in chapter 1.7. □: Spontaneous refolding; ○: GroEL only; ●: GroEL and GroES; ▼: DnaK/DnaJ/GrpE.

In contrast to recent studies describing an efficient *trans* mechanism for folding of large proteins by GroEL (Chaudhuri et al., 2001; Farr et al., 2003), SYT was found to reach only relatively small refolding yields with the help of GroEL. Neither was there an observable difference between refolding temperatures of 25 °C and 37 °C.

It appears that DnaK and GroEL can efficiently re-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)

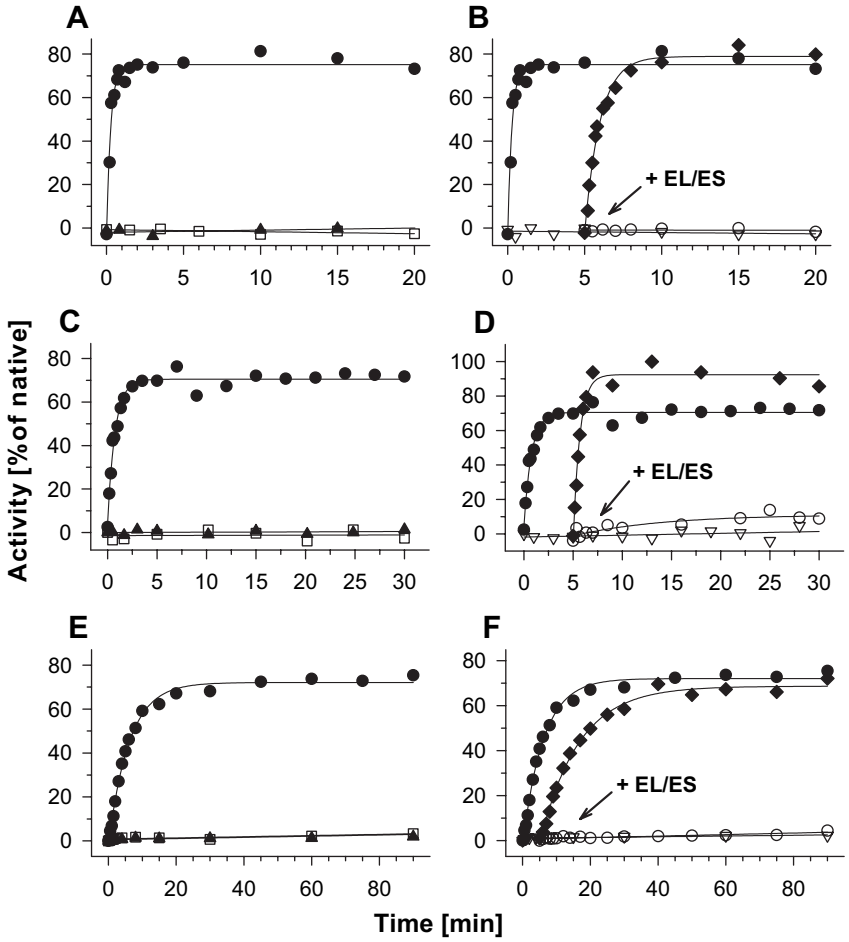


Figure 27:*In vitro* refolding of METF, METK and DAPA

GroEL and GroES are stringently required for refolding. DnaK/DnaJ/GrpE cannot fold METF (A, B), METK (C, D) or DAPA (E, F), but can keep the proteins in a folding competent state, allowing folding upon subsequent transfer to GroEL. Denatured proteins were diluted 100-fold into buffer. Refolding reactions were stopped at indicated time points and enzymatic activity was determined. Protein concentrations used are given in chapter 1.7. □: Spontaneous refolding; ●: GroEL and GroES; ▲: GroEL only; ○: Spontaneous and GroEL/GroES addition after 5 minutes incubation;▽: DnaK/DnaJ/GrpE; ◆: DnaK/DnaJ/GrpE and GroEL/GroES after 5 minutes incubation.

4.3.3. Class III: GroEL-dependent refolding

For proteins METF, METK and DAPA, only the complete GroEL/GroES system is able to mediate efficient refolding at 37°C (Figure 27). Intriguingly, METF ($t_{½}$ ~10 s) and METK ($t_{½}$ ~30 s), were folded by GroEL/GroES at relatively

fast rates, as compared to DAPA ($t_{½}$ ~4 min) and the model substrates mitochondrial rhodanese (a monomer of 33 kDa; $t_{½}$ ~5 min) and bacterial rubisco (a homodimer of 50 kDa subunits; t½ ~2.5 min, (Brinker et al., 2001). This suggests that the contribution of GroEL to protein folding in *E. coli* may be greater than what has previously been estimated with heterologous substrates (Lorimer, 1996) due to observed higher turnover rates.

Importantly, the DnaK system could not mediate the refolding of METF, METK and DAPA, consistent with their stringent GroEL/GroES dependence in depletion experiments. However, DnaK could bind and thereby stabilize aggregation-prone, nonnative forms of these substrates and efficiently transfer them to GroEL for subsequent folding (Figure 27). The successive action of DnaK and GroEL in protein folding was first observed with mitochondrial rhodanese (Langer et al., 1992a). This transfer may 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 (Figure 27), 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) (Mogk et al., 1999). Thus, DnaK appears to function as a substrate reservoir for GroEL, facilitating the efficient capture of nascent chains and stress denatured proteins. 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). This finding, together with the 10 other essential proteins identified as stringent GroEL substrates by mass spectrometry (chapter 4.2.3. and table 5) explains the essential nature of GroEL and GroES.

4.3.4. Substrate selection by GroEL

In vitro competition experiments for binding of model substrates to GroEL revealed a preference for GroEL to bind to stringent GroEL substrates, attributed to class III. Mixtures of denatured class I protein ENO with denatured class III proteins DAPA or METK and of denatured class II proteins DCEA or GATD with denatured DAPA or METK were diluted 100 fold into buffer containing GroEL at stoichiometric amounts. Subsequent separation by size exclusion chromatography and analysis by Westen Blotting of GroEL containing fractions revealed that GroEL preferentially binds to DAPA and METK even when class I or class II proteins are present in four fold excess of over the class III substrate proteins (Figure 28). This indicates that GroEL specifically recognizes folding intermediates of its stringent substrates and discriminates against proteins which can also fold either spontaneously or with other chaperone systems even in a direct competition situation. The affinity of GroEL to DAPA and METK folding intermediates is considerably higher than to folding intermediates of the tested class I and class II proteins, since even a four fold excess of these proteins did not lead to a significant shift of substrate – chaperone interaction towards these less stringent GroEL substrates.

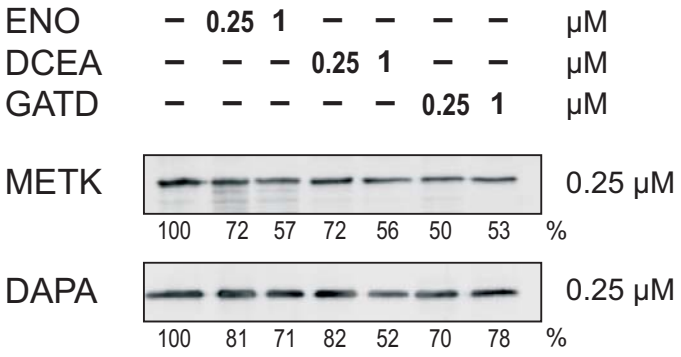


Figure 28: Competition of class I and class II proteins with class III proteins.

Class III proteins (METK, DAPA) out-compete class I (ENO) and class II (DCEA, GATD) for GroEL binding. Denatured proteins were mixed and diluted into buffer containing 0.25 µM GroEL at 37°C to the final concentrations indicated. GroEL complexes were isolated by size exclusion chromatography and analyzed by immunoblotting for METK and DAPA.

4.4. Chaperone-substrate co-expression

To confirm mass spectrometric data and *in vitro* refolding experiments and to investigate the chaperone dependence of individual GroEL substrates *in vivo*, chaperone-substrate co-expression experiments were undertaken. Identified GroEL interacting proteins were overexpressed in *E. coli* cells together with either wild-type or elevated levels of chaperonins.

Overexpression of substrate proteins was designed to exceed the capacity of the available chaperones in wild type cells so that only a limited amount of protein was produced in soluble form. Elevating the levels of GroEL/GroES about 5 fold allowed specific assessment of chaperonin contribution to folding. Co-expressions were performed in BL21(DE3) Gold cells, deficient in the LON protease, the major protease for unfolded proteins in *E. coli* (Goldberg et al., 1994). Hence, all protein synthesized either remains soluble or it aggregates; and degradation can be neglected, facilitating a comparison of different substrate proteins. Different overall levels of substrate proteins (expressed from the same promoter) are therefore due to different codon usage, mRNA stability and residual degradation by other cellular proteases. Solubility alone however is not a sufficient criterion for functionality of a given protein and different cellular factors including other chaperone systems or co-factors might be required by these proteins for folding to the native state. Nevertheless, differences in solubility observed with different chaperonin backgrounds can allow an estimate of the degree of GroEL dependence for the tested proteins.

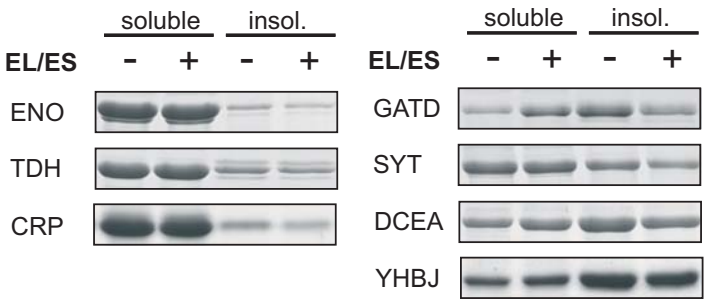


Figure 29:Solubility of GroEL substrates upon co-expression in *E. coli* with chaperonins

Coomassie blue strained SDS PAGE gels. *E. coli* cells were grown at 37 °C to exponential phase in LB-medium containing 0.4% glucose. Chaperones were induced (+) from a pBAD promoter with 0.4% arabinose or further repressed (-) with glucose. Substrates were subsequently induced at 37 °C 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.

4.4.1. GroEL-independent folding

CRP, ENO and TDH were highly soluble when overexpressed with wild type chaperonin levels and under co-expression conditions, indicating GroEL independence for folding *in vivo* (Figure 29). Comparison with in data from *in vitro* refolding allowed grouping of these proteins to class I. Proteins DCEA and GATD, SYT and YHBJ were expressed to a lower level, and roughly 50% were insoluble when chaperones were not co-expressed. Overexpression of GroEL and GroES did not have strong effects on solubility. For these proteins, the chaperonins play a beneficial role for folding. However, they can also fold by alternative means, when no chaperones are co-expressed and GroEL capacity is limited. These findings confirmed the data on chaperone dependence of class II proteins from *in vitro* refolding experiments. The proteins DCEA, GATD, SYT and YHBJ are therefore grouped into this GroEL substrate class.

4.4.2. GroEL dependent folding

Proteins ADD, DAPA, END4, HEM2, LTAE, METF, METK, NANA, TYPH, XYLA and YAJO are not as soluble as proteins folding in a GroEL independent manner. Overexpression of GroEL and GroES can lead to a

significant increase of solubility, indicating a beneficial role of GroEL and GroES on folding of these proteins. The limited GroEL capacity, given when the substrates alone are overexpressed without chaperones, is now not sufficient to keep newly synthesized proteins in a soluble form.

Proteins ALR2, GATY and LLDD show an even more drastic behavior: When overexpressed, they are virtually insoluble with wild type chaperonin levels, and only achieve about 50% solubility upon GroEL and GroES overexpression (Figure 30). GATY and GATZ were expressed together for this set of experiments, since published data suggest that GATY is stabilized and its activity enhanced by interaction with GATZ (Brinkkötter et al., 2002).

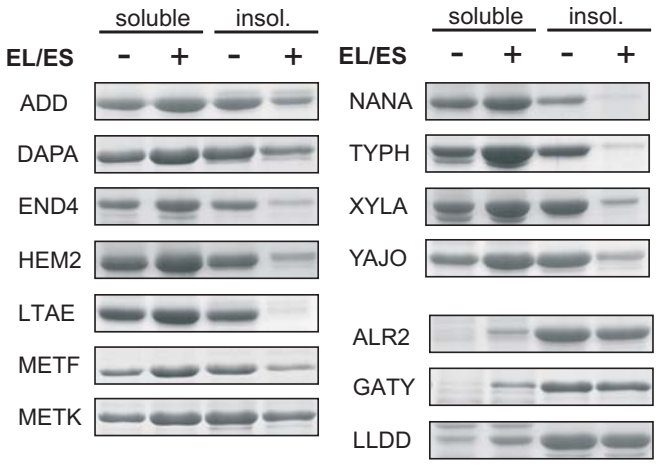


Figure 30:Solubility of GroEL substrates upon co-expression in *E. coli* with chaperonins

Coomassie blue strained SDS PAGE gels. *E. coli* cells were grown at 37 °C to exponential phase in LB-medium containing 0.4% 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.

4.4.3. Correlation with proteomic and refolding data

Practically all proteins analyzed with enrichment factors of 50 or greater show stringent dependence on GroEL for folding. Overexpression of GroEL and GroES increase substrate solubility significantly. The analyzed proteins with predicted GroEL dependence were predominantly TIM barrels and all of them showed the expected behavior for solubility and insolubility when co-expressed with GroEL and GroES. Some proteins from this group were tested for GroEL/ES dependence in refolding experiments. They all show stringent dependence for GroEL and GroES and fall into class III. Class I proteins witch low enrichment factors and GroEL-independent re-folding behavior are soluble to a high degree. Class II proteins show a higher tendency to aggregate than class I proteins. However, their solubility only mildly increases when GroEL is overepxressed, since they can also fold by the DnaK/DnaJ/GrpE system and probably only a relatively small fraction reaches GroEL *in vivo*, even when the chaperonin is overexpressed.

4.5. GroEL/GroES depletion

The co-expression experiments gave some first information on chaperonin usage of distinct proteins. It is however arguable to which extent overexpression of substrate proteins and chaperones reflect *in vivo* conditions. An alternative approach to verify GroEL dependence for productive folding of identified model substrates *in vivo* at their physiological levels was therefore established.

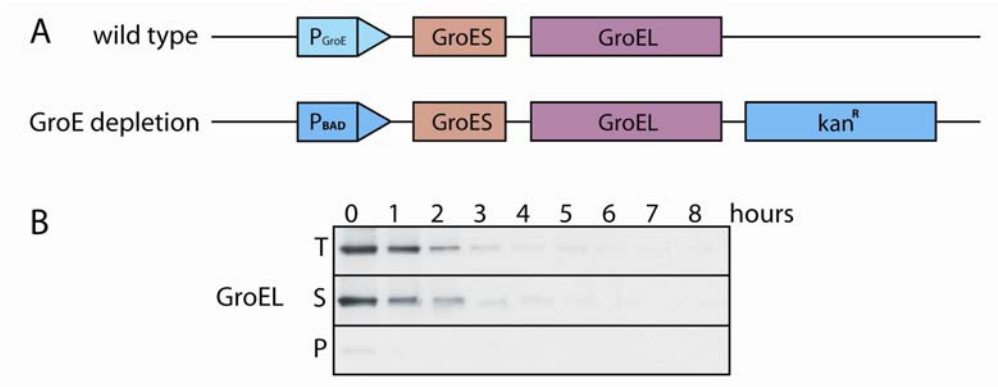


Figure 31: Experimental setup for GroEL/GroES depletion experiments

A: Schematic diagram of the chromosomal GroE region of WT *E. coli* and the GroE depletion strain. The wild type promoter was replaced with an arabinose inducible and glucose repressible P_{BAD} promoter. *groE* genes are flanked by a kanamycin resistance cassette. B: Immunoblots of total (T), soluble (S) and insoluble (P) fractions of *E. coli* cells grown under GroE depletion conditions for the indicated times. Equal amounts of sample were separated by SDS PAGE and blotted with a GroEL antibody.

4.5.1. *E. coli* GroEL depletion strain

An *E. coli* strain was kindly provided by Dr. Costa Georgopoulos (Université de Genève) in which the chromosomal GroE promoter was exchanged by the arabinose controlled araBAD (P_{BAD}) promoter. GroEL levels decreased by more than 90% within 3 hours upon a shift from arabinose- to glucose-containing growth medium (Figure 31) and after 6-8 hours cell density in liquid culture started to decrease, as cells lysed. A reduction of the cellular GroEL concentration down to 25% of the original levels is known to be tolerated without loss of cell viability (Kanemori et al., 1994). Complete shut off of GroEL and GroES expression and a drop below 25% of their original levels cannot be compensated by alternative cellular mechanisms and leads to cell death. In this study, the expression pattern of total, soluble and insoluble material of substrate proteins of interest was followed over a time course of GroEL/GroES depletion.

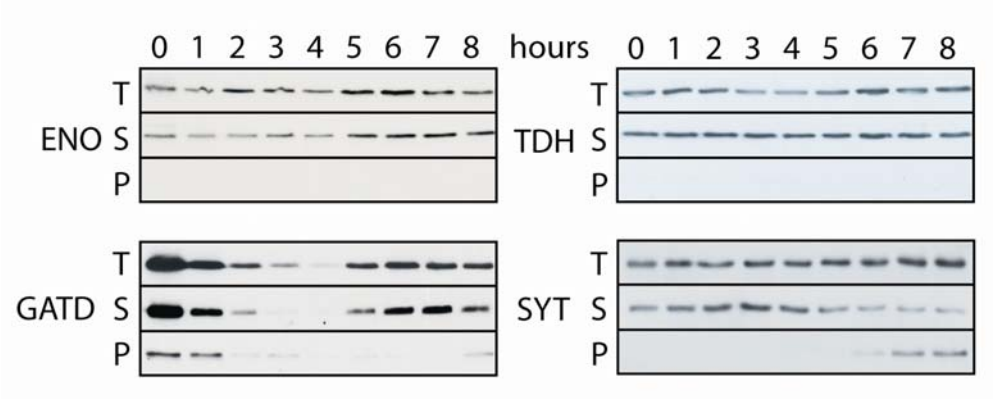


Figure 32: Solubility of GroEL substrates upon depletion of GroEL and GroES I

Immunoblots of proteins largely unaffected by GroEL/GroES depletion. Proteins ENO and TDH are unaffected by GroE depletion. SYT shows minor appearance of insoluble material towards the end of the time course. GATD expression is glucose repressed, hence the observed inconsistent band pattern. After five to six hours, the sugar regulatory mechanism on GATD is negligible, possibly due to secondary effects on regulatory proteins, and GATD still appears soluble.

4.5.2. Proteins not differentially affected by GroEL depletion

Proteins ENO and TDH remained soluble throughout GroEL/GroES depletion (Figure 32) indicating their independence of chaperonins for folding, as already shown by chaperone co-expressions and in vitro refolding experiments. Solubility of GATD was also not affected upon GroE depletion; however this protein displayed an irregular expression pattern. At the beginning and towards the end of the experiment GATD can be observed in soluble form, whereas after 3 to 4 hours of depletion, it is practically absent from the cell lysate. The shift from arabinose to glucose containing growth medium at the beginning (Nobelmann and Lengeler, 1996), and the lack of a negative regulator suppressing expression of GATD towards the end of the experiment probably account for the observed variations. For data interpretation, comparable regulatory mechanisms have to be considered for all tested substrates. SYT, a 74 kDa protein too large to be encapsulated, was expressed constitutively throughout the time course of the experiment and was only partially insoluble upon prolonged chaperonin depletion. The observed influence of GroE depletion

on protein solubility correlates very well with the above findings from co-expression experiments and *in vitro* refolding experiments.

4.5.3. GroEL-dependent proteins

Next, several proteins with high enrichment factors on GroEL were tested. All proteins from this class showed an absolute requirement for GroEL and GroES. METK maintained stable cellular levels throughout the time course of depletion, but disappeared from the soluble fraction and began to accumulate as aggregates after three hours of GroEL/GroES depletion.

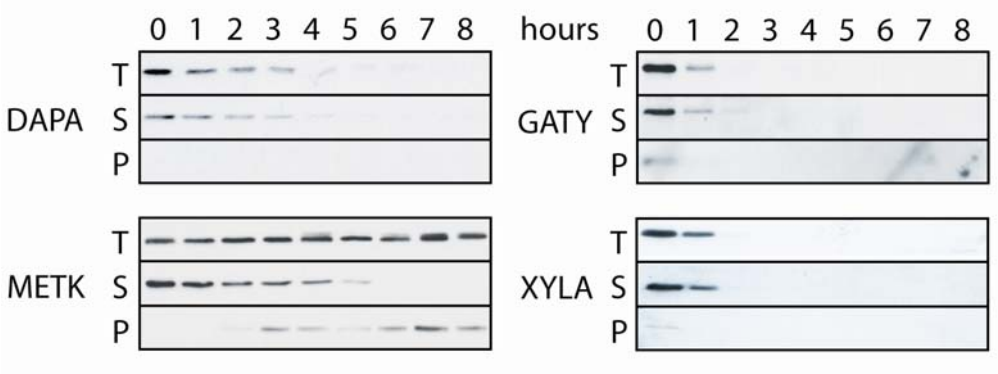


Figure 33: Solubility of GroEL substrates upon depletion of GroEL and GroES II

Immunoblots of proteins affected by GroEL and GroES depletion. Proteins DAPA and METF disappear from total and soluble fraction after two to four hours of GroE depletion. Misfolded or unfolded protein does not aggregate, but is subject to degradation. METK accumulates as aggregated material upon GroE depletion. XYLA also disappears entirely from total and soluble fraction. This can be attributed to regulatory secondary effects related with sugar supply, which overlay the effects of GroE depletion.

DAPA disappeared from the total and soluble fraction without accumulating in the insoluble fraction, suggesting that this protein is efficiently degraded when unable to fold and aggregates only upon overexpression. It has been 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.

Rapid disappearance from the cell lysate was also observed for GATY, suggesting a pronounced chaperonin dependence of this protein. In these experiments, GATY is co-expressed with GATD as the two proteins are encoded on the same operon (Nobelmann and Lengeler, 1996).

Depletion experiments were carried out in an *E. coli* MC4100 strain which is wild type for the LON protease, the major protease for unfolded proteins in *E. coli* (Goldberg et al., 1994). In contrast, the co-expression experiments mentioned above (4.3) were carried out in BL21 cells lacking LON. This difference might explain why under co-expression conditions, GATY and DAPA preferentially aggregate, whereas in depletion experiments, complete disappearance of misfolded proteins from cell lysates was observed.

XYLA also disappeared from both total and soluble fractions and does not appear as insoluble material over the time course of the depletion experimet. Tests with wild type *E. coli* MC4100 cells switched from arabinose-containing to glucose-containing medium revealed however that changes in XYLA levels are a direct effect of the sugar switch. All other proteins tested showed an expression pattern independent of sugar supply (data not shown).

The behavior of METK, DAPA and GATY in GroEL depletion experiments is consistent with their absolute GroEL/ES requirement observed in refolding experiments. This allows the validation of the established substrate class III, comprising stringent GroEL/ES dependent proteins.

4.5.4. Other effects of GroEL depletion on *E. coli* cells

Several studies (including this one) have noted that GroEL/GroES depletion or inactivation, results in an increase in the levels of numerous other proteins, including DnaK, ClpB and METE. Intriguingly, while DnaK and ClpB levels are raised ~2-4 fold, presumably to assist in disaggregation 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, which is known to require both functional METF and METH. The product of the enzymatic reaction of METF, N5-methyl-H4-folate, assists in forming a METH-B12 complex and METH is a B12 dependent methyl

transferase for METE involved in its repression. Also in part the METJ repressor protein and its co-repressor S-adenosyl methionine, the product formed by METK is responsible for METE repression (Cai et al., 1992). Therefore, 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., which were both identified as stringent GroEL substrates.

4.5.5. Co-expression of GroEL/GroES and substrates in *S. cerevisiae*

Co-expression experiments of GroEL, GroES and *E. coli* GroEL substrates in *S. cerevisiae*, a heterologous eukaryotic host, and subsequent analysis of solubility of the substrate proteins confirmed the above findings from co-expression experiment in E. coli, GroEL depletion experiments and refolding experiments.

The eukaryotic cytosol does not allow the folding of recombinantly expressed stringent GroEL substrates (Figure 25). It was shown that, while ENO was highly soluble, proteins with intermediate enrichment factors were soluble but aggregated in the absence of Ydj1, a yeast DnaJ homolog. This finding supported the consideration that those proteins are chaperone dependent, but can use the Hsp70 system for folding and do not require GroEL. On the other hand, highly enriched proteins on GroEL were virtually insoluble, and no degradation of the aggregated material was detectable. Only upon overexpression of *E. coli* GroEL and GroES together with the GroEL substrates an increase in solubility was observed (Figure 34).

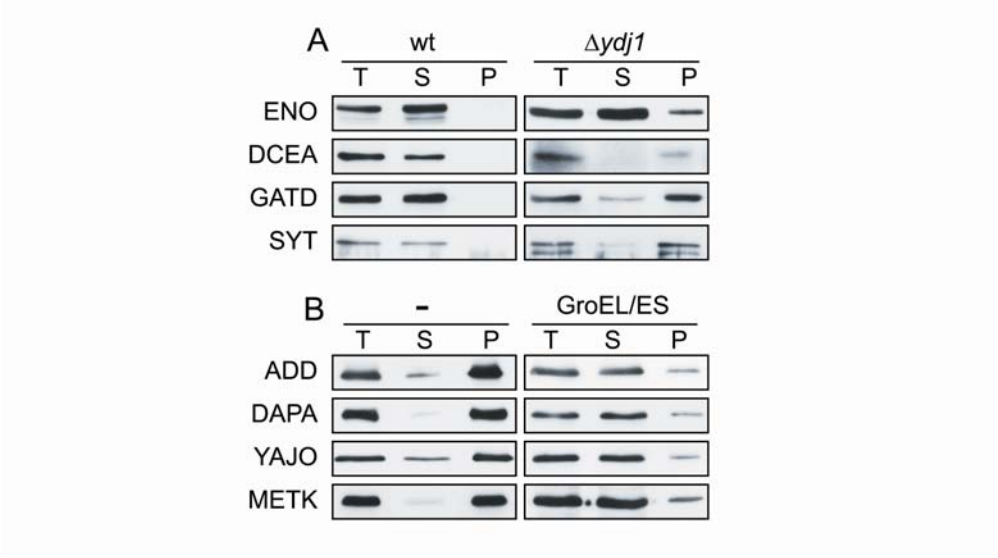


Figure 34:Coexpressions of *E. coli* GroEL, GroES and substrate proteins in yeast

Analysis of total (T), soluble (S) and insoluble (P) material after coexpression of *E. coli* substrate proteins and *E. coli* GroEL and GroES in *S. cerevisiae* by immuno-blotting. A: Solubility of weakly enriched GroEL substrates in WT and Ydj1p deficient cells. B: Solubility of highly enriched GroEL substrates in WT and GroEL and GroES overexpressing yeast cells.

5. Discussion

In this study, nearly all GroEL interacting proteins were identified. A novel mass spectrometric approach was used to quantify GroEL interacting proteins. They were analyzed with respect to functional categories and fold types. Predominantly low abundant proteins in the cell were enriched on GroEL as well as proteins with the TIM barrel fold, when compared with an *E. coli* lysate. GroEL dependency could not be attributed to distinct functional categories.

Further, *in vitro* and *in vivo* experiments allowed a categorization of chaperone dependence for selected substrates. The solubility of GroEL interacting proteins was analyzed upon co-expression with chaperones as well as in a GroEL/GroES depletion strain. *In vitro* refolding studies allowed a detailed analysis of chaperone requirements for selected substrates. In addition to GroEL and GroES usage, the role of the other main chaperone system in *E. coli*, the Hsp70 system comprising DnaK, DnaJ and GrpE proteins, was investigated in refolding experiments *in vitro*.

Based on *in vivo* and *in vitro* findings, an attempt has been made to classify GroEL substrates according to their chaperone dependency for folding and subsequently extend this classification to the complete set of GroEL interacting proteins determined by mass spectrometry. Considerations on the role of the *E. coli* chaperonin system for protein folding in the cell, on its essentiality, as well as the establishment of hypotheses on evolution and a chaperone network have become possible based on the data obtained.

5.1. Classes of GroEL substrates

Data from chaperone co-expression, depletion experiments and *in vitro* refolding allowed the categorization of GroEL interacting proteins into three classes with respect to their chaperone dependence for productive folding, as had been proposed previously (Ewalt et al., 1997).

5.1.1. Class I proteins

This class comprises proteins CRP, ENO and TDH. Class I proteins are largely chaperone-independent *in vitro* and are able to fold spontaneously. However, their refolding yield can be optimized by chaperone interaction (Figure 23). *In vivo*, class I proteins were shown to be independent of the chaperonin system. Recombinant co-expression of class I proteins with GroEL and GroES did not result in any detectable change in solubility (Figure 29). Furthermore, class I substrates remain completely soluble upon depletion of GroEL/GroES in *E. coli* (Figure 32). Identified proteins of class I represent highly abundant proteins of the *E. coli* cytosol (Figure 22). It seems plausible to assume that additional 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 have the opportunity to fold before reaching GroEL. Thus, only highly abundant proteins of class I are experimentally found to interact with GroEL in the cell. The low relative amount of the total mass of a class I protein associated with GroEL does not allow detection of low abundance class I proteins bound to GroEL *in vivo*.

5.1.2. Class II proteins

This class comprises proteins such as DCEA, GATD, SYT and YHBJ. 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 25) and *in vivo* (Figures 29. 32). Class II proteins therefore should 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 non-native states. Hence, GroES dependence during *in vitro* GroEL-mediated folding of class II proteins is variable. In some cases (Figure 24) low temperature allows GroEL to mediate class II protein folding in

the absence of GroES, with GroEL merely acting as a general binding and releasing chaperone in this case. However, this mechanism is less efficient when compared to that of DnaK. It can only promote successful folding under 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 limited. 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.

In vivo, class II proteins have a higher aggregation propensity than class I proteins. Overexpression of the chaperonin system does not lead to increased solubility of class II substrates (Figure 29). Limitation of GroEL capacity, as in depletion of GroEL (Figure 32) resulted in inconsistent behavior, either showing unperturbed solubility levels or a slight decrease in solubility, probably due to their general tendency to aggregate.

5.1.3. Class III proteins

Class III GroEL substrates constitute a group of proteins which are absolutely dependent on assistance by the chaperonin system for folding to their native state, both *in vitro* and *in vivo*. Unfolded class III substrates are highly aggregation prone and are unable to fold spontaneously. The DnaK system is able to bind unfolded class III proteins effectively and thus to 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 (Figure 27). GroEL-mediated folding of class III substrates is absolutely GroES dependent and encapsulation inside GroEL/GroES is an essential feature of chaperonin-mediated class III substrate folding (Figure 27). *In vivo* results confirmed the observed dependence of these proteins on GroEL and GroES. Co-expression of class III substrates with the complete chaperonin system enhanced their solubility in all cases (Figure 30). Three proteins, ALR2, GATY and LLDD reached detectable levels of soluble protein only upon co-expression with the full chaperonin system (Figure 30.

Depletion of GroEL/GroES leads to aggregation or degradation of endogenous class III substrates (Figure 33). Thus, no other chaperone system of *E. coli* is able to substitute for the function of GroEL in class III protein folding.

5.1.4. GroEL substrates expressed in *S. cerevisiae*

Heterologous expression of GroEL substrates in the eukaryotic cytosol, which lacks a bacterial-type chaperonin, provided a stringent system to independently test the validity of the classification of newly synthesized GroEL substrates. Class I protein ENO and class II proteins DCEA, GATD and SYT were soluble independently of co-expression of chaperones. Substantial aggregation of the class II proteins was observed in a mutant strain which lacks the yeast Hsp40 homolog Ydj1p, supporting the conclusion that class II proteins are chaperone dependent but can utilize either the Hsp70 system or GroEL/GroES for folding (Figure 34). Class III proteins ADD, DAPA, YAJO and METK were moderately expressed in different wt *S. cerevisiae* strains from galactose-inducible promoters. Remarkably, all of these proteins accumulated in the insoluble fraction, but were essentially soluble when both GroEL and GroES were expressed in addition (Figure 34). Thus, the requirement of the class III proteins for GroEL/GroES is specific and independent of the bacterial machinery of protein synthesis.

It is interesting that the general ability of the eukaryotic cytosol to fold multi-domain proteins and the presence of the type II chaperonin TRiC are not sufficient to compensate for the lack of GroEL during the folding of *E. coli* 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. Hence, the two chaperonin systems have rather evolved to meet their specific needs.

5.2. The GroEL interactome

The 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 studies have mostly been limited to a qualitative identification of chaperone substrates. Putative substrates of TF and DnaK have been identified by 2D-gel analysis and mass spectrometry (MS) of proteins that aggregated 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 ΔtigΔdnaKJ* cells (Vorderwülbecke et al., 2004). A subset of interacting proteins of the chaperonin GroEL has been identified by GroEL co-immunoprecipitation under nucleotide free conditions 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.

The mass spectrometric approach used here provides in depth qualitative and quantitative information on chaperonin substrates and serves as an example of state-of-the-art proteomic analysis in general.

5.2.1. Quality of the dataset

Several observations suggest that the identified set of 252 GroEL interacting proteins is virtually complete. The amount of identified substrates did not increase in repeated experiments, both in multiple rounds of GroEL complex purification and in multiple LC-MS/MS experiments. Also, the detection threshold seemed not to put constraints on the number of identified proteins. 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. Furthermore, GroEL dependent

proteins are highly enriched on GroEL, as compared to proteins which can also use other chaperone systems for folding and are therefore preferentially detected by MS. GroEL interactors which might have escaped detection would thus most likely not belong to the fraction of stringently dependent GroEL substrates, but rather to a group which plays a minor role among GroEL interacting proteins.

Roughly 50% of proteins identified in an earlier study on GroEL substrates in our laboratory (Houry et al., 1999) overlap with the present study. Differences in isolation procedures of GroEL associated proteins and in the subsequent analysis as well as methodological constraints prevent a more detailed comparison of the two studies.

5.2.2. Methodological constraints

The use of MmES-$(His)_6$ for isolation of GroEL/GroES/substrate complexes had a beneficial effect on complex stability. This allowed stringent washing conditions and therefore purification of GroEL with bound substrates without much contamination by unspecifically detected proteins (Figure 12). As a result, only seven proteins were found to bind non-specifically to the purification matrix. Use of EcES-$(His)_6$ resulted in a reduced yield after purification (data not shown). Since the apical domains of GroEL are responsible for substrate capture and substrate isolation was carried out quickly after MmES-$(His)_6$ expression, it seems reasonable to assume that the use of MmES did not lead to a different substrate spectrum or substrate concentrations in the complexes when compared with complexes formed with *E. coli* GroES. This was confirmed by comparing levels of selected proteins in GroEL complexes by Western blotting and quantification analysis (data not shown). In addition, the experimental approach used in the present study ensured that proteins were captured by GroEL *in vivo* and did not exchange after cell lysis and during purification.

5.3. Properties of GroEL interactors

5.3.1. Size distribution of proteins associated with GroEL

As expected, GroEL interacting proteins showed a different size distribution when compared to all lysate proteins (Figure 14). The GroEL/GroES cavity has a volume of ~85 000 Å^3, which would be able to accommodate an unfolded polypeptide of a ~60 kDa protein (Sigler et al., 1998). 77% of all GroEL substrate proteins were found to be between 20 and 60 kDa in size as compared to 64% of lysate proteins. 82% of substrate proteins with an enrichment factor above 50, indicating high GroEL dependence, were within size limits for encapsulation. Virtually no proteins below 10 kDa were determined to interact with GroEL, whereas in the total lysate, 5% of proteins were smaller than 10 kDa (Figure 14). Such small proteins are thought to fold rapidly and spontaneously before reaching GroEL.

5.3.2. Substrates too large to fit inside the GroEL/GroES cavity

GroEL interacting proteins larger than ~60 kDa exceed the size limit of the chaperonin cavity (Sigler et al., 1998). They form a special class of GroEL interacting proteins, since they can bind to GroEL in *trans*, at the ring opposite to bound GroES, and do not require encapsulation for productive folding (Chaudhuri et al., 2001; Farr et al., 2003). Among 252 GroEL interacting proteins, 30 were larger than 60 kDa.

SYT, an endogenous substrate too large for GroEL encapsulation was studied further, both *in vivo* and *in vitro*. This homodimeric protein of 74 kDa subunits behaves like a typical class II substrate *in vivo*. GroEL/GroES co-expression enhanced its solubility (Figure 29), but a reduced GroEL capacity resulting from GroEL/GroES depletion did not lead to significant aggregation (Figure 32). *In vitro*, DnaK mediates folding of SYT much more efficiently than the GroEL system and it is thus believed that SYT interaction with GroEL merely depends on a DnaK-like unspecific binding and release mechanism without requiring encapsulation for productive folding (Figure 26).

Six large proteins have been identified with an enrichment factor greater than 50. They seem to depend stringently on GroEL for productive folding, without requiring encapsulation. Here, simple binding and release of these proteins by GroEL without requiring GroES, in analogy to the DnaK mechanism, is probably sufficient for productive folding.

5.4. Structures of GroEL substrates

Analysis of fold types among GroEL interacting proteins revealed a broad spectrum of different final structures, whose intermediate forms during folding are recognized by GroEL. Only the ten most common fold classes were analyzed further in this study (Figure 19). Chaperonin interacting proteins show a bias towards the TIM barrel fold. This fold is even more favored among the highly GroEL dependent class III proteins, where TIM barrel proteins are enriched by 3.5 times, compared to lysate proteins (Figure 20).

5.4.1. The TIM barrel fold

The TIM barrel is a ubiquitous topology and many representatives of this fold are found in all types of cells as well as among GroEL interacting proteins. Thus, a particular structural feature exhibited in the nascent, unfolded form or in intermediate states during the folding pathway must exist that distinguishes chaperonin independent, and even spontaneously folding TIM barrels, from GroEL dependent ones. Our analysis could not reveal which specific structural feature of TIM barrels is responsible for the stringent GroEL dependence found for a subset of these proteins.

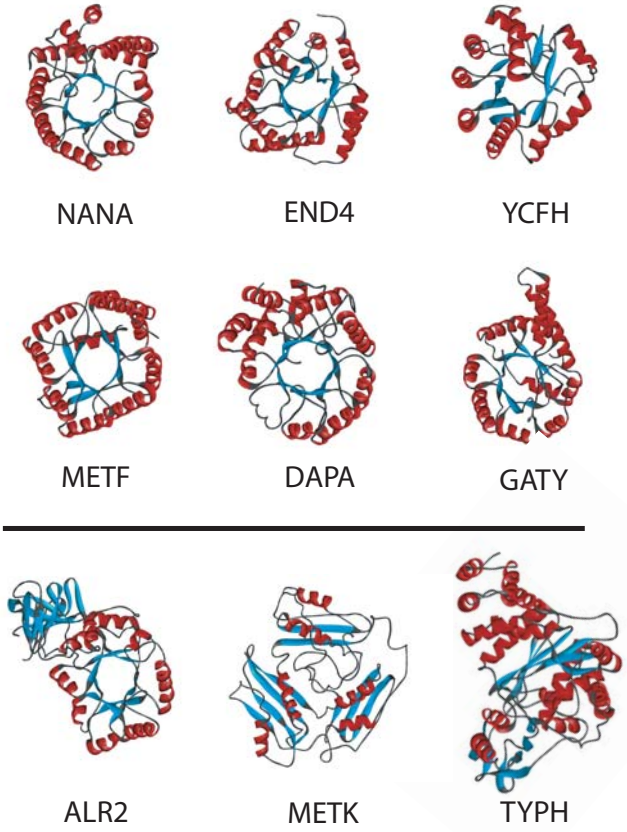


Figure 35: Exemplary structures of GroEL interacting proteins.

Ribbon displays of examples for stringent GroEL substrates. NANA (1HL2), END4 (1QTW), YCFH (1J6O (HSSP)), METF (1B5T), DAPA (1DHP) and GATY (1GVF (HSSP)) are examples for TIM barrel proteins. ALR2 (1BD0 (HSSP)) is a TIM barrel protein with an additional β-barrel domain extension, METK (1MXA) and TYPH (1OPT) are enriched proteins on GroEL with other fold types. Codes in brackets are PDB codes. HSSP: homology derived secondary structure of proteins. Structures were edited with ViewerPro software.

Folding of TIM β/α barrel proteins has been previously studied in detail. For example, triosephosphate isomerase (Rietveld and Ferreira, 1998), the name-giving protein of this fold class, and aldolase (Rudolph et al., 1992) exhibit apparent two-state folding mechanisms, whereas the alpha subunit of tryptophan synthase (Wu and Matthews, 2002), indole-3-glycerol phosphate synthase (Sanchez del Pino and Fersht, 1997) and imidazole glycerol phosphate

synthase subunit hisF (Höcker et al., 2001) feature apparent non two-state folding mechanisms. The latter are supposed to fold in partial fragments, which serve as autonomous folding units that provide an initial scaffold for the formation of the complete native structure. Such partial structure acquisition can accelerate folding, but might also result in the population of 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 triose phosphate isomerase and the tryptophan synthase alpha subunit were detected in the *E. coli* lysate. 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 stringently dependent on GroEL for productive folding identified in this study.

These TIM barrel proteins (for representative structures see Figure 35) thus probably constitute examples of proteins with a particular tendency to accumulate inactive, aggregation-prone intermediates. The energy barriers towards the native fold may be overcome by confinement through encapsulation inside the GroEL/GroES cage. Detailed comparison of identified TIM barrel proteins in collaboration with D. Frishman, GSF, Neuherberg, Germany and with A. Lupas and J. Soeding, MPI for Developmental Biology, Tübingen, Germany did not reveal detectable structural differences between GroEL dependent and independent TIM barrel proteins. The features determining a protein to be GroEL dependent for folding are probably exposed only during the folding process and are therefore not identifiable in their final structure, especially since native proteins naturally are not interacting with GroEL.

5.4.2. Other folds and substrate orthologs in other organisms

About 25% of highly enriched proteins on GroEL (REF >50) adopt the TIM barrel fold. Hence, a significant number of proteins folding via GroEL display different topologies (Figure 18). No common feature among them could be identified that would explain GroEL dependence. For a more detailed analysis, larger data sets, possibly from different organisms, are required.

Fold types identified with lower frequency on GroEL as compared to the *E. coli* lysate are the flavodoxin like fold (c.23) and the P-loop containing nucleotide triphosphate hydrolases (c.37). Again, the obtained data set is too small to allow statistical conclusions on chaperone usage of these fold types.

Recent sequencing efforts revealed the existence of GroEL deficient organisms (Wong and Houry, 2004). Orthologs of highly enriched *E. coli* GroEL substrates are less abundant in these organisms (15—20%) than expected, based on homology of their genomes (25-40%). Nevertheless, some orthologs of highly enriched *E. coli* GroEL substrates, such as the TIM barrel proteins YCFH, GATY, and END4 were identified. These proteins must therefore have evolved to fold in a GroEL-independent fashion. Detailed structural analysis of these orthologs, as well as the investigation of individual folding pathways might reveal further insight into the folding properties that determine strict chaperonin dependence.

5.5. Classification of GroEL interactors

5.5.1. Extension of the classification to all GroEL interacting proteins

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. Class I proteins were found to have less than 0.05% (enrichment factor <1) of their total cytosolic amount associated with GroEL. Of the 252 identified GroEL substrates, this group comprises 41 proteins. However, it is likely that not all GroEL interacting proteins of class I were identified in this study, since detection limits of the analytical techniques applied might lead to a failure in identification. The omitted proteins are low abundant proteins in the cell with a very low enrichment on GroEL for which the chaperonin does not contribute significantly to folding.

Many identified class III proteins had enrichment factors of or greater than 100. Generally an enrichment factor above 50 (>3% of a protein interacts with GroEL at any given time) was considered a good cut off for class III proteins. This group is thus made up of 78 out of 252 proteins. Class II proteins showed relative concentrations on GroEL between those of class I and III. Nearly half of all identified proteins fall into this class (133 proteins).

Name	**Function**	**MW**	**Determined *in vitro***	**Determined from REF**
ADD	Adenosine deaminase	36.4	III	III
ALR2	Alanine racemase, catabolic	38.8	III	III
CRP	Catabolite gene activator	23.6	I	III
DAPA	Dihydrodipicolinate synthase	31.3	III	n.d.
DCEA	Glutamate decarboxylase alpha	52.7	II	n.d.
END4	Endonuclease IV	31.5	III	II
ENO	Enolase	45.6	I	n.d.
GATD	Galactitol-1-phosphate	37.4	II	II
GATY	Tagatose-1,6-bisphosphate aldolase	30.8	III	III
HEM2	Delta-aminolevulinic acid dehydratase	35.5	III	III
LLDD	L-lactate dehydrogenase (Cytochrome)	42.7	III	III
LTAE	Low-specificity L-threonine aldolase	36.5	III	III
METF	5,10-methylenetetrahydrofol. reductase	33.1	III	III
METK	S-adenosylmethionine synthetase	41.8	III	II
NANA	N-acetylneuraminate lyase	32.5	III	n.d.
SYT	Threonyl-tRNA synthetase	74.0	II	II
TDH	L-threonine 3-dehydrogenase	37.2	I	n.d.
TYPH	Thymidine phosphorylase	47.2	III	III
XYLA	Xylose isomerase	49.7	III	III
YAJO	Hypothetical oxidoreductase yajO	36.4	III	III
YHBJ	Hypothetical UPF0042 protein yhbJ.	32.5	III	III

Table 7: GroEL interacting proteins sorted into substrate classes

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 28). Nearly 90% of all polypeptide chains associated with GroEL are members of class III; less than 10% belong to class II; and class I substrates comprise about 1% of all GroEL interacting polypeptide chains. This high concentration of class III proteins on GroEL is in great contrast to their relatively low cellular concentrations (Figure 22). Class III proteins are generally proteins of very low abundance in the *E. coli* cytosol, whereas class I proteins represent the most abundant soluble proteins of *E. coli*.

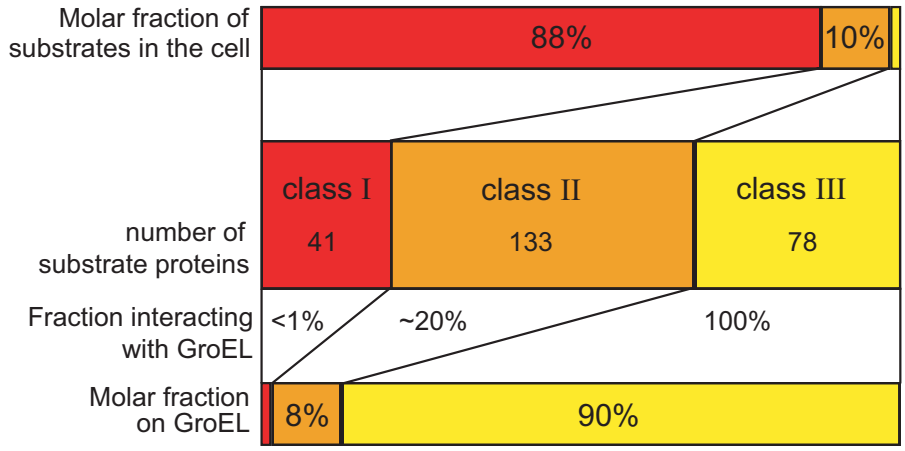


Figure 36: Classification of all GroEL substrates

Substrate classes predicted from *in vivo* and *in vitro* experiments with selected substrates were extended to all GroEL interacting proteins. A small molar fraction of cellular proteins is highly enriched on GroEL and occupies the major part of the GroEL capacity.

5.5.2. Calculations on GroEL transit of substrate proteins

An attempt was made to calculate protein transit through GroEL in living *E. coli* cells based on the experimentally determined distribution of substrate proteins on GroEL.

GroEL concentration in the cell is thought to be about 3 µM (Ellis and Hartl, 1996; Mogk et al., 1999). The average time a protein needs to fold on GroEL is assumed to be around 60 s (Ewalt et al., 1997). Therefore, in one minute 3 µM of substrate protein potentially transit GroEL.

The doubling time of an *E. coli* cell under the experimental conditions applied in this study is about 40 minutes. Therefore 40 x 3 µM substrate = 120 µM substrate protein transits GroEL during a cell cycle.

The average size of an *E. coli* protein is about 35 000 Da = 35 000 g/ mol = 35 mg/ µmol. Consequently:

$$\frac{35\,mg \times 120\,\mu mol}{1\,\mu mol \times l} = 4200\,{}^{mg}\!/_{l} = 4.2\,{}^{g}\!/_{l} \qquad (4)$$

of protein transits GroEL in every cell cycle. Since in one cell cycle, the complete protein inventory of a cell must be doubled (newly synthesized) and the overall protein amount in *E. coli* cells equals 200 g/ l

$$\frac{4.2\,g \times l}{200\,g \times l} = 0.021 \equiv \underline{\underline{2.1\%}} \qquad (5)$$

of all newly synthesized proteins by mass transits GroEL.

By definition, 100% of class III substrate protein has to transit GroEL for productive folding. 2.7 µM class III substrate protein is bound to GroEL at any given time, since class III proteins account for 90% of protein mass interacting with GroEL (3 µM, Figure 22). Extrapolated to a doubling time (40 folding events),

$$\frac{35\,mg \times 108\,\mu mol}{1\,\mu mol \times l} = 3780\,{}^{mg}\!/_{l} = 3.78\,{}^{g}\!/_{l} \qquad (6)$$

is the theoretical concentration of class III proteins in *E. coli*. Division by the total amount of protein in the cell results in

$$\frac{3.78\,g \times l}{200\,g \times l} = 0.0189 \equiv \underline{\underline{1.89\%}} \qquad (7)$$

of the total protein mass in the cell can stringently depend on GroEL and transit the chaperonin for productive folding.

This calculated result corresponds well to the experimentally determined fraction of stringent GroEL class III substrate protein of 1.8% (Figure 22).

5.5.3. Chaperone networks in *E. coli*

High enrichment of class III substrates on GroEL in the cell is presumably a consequence of mainly two factors. Firstly, unfolded class III substrates have been shown to exhibit higher affinity to GroEL than class I or class II substrates (Figure 28). Class III substrates therefore preferentially bind to GroEL when competing with class I or II substrates.

Secondly, 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 trigger factor (TF) upon synthesis at the ribosome might already be sufficient for correct folding of most class I proteins. Since TF binds at the ribosome and receives nascent polypeptides directly at the exit tunnel, it has an advantage over GroEL to interact with newly synthesized proteins.

DnaK is about 10 times more abundant in the cell than GroEL and contributes significantly to the folding of proteins which have a high chaperone dependency but do not dependent on encapsulation by GroEL and GroES. Consequently, this chaperone system promotes folding of almost all class I and many class II substrates. Mostly proteins which interact unproductively with DnaK are transferred to the chaperonin, such as class III proteins, which are stabilized by DnaK against aggregation but do not fold with this chaperone machinery. 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 (Figure 9).

The chaperone pathway model is further supported by the identification of GroEL substrates from *E. coli* cells lacking both DnaK and TF. Cells missing only one of either of these chaperones do not exhibit a significantly different GroEL substrate spectrum from wild type *E. coli* cells, since TF and the DnaK system have overlapping function (Deuerling et al., 2003; Teter et al., 1999). However, combined deletion of the genes encoding TF and DnaK considerably increases the number of identifiable GroEL substrates. The newly identified proteins from *ΔtigΔdnaKdnaJ* had not been identified as specific GroEL substrates

previously. These proteins 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.

Further support for a co-operation of the DnaK system and GroEL in folding is given by reviewing published DnaK and TF substrates. Analysis of aggregating proteins in an *E. coli* strain lacking DnaK, DnaJ and TF (Vorderwülbecke et al., 2004) sugests a substrate spectrum shifted towards high molecular weight proteins. This can be explained, by the finding that GroEL can partially take over the function of TF and DnaK (Genevaux et al., 2004) and thereby contributes to folding of proteins which otherwise would not reach the chaperonin. The bias towards large proteins arises from the fact that GroEL can only accommodate proteins up to 60 kD size in its cavity (Figure 14). Thus, the substrate spectrum of DnaK is probably larger than previously reported.

5.5.4. The essentiality of GroEL, GroES and other chaperone systems

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, whereas the cell can not compensate for the loss of GroEL and GroES (Fayet et al., 1989; Genevaux et al., 2004). The GroEL/GroES system is the only chaperone system in *E. coli* which is essential under all growth conditions tested.

The existence of essential and stringently GroEL-dependent class III substrates (Table 5) 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, and thus are unable to fulfill their cellular functions. Hence, GroEL/GroES deficient cells cannot survive.

DnaK and TF do not have essential substrates which absolutely depend on either one of these chaperones or both of them for folding into the native state. *E. coli* mutant cell lines with either TF or DnaK and DnaJ deleted show a temperature sensitive phenotype, but are less affected in growth than cells which lack both chaperone systems. Since TF and DnaK have overlapping substrate spectra, they can compensate for deletion of either one by taking over its role in folding *in vivo* (Deuerling et al., 1999; Teter et al., 1999). However, a combined

deletion of both the genes encoding for TF and DnaK results in high levels of protein aggregation and impaired growth (Genevaux et al., 2004). Notably, GroEL is overexpressed in this strain to compensate for the loss of TF and DnaK. Consequently, GroEL is thought to be able to fold at least all essential DnaK and TF substrates in their absence.

5.6. Evolutionary considerations

Stringent GroEL substrates are enriched in proteins which contain more distinct SCOP superfamily domains than the average *E. coli* lysate protein (D. Frishman, personal communication). It can be speculated that GroEL functions as a capacitor to facilitate the evolution of structurally more diverse protein families. A reciprocal approach in which GroEL was successfully mutated to increase the folding yield of the model protein GFP resulted in a diminished capacity of the chaperonin for the folding of other proteins (Wang et al., 2002).

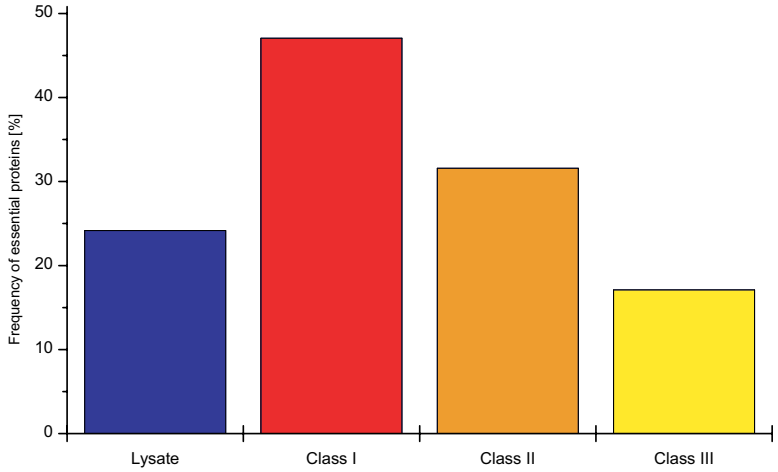


Figure 37: Essentiality of proteins by classes

Bar graph of essential proteins among the experimentally analysed *E. coli* lysate and determined classes of GroEL substrates. Class I proteins are enriched in essential proteins, correlating with their high cellular abundance.

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 are essential for cell growth, compared to 24% of the *E. coli* lysate proteins (Gerdes et al., 2003) (Figure 37). The high number of essential proteins among class I substrate

proteins correlates with the finding that in general, proteins of high concentration in *E. coli* are essential (Gerdes et al., 2003). On average however, essential proteins are not significantly enriched among GroEL substrates.

GroEL has a limited capacity for protein folding, and thus it follows that very abundant proteins in the cell have an evolutionary pressure to become or to remain GroEL independent, since excessive use of GroEL by a single, abundant protein is detrimental for cell viability. Folding of a single protein would occur at the expense of many other proteins depending on GroEL for folding. For example, the eukaryotic homolog of GroEL, TRiC/CCT has evolved as a specialized folding machinery for actin and tubulin (Lewis et al., 1996), which are by mass the most abundant proteins interacting with the eukaryotic chaperonin. The GroEL/GroES system could only evolve to become a chaperone system with a diverse substrate spectrum by granting GroEL-independent folding for highly abundant proteins.

The available GroEL capacity *in vivo* has not yet reached its limits, since GroEL can be depleted to about 25% of physiological levels without affecting cellular viability (Kanemori et al., 1994). Further evidence for additionally available GroEL capacity *in vivo* is given by the fact that stringently GroEL dependent proteins can be overexpressed and reach nativeness without affecting cellular viability. This additionally available capacity for folding could allow proteins to evolve with respect to their function without requiring equally well or better folding properties at the same time. In other words, GroEL buffers against deleterious mutations (Fares et al., 2002).

6. References

Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198-207.

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 Research*, **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. *Journal of Bacteriology*, **171**, 2748-2755.

Ang, D., Keppel, F., Klein, G., Richardson, A. and Georgopoulos, C. (2000) Genetic analysis of bacteriophage-encoded cochaperonins. *Annu Rev Genet*, **34**, 439-456.

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

Baldwin, R.L. (1995) The nature of protein folding pathways: the classical versus the new view. *J Biomol NMR*, **5**, 103-109.

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 in Biochemical Sciences*, **24**, 77-83.

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

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

Becker, J., Walter, W., Yan, W. and Craig, E. (1996) Functional interaction of cytosolic hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation in vivo. *Mol. Cell. Biol.*, **16**, 4378-4386.

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

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

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.

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

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.

Cohen, F.E. (2000) Prions, peptides and protein misfolding. *Mol Med Today*, **6**, 292-293.

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 Expression & Purification*, **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. *Molecular Microbiology*, **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 in Biochemical Sciences*, **25**, 331-339.

Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends in Biochemical Sciences*, **24**, 329-332.

Dobson, C.M., Sali, A. and Karplus, M. (1998) Protein folding: a perspective from theory and experiment. *Angewandte Chemie (International Edition in English)*, **37**, 868-893.

Ellis, R.J. (1994) Roles of molecular chaperones in protein folding. *Curr Opin Struct Biol*, **4.**, 117-122.

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

Ellis, R.J. (1997) Molecular chaperones: avoiding the crowd. *Curr Biol*, **7**, R531-533.

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

Ellis, R.J. and Hemmingsen, S.M. (1989) Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem Sci*, **14**, 339-342.

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. (1997) GroEL-mediated protein folding. *Protein Sci*, **6**, 743-760.

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. *Journal of Molecular Biology*, **286**, 1597-1608.

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.

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 Engineering*, **15**, 169-183.

Frishman, D., Mokrejs, M., Kosykh, D., Kastenmüller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., Volz, A., Wagner, C., Fellenberg, M., Heumann, K. and Mewes, H.W. (2003) The PEDANT genome database. *Nucleic Acids Research*, **31**, 207-211.

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., Oltvai, Z.N. and Osterman, A.L. (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J. Bacteriol.*, **185**, 5673-5684.

Gething, M.J. and Sambrook, J. (1992) Protein folding in the cell. *Nature*, **355**, 33-45.

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.

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 U S A*, **93**, 4437-4441.

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. *Nature Structural Biology*, **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.

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. *PNAS*, **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.

Kanemori, M., Mori, H. and Yura, T. (1994) Effects of reduced levels of GroE chaperones on protein metabolism: enhanced synthesis of heat shock proteins during steady-state growth of Escherichia coli. *J Bacteriol*, **176**, 4235-4242.

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

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., Muller, 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.

Knapp, S., Schmidt-Krey, I., Hebert, H., Bergman, T., Jornvall, H. and Ladenstein, R. (1994) The molecular chaperonin TF55 from the Thermophilic archaeon Sulfolobus solfataricus. A biochemical and structural characterization. *J Mol Biol*, **242**, 397-407.

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.

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.

Lewis, S.A., Tian, G., Vainberg, I.E. and Cowan, N.J. (1996) Chaperonin-mediated folding of actin and tubulin. *J Cell Biol*, **132**, 1-4.

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.

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. (1980) S-Adenosylmethionine synthetase from Escherichia coli. *J Biol Chem*, **255**, 9082-9092.

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.

Mogk, A. and Bukau, B. (2004) Molecular chaperones: structure of a protein disaggregase. *Current Biology*, **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 Journal*, **18**, 6934-6949.

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.

Netzer, W.J. and Hartl, F.U. (1998) Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. *Trends Biochem Sci*, **23**, 68-73.

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. *Molecular & Cellular 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). *Journal of Proteome Research*, **2**, 173-181.

Onuchic, J.N. and Wolynes, P.G. (2004) Theory of protein folding. *Curr Opin Struct Biol*, **14**, 70-75.

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 U S A*, **91**, 10422-10425.

Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*, **405**, 837-846.

Privalov, P.L. (1996) Intermediate states in protein folding. *Journal of Molecular Biology*, **258**, 707-725.

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

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

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. *Analytical Chemistry*, **75**, 663-670.

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

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.

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

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

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

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

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. *Annual Review of Biochemistry*, **67**, 581-608.

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

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

Tanford, C. (1962) Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *Journal of the American Chemical Society*, **84**, 4240-4247.

Taniuchi, H. and Anfinsen, C.B. (1969) An experimental approach to the study of the folding of staphylococcal nuclease. *J. Biol. Chem.*, **244**, 3864-3875.

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 U S A*, **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 U S A*, **76**, 4350-4354.

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 Letters*, **559**, 181-187.

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

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

Wanker, E.E. (2000) Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations. *Biol Chem*, **381**, 937-942.

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., 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.

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

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

Wu, Y. and Matthews, C.R. (2002) A cis-prolyl peptide bond isomerization dominates the folding of the alpha subunit of trp synthase, a TIM barrel protein. *Journal of Molecular Biology*, **322**, 7-13.

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.

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.

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 Science*, **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.

7. Supplementary Material

Supplementary Table S1: Substrates of GroEL, identified by LC-MS/MS analysis of GroEL/GroES/substrate complexes.

For SCOP fold abbreviations, see http://scop.mrc-lmb.cam.ac.uk/scop/parse/dir.des.scop.txt_1.65. Essentiality: 1 = essential protein, 0 = not essential protein. COG functional categories: J: Translation, ribosomal structure and biogenesis K: Transcription L: DNA replication, recombination and repair 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 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 R: General function prediction only S: Function unknown

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
thi2_ecoli	P33636	Thioredoxin 2 (EC 1.8.1.8) (Protein-disulfide reductase) (Disulfide reductase) (Trx2).	1	15.6	c.47	1		Cytoplasmic	O
tpx_ecoli	P37901	Thiol peroxidase (EC 1.11.1.-) (Scavengase P20).	1	17.7	c.47	0		Periplasmic	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		Cytoplasmic	G

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
faba_ecoli	P18391	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.60) (Beta-hydroxydecanoyl thioester dehydrase).	1	18.8	d.38	1	Homodimer	Cytoplasmic	I
ipyr_ecoli	P17288	Inorganic pyrophosphatase (EC 3.6.1.1) (Pyrophosphate phospho- hydrolase) (PPase).	1	19.6	b.40	1	Homohexamer	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	Homodimer (By similarity)		O
rrf_ecoli	P16174	Ribosome recycling factor (Ribosome releasing factor) (RRF).	1	20.6	d.67	1		Cytoplasmic	J
grpe_ecoli	P09372	GrpE protein (HSP-70 cofactor) (Heat shock protein B25.3) (HSP24).	1	21.8	b.73	1			O
deod_ecoli	P09743	Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase) (PNP).	1	25.8	c.56; c.48	0	Homohexamer		F
thig_ecoli	P30139	Thiazole biosynthesis protein thiG.	1	26.9	c.1			Cytoplasmic	F
deoc_ecoli	P00882	Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase) (DERA).	1	27.7	c.1	0	Monomer and homodimer	Cytoplasmic	F

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
panb_ecoli	P31057	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11) (Ketopantoate hydroxymethyltransferase).	1	28.2	c.78; c.1	0	Hexamer (Potential)		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	Homodimer		G
efts_ecoli	P02997	Elongation factor Ts (EF-Ts).	1	30.3	d.43; a.5	1	Heterotetramer composed of two EF-Ts.EF-Tu dimer complex.	Cytoplasmic	J
rbsb_ecoli	P02925	D-ribose-binding periplasmic protein precursor.	1	31.0	c.93	0		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				M
g3p1_ecoli	P06977	Glyceraldehyde 3-phosphate dehydrogenase A (EC 1.2.1.12) (GAPDH-A).	1	35.4	c.2; d.81	1	Homotetramer	Cytoplasmic	G

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional 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		Homodimer.		K
ompa_ecoli	P02934	Outer membrane protein A precursor (Outer membrane protein II*).	1	37.2	f.4	0	Monomer (Probable)	Integral membrane protein. Outer membrane.	M
ynce_ecoli	P76116	Hypothetical protein yncE precursor.	1	38.6	b.69; b.70; b.68	1			S
ompc_ecoli	P06996	Outer membrane protein C precursor (Porin ompC) (Outer membrane protein 1B).	1	40.4	f.4	0	Homotrimer	Integral membrane protein. Outer membrane.	M
pgk_ecoli	P11665	Phosphoglycerate kinase (EC 2.7.2.3).	1	41.0	c.86	1	Monomer	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	Homodimer	Cytoplasmic	I
acka_ecoli	P15046	Acetate kinase (EC 2.7.2.1) (Acetokinase).	1	43.3	c.55	1	Homodimer	Cytoplasmic	C

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
sera_ecoli	P08328	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95) (PGDH).	1	44.0	c.2; d.58; c.23	0	Homotetramer		E
glya_ecoli	P00477	Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase) (SHMT).	1	45.3	c.67	1	Homotetramer	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		Homodimer	Cytoplasmic	G
pura_ecoli	P12283	Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase) (AdSS) (AMPSase).	1	47.2	c.37	0	Homodimer	Cytoplasmic	F
tig_ecoli	P22257	Trigger factor (TF).	1	48.2	d.26	0	Homodimer and monomer		O
kpy1_ecoli	P14178	Pyruvate kinase I (EC 2.7.1.40) (PK-1).	1	50.7	b.58; c.49; c.1	0	Homotetramer		G
6pgd_ecoli	P00350	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).	1	51.5	c.2; a.100	0			G
syn_ecoli	P17242	Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase) (AsnRS).	1	52.4	b.40; d.104	1	Homodimer	Cytoplasmic	J
oppa_ecoli	P23843	Periplasmic oligopeptide-binding protein precursor.	1	60.9	c.94	0		Periplasmic	E

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
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	24-polypeptide structural core with octahedral symmetry.		C
cira_ecoli	P17315	Colicin I receptor precursor.	1	73.9	f.4	0		Outer membrane	P
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	Homotrimer	Cytoplasmic	J
efg_ecoli	P02996	Elongation factor G (EF-G).	1	77.5	d.14; c.37; d.58; b.43	1		Cytoplasmic	J
odp1_ecoli	P06958	Pyruvate dehydrogenase E1 component (EC 1.2.4.1).	1	99.5	c.48; c.36	0	Homodimer		C
muli_ecoli	P02937	Major outer membrane lipoprotein precursor (Murein-lipoprotein).	1 or 2	8.3	-			Attached to the outer membrane by a lipid anchor	N

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
if3_ecoli	P02999	Translation initiation factor IF-3.	1 or 2	20.6	d.15; d.68	1	Monomer	Cytoplasmic	J
yhgi_ecoli	P46847	Protein yhgI.	1 or 2	21.0	-	0			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	Homodimer	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		Cytoplasmic	O
yohl_ecoli	P76424	Hypothetical protein yohL.	2	10.1	-	1			S
ydhd_ecoli	P37010	Protein ydhD.	2	12.9	c.47	1			O
nikr_ecoli	P28910	Nickel responsive regulator.	2	15.1	-	0	Homotetramer		K
yjbq_ecoli	P32698	Hypothetical protein yjbQ.	2	15.7	-	0			S
uspg_ecoli	P39177	Universal stress protein G.	2	15.9	c.29	0	Interacts with groEL.		T
moac_ecoli	P30747	Molybdenum cofactor biosynthesis protein C.	2	17.3	d.58	0	Homohexamer		H
yfhp_ecoli	P77484	Hypothetical protein yfhP.	2	17.3	-	0			K
grea_ecoli	P21346	Transcription elongation factor greA (Transcript cleavage factor greA).	2	17.6	a.2; d.26	0			K

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
ppib_ecoli	P23869	Peptidyl-prolyl cis-trans isomerase B (EC 5.2.1.8) (PPIase B) (Rotamase B).	2	18.2	b.62	1		Cytoplasmic	O
moab_ecoli	P30746	Molybdenum cofactor biosynthesis protein B.	2	18.5	c.57	0			H
dps_ecoli	P27430	DNA protection during starvation protein.	2	18.6	a.25	0	Complex of 12 subunits forming two stacked hexameric rings.		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		Composed of 13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		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		Cytoplasmic (Potential)	L
luxs_ecoli	P45578	S-ribosylhomocysteinase (EC 3.13.1.-) (Autoinducer-2 production protein LuxS) (AI-2 synthesis protein).	2	19.3	d.185		Homodimer (By similarity)		T
arok_ecoli	P24167	Shikimate kinase I (EC 2.7.1.71) (SKI).	2	19.4	c.37	1		Cytoplasmic (Probable)	E

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
yfbu_ecoli	P76492	Protein yfbU.	2	19.5	-	0			S
seqa_ecoli	P36658	SeqA protein.	2	20.3	-	1			L
nusg_ecoli	P16921	Transcription antitermination protein nusG.	2	20.4	-	0			K
rimm_ecoli	P21504	16S rRNA processing protein rimM (21K).	2	20.6	-	1		Cytoplasmic (Potential)	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		Cytoplasmic	J
nudh_ecoli	Q46930	(Di)nucleoside polyphosphate hydrolase (EC 3.6.1.-) (Ap5A pyrophosphatase).	2	20.8	d.113	0	Monomer		L
pth_ecoli	P23932	Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH).	2	21.1	c.56	1	Monomer	Cytoplasmic	J
hemg_ecoli	P27863	Protoporphyrinogen oxidase (EC 1.3.3.4) (PPO).	2	21.2	c.23	1	Belongs to a multi-protein complex.		C
tehb_ecoli	P25397	Tellurite resistance protein tehB.	2	22.5	c.66	0		Cytoplasmic (Potential)	Q
yihx_ecoli	P32145	Hypothetical protein yihX.	2	22.7	c.108	0			R

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
thie_ecoli	P30137	Thiamine-phosphate pyrophosphorylase (EC 2.5.1.3) (TMP pyrophosphorylase) (TMP-PPase) (Thiamine-phosphate synthase).	2	23.0	c.1	0			H
ycio_ecoli	P45847	Protein yciO.	2	23.2	d.115	0			J
engb_ecoli	P24253	Probable GTP-binding protein engB.	2	23.6	c.37	1			D
rcsb_ecoli	P14374	Capsular synthesis regulator component B.	2	23.7	c.23; a.4	0			T
ycbl_ecoli	P75849	Hypothetical protein ycbL.	2	23.8	d.157	0			R
uvry_ecoli	P07027	Response regulator uvrY.	2	23.9	c.23; a.4	0		Cytoplasmic (Probable)	T
glr2_ecoli	P39811	Glutaredoxin 2 (Grx2).	2	24.4	c.47; a.45	0			O
yoda_ecoli	P76344	Protein yodA.	2	24.8	-	0			R
yadf_ecoli	P36857	Protein yadF.	2	25.1	c.53	1			P
ygea_ecoli	P03813	Hypothetical protein ygeA.	2	25.2	c.78	0			M
inaa_ecoli	P27294	Protein inaA.	2	25.3	-	0			S
trmh_ecoli	P19396	tRNA (Guanosine-2'-O-)-methyltransferase (EC 2.1.1.34) (tRNA [GM18] methyltransferase).	2	25.3	c.116	0		Cytoplasmic (Potential)	J

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
arad_ecoli	P08203	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose isomerase).	2	25.5	c.74	0			G
phop_ecoli	P23836	Transcriptional regulatory protein phoP.	2	25.5	c.23; a.4	0		Cytoplasmic (Probable)	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			Q
proq_ecoli	P45577	ProP effector.	2	25.9	a.136			Cytoplasmic (Potential)	T
cpxr_ecoli	P16244	Transcriptional regulatory protein cpxR.	2	26.3	c.23; a.4	0		Cytoplasmic (Probable)	T
pyrf_ecoli	P08244	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase) (OMPDCase) (OMPdecase).	2	26.4	c.1	1	Homodimer		F
ompr_ecoli	P03025	Transcriptional regulatory protein ompR.	2	27.4	c.23; a.4	0	Monomer and multimer	Cytoplasmic	T
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	Homotrimer	Cytoplasmic	M

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
ybff_ecoli	P75736	Putative esterase/lipase ybfF (EC 3.1.-.-).	2	28.4	c.69	0			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	Two ATP-binding proteins (pstB), two transmembrane proteins (pstC and pstA) and a solute-binding protein (pstS) (Probable).	Inner membrane-associated	P
yjjv_ecoli	P39408	Putative deoxyribonuclease yjjV (EC 3.1.21.-).	2	28.9	c.1	1			L
cyse_ecoli	P05796	Serine acetyltransferase (EC 2.3.1.30) (SAT).	2	29.3	b.81	1	Homohexamer. Dimer of a homotrimer.	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	Homotrimer	Cytoplasmic	M
yffs_ecoli	P76550	Hypothetical protein yffS.	2	31.0	-	1			S
ypt2_ecoli	Q99390	Hypothetical 31.7 kDa protein in TRAX-FINO intergenic region (ORFC).	2	31.7	c.69				R

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
ybbn_ecoli	P77395	Protein ybbN.	2	31.8	a.118; c.47	0			O
yijo_ecoli	P32677	Hypothetical transcriptional regulator yijO.	2	32.1	a.4	0			K
ynia_ecoli	P77739	Hypothetical protein yniA.	2	32.5	d.144				S
hslo_ecoli	P45803	33 kDa chaperonin (Heat shock protein 33) (HSP33).	2	32.5	d.193			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			J
rob_ecoli	P27292	Right origin-binding protein.	2	33.1	d.60; a.4	0			K
ycjz_ecoli	P77333	Putative HTH-type transcriptional regulator ycjZ.	2	33.5	c.94; a.4	0			K
yadb_ecoli	P27305	Hypothetical protein yadB.	2	33.6	c.26	0			J
ydhf_ecoli	P76187	Hypothetical oxidoreductase ydhF (EC 1.-.-.-).	2	33.6	c.1	0			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		Cytoplasmic	F

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
oxyr_ecoli	P11721	Hydrogen peroxide-inducible genes activator (Morphology and auto- aggregation control protein).	2	34.3	c.94; a.4	0	Homodimer and homotetramer		K
yeat_ecoli	P76250	Putative HTH-type transcriptional regulator yeaT.	2	34.6	c.94; a.4				K
ybib_ecoli	P30177	Hypothetical protein ybiB.	2	35.0	a.46	0			E
acca_ecoli	P30867	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (EC 6.4.1.2).	2	35.1	-	1	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.		I
cysb_ecoli	P06613	HTH-type transcriptional regulator cysB (Cys regulon transcriptional activator).	2	36.2	c.94; a.4	0	Homotetramer (By similarity)	Cytoplasmic	K
dusa_ecoli	P32695	tRNA-dihydrouridine synthase A (EC 1.-.-.-).	2	36.8	c.1	0			J
mreb_ecoli	P13519	Rod shape-determining protein mreB.	2	37.0	c.55	1			D
moaa_ecoli	P30745	Molybdenum cofactor biosynthesis protein A.	2	37.3	-	0			H

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
gatd_ecoli	P37190	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251).	2	37.4	c.2; c.66; b.35	1			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			J
inh5_ecoli	P76071	Transposase insH for insertion sequence element IS5Y.	2	37.8	-	1			L
yihe_ecoli	P32127	Hypothetical protein yihE.	2	38.1	d.144	0			R
tas_ecoli	Q46933	Tas protein.	2	38.5	c.1	0			C
pyrc_ecoli	P05020	Dihydroorotase (EC 3.5.2.3) (DHOase).	2	38.7	c.1	0	Homodimer		F
yghz_ecoli	Q46851	Hypothetical protein yghZ.	2	38.8	c.1	0			C
alf_ecoli	P11604	Fructose-bisphosphate aldolase class II (EC 4.1.2.13) (FBP aldolase).	2	39.0	c.1	1	Homodimer		G
dcup_ecoli	P29680	Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).	2	39.2	c.1	1		Cytoplasmic (Probable)	H
insh_ecoli	P03837	Transposase insH for insertion sequence element IS5.	2	39.3	-	0			L

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
ompf_ecoli	P02931	Outer membrane protein F precursor (Porin ompF) (Outer membrane protein 1A) (Outer membrane protein IA) (Outer membrane protein B).	2	39.3	f.4	0	Homotrimer	Integral membrane protein. Outer membrane.	M
serc_ecoli	P23721	Phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).	2	39.7	c.67	0	Homodimer	Cytoplasmic	H
ybbb_ecoli	P33667	Hypothetical protein ybbB.	2	41.1	c.46	0			R
entc_ecoli	P10377	Isochorismate synthase entC (EC 5.4.99.6) (Isochorismate mutase).	2	42.9	d.161	0	Monomer		H
argm_ecoli	P77581	Succinylornithine transaminase (EC 2.6.1.-) (Succinylornithine aminotransferase) (Carbon starvation protein C).	2	43.7	c.67	0			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	24-polypeptide structural core with octahedral symmetry.		C
iscs_ecoli	P39171	Cysteine desulfurase (EC 4.4.1.-) (ThiI transpersulfidase) (NifS protein homolog).	2	45.1	c.67	1			E

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional 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	glpA/B dimer and membrane bound glpC.	Loosely bound to the cytoplasmic membrane often occurring in vesicles associated with fumarate reductase	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	Homodimer	Cytoplasmic (Potential)	H
yade_ecoli	P31666	Hypothetical protein yadE precursor.	2	46.3	-	0			G
avta_ecoli	P09053	Valine--pyruvate aminotransferase (EC 2.6.1.66) (Transaminase C) (Alanine--valine transaminase).	2	46.7	c.67	0	Homodimer (By similarity)	Cytoplasmic (By similarity).	E
rhlb_ecoli	P24229	ATP-dependent RNA helicase rhlB (EC 3.6.1.-).	2	47.0	c.37	0	Component of the degradosome complex. Binds to RNase E and PNPase. Forms multimers.		L

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
rho_ecoli	P03002	Transcription termination factor rho.	2	47.0	c.37; a.140; b.40		Homohexamer		K
dhna_ecoli	P00393	NADH dehydrogenase (EC 1.6.99.3).	2	47.2	c.4; c.3	0		Membrane	C
cisy_ecoli	P00891	Citrate synthase (EC 2.3.3.1).	2	48.0	a.103	0	Homohexamer		C
ygaf_ecoli	P37339	Hypothetical protein ygaF.	2	48.6	c.3				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			H
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	13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		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		heterohexamer		I
yegd_ecoli	P36928	Hypothetical chaperone protein yegD.	2	49.4	c.55	0			O
ycaj_ecoli	P45526	Hypothetical protein ycaJ.	2	49.6	c.37	0			L

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
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	Homooligomer; probable homooctamer	Cytoplasmic	C
dcea_ecoli	P80063	Glutamate decarboxylase alpha (EC 4.1.1.15) (GAD-alpha).	2	52.7	c.67		Homohexamer		E
tnaa_ecoli	P00913	Tryptophanase (EC 4.1.99.1) (L-tryptophan indole-lyase) (TNase).	2	52.8	c.67	0	Homotetramer		E
ydcr_ecoli	P77730	Hypothetical protein ydcR.	2	52.8	a.4; c.67	0			K
pcnb_ecoli	P13685	Poly(A) polymerase (EC 2.7.7.19) (PAP) (Plasmid copy number protein).	2	54.7	-	0	Monomer.		J
mgla_ecoli	P23199	Galactoside transport ATP-binding protein mglA.	2	56.4	c.37	0		Inner membrane-associated (Potential)	G
typa_ecoli	P32132	GTP-binding protein typA/BipA (Tyrosine phosphorylated protein A).	2	65.4	c.37; d.58; b.43	0			N

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
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	Composed of 13 different subunits. Subunits nuoCD, E, F, and G constitute the peripheral sector of the complex.		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			O
gida_ecoli	P17112	Glucose inhibited division protein A.	2	69.5	c.2; c.4; c.3	0			D
ydcp_ecoli	P76104	Putative protease ydcP precursor (EC 3.4.-.-).	2	72.7	-	0			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	Homodimer	Cytoplasmic	J
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				T
lon_ecoli	P08177	ATP-dependent protease La (EC 3.4.21.53).	2	87.4	c.37	0	Homotetramer	Cytoplasmic	O

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
rnr_ecoli	P21499	Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein).	2	92.1	b.40	0	Monomer		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	Seems to form a rod shaped polymer composed of about 40 identical subunits.		C
gyra_ecoli	P09097	DNA gyrase subunit A (EC 5.99.1.3).	2	97.0	e.11	1	forms an A2B2 tetramer with GyrB.		L
if2_ecoli	P02995	Translation initiation factor IF-2.	2	97.4	c.37; a.114; c.20; b.43	1		Cytoplasmic	J
glnd_ecoli	P27249	[Protein-PII] uridylyltransferase (EC 2.7.7.59) (PII uridylyl- transferase) (Uridylyl removing enzyme) (UTase).	2	102.4	-	1			O
odo1_ecoli	P07015	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha- ketoglutarate dehydrogenase).	2	105.1	c.36	1	Homodimer		C

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
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	Monomer	Cytoplasmic	F
hrpa_ecoli	P43329	ATP-dependent helicase hrpA.	2	149.0	c.37	0			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	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.		K
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	See rpob		K
ybak_ecoli	P37175	Protein ybaK.	3	17.1	d.116	0			S

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
hlpa_ecoli	P11457	Histone-like protein HLP-1 precursor (DNA-binding 17 kDa protein).	3	17.7	-	0	Homotetramer	Either in the nucleoid (chromatin) or in the outer membrane	M
ssrp_ecoli	P32052	SsrA-binding protein (Small protein B).	3	18.1	b.111	0		Cytoplasmic (Potential)	O
rsd_ecoli	P31690	Regulator of sigma D.	3	18.2	-	1			K
ubic_ecoli	P26602	Chorismate--pyruvate lyase (EC 4.-.-.-).	3	18.6	d.190	0	Monomer	Cytoplasmic	H
yqab_ecoli	P77475	Hypothetical protein yqaB.	3	20.8	c.108	0			R
ycfp_ecoli	P75950	Hypothetical protein ycfP.	3	21.2	-	1			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	Homodimer (By similarity)		M
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		Cytoplasmic	J
yajb_ecoli	P21515	Hypothetical protein yajB.	3	23.0	-	0			S
yqji_ecoli	Q46872	Hypothetical protein yqjI.	3	23.4	-	0			K

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
crp_ecoli	P03020	Catabolite gene activator (cAMP receptor protein) (cAMP-regulatory protein).	3	23.6	b.82; a.4		Binds DNA as a dimer		T
ftse_ecoli	P10115	Cell division ATP-binding protein ftsE.	3	24.4	c.37	1			D
gch1_ecoli	P27511	GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).	3	24.7	d.96	1	Homodecamer, composed of a dimer of pentamers.		H
trmb_ecoli	P32049	tRNA (guanine-N(7)-)-methyltransferase (EC 2.1.1.33) (tRNA(m7G46)- methyltransferase).	3	27.3	c.66	0	Monomer		J
fucr_ecoli	P11554	L-fucose operon activator.	3	27.4	a.4; c.35	0			K
pfla_ecoli	P09374	Pyruvate formate-lyase 1 activating enzyme (EC 1.97.1.4) (PFL- activating enzyme).	3	28.1	-	0		Cytoplasmic	O
trmd_ecoli	P07020	tRNA (Guanine-N(1)-)-methyltransferase (EC 2.1.1.31) (M1G- methyltransferase) (tRNA [GM37] methyltransferase).	3	28.4	-	1	Monomer	Cytoplasmic (Potential)	J
glcc_ecoli	P52072	Glc operon transcriptional activator.	3	28.8	a.4	0			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	Monomer		G
ycfh_ecoli	P37346	Putative deoxyribonuclease ycfH (EC 3.1.21.-).	3	29.8	c.1	0			L

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
yafd_ecoli	P30865	Hypothetical protein yafD.	3	30.0	d.151	0		Cytoplasmic (Potential)	S
gaty_ecoli	P37192	Tagatose-1,6-bisphosphate aldolase gatY (EC 4.1.2.-) (TBPA).	3	30.8	c.1	0			G
dapa_ecoli	P05640	Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDPS).	3	31.3	c.1	1	Homotetramer	Cytoplasmic	E
amia_ecoli	P36548	Probable N-acetylmuramoyl-L-alanine amidase amiA precursor (EC 3.5.1.28).	3	31.4	-	0			M
end4_ecoli	P12638	Endonuclease IV (EC 3.1.21.2) (Endodeoxyribonuclease IV).	3	31.5	c.1	0	Monomer		L
ypt1_ecoli	P29368	Hypothetical 31.7 kDa protein in TRAX-FINO intergenic region.	3	31.8	c.69				R
yneb_ecoli	P76143	Putative aldolase yneB (EC 4.2.1.-).	3	31.9	c.1	0			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	Homotetramer	Cytoplasmic	E
yhbj_ecoli	P33995	Hypothetical UPF0042 protein yhbJ.	3	32.5	c.37	0			R

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
metf_ecoli	P00394	5,10-methylenetetrahydrofolate reductase (EC 1.7.99.5).	3	33.1	c.1	0	Homotetramer		E
arac_ecoli	P03021	Arabinose operon regulatory protein.	3	33.4	b.82; a.4	0	Homodimer	Cytoplasmic	K
icia_ecoli	P24194	Chromosome initiation inhibitor (OriC replication inhibitor).	3	33.5	c.94; a.4	0	Behaves as an homodimer in solution.		K
dusc_ecoli	P33371	tRNA-dihydrouridine synthase C (EC 1.-.-.-).	3	35.2	c.1	0			J
hem2_ecoli	P15002	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen synthase) (ALAD) (ALADH).	3	35.5	c.1	1	Homooctamer		H
dusb_ecoli	P25717	tRNA-dihydrouridine synthase B (EC 1.-.-.-).	3	35.9	c.1	0			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			J
lipa_ecoli	P25845	Lipoic acid synthetase (Lip-syn) (Lipoate synthase).	3	36.1	-	0	Monomer or homodimer	Cytoplasmic	H
add_ecoli	P22333	Adenosine deaminase (EC 3.5.4.4) (Adenosine aminohydrolase).	3	36.4	c.1	0			F
yajo_ecoli	P77735	Hypothetical oxidoreductase yajO (EC 1.-.-.-).	3	36.4	c.1	0			C

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
ltae_ecoli	P75823	Low-specificity L-threonine aldolase (EC 4.1.2.5) (Low-specificity L- TA).	3	36.5	c.67	0	Homotetramer (Probable)		E
nagz_ecoli	P75949	Beta-hexosaminidase (EC 3.2.1.52) (N-acetyl-beta-glucosaminidase) (Beta-N-acetylhexosaminidase).	3	37.6	c.1	0	Monomer (Potential)	Cytoplasmic	G
yfif_ecoli	P33635	Hypothetical tRNA/rRNA methyltransferase yfiF (EC 2.1.1.-).	3	37.8	c.116	0			J
alf1_ecoli	P71295	Fructose-bisphosphate aldolase class I (EC 4.1.2.13) (FBP aldolase).	3	38.0	-	0	Homooctamer or homodecamer	Cytoplasmic (Probable)	G
ybjs_ecoli	P75821	Hypothetical protein ybjS.	3	38.1	c.2	1			M
alr2_ecoli	P29012	Alanine racemase, catabolic (EC 5.1.1.1).	3	38.8	c.1; b.49	0			M
yjju_ecoli	P39407	Hypothetical protein yjjU.	3	39.8	-	0			R
dhas_ecoli	P00353	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA dehydrogenase) (ASADH).	3	40.0	c.2; d.81	1	Homodimer		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		Cytoplasmic	E

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
phol_ecoli	P77349	PhoH-like protein.	3	40.7	-	0		Cytoplasmic (Potential)	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	Tetramer of two alpha chains and two beta chains	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	Homodimer	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	Homodimer		H
metk_ecoli	P04384	S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine adenosyltransferase) (AdoMet synthetase) (MAT).	3	41.8	d.130	1	Homotetramer	Cytoplasmic	H
trma_ecoli	P23003	tRNA (Uracil-5-)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)- methyltransferase) (RUMT).	3	42.0		0			J

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
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			J
arge_ecoli	P23908	Acetylornithine deacetylase (EC 3.5.1.16) (Acetylornithinase) (AO) (N-acetylornithinase) (NAO).	3	42.3	d.58; c.56		Homodimer	Cytoplasmic (Probable)	E
lldd_ecoli	P33232	L-lactate dehydrogenase (Cytochrome) (EC 1.1.2.3).	3	42.7	c.1	0			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	Homodimer		I
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		Cytoplasmic	E
thih_ecoli	P30140	Thiazole biosynthesis protein thiH.	3	43.2	-				H
ints_ecoli	P37326	Putative prophage CPS-53 integrase.	3	44.1	d.163				L
csdb_ecoli	P77444	Selenocysteine lyase (EC 4.4.1.16) (Selenocysteine reductase) (Selenocysteine beta-lyase) (SCL).	3	44.4	c.67	0	Homodimer		E

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
yfbq_ecoli	P77727	Probable aminotransferase yfbQ (EC 2.6.1.-).	3	45.5	c.67		Homodimer (By similarity)	Cytoplasmic (By similarity)	E
rspa_ecoli	P38104	Starvation sensing protein rspA.	3	46.0	d.54; c.1	0			H
gatz_ecoli	P37191	Putative tagatose 6-phosphate kinase gatZ (EC 2.7.1.144).	3	47.1	-	0			G
typh_ecoli	P07650	Thymidine phosphorylase (EC 2.4.2.4) (TdRPase).	3	47.2	d.41; a.46; c.27	0	Homodimer		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	Heterodimer of a small and a large subunit	Inner membrane-bound	E
pmba_ecoli	P24231	PmbA protein (TldE protein).	3	48.4	-	0		Cytoplasmic	O
eutb_ecoli	P19635	Ethanolamine ammonia-lyase heavy chain (EC 4.3.1.7) (Ethanolamine ammonia-lyase large subunit).	3	49.4	-	0	Heterodimer of two nonidentical chains		C
xyla_ecoli	P00944	Xylose isomerase (EC 5.3.1.5) (D-xylulose keto-isomerase).	3	49.7	c.1	0	Homotetramer	Cytoplasmic	G
rhle_ecoli	P25888	Putative ATP-dependent RNA helicase rhlE.	3	50.0	c.37	0	Interacts with pcnB.		L

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional 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			E
tldd_ecoli	P46473	TldD protein.	3	51.4	-	0			O
uxac_ecoli	P42607	Uronate isomerase (EC 5.3.1.12) (Glucuronate isomerase) (Uronic isomerase).	3	54.0	c.1				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	Homohexamer		E
araa_ecoli	P08202	L-arabinose isomerase (EC 5.3.1.4).	3	56.1	-	0			G
aldb_ecoli	P37685	Aldehyde dehydrogenase B (EC 1.2.1.22) (Lactaldehyde dehydrogenase).	3	56.3	c.82				C
dhsa_ecoli	P10444	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1).	3	64.4	a.7; c.3; d.168	0	Part of an enzyme complex containing four subunits: a flavoprotein, an iron-sulfur, cytochrome b-556, and an hydrophobic anchor protein.		C

SwissProt Entry Name	Swiss Prot Accession Number	Protein Description	Predicted GroEL Substrate Class	Molecular Mass [kDa]	SCOP Fold	Essen-tiality	Oligomeric state (Swiss Prot Entry)	Subcellular Localization (SwissProt Entry)	COG Func-tional Category
frda_ecoli	P00363	Fumarate reductase flavoprotein subunit (EC 1.3.99.1).	3	65.8	a.7; c.3; d.168	0	forms part of an enzyme complex containing four subunits: a flavoprotein, an iron-sulfur, and two hydrophobic anchor proteins.		C
dead_ecoli	P23304	Cold-shock DEAD-box protein A (ATP-dependent RNA helicase deaD).	3	70.4	c.37	0		Cytoplasmic (Probable)	L
ycby_ecoli	P75864	Hypothetical protein ycbY.	3	78.9	c.66	0			L
parc_ecoli	P20082	Topoisomerase IV subunit A (EC 5.99.1.-).	3	83.8	e.11	1	Composed of two subunits: parC and parE.	Membrane-associated	L
phsm_ecoli	P00490	Maltodextrin phosphorylase (EC 2.4.1.1).	3	90.3	c.87	0	Homodimer		G

Lebenslauf

Persönliche Daten

Name	Tobias Maier
Anschrift	Helfenbergstr. 13
	74182 Obersulm
E-mail	TMaier@biochem.mpg.de
Geburtsdatum	9. Januar 1976
Geburtsort	Heilbronn
Familienstand	Ledig, kinderlos
Staatsbürgerschaft	Deutsch

Werdegang

1982-1995	Schulausbildung an der Grundschule Affaltrach und am Justinus-Kerner-Gymnasium in Weinsberg. Allgemeine Hochschulreife.
1995-1997	Grundstudium Diplombiologie an der Carl-von-Ossietzky-Universität Oldenburg. Diplomvorprüfung.
1997-1998	Studiengang Mikrobiologie and der University of East Anglia in Norwich/ Grossbritanien. Bachelor in Mikrobiologie.
1998-2000	Hauptstudium Diplombiologie an der Universität Konstanz. Diplomprüfung.
2000-2001	Diplomarbeit bei Professor Michael Ehrmann: „Periplasmic proline isomeraes in *Escherichia coli*“ an der Cardiff University in Cardiff/ Grossbritanien.
2002-2005	Dissertation bei Professor F. Ulrich Hartl: „Proteome-Wide Analysis of Chaperonin-Dependent Protein Folding in *Eschrichia coli*“ am Max-Planck-Institut für Biochemie in Martinsried.

Publikationen

M.J. Kerner, D.J. Naylor, Y. Ishihama, **T. Maier**, H.C. Chang, A.P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann and F.U. Hartl. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell, Vol. 120 pp. 209-220 (Juli 2005)

T. Maier, M.J. Kerner, D.J. Naylor, Y. Ishihama, A.P. Stines, M. Mann, D. Frishman, M. Hayer-Hartl, and F.U. Hartl. *In viv*o susbtrates of the GroEL/GroES chaperonin system. Poster Präsentation beim 56. Mosbacher Kolloquium der Deutschen Gesellschaft für Biochemie „Molecular Machines" (März 2005).

M.J. Kerner, **T. Maier**, Y.C. Tang, A. Hirtreiter, B. Haas, L. Figueiredo, D. Klunker, D.J. Naylor, A.P. Stines, M. Hayer-Hartl und F.U. Hartl. Mechanism of chaperonin assisted protein folding, Symposium des SFB 594 in München (November 2003).

T. Maier, M.J. Kerner, D.J. Naylor, Y. Ishihama, A.P. Stines, M. Mann, D. Frishman, M. Hayer-Hartl, and F.U. Hartl. *In viv*o substrate proteins of the *E. coli* GroEL/GroES chaperonin system. Poster Präsentation bei der EURESCO und FEBS Konferenz „Biology of Molecular Chaperones" in Tomar/ Portugal (August 2003).

Danksagung

Die vorliegende Arbeit wurde in der Zeit von Februar 2002 bis Oktober 2005 am Max-Planck-Institut für Biochemie in Martinsried angefertigt.

Mein besonderer Dank gilt Professor Dr. F. Ulrich Hartl für seine professionelle Betreuung, die immer währende Diskussionsbereitschaft und für die Bereitstellung eines interessanten Themas und hervorragender Arbeitsbedingungen. Bei Dr. Manajit Hayer-Hartl möchte ich mich für ihre freundliche, kollegiale Art bedanken und für die Unterstützung bei der Konzeption einiger Experimente.

Professor Dr. Matthias Mann, Dr. Yasushi Ishihama und Morten Kirkegaard danke ich für die intensive und sehr zeitaufwendige massenspektrometrische Analyse der Proben.

Ich bedanke mich weiter bei Michael Kerner und bei Dr. José M. Barral für die mehr als dreijährige sehr gute und kameradschaftliche Zusammenarbeit und die gründliche Prüfung und Korrektur des Manuskripts der Doktorarbeit. Beide waren nicht nur gute Kollegen, sondern sind auch gute Freunde geworden.

Weiter gilt mein Dank vielen Kollegen in der Arbeitsgruppe. Besonders Christian Behrends, Angela Hirtreiter und Luis Figueiredo sind mir nicht nur beruflich sonder auch privat verbunden.

Besonderer Dank gilt meinen Eltern Barbara Maier und Albrecht Maier für ihre fortwährende Unterstützung in allen Lebenslagen.

Schließlich möchte ich mich auch bei meiner Partnerin Sara Tomás bedanken, die fast vier Jahre auf mich in Spanien gewartet hat und mit der ich jetzt einen neuen Lebensabschnitt in Barcelona beginnen werde.