Cell, Vol. 122, 209–220, July 29, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.05.028

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

Michael J. Kerner,[1,6,7] Dean J. Naylor,[1,6,8] Yasushi Ishihama,[2,6,9] Tobias Maier,[1,6] Hung-Chun Chang,[1] Anna P. Stines,[1] Costa Georgopoulos,[3] Dmitrij Frishman,[4,5] Manajit Hayer-Hartl,[1] Matthias Mann,[2,10] and F. Ulrich Hartl[1,*]

[1]Department of Cellular Biochemistry
Max Planck Institute of Biochemistry
Am Klopferspitz 18
D-82152 Martinsried
Germany

[2]Center for Experimental BioInformatics
Department of Biochemistry and Molecular Biology
University of Southern Denmark–Odense
Campusvej 55
DK-5230 Odense M
Denmark

[3]Département de Microbiologie et Médecine Moléculaire
Centre Médicale Universitaire
1 rue Michel-Servet
CH-1211 Geneva
Switzerland

[4]Institute for Bioinformatics
GSF-National Research Center for Environment and Health
Ingolstädter Landstraße 1
D-85764 Neuherberg
Germany

[5]Department of Genome-Oriented Bioinformatics
Wissenschaftszentrum Weihenstephan
Technische Universität München
Am Forum 1
D-85354 Freising
Germany

Summary

The *E. coli* chaperonin GroEL and its cofactor GroES promote protein folding by sequestering nonnative polypeptides in a cage-like structure. Here we define the contribution of this system to protein folding across the entire *E. coli* proteome. Approximately 250 different proteins interact with GroEL, but most of these can utilize either GroEL or the upstream chaperones trigger factor (TF) and DnaK for folding. Obligate GroEL-dependence is limited to only ~85 substrates, including 13 essential proteins, and occupying more than 75% of GroEL capacity. These proteins appear to populate kinetically trapped intermediates during folding; they are stabilized by TF/DnaK against aggregation but reach native state only upon transfer to GroEL/GroES. Interestingly, substantially enriched among the GroEL substrates are proteins with $(\beta\alpha)_8$ TIM-barrel domains. We suggest that the chaperonin system may have facilitated the evolution of this fold into a versatile platform for the implementation of numerous enzymatic functions.

*Correspondence: uhartl@biochem.mpg.de
[6]These authors contributed equally to this work.
[7]Present address: Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark, Kemitorvet 208, DK-2800 Lyngby, Denmark.
[8]Present address: CBio Ltd., Brisbane Technology Park, 85 Brandl Street, Eight Mile Plains, QLD 4113, Australia.
[9]Present address: Eisai Co., Ltd., 5-1-3 Tokodai, Tsukuba, Ibaraki 300-2635, Japan.
[10]Present address: Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany.

Introduction

Many newly synthesized proteins rely on assistance by molecular chaperones to reach their native states efficiently and at a biologically relevant timescale. Molecular chaperones protect newly synthesized or stress-denatured polypeptides from misfolding and aggregating in the highly crowded cellular environment, often in an ATP-driven process (Frydman, 2001; Hartl and Hayer-Hartl, 2002). While the basic mechanisms of several major chaperone classes are well understood, the biological role of this machinery at a proteome-wide level remains to be defined. How many proteins in a cell have an absolute chaperone requirement for de novo folding? To what extent are these proteins dependent on a specific chaperone mechanism, and is such dependence linked to structural properties? The answers to these questions will help to define the contribution of molecular chaperones to overall protein biogenesis and whether they play a role in the structural evolution of their substrates.

In *E. coli*, trigger factor (TF) and the Hsp70 member DnaK have overlapping functions in stabilizing a wide range of translating polypeptides in a nonaggregated, folding-competent state. Neither component is absolutely essential for viability, but their combined deletion causes synthetic lethality at >30°C (Deuerling et al., 1999; Teter et al., 1999). Alternate chaperones, including GroEL/GroES, can partially compensate for the combined loss of TF and DnaK (Genevaux et al., 2004; Ullers et al., 2004; Vorderwülbecke et al., 2004). Indeed, the chaperonin GroEL (and its cofactor GroES) is the only *E. coli* chaperone that is essential for viability under all growth conditions tested (Fayet et al., 1989; Horwich et al., 1993). GroEL has been shown to act downstream of TF and DnaK in the posttranslational folding of ~10% of cytosolic proteins (Ewalt et al., 1997; Houry et al., 1999). We have previously identified ~50 of these proteins in GroEL immunoprecipitates (Houry et al., 1999), but the extent to which they require GroEL for folding has not yet been determined.

GroEL is an ~800 kDa complex with ATPase activity consisting of two stacked heptameric rings of 57 kDa subunits (Braig et al., 1994). Each ring provides a central cavity for the binding of nonnative protein via multiple interactions with hydrophobic surfaces on the

apical GroEL domains. GroES, a ring of seven 10 kDa subunits, associates with ATP bound GroEL and forms a lid on the GroEL cavity, thus causing the displacement of nonnative protein into a cage-like compartment (the *cis* cavity). Because GroES binding results in the burial of the hydrophobic surfaces of GroEL (Xu et al., 1997), enclosed polypeptide is then free to fold unimpaired by aggregation (Mayhew et al., 1996; Weissman et al., 1996). Nonnative proteins of up to ~60 kDa in size can be encapsulated, and their confinement in the GroEL/GroES cage may result in accelerated folding (Brinker et al., 2001). GroES dissociates from GroEL every 10–15 s in a reaction dependent on the GroEL ATPase, thus allowing for the release of folded substrate and the recapture of incompletely folded protein (reviewed in Fenton and Horwich, 2003; Hartl and Hayer-Hartl, 2002). Some proteins too large to be encapsulated can nevertheless utilize GroEL for folding by cycling on and off the GroEL ring in *trans* to bound GroES (Chaudhuri et al., 2001).

Here we describe the characterization of the GroEL-substrate proteome by a combination of biochemical analyses and quantitative proteomics. Approximately 250 of the ~2400 cytosolic *E. coli* proteins interact with GroEL in wild-type cells, and this number increases substantially in cells lacking the upstream chaperones TF and DnaK. However, only ~85 substrates exhibit an obligate dependence on GroEL for folding under normal growth conditions, occupying 75%–80% of the GroEL capacity. Proteins with $(\beta\alpha)_8$ triosephosphate isomerase (TIM) barrel domains are highly enriched among these substrates, suggesting a role for the chaperonin in the structural evolution of this widely distributed enzyme fold. On the other hand, the restriction of obligate GroEL dependence to less than 5% of cytosolic proteins indicates a high degree of folding robustness for the *E. coli* proteome, presumably resulting from an extensive overlap among chaperone functions.

Results

The set of ~50 previously identified GroEL interactors (Houry et al., 1999) contains a number of proteins for which functional enzyme assays are available (see Table S1 in the Supplemental Data available with this article online). We studied the GroEL requirement of these proteins for acquisition of the enzymatically active state in refolding experiments in vitro. Dihydrodipicolinate synthase (DAPA) was included in this analysis since it was independently suggested to be a GroEL substrate (McLennan and Masters, 1998). The selected proteins could be tentatively grouped into three classes with an increasing requirement for GroEL.

Class I and II Substrates Are Only Partially Chaperonin Dependent

The abundant enzyme enolase (ENO; 46 kDa) exhibited a low propensity to aggregate upon dilution from denaturant and, consequently, only a partial chaperone requirement for refolding in vitro. A similar behavior was found for glyceraldehyde-3-phosphate dehydrogenase (G3P1; 35 kDa), another abundant glycolytic enzyme previously identified as a GroEL interactor (Houry et al., 1999). These proteins were included in class I. As shown representatively for ENO, upon dilution from denaturant under standard conditions (37°C, 0.5 μM final concentration), ~55% of enzyme activity was recovered within 1 min in the absence of chaperones (Figure 1A). Nearly 100% of enzyme activity was regained when either the DnaK system (DnaK, DnaJ, GrpE) or GroEL/GroES or GroEL alone was added with Mg-ATP (Figure 1A).

In contrast, a second group of GroEL-interacting proteins (designated as class II), including glutamate decarboxylase α (DCEA; 53 kDa) and galactitol-1-phosphate 5-dehydrogenase (GATD; 37 kDa), failed to refold spontaneously under standard conditions (Figure 1B; Figure S1) due to their rapid aggregation (data not shown). The presence of both GroEL and GroES, but not GroEL alone, was necessary in assisting the refolding of these proteins at 37°C (Figure 1B). However, GroES was not absolutely required for refolding at 25°C where substantial spontaneous refolding was observed (data not shown), suggesting that DCEA and GATD are not obligate GroEL/GroES substrates. Indeed, the DnaK system was as efficient in mediating refolding at 37°C (Figure 1B), and an additive effect of both the DnaK and the GroEL system was observed with DCEA (Figure 1B) but not with GATD (Figure S1A).

In the case of threonyl-tRNA synthetase (SYT; 74 kDa), a protein too large to be encapsulated in the GroEL cavity, GroEL/GroES-assisted refolding was only ~20% efficient (Figure 1C). In contrast, the DnaK system supported 70% refolding at 37°C, without an additional increase in yield upon combining DnaK and GroEL/GroES (Figure 1C). Thus, it is likely that DnaK and GroEL share a number of substrates mainly in the preferred size range of GroEL (up to ~60 kDa), whereas larger proteins may generally be more adapted for folding by the DnaK system.

Class III Substrates Have an Obligate Requirement for GroEL

A third group of GroEL interactors was found to be stringently chaperonin dependent (class III proteins), including 5,10-methylenetetrahydrofolate reductase (METF; 33 kDa), S-adenosyl methionine synthetase (METK; 42 kDa), and DAPA (31 kDa) (Figure 1D; Figure S1). While METK and METF failed to refold spontaneously under a variety of conditions known to reduce aggregation, slow but efficient spontaneous refolding was observed for DAPA in the presence of 0.5 M arginine. GroEL/GroES accelerated this folding reaction ~10-fold (data not shown), similar to the GroEL model substrate bacterial RuBisCo (Brinker et al., 2001). Importantly, DAPA and METK are essential gene products (Gerdes et al., 2003), and disruption of the *metF* gene leads to methionine auxotrophy. An additional GroEL interactor, tagatose-1,6-bisphosphate aldolase (GATY, 31 kDa), could be assigned to class III by experiments in vivo (see below).

A distinct feature of METF, METK, and DAPA is that the DnaK system alone failed to mediate their refolding (Figure 1D; Figure S1). However, DnaK was able to efficiently bind and stabilize aggregation-prone, nonnative forms of these substrates and transfer them to GroEL

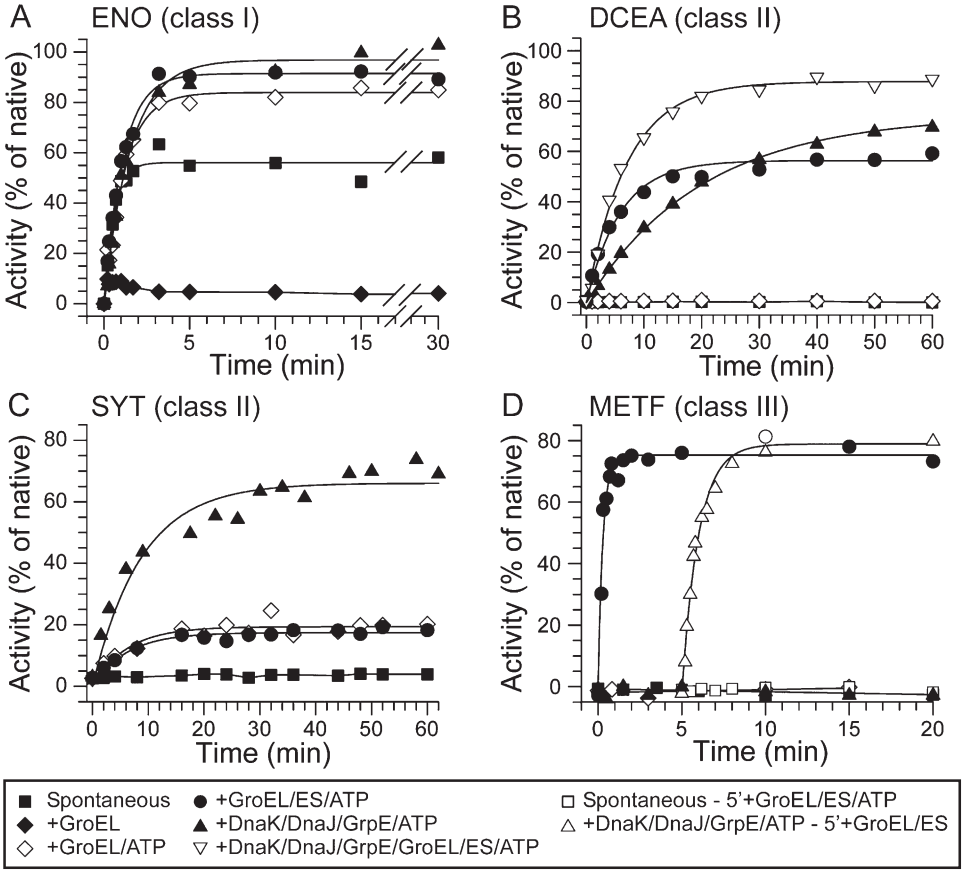

Figure 1. Spontaneous and Assisted Refolding of GroEL-Interacting Proteins

In vitro refolding of ENO (class I) (A), DCEA (B) and SYT (C) (class II), and METF (D) (class III) was analyzed upon dilution from denaturant at 37°C into buffer containing various combinations of chaperones and 5 mM ATP, as indicated, and was followed by measuring enzymatic activity. Refolding of METF was also analyzed upon dilution of the denatured protein into buffer containing DnaK/DnaJ/GrpE, followed by addition of GroEL/GroES after 5 min. The stoichiometry of components was 1 substrate:2 GroEL (14-mer):4 GroES (7-mer):5 DnaK:2.5 DnaJ:5 GrpE. The enzymatic activity of an equivalent amount of native substrate protein is set to 100%.

for subsequent folding. In contrast, upon dilution from denaturant into buffer lacking chaperones, these proteins lost their competence for GroEL-assisted folding within minutes due to aggregation (Figure 1D). These findings suggest that METF, METK, and DAPA populate aggregation-sensitive folding intermediates and require the specific folding environment provided by GroEL/GroES to progress to their native state. Binding to DnaK may both function as a reservoir for these substrates and facilitate their efficient capture by GroEL.

Dependence of Substrates on GroEL In Vivo

We next sought to confirm the validity of our GroEL-substrate classification in vivo. Proteins were overexpressed at 37°C in *E. coli* cells containing wild-type (wt) or ~5-fold elevated levels of chaperonin (Figure S2). ENO and G3P1 (class I) were essentially soluble in wt cells, consistent with their chaperone independence in vitro. In contrast, DCEA and GATD (class II) and METK, METF, DAPA, and GATY (class III) were 60%–70% insoluble. Elevating the levels of GroEL/GroES caused a ~2- to 3-fold increase in solubility for these proteins, whereas overexpression of GroEL alone, equivalent to a relative depletion of GroES, tended to reduce solubility. The 74 kDa class II protein SYT was also partially insoluble in wt cells but was unaffected by GroEL/GroES expression, consistent with the limited efficiency of GroEL/GroES to assist SYT refolding in vitro.

The fate of these proteins at their endogenous levels was examined upon GroEL/GroES depletion, employing cells in which the *groE* promoter was exchanged by the arabinose-controlled *pBAD* promoter (McLennan and Masters, 1998). Upon shifting these cells from arabinose to glucose, GroEL levels decreased by ~90% within 3 hr, while cell growth continued for ~8 hr. Class I proteins remained soluble throughout GroEL/GroES depletion, as shown for ENO (Figure 2). Similarly, class II substrates GATD and DCEA were not affected in their solubility (Figure 2 and data not shown). GATD showed a nonuniform expression behavior during the time course of the experiment, a phenomenon linked to the change of media (Nobelmann and Lengeler, 1996). SYT was expressed uniformly and was only partially insoluble upon prolonged chaperonin depletion (Figure 2). In contrast, class III substrates showed an absolute chap-

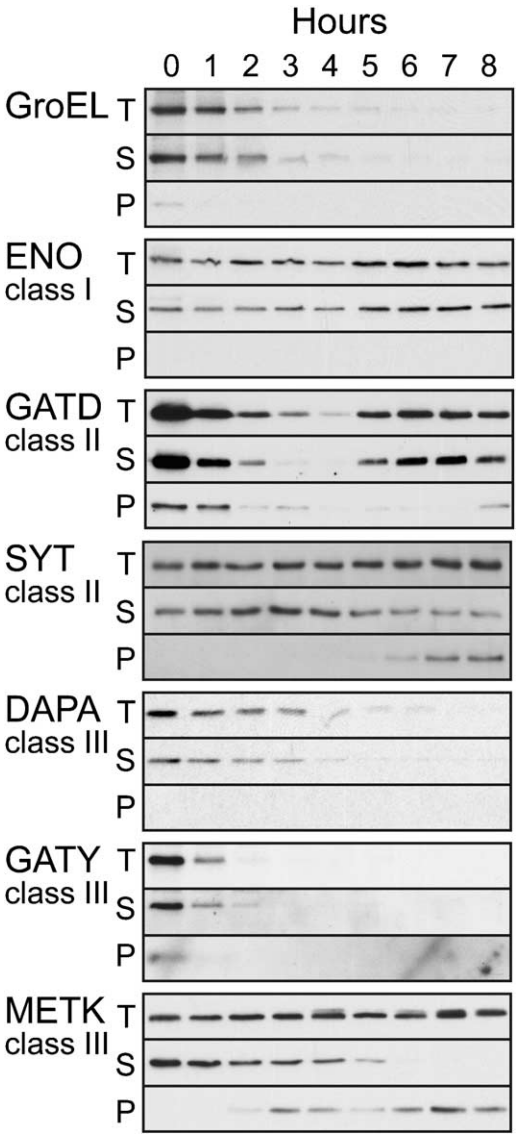

Figure 2. Dependence of Substrates on GroEL for Folding In Vivo

GroEL/GroES was depleted in *E. coli* MC4100 cells carrying the groE operon under an arabinose-regulated promoter. Cells grown at 37°C in arabinose medium were shifted to glucose-containing medium (t = 0 hr). At the indicated times, equivalent amounts of cells were taken for preparation of total (T), soluble supernatant (S), and insoluble pellet (P) fractions. Proteins were detected by immunoblotting.

eronin requirement. As a consequence of GroEL/GroES depletion, DAPA disappeared from the total and soluble fraction without accumulating in the insoluble fraction (Figure 2), suggesting that, at endogenous levels, this protein can be rapidly degraded when unable to fold. Rapid disappearance was also observed for GATY. A different behavior was noted for METK, which accumulated as aggregates upon GroEL/GroES depletion (Figure 2). These observations validate the classification of GroEL-interacting proteins based on their in vitro refolding properties.

Isolation of GroEL/GroES Complexes with Encapsulated Substrates

A comprehensive identification and characterization of GroEL interactors was undertaken to extend the classification of GroEL substrates to the entire *E. coli* proteome. GroEL-associated proteins were trapped within the folding-active *cis* cavity of GroEL under the lid of a fully functional, C-terminally His_6-tagged GroES (Figure 3A). The GroEL/GroES-His_6 complexes were fixed in the ADP bound state upon lysis of live spheroplasts in the presence of glucose and hexokinase to rapidly (in <3 s) convert cellular ATP to ADP, followed by isolation by immobilized metal affinity chromatography (IMAC). GroEL interactors contained in slices of one-dimensional SDS gels were digested with trypsin, and the resulting peptides were separated by liquid chromatography coupled to Q-TOF tandem mass spectrometry (LC-MS/MS) for identification (Lasonder et al., 2002). GroEL/GroES complexes formed with *E. coli* GroES-His_6 proved to be of limited stability during isolation, raising the possibility of postlysis loss or exchange of substrates. However, efficient recovery of GroEL complexes was achieved upon short-term expression of the highly similar GroES of *Methanosarcina mazei* (Mm) (Figure 3B). MmGroES can functionally replace *E. coli* GroES in vivo but was found to bind more stably to GroEL in the presence of ADP (Figueiredo et al., 2004; Klunker et al., 2003).

Around 300 different proteins were repeatedly identified in the isolated GroEL/GroES complexes (Figure 3B, lanes 1 and 2). In cells not expressing GroES-His_6, no GroEL was isolated (Figure 3B, lane 3), but LC-MS/MS identified seven proteins to be nonspecifically bound to IMAC beads (Table S2). To identify proteins interacting with GroEL during and after cell lysis, cells expressing MmGroES-His_6 were mixed with wt cells that had been isotope labeled with leucine-D3 (see SILAC below) and lysed together in the presence of glucose and hexokinase. Upon isolation of GroEL/GroES and LC-MS/MS, a total of 32 Leu-D3-labeled proteins could be identified as nonspecific GroEL interactors, including 6 of the 7 proteins found to bind to IMAC beads and 20 ribosomal proteins (Table S2). For this reason, ribosomal proteins were excluded from further analysis.

Proteins enclosed in the *cis* ring of GroEL are protected from externally added Proteinase K (PK) by GroES, whereas proteins bound to the *trans* ring are degraded, as are the flexible C-terminal tails of the GroEL subunits in that ring (Figure 3A) (Mayhew et al., 1996; Weissman et al., 1995). As expected, when isolated GroEL/GroES complexes were treated with PK, only half of the C termini of GroEL were cleaved (Figure 3C). Western blot analysis revealed that all substrates in the original test set of <60 kDa were protease protected, indicating efficient encapsulation under the GroES lid, whereas the same substrates in their native states, not bound to GroEL, were either partially or completely proteolyzed (Figure 3C and data not shown). In contrast, GroEL-associated proteins of >60 kDa such as SYT (74 kDa) were degraded (Figure 3C), indicative of binding to the *trans* GroEL ring.

Substantial amounts of the chaperones DnaK and

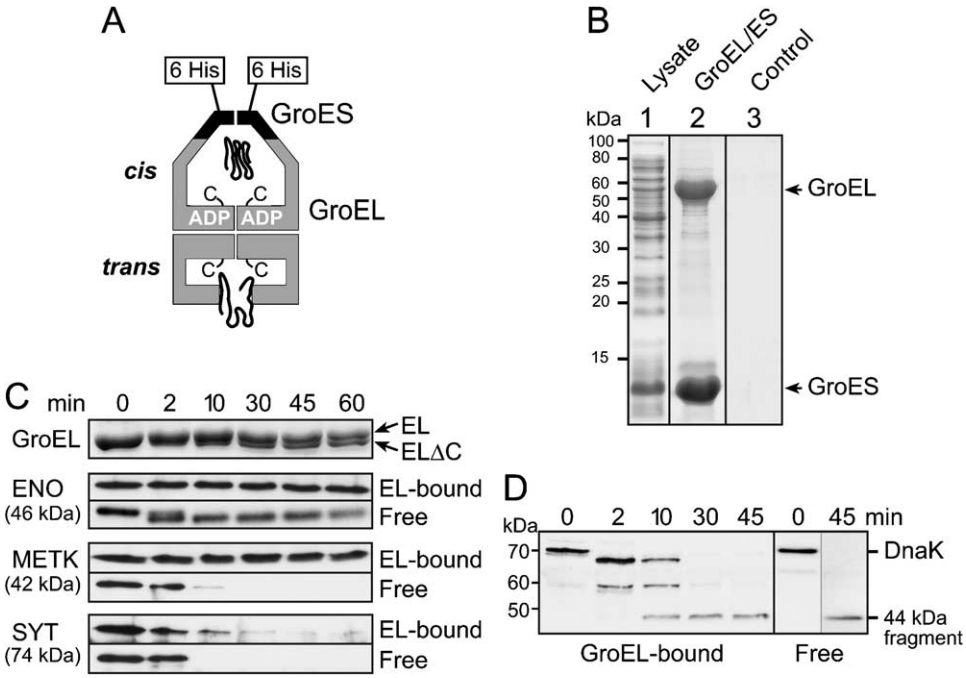

Figure 3. Isolation of In Vivo GroEL/GroES Substrates

(A) Schematic depiction of the capture of substrate proteins within the *cis* cavity of GroEL/GroES stabilized in the ADP bound state. Proteins too large to be encapsulated are expected to be bound to the GroEL ring in *trans* to GroEL. GroEL/GroES/substrate complexes were isolated utilizing C-terminal His_6 tags on GroES.

(B) Purification of GroEL/GroES complexes. *E. coli* MC4100 spheroplasts expressing *M. mazei* GroES-His_6 were lysed in presence of glucose and hexokinase to rapidly convert all cellular ATP to ADP (see Experimental Procedures and Supplemental Experimental Procedures). GroEL/GroES/substrate complexes contained in a soluble cell extract (lane 1) were bound to an IMAC column and eluted with imidazole (lane 2). To identify proteins interacting nonspecifically with the IMAC beads, GroEL/GroES complexes were prepared from cells expressing nontagged GroES (lane 3). Fractions were subjected to 16% SDS-PAGE, followed by Coomassie blue staining.

(C and D) Encapsulation of proteins in the GroEL *cis* cavity and binding to the GroEL *trans* ring. GroEL/GroES/substrate complexes and native control proteins not bound to GroEL (free) were incubated with Proteinase K (PK) at 25°C for the times indicated, followed by SDS-PAGE and silver staining (GroEL) or immunoblotting for the proteins indicated in (C) and (D).

DnaJ were also specifically associated with the GroEL complexes. Incubation of GroEL/GroES/substrate complexes with PK resulted in the production of the 44 kDa ATPase domain of DnaK, similar to native DnaK not associated with GroEL (Figure 3D). This indicates that DnaK (69 kDa) is not encapsulated in the GroEL *cis* ring but rather interacts, as a functional chaperone, with unfolded substrates bound to GroEL in *trans*, consistent with its ability to stabilize certain proteins for subsequent interaction with GroEL.

Overview of the Proteomic Data Sets

A total of ~250 proteins were reproducibly identified as specifically associated with GroEL at 30°C and 37°C (Table S3 and http://pedant.gsf.de). LC-MS/MS also identified 1132 proteins out of the ~2400 proteins predicted to be present in the soluble cell lysates (Frishman et al., 2003). These proteins vary as much as 10,000-fold in abundance, as indicated by their exponentially modified protein abundance index (emPAI) (Figure 4B). This value is based on the number of different peptides of a specific protein identified by MS and provides an estimate for abundance comparable or superior to conventional staining techniques (Rappsilber et al., 2002 and Supplemental Experimental Procedures). Assuming that the ~1300 undetected lysate proteins have very low abundance values (<50 ppm), these proteins would contribute less than 7% to total soluble protein by mass. The data set of GroEL substrates was estimated to be essentially complete based on the following criteria: (1) the number of identified proteins did not increase in repeated analyses; (2) analysis by more sensitive FT-MS did not significantly increase the number of GroEL substrates identified (data not shown); and (3) an additional ~150 GroEL interactors were identified by the same experimental protocol in cells lacking TF and DnaK, indicating sufficient sensitivity of the method in wt cells.

The identified GroEL interactors are between 10 and 150 kDa in size and contain all the proteins of the initial test set selected for in vitro analysis (Houry et al., 1999). They are almost exclusively cytosolic, except for eight proteins of the periplasm and outer membrane (OmpA and OmpC), and comprise proteins of all major functional categories (Figure S3A and Table S3). No predicted membrane-spanning proteins of the inner membrane but several membrane-associated proteins were

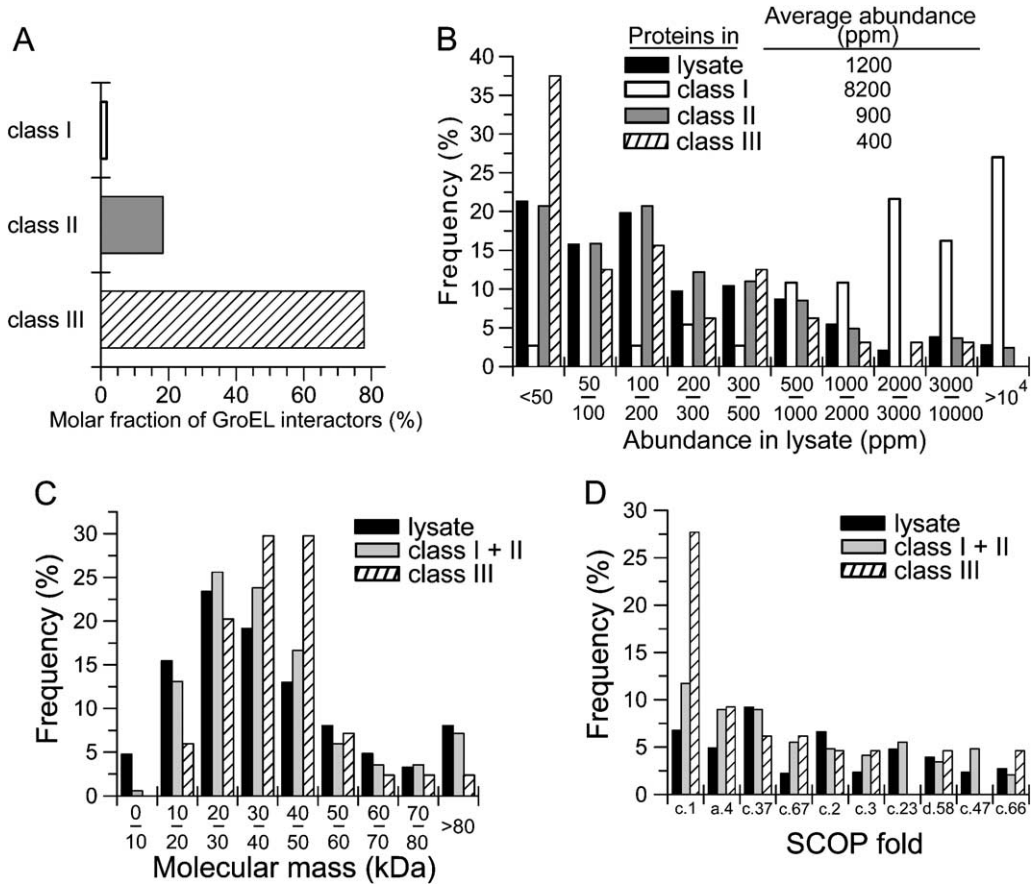

Figure 4. Properties of the GroEL-Interacting Proteins

(A) Molar fraction of GroEL interactors of classes I–III in isolated GroEL/GroES complexes, as based on cumulative abundance values of GroEL interactors determined by MS.

(B) Abundance distribution of total *E. coli* soluble proteins and GroEL-interacting proteins of classes I–III determined in soluble lysate.

(C and D) Distribution of molecular mass (C) and SCOP fold (D) in *E. coli* lysate proteins (lysate) and classes of GroEL-interacting proteins as indicated. SCOP fold abbreviations: c.1, TIM β/α barrel; a.4, DNA/RNA binding 3-helical bundle; c.37, P loop containing nucleotide triphosphate hydrolases; c.67, PLP-dependent transferases; c.2, NAD(P) binding Rossmann fold domains; c.3, FAD/NAD(P) binding domain; c.23, flavodoxin-like; d.58, ferredoxin-like; c.47, thioredoxin fold; c.66, S-adenosyl-L-methionine-dependent methyltransferases.

detected, including subunits of the peripheral sector of NADH-quinone oxidoreductase. The set of GroEL interactors includes 67 from a total of 620 essential *E. coli* proteins (Gerdes et al., 2003) (Figure S3B).

Enrichment of Obligate Substrates among GroEL Interactors

Based on quantitative proteomic analysis by SILAC (*s*table *i*sotope *l*abeling by *a*mino acids in *c*ell culture) (Ong et al., 2002), we deduced criteria to assign the ~250 GroEL interactors to the three substrate classes found in the initial set of test proteins. The validated substrates described above were used as standards. Under normal growth conditions, the cytosolic concentration of GroEL complexes is limited to approximately one-tenth of ribosomes (Ellis and Hartl, 1996). Thus, it seemed likely that class III substrates would be enriched in GroEL complexes relative to proteins of classes I and II. To test this hypothesis, cell lysates from Arg-^{13}C6-labeled cells were mixed with known amounts of isolated, unlabeled GroEL/GroES/substrate complexes. The GroEL-associated fraction of the substrates relative to total was derived from the intensity ratio of pairs of unlabeled and labeled peptides measured by MS (Ong et al., 2002) (Figure S4). The validated class III proteins METK, METF, and GATY were among the most highly enriched substrates, with 3%–6% of their total cellular content being GroEL associated. More than 3% of a protein is expected to be GroEL associated when all folding must proceed via GroEL (assuming a doubling time of *E. coli* of 30–40 min and an average half-time ($t_{1/2}$) of GroEL-assisted folding of ~60 s; see Ewalt et al., 1997). An additional 80 proteins accumulated in GroEL/GroES complexes to 4% or more of total and thus were assigned to class III (Table S3). Several of these proteins (ADD, END4, HEM2, NANA, XYLA, YAJO, LTAE, and TYPH) were analyzed functionally, confirming their GroEL dependence (see Figure 5A below and data not shown). Notably,

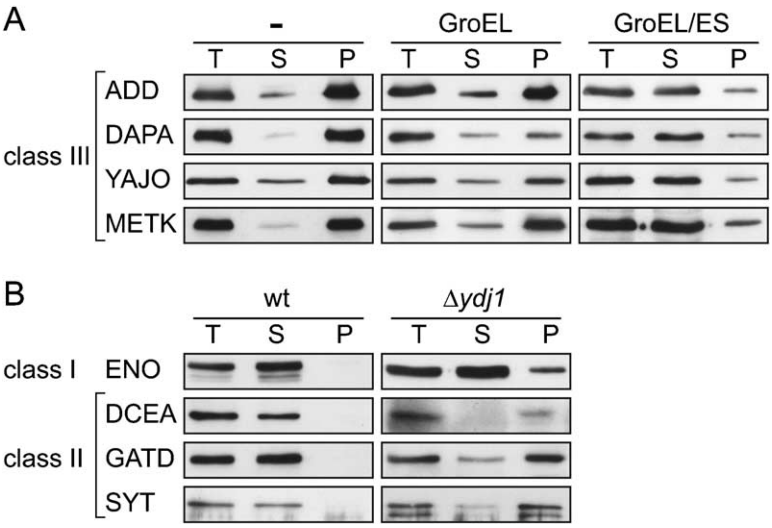

Figure 5. Solubility of *E. coli* GroEL Substrates upon Heterologous Expression in *S. cerevisiae*

(A) Solubility of class III proteins ADD, DAPA, YAJO, and METK, expressed at 30°C under galactose control in *S. cerevisiae* YPH499 wt cells either without (–) or with expression of GroEL or GroEL/GroES (see Experimental Procedures). Total (T), soluble (S), and insoluble pellet fractions (P) were analyzed by immunoblotting.

(B) Solubility of class I and II proteins in wt and Ydj1p-deficient yeast cells. ENO (class I) and DCEA, GATD, and SYT (class II) were expressed in strains DS10 (wt) and Δ*ydj1*. Cell fractions were analyzed as above.

based on their cumulative abundance values in GroEL complexes, the predicted class III substrates together contribute 75%–80% of the total mass of GroEL interactors (Figure 4A), despite being of low to intermediate overall abundance in the cytosol (Figure 4B). They include 13 essential enzymes of diverse function in amino acid and sugar metabolism, cell-wall synthesis, and other cellular pathways (Figure S3B and Table S3).

In contrast, for validated class I proteins, less than 0.02% of total was found to be GroEL associated (Figure S4), indicating that their folding is essentially GroEL independent. The substrate set contained ~40 other proteins sharing this property (Table S3). Most of these proteins are very abundant in the cytosol, but collectively they make up only 1%–3% of all GroEL interactors by mass (Figures 4A and 4B). The remaining ~130 GroEL interactors were tentatively grouped into class II. For these proteins, including GATD and SYT, between 0.1% and 2.6% of total was recovered on GroEL, indicative of partial GroEL dependence. These proteins are of average abundance and together occupy ~20% of the GroEL capacity (Figures 4A and 4B).

The molecular mass distribution of the predicted class III substrates is shifted to larger sizes compared to that of total lysate proteins and shows a sharp cutoff toward proteins of >50 kDa (Figure 4C), consistent with a dependence on the encapsulation mechanism for the vast majority of class III substrates. The other GroEL interactors do not show such a pronounced preference for a size that fits the GroEL/GroES cavity (Figure 4C).

The isoelectric point (pI) distribution of the predicted class III substrates deviates from that of total lysate proteins, with a greater fraction of the former exhibiting pI values between 5.5 and 6.5 (Figure S5A). Thus, at physiological pH, many class III proteins have a lower net charge than the bulk of cytosolic proteins, a property known to enhance the tendency of proteins to aggregate upon attempted refolding (Chiti et al., 2002). Compared to the other GroEL interactors, class III substrates show no apparent enrichment of hydrophobic amino acid residues and no significant sequence similarities.

Preference for the TIM-Barrel Fold among GroEL Substrates

Does the dependence on GroEL for folding correlate with a specific type of protein fold? To address this question, we performed a homology-based fold assignment for all GroEL interactors by querying the protein sequences against the SCOP database of structural domains (Lo Conte et al., 2002). The fold distribution of the experimentally identified lysate proteins was virtually identical to that of the genome-based *E. coli* proteome (Figure S5B). The GroEL interactors were found to be significantly enriched in the $(\beta\alpha)_8$ TIM-barrel fold (SCOP class c.1) (Figure 4D). This fold is shared by 6.8% of all lysate proteins with an identifiable structural homology (55 out of 814 lysate proteins) and by 7.6% of all proteins in the preferred size range of class III proteins (45 out of 595 lysate proteins). The complete set of GroEL interactors contains 17% protein sequences with TIM-barrel fold (35 out of 210), and the predicted class III substrates are further enriched in TIM-barrel proteins to 28% (18 out of 65 proteins with identifiable structure). Based on their cumulative abundance in GroEL complexes, TIM-barrel proteins contribute ~35% to the total mass of all GroEL substrates but only ~6% to overall protein mass in the cytosol. These results indicate a pronounced dependence of a subset of $(\beta\alpha)_8$ barrel proteins on the chaperonin system for effective folding.

The GroEL-interacting TIM-barrel representatives are between 23 and 54 kDa in size (30–54 kDa for predicted class III TIM barrels). Most of them are homo-oligomeric (dimeric to octameric) enzymes. They comprise 10 of the 26 known SCOP superfamilies of $(\beta\alpha)_8$ barrel proteins and consist mainly of the TIM-barrel domain with various small appendages and insertions (Figure S6) but share little or no sequence identity. Among these proteins are the validated class III substrates METF, DAPA, and GATY. GroEL dependence in vivo was experimentally demonstrated for several additional TIM-barrel representatives grouped as class III (see Figure 5A and data not shown). No other fold type was significantly enriched in the set of predicted class III sub-

strates (Figure 4D). Almost all of the non-TIM-barrel representatives contain α/β (SCOP class c; 49 proteins) or α+β domains (SCOP class d; 18 proteins), often in combination.

GroEL Requirement of Class III Proteins Is Independent of the Bacterial Folding Environment

Heterologous expression in the eukaryotic cytosol, which lacks a bacterial-type chaperonin, provides a stringent system to independently test the validity of the classification of newly synthesized GroEL substrates. A set of class III proteins, including METK and ten TIM-barrel substrates, 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 5A and data not shown). Thus, the requirement of the class III proteins for GroEL/GroES is specific and independent of the bacterial machinery of protein synthesis. In contrast, ENO (class I) as well as three class II proteins tested was soluble upon expression in wt yeast (Figure 5B). Substantial aggregation of the class II proteins was observed in the mutant strain *Δydj1* that lacks the yeast Hsp70 cofactor Ydj1p (Figure 5B), supporting the conclusion that class II proteins are chaperone dependent but can utilize either the Hsp70 system or GroEL/GroES for folding.

Mechanisms for Substrate Selection by GroEL

Since the TIM-barrel fold is widely distributed (Nagano et al., 2002), it cannot per se be the sole criterion for the GroEL/GroES dependence of a protein. Indeed, the abundant TIM-barrel protein ENO (class I) folds robustly in the absence of chaperonin (see Figure 1A). An extensive search for a more detailed common structural feature of the class III TIM-barrel substrates remained unproductive. This may suggest that the folding intermediates of these proteins, rather than their final structures, share characteristic features that confer GroEL dependence. Evidence in support of this hypothesis was obtained by competition GroEL binding experiments. We found that ENO bound efficiently to GroEL upon dilution from denaturant, based on the ability of GroEL to prevent spontaneous ENO refolding in the absence of ATP (see Figure 1A). However, even a 4-fold excess of ENO or the class II proteins DCEA and GATD resulted in only a minor reduction of GroEL binding for the class III TIM barrel DAPA, as observed following reisolation of GroEL complexes by gel filtration (Figure 6A). In the case of the non-TIM barrel class III substrate METK, competition for GroEL binding by ENO and the class II proteins was only slightly more effective (Figure 6A). Thus, proteins with an obligate chaperonin dependence populate nonnative states during folding with high affinity for GroEL, providing the basic mechanism for their enrichment in GroEL complexes.

Given the high cytosolic abundance of class I/II proteins relative to class III proteins (Figure 4B), we considered the possibility that the upstream chaperones TF and DnaK may facilitate the preferential selection of class III substrates by GroEL. To test this possibility, we identified the GroEL-interacting proteins in cells lacking either the DnaK system (Δ*dnaKJ*), TF (Δ*tig*), or both (Δ*dnaKJ*Δ*tig*) (Genevaux et al., 2004) at 30°C. The number and composition of GroEL substrates in Δ*dnaKJ* and Δ*tig* cells was similar to that found in wt cells (data not shown), consistent with the known functional over-

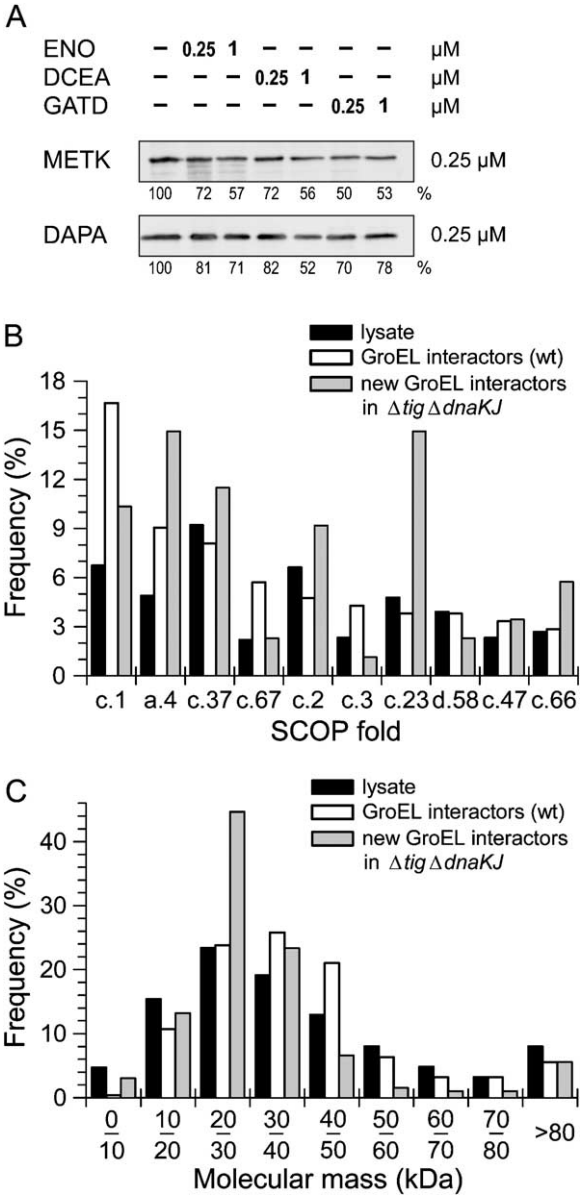

Figure 6. Mechanisms of Substrate Selectivity by GroEL

(A) Class III proteins (METK, DAPA) out-compete class I (ENO) and class II proteins (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 (see Supplemental Experimental Procedures).

(B and C) Distribution of SCOP fold (B) and molecular mass (C) in *E. coli* lysate proteins (lysate), GroEL interactors from wt cells (~250 proteins), and new GroEL interactors in Δ*tig*Δ*dnaKJ* cells (~150 proteins). See Figure 4D for SCOP fold abbreviations.

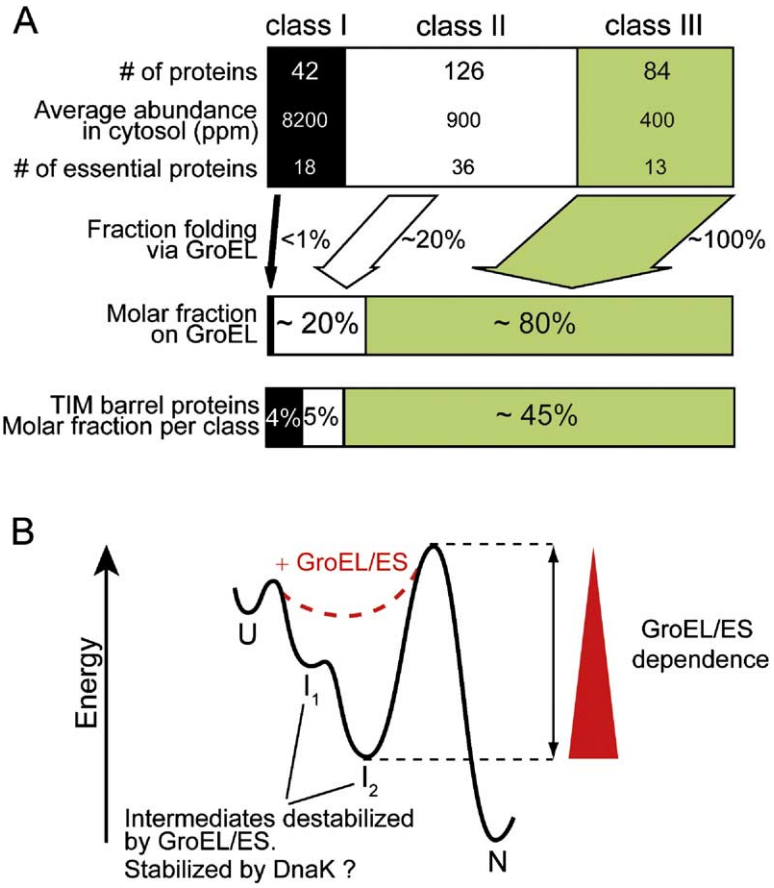

Figure 7. Contribution of GroEL/GroES to Overall Protein Folding

(A) Properties of GroEL-interacting proteins of classes I–III.

(B) Energy diagram for the folding of a hypothetical class III protein, illustrating the proposal that these proteins populate kinetically trapped folding intermediates that require GroEL/GroES to proceed to the native state. U, unfolded; I1 and I2, kinetically trapped intermediates; N, native state. See Discussion for details.

lap between TF and DnaK. However, an additional ~150 GroEL interactors were identified in Δ*dnaKJ*Δ*tig* cells. Proteins with TIM-barrel domains were not significantly enriched among these proteins (Figure 6B). We also found that several class III substrates partially aggregated in Δ*dnaKJ*Δ*tig* cells while being fully soluble in wt cells (data not shown), suggesting that, in the absence of TF/DnaK, these proteins fail to interact with GroEL effectively. The size distribution of GroEL substrates isolated from Δ*dnaKJ*Δ*tig* cells is shifted toward smaller sizes, similar to the distribution of total lysate proteins (Figure 6C). The additional GroEL interactors in Δ*dnaKJ*Δ*tig* cells are mostly of average abundance, comparable to class II substrates in wt cells. Thus, in the absence of TF and DnaK, the substrate selectivity of GroEL is reduced; GroEL assumes a more general role in folding, with an additional set of proteins competing with stringently GroEL-dependent substrates.

Discussion

Contribution of GroEL/GroES to Protein Folding

About 250 different proteins interact with GroEL under normal growth conditions of *E. coli*, consistent with previous estimates (Ewalt et al., 1997; Houry et al., 1999). However, 75%–80% of the available GroEL capacity is occupied by only ~85 substrates (Figure 7A). These proteins are mostly of low to intermediate abundance in the cytosol and are stringently GroEL dependent for folding (class III), whereas the other GroEL interactors are more abundant but have only a partial requirement for GroEL (classes I and II). The obligate GroEL substrates include at least 13 essential proteins, explaining why the chaperonin system is indispensable for *E. coli* viability. The limited set of class III proteins (less than 5% of total) probably defines the core cytosolic proteins with an obligate dependence on a specific chaperone mechanism, suggesting a high degree of folding robustness of the *E. coli* proteome as a result of an extensive functional redundancy among chaperone classes. Proteins with the $(\beta\alpha)_8$ TIM-barrel fold are highly enriched among class III substrates.

Upon GroEL depletion of *E. coli*, class III substrates either disappeared from the cells or accumulated in aggregates, equivalent to a loss of biological function. Between 3% and 6% of the total amount of a class III protein is associated with GroEL in growing cells. This finding strongly supports the conclusion that these proteins fold essentially completely via GroEL, based on the following considerations. The half-time of assisted folding for several class III proteins in vitro (30–60 s) (Figure 1) is in good agreement with the transit time through GroEL for the bulk of GroEL interactors in vivo (Ewalt et al., 1997). At this folding speed and at a doubling time of *E. coli* of 30–40 min, at least 3% of a given class III protein must be in the process of folding. This fraction should increase for proteins with slower folding

rates or for proteins that must return to GroEL during their lifetime (Houry et al., 1999). In contrast, on average, only ~20% of newly synthesized class II substrates and less than 1% of class I proteins fold via GroEL, based on the respective fraction of GroEL-associated protein (Figure 7A).

Properties of Obligate GroEL Substrates

Our analysis of the GroEL-substrate proteome allows us to deduce certain features of chaperonin-dependent proteins. Based on a large body of evidence, the degree of exposure of hydrophobic amino acids is considered the major determinant for the binding of nonnative proteins to GroEL (reviewed in Fenton and Horwich, 2003). The apical GroEL domains are known to bind extended hydrophobic β strands and amphiphilic α helices, but these redundant features are not significantly enriched among class III substrates compared to class I/II proteins of similar size. Neither do class III substrates contain an increased number of motifs resembling the mobile loop segment of GroES, which plays a role in displacing bound protein from the apical GroEL domains. Yet, upon dilution from denaturant, class III substrates out-compete other proteins for binding to GroEL. Thus, obligate substrates must expose hydrophobic GroEL recognition elements for longer periods during folding, and this would explain their pronounced tendency to aggregate, particularly when hydrophobic β strand regions are frequent. These considerations suggest that many class III substrates exhibit energetically frustrated folding pathways, i.e., they fold along rugged energy landscapes populating kinetically trapped intermediates that still expose substantial hydrophobic regions (Dobson et al., 1998) (Figure 7B).

Class III proteins are relatively large and have complex α/β and α+β domain topologies. Such proteins, including the $(\beta\alpha)_8$ TIM barrels, are stabilized by many long-range contacts and are predicted to have a marked propensity to populate kinetic intermediates during folding (Gromiha and Selvaraj, 2004). A relevant example is bacterial RuBisCo (Thirumalai et al., 2003), a well-studied GroEL model substrate with a TIM-barrel fold. On the other hand, proteins with very similar structures may nevertheless fold along substantially different pathways (Ferguson et al., 1999), and, therefore, it would appear that the degree of energetic frustration during folding, rather than the specific final structure, determines the GroEL dependence of a protein (Figure 7B). Our results suggest that only a subset of the *E. coli* TIM barrels populate off-pathway species that result in severe kinetic trapping during folding. We note that, due to their predominantly oligomeric nature, many of these proteins must fold into subunits still exposing substantial hydrophobic interfaces, and this would likely add to the ruggedness of their folding pathways.

Mechanism of GroEL/GroES in Class III Protein Folding

How does the chaperonin promote the folding of its obligate substrates, and why is the DnaK system unable to do so? GroEL and GroES provide a mechanism for the concerted release of bound substrate from multiple attachment sites into an enclosed cage in which the protein is free to fold, unimpaired by aggregation. In addition to this effect, the physical environment of the cage may promote the folding of many class III proteins by smoothing their folding energy landscape (Figure 7B). As shown for RuBisCo (50 kDa), enclosure of the nonnative protein inside GroEL/GroES results in not only aggregation prevention but also a substantial acceleration of folding (Brinker et al., 2001). This effect has been attributed to the entropic destabilization of certain kinetic folding intermediates inside the spatially confined environment of the cage (Brinker et al., 2001; Lin and Rye, 2004). Given the limited volume of this compartment, the "catalytic" effect of confinement on folding may be pronounced for many substrates in the size range of the class III TIM barrels (30–54 kDa), provided that their major transition state of folding is close to the native state in compactness and their local minima in the energy landscape correspond to more expanded conformations (Jewett et al., 2004; Thirumalai et al., 2003). In contrast to GroEL, the DnaK system does not provide a confined folding environment. Moreover, there is no mechanism for DnaK molecules to release bound peptide segments of a substrate protein in a coordinated fashion. These features would explain why the DnaK system is ineffective in promoting the folding of class III proteins (Figure 7B).

Interplay between the GroEL and the TF/DnaK Chaperone Systems

Our results indicate that TF and the DnaK system contribute to achieving the high degree of substrate selectivity by GroEL, in conjunction with the intrinsic folding properties of GroEL substrates. DnaK can effectively serve as a substrate filter; it stabilizes class III proteins in a nonaggregated state for productive interaction with GroEL while promoting the folding of class I/II proteins. As a consequence, in cells lacking both TF and DnaK, the number of GroEL interactors increases substantially, and the enrichment of class III substrates on GroEL is reduced, with several of these proteins aggregating partially. This result is consistent with the view that, in wt cells, TF and DnaK prevent an overloading of GroEL with class I /II proteins, thus ensuring that class III substrates reach GroEL efficiently. On the other hand, in the absence of TF/DnaK, GroEL/GroES assumes a broader role in folding.

Evolutionary Considerations

GroEL is largely devoted to assisting the folding of a rather small number of obligate substrates, with TIM-barrel proteins contributing ~45% by mass (Figure 7A). This surprising finding suggests that the chaperonin and its major substrates have mutually adapted during evolution. In analogy to this proposed process, GroEL/GroES was successfully optimized by in vitro mutagenesis to promote the folding of the heterologous green fluorescent protein, but this resulted in diminished capacity to assist the folding of other model substrates (Wang et al., 2002). The notion of coevolution of GroEL and its substrates is supported by our finding that *E. coli* class III proteins maintain their GroEL dependence when expressed in *S. cerevisiae*. The eukaryotic cytosol lacks GroEL and instead contains the distantly

related chaperonin TRiC/CCT, which proved to be inactive in mediating the folding of several of the bacterial class III substrates tested (H.-C.C, unpublished data). Interestingly, a number of these proteins have cytosolic orthologs in yeast and thus must have lost their GroEL dependence, presumably by adapting to the eukaryotic folding machinery. In contrast, the streamlined genomes of GroEL-deficient bacteria, such as *Mycoplasma* and *Ureaplasma*, encode orthologs for only 15%–20% of the *E. coli* class III substrates (12–16 of the 85 proteins) while sharing 25%–40% orthologous proteins with *E. coli* in general.

Finally, we note that most obligate GroEL substrates belong to the fold classes displaying a greater number of structural superfamilies than those found for GroEL-independent *E. coli* proteins (data not shown). This trend may suggest a role of GroEL in facilitating the structural diversification of certain protein folds during evolution, perhaps by buffering mutations that would otherwise cause severe energetic frustration during folding. While increasing the general adaptability of *E. coli* to various environmental conditions, such a role may have been important in evolving the TIM-barrel fold into one of the most versatile structural platforms for the implementation of enzymatic functions (Nagano et al., 2002).

Experimental Procedures

Strains, Plasmids, and Proteins

A detailed listing of bacterial and *S. cerevisiae* strains as well as the proteins used in this study is provided in Supplemental Experimental Procedures together with a description of the cloning strategies used.

Protein Refolding

In vitro refolding of GroEL-substrate proteins (Houry et al., 1999) was analyzed at 25°C and 37°C upon 100-fold dilution of the respective protein from 6 M guanidinium-HCl in buffer A (20 mM MOPS-KOH [pH 7.5], 100 mM KCl, 10 mM $MgCl_2$), containing 10 mM DTT, to a final protein concentration of 0.5 μM. Molecular chaperones were present when indicated, and refolding was monitored by enzymatic assays (see Supplemental Experimental Procedures).

Coexpression of Substrates with Chaperones and GroE Depletion

E. coli experiments were performed in BL21 (DE3) cells containing elevated levels of GroEL/GroES or GroEL expressed for 1 hr from arabinose-controlled plasmids. Substrate proteins were induced by IPTG (*T7* promoter) for 1 hr after shifting cells from arabinose to glucose medium to switch off chaperone expression. Amounts of soluble and insoluble protein were determined as described (Agashe et al., 2004). The fate of endogenous GroEL substrates was also analyzed in a strain in which the *groE* promoter was replaced with the *araC* gene and the *pBAD* promoter (McLennan and Masters, 1998), thus allowing depletion of GroEL/GroES upon shift from arabinose to glucose growth medium. Solubility of GroEL-substrate proteins in *S. cerevisiae* was assayed analogously (Agashe et al., 2004 and Supplemental Experimental Procedures).

In Vivo Capture of GroEL Substrates

GroEL/GroES/substrate complexes were isolated from live *E. coli* spheroplasts (Ewalt et al., 1997) expressing C-terminally His_6-tagged GroES from *M. mazei*. These spheroplasts are fully active in protein synthesis and GroEL-assisted protein folding (Ewalt et al., 1997; Houry et al., 1999). Complexes were fixed upon cell lysis by rapidly converting ATP to ADP with glucose/hexokinase and purified on IMAC resin (see Supplemental Experimental Procedures).

Protein Identification by Mass Spectrometry

Isolated GroEL/GroES/substrate complexes were separated by SDS-PAGE, and proteins contained in gel slices were identified by LC-MS/MS (Lasonder et al., 2002). Amounts of substrates bound to GroEL relative to total cell lysate were quantified using cell lysates prepared from cells grown in SILAC media containing Arg-^{13}C6 (Ong et al., 2002). Proteomics methods and data-analysis procedures are described in detail in Supplemental Experimental Procedures.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, Supplemental References, three tables, and six figures and can be found with this article online at http://www.cell.com/cgi/content/full/122/2/209/DC1/.

Acknowledgments

We thank B. Laber (Aventis, Frankfurt, Germany), R.E. Viola, and R. Moore (Toledo, Ohio) for their generous gifts of DL-aspartate β-semialdehyde; P.A. Lund (Birmingham, United Kingdom) for the GroEL depletion strain; and C. Dobson, A. Lupas, S. Radford, and L. Serrano for discussion. The expert technical assistance by M. Kirkegaard is acknowledged. This work was supported by the Integrated Project "Interaction Proteome" of the European Commission, the Deutsche Forschungsgemeinschaft (SFB 594), a grant from the Swiss National Foundation, and an EMBO fellowship to D.J.N.

Received: February 19, 2005
Revised: May 15, 2005
Accepted: May 27, 2005
Published: July 28, 2005

References

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

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 Å. 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.

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.

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

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.

Dobson, C.M., Sali, A., and Karplus, M. (1998). Protein folding: a perspective from theory and experiment. Angew. Chem. Int. Ed. Engl. *37*, 868–893.

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

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

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

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

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 Im91. J. Mol. Biol. *286*, 1597–1608.

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.

Frishman, D., Mokrejs, M., Kosykh, D., Kastenmüller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., et al. (2003). The PEDANT genome database. Nucleic Acids Res. *31*, 207–211.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu. Rev. Biochem. *70*, 603–647.

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., et al. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J. Bacteriol. *185*, 5673–5684.

Gromiha, M.M., and Selvaraj, S. (2004). Inter-residue interactions in protein folding and stability. Prog. Biophys. Mol. Biol. *86*, 235–277.

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

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.

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

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

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.

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.

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.

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

Nagano, N., Orengo, C.A., and Thornton, J.M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. *321*, 741–765.

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

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

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

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

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

Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C., and Genevaux, P. (2004). SecB is a bona fide generalized chaperone in Escherichia coli. Proc. Natl. Acad. Sci. USA *101*, 7583–7588.

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

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.

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

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

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.