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

Mechanisms of *De Novo* Multi-domain Protein Folding in Bacteria and Eukaryotes

Hung-Chun Chang

aus
Kaohsiung
Taiwan, R.O.C.

2006

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfen erarbeitet.

München, am .....

.....
Hung-Chun Chang

Dissertation eingereicht am	07. December 2006
1. Gutachter:	Professor Dr. F. Ulrich Hartl
2. Gutachter:	Professor Dr. Dr. Walter Neupert
Mündliche Prüfung am	05. February 2007

Acknowledgements

First of all, I would like to express my deepest gratitude to Prof. Dr. F. Ulrich Hartl for giving me the opportunity to study the extremely interesting subject in his laboratory. I would like to thank him for his encouragement and his continual support throughout the entire period of my study. And also many thanks to Dr. Manajit Hayer-Hartl for her numerous helpful discussions and advices.

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

Uncountable thanks go to Dr. José M. Barral for his helpful supervision of my work. His great knowledge and enthusiasm toward science contributed to an invaluable source of inspiration as I worked through the challenges of my project. I would like to thank Dr. Christian M. Kaiser for fruitful cooperation and many insightful discussions. Their friendships and the good working atmosphere became the main basis for the success of this work.

I thank colleagues in the department of cellular biochemistry for providing accommodative environment to a foreigner like me and many technical helps. In particularly, I would like to thank Dr. Peter Breuer, Dr. Gregor Schaffar and Dr. Andreas Bracher for generously sharing their speciality opinions.

Special thanks to Dr. Ramunas M. Vabulas for his willingness to engage in frequent discussions and his friendship.

I also would like to thank Prof. Dr. F. Ulrich Hartl and Dr. José M. Barral for their intensive review and beneficial feedback on my English writing.

The deepest thanks go to my wife, Yun-Chi Tang, not only for her enormous support and patience, but also beneficial discussions with her. The same deep thanks belong to my parents and my senior brother in Taiwan for their understanding and support.

Contents

1 Summary **1**

2 Introduction **4**

2.1 Protein folding 4

2.1.1 Protein structure 4

2.1.2 The protein folding problem 6

2.1.3 Protein folding mechanisms 7

2.2 Protein folding in the cell 10

2.2.1 Molecular chaperones and *de novo* protein folding 10

2.2.2 Human diseases caused by protein misfolding 12

2.3 Molecular chaperone systems 15

2.3.1 Overview of the substrate flux through chaperone networks in the cytosol 15

2.3.2 Ribosome-associated chaperones 17

2.3.3 The Hsp70 system 19

2.3.4 The chaperonins 24

2.4 Aim of the study 29

3 Materials and Methods **30**

3.1 Materials 30

3.1.1 Chemicals 30

3.1.2 Enzymes 32

3.1.3 Materials 32

3.1.4 Instruments 33

3.1.5 Media 34

3.1.6 Antibiotic stock solutions 35

3.2 Bacterial strains and plasmids 35

3.2.1 *E.coli* strains 35

3.2.2 *S. cerevisiae* strains 36

3.2.3 Plasmids 37

3.2.3.1 Construction of vectors for expression of fusion proteins in *E. coli* 37

3.2.3.2 Construction of vectors for expression of fusion proteins in *S. cerevisiae* 40

3.3 Molecular cloning methods 41

3.3.1 Preparation and transformation of *E. coli* competent cells 41

3.3.2 Preparation and transformation of *S. cerevisiae* competent cells 42

3.3.3 Plasmid purification 42
3.3.4 PCR amplification 43
3.3.5 DNA restriction and ligation 44
3.3.6 DNA analytical methods 44
3.3.7 Gene disruption in *S. cerevisiae* 44
3.4 Protein purification 45
3.5 Protein analytical methods 46
3.5.1 Determination of protein concentration 46
3.5.2 SDS-PAGE (sodium-dodecylsufate polyacryamide gel electrophoresis) 46
3.5.3 Western-blotting 47
3.6 Expression and assessment of protein solubility 48
3.6.1 Expression and assessment of protein solubility in *E. coli* 48
3.6.2 Expression and assessment of protein solubility in *S. cerevisiae* 49
3.7 *In vivo* experiments 50
3.7.1 Quantitation of rates of accumulation of luciferase activity in *E. coli* and *S. cerevisiae* 50
3.7.2 Determination of enzyme activity and solubility *in vivo* 50
3.7.3 Determination of folding kinetics *in vivo* 51
3.7.4 *De novo* folding of firefly luciferase in *S. cerevisiae* Δ*fes1* strain 52
3.7.5 Expression of GroEL substrates in *S. cerevisiae* 53
3.7.6 Fluorescence microscopy 54
3.8 *In vitro* protein assays 54
3.8.1 *In vitro* refolding assays 54
3.8.2 Translation and determination of enzyme activity *in vitro* 55
3.8.3 Post-translational folding assay 56
3.8.4 Ribosome binding of TF 56
3.8.5 Ribosome recruitment assay 57
3.8.6 Kinetic simulation 57
3.8.7 Prediction of DnaK binding sites 58

4 Results 59
4.1 *De novo* folding of multi-domain GFP fusion proteins in *E. coli* and *S. cerevisiae* 59
4.1.1 *De novo* folding of GFP fusions is inefficient in *E. coli* yet efficient in *S. cerevisiae* 61
4.1.2 High folding efficiency in yeast is independent of expression levels 67
4.1.3 Misfolding of GFP fusions in *E. coli* is due to an intra- 69

molecular misfolding event
4.1.4 The rate of production of folded protein *per* ribosome is higher in yeast than in bacteria 73
4.2 The influence of nascent chain binding chaperones on the folding of naturally occurring multi-domain proteins 76
4.2.1 Effect of trigger factor and DnaK on the folding yields of firefly luciferase and β-galactosidase in *E. coli* 78
4.2.2 Distinct folding kinetics of firefly luciferase in bacteria *vs.* yeast 82
4.2.3 TF and DnaK act co-translationally to cause a shift in folding mechanism 88
4.2.4 Additional TF molecules are recruited to translating ribosomes 90
4.3 Efficient folding of multi-domain proteins in yeast is supported by chaperones 92
4.3.1 Nucleotide exchange factor Fes1p in yeast facilitates the folding of firefly luciferase *in vivo* 92
4.3.2 FL folding in yeast strains lacking nascent chain binding chaperones 98
4.3.3 The yeast cytosol can support folding of recombinantly expressed bacterial unless these are strictly dependent on the specialized bacterial chaperonin GroEL 100

5 Discussion 105
5.1 Folding complexity arising from domain fusion impeded the folding of GFP fusions in bacteria 105
5.2 Co- *vs*. post-translational folding: effect of nascent chain binding chaperones on multi-domain protein folding 107
5.2.1 Default folding *versus* chaperone-assisted folding 108
5.2.2 Mechanism of delayed folding 110
5.2.3 Structure of trigger factor and current model of its function 111
5.3 Factors that contribute to efficient co-translational folding in yeast (and their limitations) 117
5.3.1 Inefficient FL folding in yeast strains devoid of nascent chain binding chaperones and the nucleotide exchange factor Fes1p 119
5.3.2 Inability of the yeast cytosol to support the folding of bacterial proteins that are GroEL dependent 121
5.4 Perspectives 123

6 References 124

7 Appendices 136
7.1 Abbreviations 136
7.2 Publications 139
7.3 *Curriculum vitae* 140

1. Summary

Eukaryotic genomes encode a considerably higher fraction of multi-domain proteins than their prokaryotic counterparts. It has been postulated that efficient co-translational and sequential domain folding has facilitated the explosive evolution of multi-domain proteins in eukaryotes by recombination of pre-existent domains. In the present study, we tested whether eukaryotes and bacteria differ in the folding efficiency and mechanisms of multi-domain proteins in general. To this end, a series of recombinant proteins comprised of GFPuv fused to four different robustly folding proteins through six different linkers were generated, and their folding behavior upon expression in *E. coli* and the yeast *S. cerevisiae* was compared. Unlike yeast, bacteria were found to be remarkably inefficient at folding these fusion proteins. By following the accumulation of enzymatically active fusion proteins, we found that the rate of appearance of correctly folded fusion protein *per* ribosome is indeed considerably higher in yeast than in bacteria.

Increasing evidence suggests that elongating polypeptide chains on ribosomes interact substantially with nascent chain binding chaperones to facilitate their folding. Our observations regarding the low efficiency of multi-domain protein folding in bacteria prompted us to investigate the possible roles of trigger factor and DnaK, the two major nascent chain binding chaperones in *E. coli*, in determining the folding fate of multi-domain nascent polypeptides. For our experiments, we utilized living bacterial strains carrying null deletions of trigger factor and DnaK as well as an *in vitro* bacterial translation lysate. We found that upon expression under chaperone-depleted conditions, multi-domain proteins such as bacterial β-galactosidase and eukaryotic firefly luciferase fold by a rapid but inefficient default-pathway tightly coupled to translation. Although trigger factor and DnaK improve the folding yield both *in vivo* and *in vitro*, these chaperones markedly delay the folding process. This delay in folding requires the dynamic recruitment of additional trigger

factor molecules to translating ribosomes. While β-galactosidase uses this chaperone mechanism effectively, firefly luciferase folding in *E. coli* remains inefficient. The efficient co-translational domain folding of firefly luciferase observed in the eukaryotic system is not compatible with the bacterial chaperone system. These findings suggest important differences in the coupling of translation and folding between bacterial and eukaryotic cells, thus explaining the higher folding yield of multi-domain proteins in the eukaryotic cytosol.

Additional experiments revealed that efficient protein folding in the eukaryotic model organism *S. cerevisiae* is supported by a complex chaperone system. Fes1p was structurally and functionally demonstrated to be a nucleotide exchange factor for the yeast cytosolic hsp70 homolog Ssa1p. We first tested the folding of luciferase in a yeast *FES1*-deletion strain and found that the specific activity of luciferase expressed at elevated temperatures was decreased ~50% compared to the wild-type control. Thus, the folding of luciferase in yeast is dependent on the involvement of Ssa1p in the cytosol. Indeed, without Fes1p, a larger molecular weight species of luciferase could be isolated owing to a longer association of the folding intermediate with Ssa1p and Ydj1p. Additional evidence supporting the notion that the yeast cytosol contains a versatile chaperone network highly efficient in supporting correct protein folding came from the analysis of the fate of recombinantly expressed proteins of bacterial origin in yeast. Our laboratory has previously classified a large number of *E. coli* proteins based on their chaperone-dependency for folding both *in vivo* and *in vitro*. Class I substrates, such as enolase, exhibit low chaperone dependency and accordingly showed only a minor folding deficiency in a yeast *YDJ1*-deletion strain. Class II substrates, such as DCEA, GATD and SYT, utilize either the DnaK or the GroEL/ES systems for folding and showed a strong inability to fold in the same *YDJ1*-deletion strain. Consistent with their stringent requirement for GroEL/ES, class III substrates showed major folding deficiencies even in the wild-type yeast background. Therefore, although the yeast cytosol is remarkably efficient at folding a wide range of proteins, it cannot provide

assistance to proteins with stringent requirements for chaperones not normally found in eukaryotes.

In summary, the present study revealed that the eukaryotic cytosol is capable of folding multi-domain proteins with much higher efficiency than the bacterial cytosol. This can be explained by a presented post-translational folding pathway in bacteria that is enforced by chaperones and is incompatible with co-translational folding of eukaryotic proteins. Thus, a post-translational shift imposed by TF and DnaK on the folding mechanism of multi-domain proteins in bacteria may have profound consequences for the heterologous expression of eukaryotic proteins in bacterial systems in general.

2. Introduction

Proteins are fundamental to most biological processes. Nearly all the molecular transformations that define cellular metabolism are mediated by protein catalysts. There are structural proteins (molecules of the cytoskeleton, epidermal keratin, viral coat proteins); catalytic proteins (enzymes); transport and storage proteins (hemoblobin, myoglobin, ferritin); regulatory proteins, (including hormones, many kinases and phosphatases, and proteins that control gene expression); and proteins of the immune system and the immunoglobulin superfamily (including antibodies, and proteins involved in cell-cell recognition and signaling). Proteins also perform regulatory roles, monitoring extracellular and intracellular conditions and relaying information to other cellular components. A complete list of known protein functions would contain many thousands of entries, including proteins that transport other molecules and proteins that generate mechanical and electrochemical forces. And such a list would not account for the thousands of proteins whose functions are not yet fully characterized or, in many cases, are completely unknown. Clearly, there is considerable validity to the statement that proteins are the “building blocks” of life.

2.1. Protein folding

2.1.1. Protein structure

In order to cover such a variety of functions, most proteins have to adopt specific and unique three-dimensional structures. Like all polymeric molecules, proteins can be described in terms of levels of organization, in this case, their primary, secondary, tertiary and quaternary structures. A protein’s primary structure is the amino acid sequence of its polypeptide chain, namely, the chemical bonding of the atoms in that protein. The secondary

structure is the local spatial arrangement of a polypeptide's backbone atoms without regard to the conformations of its side chains. In most cases, it is referring the formation of helices and sheets from particular regions of a protein. The tertiary structure refers to the three-dimensional structure of an entire polypeptide, which is described by the way that helices and sheets are organized and interact in space. As many proteins are composed of two or more polypeptide chains, a protein's quaternary structure refers to the spatial arrangement of its subunits.

Compared to the thousands of protein tertiary structures that can be found, and the overall uniqueness of conformation of each protein, the protein secondary structures are surprisingly simpler, which include folding patterns such as helices, sheets and turns. The α-helix and the β-sheet are such elements that not only can keep the main chain in an unstrained conformation, but also satisfy the hydrogen-bonding potential of the main-chain N-H and C=O groups. Both patterns were discovered 50 years ago from studies of hair and silk. The first folding pattern to be discovered, the α-helix, was found in the protein α-keratin, which is abundant in skin and its derivatives, such as hair, nails, and horns (Pauling and Corey, 1951a). Within a year of the discovery of the α-helix, a second folded structure, the β-sheet, was found in the protein fibroin, the major constituent of silk (Pauling and Corey, 1951b). These two patterns are particularly common because they result from hydrogen bonding between the N-H and C=O groups in the polypeptide backbone, without involving the side chains of the amino acids. Thus, they can be formed by many different amino acid sequences. In each case, the protein chain adopts a regular, repeating conformation.

2.1.2. The protein folding problem

The conformation of a single amino acid in a protein can be described by a pair of angles, psi (ψ) and phi (φ) (Figure 1). The peptide bond itself is planar, a consequence of its partial double bond character. ψ describes the angle formed by rotation around the axis through C_α and the carboxyl carbon; φ describes rotation around C_α and the amino group. Because of steric collisions between atoms within each amino acid, most pairs of ψ and φ angles do not actually occur. G. N. Ramachandran calculated the energy contained in various pairs of ψ and φ angles and found two most stable pairs, the so called α and β conformations (Ramachandran and Sasisekharan, 1968). These two pairs of angles are found to almost exclusively occur naturally in folded proteins, including the two most prominent examples of secondary structure: α-helix and β-strand.

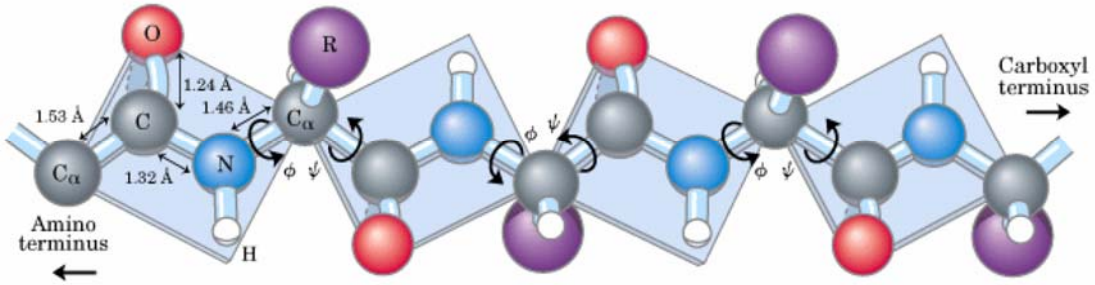

Figure 1. Rotation about bonds in a polypeptide

The structure of each amino acid in a polypeptide can be adjusted by rotation about two single bonds. Phi (Φ) is the angle of rotation about the bond between the nitrogen and the α-carbon atoms, whereas psi (Ψ) is the angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms. The peptide bond is planar as represented in blue shading. Adapted from (Lehninger *et al*., 2000).

Studies of protein stability and renaturation suggest that protein folding is directed largely by the residues that occupy the interior of the folded protein. But how does a protein fold to its native conformation? One might assume that this process occurs through the protein's random exploration of all the conformations available to it until it eventually stumbles onto the correct one. A simple calculation was first made by Cyrus Levinthal, however, convincingly demonstrates that this cannot possibly be the case: Assuming that every residue had only two possible states α_R and β, a 100-residue peptide would have $2^{100} \sim 10^{30}$ possible conformations. If the rate of interconversion between conformations is $\sim 10^{-13}$ seconds, then the 100-residue polypeptide would require $10^{30} \times 10^{-13}$ s = 10^{17} s ~ 10^{9} years to explore its conformational space, which is well beyond the range of any biological process. Clearly, proteins cannot fold by randomly groping in the dark for the native state (Levinthal, 1968). In fact, many proteins fold to their native conformations in less than a few seconds. This is because proteins fold to their native conformations *via* directed pathways rather than by performing random conformational searches.

2.1.3. Protein folding mechanism

Protein folding thus appears 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, 1999a; Baldwin and Rose, 1999b). 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.

Indeed, experimental observations indicate that protein folding begins with the formation of local segments of secondary structure (α helices and β sheets) (Karplus and Weaver, 1976). This early stage of protein folding is extremely rapid, with much of the native secondary structure in small proteins appearing within 5 ms of the initiation of folding. Since native proteins contain compact hydrophobic cores, it is likely that the driving force in protein folding is what has been termed a hydrophobic collapse (Baldwin, 1989). The collapsed state is known as a molten globule, a species that has much of the secondary structure of the native protein but little of its tertiary structure. Over the next 5 to 1000 ms, the secondary structure becomes stabilized and tertiary structure begins to form. During this intermediate stage, the native-like elements are thought to take the form of subdomains that are not yet properly docked to form domains. In the final stage of folding, which for small single-domain proteins occurs over the next few seconds, the protein undergoes a series of complex motions in which it attains its relatively rigid internal side chain packing and hydrogen bonding while it expels the remaining water molecules from its hydrophobic core (Daggett and Fersht, 2003).

In multi-domain and multi-subunit proteins, the respective units then assemble in a similar manner, with slight conformational adjustments being required to produce the protein's native tertiary or quaternary structure. Thus, proteins appear to fold in a hierarchical manner, with small local elements of simpler structure forming and then joining them together to yield larger complex elements, which coalesce with one another to form yet larger elements, *etc*.

Folding, like denaturation, is a cooperative process, with small elements of structure accelerating the conversion of a high-energy, high-entropy state to a low-energy, low-entropy state. This energy-entropy relationship is diagrammed in Figure 2, a so called folding funnel (Dill and Chan, 1997). An unfolded polypeptide has many possible conformations (high entropy). As it folds, thereby ever decreasing the number of possible

conformations, its entropy and free energy decrease. The energy-entropy landscape is not generally smooth but can be rugged. Minor clefts and gullies represent conformations that are temporarily trapped until, through random thermal activation, they overcome an "uphill" free energy barrier and can then proceed through an accessible trajectory toward a lower energy conformation.

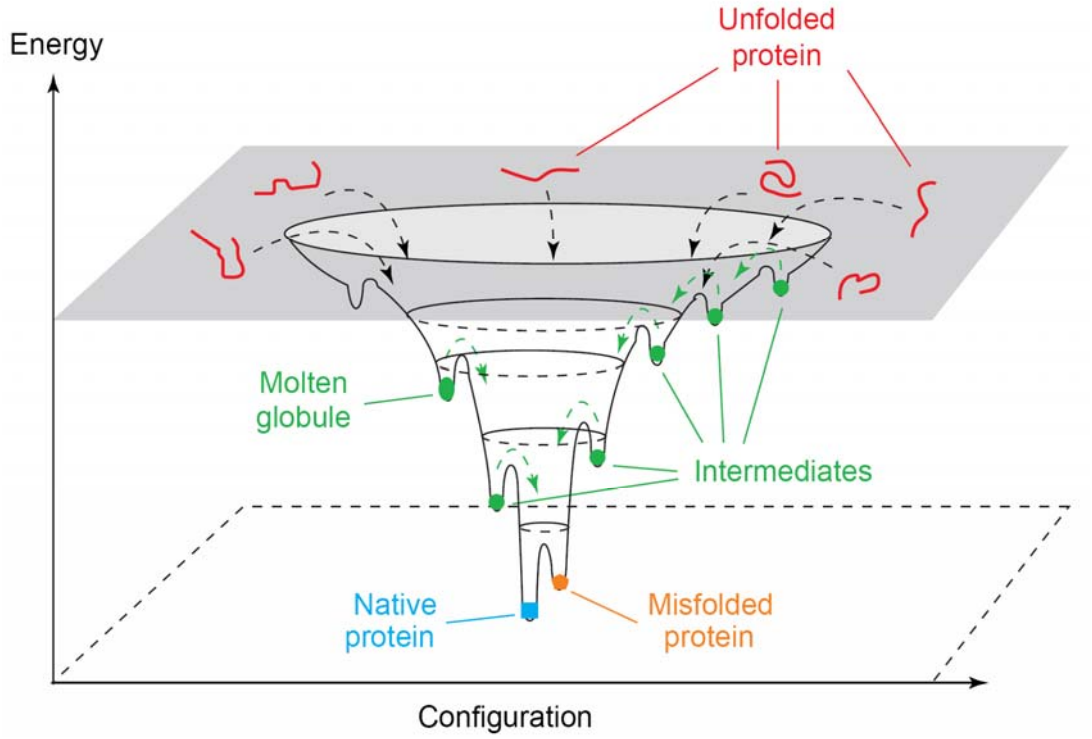

Figure 2. Schematic of the folding energy landscape

The multiple states of the unfolded protein located at the top fall into a folding funnel consisting of an almost infinite number of local minima, each of which describes possible folding arrangements in the protein. Most of these states represent transient folding intermediates in the process of attaining the correct native fold. Some of these intermediates retain a more stable structure such as the molten globule, whereas other local minima act as folding traps irreversibly capturing the protein in a misfolded state. For proteins that fold without populating intermediates, the surface of the funnel would be smooth. Adapted from (Radford, 2000).

2.2. Protein folding in the cell

Although the final native conformations of polypeptides are encoded in their linear sequences and the unassisted folding of isolated proteins has been carefully studied *in vitro* (Anfinsen, 1973; Dobson and Karplus, 1999), *de novo* folding in living cells is considerably more complex. For instance, in the extremely crowded milieu inside cells, the smooth process of protein folding as observed in diluted solution is threatened by aggregation. Furthermore, the complete folding information is not available for the nascent chain since its more C-terminal elements are still being translated by the ribosome. Therefore, a whole molecule hydrophobic collapse to prevent the polypeptide chain from aggregation, as it occurs *in vitro* is unlikely or even impossible; rather, the incomplete chain must be protected from misfolding and aggregation during its synthesis.

The cellular machinery of molecular chaperones is designed to effectively counteract the tendency of non-native proteins to aggregate, both during *de novo* synthesis and under conditions of cellular stress that lead to the unfolding of preexistent proteins (Frydman and Hartl, 1996).

2.2.1. Molecular chaperones and *de novo* protein folding

The cellular environment is extremely crowded with high concentrations of proteins, nucleic acids, and other macromolecules, in contrast to the dilute conditions employed in refolding *in vitro*. The excluded volume effects resulting from the highly crowded nature of the cytosol (300 to 400 g/liter of proteins and other macromolecules in *Escherichia coli*) (Ellis, 2001) (Figure 3) are predicted to enhance the aggregation of nonnative protein chains substantially by increasing their effective concentrations (van den Berg *et al*., 1999). Also, crowding generally provides a nonspecific force for macromolecular compaction and

association (Minton, 2000). Molecular crowding therefore represents a challenge for protein folding in the cell.

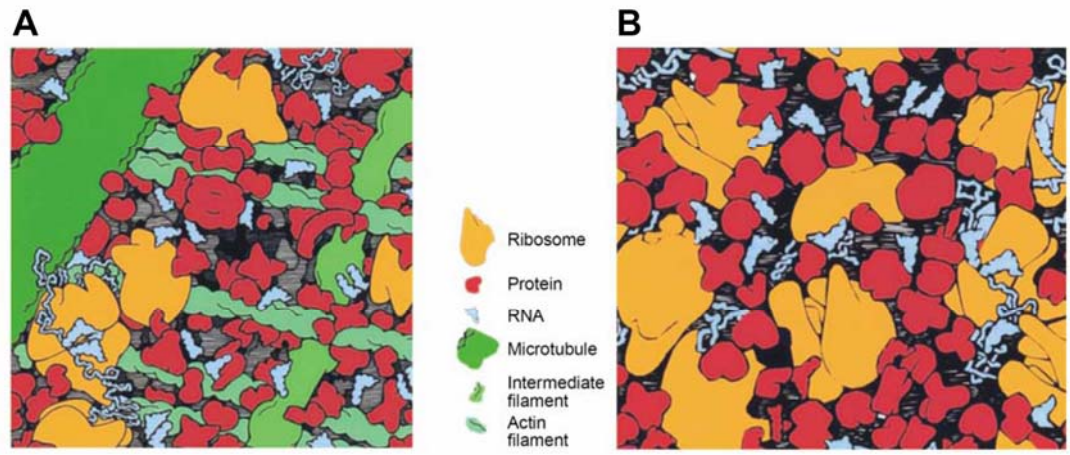

Figure 3. The crowded state of the cytoplasm

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

In addition to the off-pathway reactions that can occur due to molecular crowding, a newly synthesized protein faces further challenges in reaching its native state. Being vectorial, the protein synthesis process itself is problematic. Although some domains of a nascent chain might be capable of folding spontaneously, the completely folded structure cannot be obtained until an entire domain is synthesized (Jaenicke, 1991). This time lag increases the chance that hydrophobic sequences normally buried in the interior of the protein will be exposed, leading to protein aggregation (Dobson and Karplus, 1999).

For these reasons, protein folding in living cells must be assisted by molecular chaperones in many cases. In order to assist the many diverse types of proteins present in

living cells, chaperones function in versatile manners (Young *et al*., 2004). There are ribosome-associated chaperones that prevent possible intra- and inter-chain aggregation in close proximity to the ribosome exit tunnel. Many chaperones, such as the Hsp70s and the cylindrical chaperonins, undergo an ATP hydrolysis-dependent process to promote substrate folding by repeated cycles of binding and release. They bind to partly folded polypeptide chains and normally help them progress along an energetically favorable folding pathway. While Hsp70 shields only a short hydrophobic region of an unfolded polypeptide (Hartl and Hayer-Hartl, 2002), the chaperonin complex can provide a chamber for nonnative substrates to be enclosed and fold without interference from the difficult environment presented by the crowded cytosol. There are even chaperones, such as the members of the Clp (Hsp100) family, which can remove small protein aggregates in cooperation with Hsp70s (Ben-Zvi and Goloubinoff, 2001). In addition to shielding hydrophobic surface of unfolded polypeptides, some chaperones can as well serve as folding catalysts to support peptidyl-prolyl *cis/trans* isomerization or disulfide bond formation (Schiene and Fischer, 2000).

2.2.2. Human diseases caused by protein misfolding

In recent years, the process of protein folding has been recognized to be of considerable medical relevance. A number of human diseases are now known to result, directly or indirectly, from aberrant folding reactions (Table 1) (Barral *et al*., 2004). Classic examples include Alzheimer's disease, cystic fibrosis, and hypertrophic cardiomyopathy, to mention only a few. There are various mechanisms by which the accumulation of misfolded protein chains may cause cellular dysfunction, and often a combination of these appears to be responsible for the disease. Misfolded polypeptides not only loose their normal function, they may also form toxic species, including oligomers or larger aggregates (*e.g.* amyloid precursor protein (APP) in Alzheimer's and other neurodegenerative disease), they may be

prevented from reaching their proper cellular localization due to retention and/or degradation (*e.g.* CFTR in cystic fibrosis), or they may exert a dominant negative effect by preventing the function of interacting partners (*e.g.* myosin in hypertrophic cardiomyopathy).

Recently, a more direct involvement of molecular chaperones in human diseases of protein folding has become increasingly evident. Many studies have shown that misfolded disease proteins or their aggregates are associated with molecular chaperones, most prominently those of the Hsp70 family. For example, Hsp70 chaperones are found in association with aggregates of huntingtin and other polyQ-proteins (Cummings, 1998; Jana *et al*., 2000; Muchowski *et al*., 2000) as well as with Lewy bodies in affected human brain tissue of Parkinson's disease patients (McNaught *et al*., 2002). Additionally, Hsp70 and Hsp16 chaperones may interact with intracellular Aβ peptide (Fonte *et al*., 2002). The association of aggregates or misfolded disease proteins with chaperones implies that the cellular quality control machinery is activated in an attempt to prevent the accumulation of misfolded species (Sherman and Goldberg, 2001).

Table 1. Examples of human diseases caused by protein misfolding

Disease	**Target protein**
Formation of toxic aggregates by the disease protein	
Alzheimer's disease	APP β-peptide
Creutzfeldt–Jakob's disease	Prion protein
Polyglutamine expansion diseases	Various polyQ proteins
Familial amyloidosis	Transthyretin
Prevention of accumulation of disease protein at appropriate cellular location	
Cystic fibrosis	CFTR
Familial hypercholesterolemia	LDL receptor
Phenylketonuria	Phenylalanine hydroxylase
MCAD deficiency	Medium-chain acyl-CoA dehydrogenase
Exertion of a dominant negative effect on interacting partners	
Hypertrophic cardiomyopathy	Various sarcomeric proteins
Osteogenesis imperfecta	Collagens
Epidermolysis bullosa simplex	Keratins
Mutation in an actual or putative molecular chaperone	
Hereditary spastic paraplegia	Mitochondrial Hsp60
Desmin-related myopathy	αB-Crystallin
Sanjad–Sakati and Kenny–Caffrey	TBCE

2.3. Molecular chaperone systems

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

Cytosolic chaperones participate in *de novo* folding mainly through two distinct mechanisms. Chaperones, such as trigger factor and the Hsp70s, act by holding nascent and newly synthesized chains in a state competent for folding upon release into the medium. In contrast, the large, cylindrical chaperonin complexes provide physically defined compartments inside which a complete protein or a protein domain can fold while being sequestered from the cytosol. These two classes of chaperones are conserved in all three domains of life and can cooperate in a topologically and timely ordered manner (Figure 4).

Though the essential nature of the chaperonins has long been recognized (Fayet *et al*., 1989; Horwich *et al*., 1993), it has proved more difficult to demonstrate the essential role of nascent chain-binding chaperon in protein folding, because of considerable functional redundancy among components (Deuerling *et al*., 1999; Teter *et al*., 1999). Some of these chaperones, including trigger factor and specialized Hsp70 proteins, bind directly to the ribosome near the polypeptide exit site and are positioned to interact generally with nascent chains (Figure 4A). The majority of small proteins are thought to fold rapidly and without further assistance upon completion of synthesis and release from this first set of components (Figure 4A). Longer chains interact subsequently with members of a second class of nascent chain binding chaperones, including the classical Hsp70s and prefoldin, which do not associate directly with the ribosome (Deuerling *et al*., 1999; Teter *et al*., 1999; Thulasiraman *et al*., 1999). In addition to stabilizing elongating chains, these chaperones also assist in co- or post-translational folding, or facilitate chain transfer to downstream chaperones (Figure 4B and 4C) (Siegers *et al*., 1999). A subset of slow-folding and aggregation-sensitive proteins (10 to 15% of total) interacts with a chaperonin for folding in

both prokaryotes and eukaryotes (Ewalt *et al*., 1997; Houry *et al*., 1999). Many eukaryotic kinases and other signal transduction proteins use an additional chaperone after Hsp70: Hsp90 (Figure 4C), a specialized ATP-dependent chaperone that cooperates with ancillary factors in protein folding and regulation.

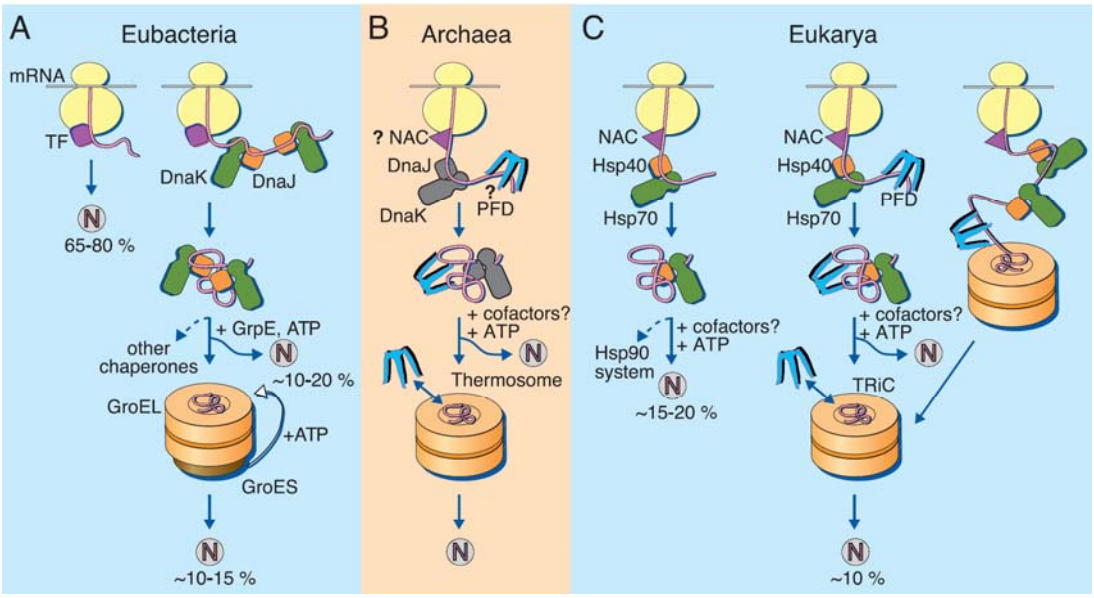

Figure 4. Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol

TF: trigger factor; N: native protein. PFD: prefoldin; NAC: nascent chain-associated complex. (A) In Eubacteria, nascent chains possibly interact generally with TF. Longer chains interact subsequently with DnaK and DnaJ and fold upon one or several cycles of ATP-dependent binding and release. About 10 to 15% of chains transit the chaperonin system (GroEL and GroES) for folding after their interaction with DnaK. (**B**) Archaea. Only some archaeal species contain DnaK/DnaJ. (**C**) Eukarya, the example of the mammalian cytosol. NAC probably interacts generally with nascent chains as TF does in bacteria. The majority of small chains may fold upon ribosome release without further assistance. About 15 to 20% of chains reach their native states in a reaction assisted by Hsp70 and Hsp40. A subset of polypeptide chains requires Hsp90 for folding. About 10% of chains are co- or post-translationally passed on to the chaperonin TRiC in a reaction mediated by prefodin (PFD). From Hartl and Hayer-Hartl (2002).

2.3.2. Ribosome –associated chaperones

During translation, ~30-40 amino acid residues of the nascent chain are protected from the cytosol by the ribosome exit tunnel (Malkin and Rich, 1967). Recent evidence indicates that as soon as a nascent chain leaves the tunnel, molecular chaperones bind to it, preventing it form aggregation with other chains. Both prokaryotic and eukaryotic chaperones have evolved to associate specifically with ribosomes and to bind to polypeptide chains that have just emerged from the tunnel. These ribosome-associated chaperones include trigger factor (TF) in bacteria and presumably nascent chain-associated complex (NAC) in eukaryotes (Hartl and Hayer-Hartl, 2002). TF is a 48 kDa protein, which binds to ribosomes at a 1:1 stoichiometry and interacts with nascent chains as short as 57 residues (Hesterkamp *et al*., 1996). The nascent chain-TF complex dissociates after chain release from the ribosome in an ATP-independent manner (Hesterkamp *et al*., 1996). Although TF exhibits peptidyl-prolyl *cis/trans* isomerase (PPIase) activity *in vitro*, recognition of target polypeptides by TF is independent of proline residues (Patzelt *et al*., 2001) and is mediated by short sequences enriched in hydrophobic (aromatic) amino acids (Patzelt *et al*., 2001). TF has an overlapping chaperone function with the main bacterial Hsp70 system, DnaK and DnaJ, in stabilizing nascent chains in a state competent for subsequent folding (Deuerling *et al*., 1999; Teter *et al*., 1999). *E. coli* cells lacking TF (Δ*tig*) or DnaK (Δ*dnaK*) exhibit no apparent folding defects at 37^{o}C. However, deletion of *dnaK* in a Δ*tig* strain at such temperature leads to a severe loss of viability. In light of this functional redundancy, the biological significance of the PPIase activity of TF remained unclear, yet this isomerase activity may allow TF to maintain nascent chains in a flexible state, poised for rapid folding upon their release. In contrast to DnaK, a role of TF in assisting folding independently of the ribosome has not yet been demonstrated, but would be consistent with the finding that only half of total TF is ribosome bound (Bukau *et al*., 2000).

The eukaryotic cytosol lacks TF but contains a ribosome-associated heterodimeric complex of α (33 kDa) and β (22 kDa) subunits, termed NAC (Wiedmann *et al*., 1994). A homolog of α-NAC appears to be present in some archaea (Leroux, 2001). Similar to TF, NAC is strategically bound to ribosomes and contacts nascent polypeptide chains as they emerge from the ribosomal exit tunnel (Shi *et al*., 1995; Wiedmann *et al*., 1994) and also dissociates upon chain release from the ribosome (Beatrix *et al*., 2000). NAC may serve to influence the fidelity of co-translational targeting of nascent chains to the endoplasmic reticulum (Wiedmann *et al*., 1994). However, a direct role for NAC in protein folding remains to be established.

In addition to NAC, nascent polypeptides are also met by other binding factors on their emergence from the ribosomal exit tunnel, as shown in the model eukaryote *S. cerevisiae*. The stable RAC (Ribosome-Associated Complex) heterodimer consists of the DnaJ-related co-chaperone Zuotin and the Hsp70-related Ssz1p/Pdr13p (Gautschi *et al*., 2001; Michimoto *et al*., 2000). Ssz1p is evolutionarily divergent from the other cytosolic forms of Hsp70 *(Gautschi et al*., 2001; Michimoto *et al*., 2000) and does not appear to interact directly with nascent chains. Also, the putative peptide-binding domain of Ssz1p is not essential for its cellular function, although it might have a secondary, regulatory role (Hundley *et al*., 2002). Zuotin contains an Hsp70-interacting J domain that is homologous to that present in bacterial DnaJ, as well as a separate domain that binds ribosomes, possibly through ribosomal RNA (Yan *et al*., 1998). The Ssb Hsp70 proteins (Ssb1p and Ssb2p) are the proposed partner chaperones of Zuotin (Yan *et al*., 1998), and they are found both associated with ribosomes and free in the cytosol (Nelson *et al*., 1992; Pfund *et al*., 1998; Siegers *et al*., 2003). Deletion of Ssz1p, or mutation of the Zuotin J domain, apparently reduces Ssb binding to nascent polypeptide chains, which indicates that RAC might recruit Ssb to assist in the folding of nascent polypeptides (Gautschi *et al*., 2002). Loss of the RAC components leads to cold sensitivity and sensitivity to translation inhibitors, which is similar

to the phenotype that is caused by deletion of *SSB* genes (Hundley *et al*., 2002; Gautschi *et al*., 2002). However, it is not yet clear how this phenotype is caused by a protein-folding defect.

Interestingly, RAC homologs in mammals were recently identified. Mpp11, the human ortholog (MIDA1 in mice) of the yeast Zuotin can be partially substituted for Zuotin by partnering with the multipurpose Hsp70 Ssa, when expressed in yeast. This suggested that in metazoans, ribosome-associated Mpp11 recruits the multifunctional soluble Hsc70 to nascent polypeptide chains as they exit the ribosome (Hundley *et al*., 2005). An independent finding showed that Mpp11 forms a stable complex with Hsp70L1, a distantly related homolog of Ssz1p (Otto *et al*., 2005). Complementation experiments indicated that mammalian ribosome-associated complex is functional in yeast (Otto *et al*., 2005). These results indicate that the cooperation of ribosome-associated chaperones with the translational apparatus is well conserved in eukaryotic cells.

2.3.3. The Hsp70 system

Hsp70s are arguably the most ubiquitous and versatile molecular chaperones. They constitute a highly conserved family of proteins, distributed ubiquitously in all bacteria and in the cellular compartments of eukaryotic organisms. Some compartments contain multiple Hsp70 homologs with distinct cellular functions (Mayer and Bukau, 2005). For instance, the cytosol of the yeast *S. cerevisiae* contains four functionally redundant Hsp70 homologs, called Ssa1p, Ssa2p, Ssa3p, Ssa4p and three specialized ribosome-associated Hsp70s, called Ssb1p, Ssb2p and Pdr13p (or Ssz1p) as described above. Genetic studies indicate that only Ssa-type function is essential for viability, and that Ssb activity cannot substitute for Ssa activity (Craig *et al*., 1994).

Most Hsp70s have a molecular mass of approximately 70 kilodaltons (kDa) and consist of two functionally coupled domains: a 44-kDa N-terminal domain, which mediates ATP binding (Flaherty *et al*., 1990), and a 27-kDa C-terminal domain that binds substrate polypeptides (Zhu *et al*., 1996). Binding and release of the substrate rely on modulation of the intrinsic peptide affinity of Hsp70 by cycles of ATP binding and hydrolysis by the N-terminal domain (Bukau and Horwich, 1998; Hartl, 1996). In the ATP-bound state, Hsp70 binds and releases substrates rapidly, whereas the ADP-bound form binds and releases substrates slowly. In *Escherichia coli*, cycling of the Hsp70 homolog, DnaK, between its different nucleotide-bound states is regulated by two cofactors, DnaJ and GrpE (Figure 5) (McCarty *et al*., 1995; Szabo *et al*., 1994). The 41-kDa DnaJ protein is itself a chaperone, which can bind to unfolded polypeptides and prevent their aggregation (Szabo *et al*., 1996). DnaJ binds to DnaK and stimulates its ATPase activity, generating the ADP-bound state of DnaK, which interacts stably with the polypeptide substrate (Suh *et al*., 1999). The 23-kDa GrpE protein acts as a nucleotide exchange factor. It binds to the ATPase domain of DnaK and, and by distorting the nucleotide binding pocket, induces release of bound ADP (Harrison *et al*., 1997). Finally, rebinding of ATP triggers dissociation of the DnaK-substrate complex.

How does Hsp70 recognize unfolded polypeptides? The peptide-binding site of Hsp70 contains a cleft that binds the peptide in an extended conformation (Figure 5) (Zhu *et al*., 1996). Analysis of the substrate specificity of Hsp70 using a number of different approaches, including phage display and synthetic peptide libraries, indicate that this chaperone recognizes linear polypeptide sequences enriched in hydrophobic amino acids (Flynn *et al*., 1991; Rudiger *et al*., 1997b). Because of its hydrophobic nature, this binding motif would typically be located in the interior of a correctly folded protein; consequently, surface exposure of such a sequence may be a distinctive feature of nonnative conformations. Such hydrophobic regions are probably present in all unfolded polypeptides,

and it has been predicted that an Hsp70-binding site occurs, on average, every 40 residues (Rüdiger *et al*., 1997). Association with Hsp70 results in the stabilization of the substrate protein in an extended conformation, thereby preventing its aggregation. *In vitro*, polypeptides can undergo multiple rounds of binding and release from Hsp70. This process is sufficient to promote folding of some model substrates, such as firefly luciferase (Szabo *et al*., 1994). However, in many instances, the Hsp70-bound substrate must be transferred to another type of chaperone, such as the chaperonin complex, for productive folding.

Elucidation of the DnaK reaction cycle has provided a paradigm for Hsp70 function. In fact, homologs of bacterial DnaJ, collectively called Hsp40 proteins, have been identified in all cellular compartments that contain an Hsp70 (Kelley, 1998). All DnaJ homologs possess a so-called J-domain, a scaffold of four α-helices and a solvent-accessible loop region that exposes a conserved tripeptide (His-Pro-Asp) essential for interaction of the J-protein with Hsp70 (Pellecchia *et al*., 1996). J-domain-containing proteins can stimulate ATP hydrolysis by Hsp70 and generate the ADP-bound state. In the eukaryotic cytosol, the DnaJ homologs, Hdj1 (also called Hsp40, or Sis1p in yeast) and Hdj2 (Ydj1p in yeast), regulate the activity of Hsp70 homologs (Lu and Cyr, 1998; Minami, 1996). In addition to the N-terminal J-domain, Ydj1p (Hdj2) and Sis1p (Hdj1) contain C-terminal chaperone domains that bind unfolded polypeptides (Cyr *et al*., 1994). In Ydj1p, this substrate binding domain contains two essential cysteine-rich, Zn^{2+}-binding domains, also found in DnaJ. The C-terminal domain of Sis1p has been crystallized and its structure reveals several patches of hydrophobic side chains that are required for substrate binding (Sha *et al*., 2000). In addition to these *bona fide* DnaJ homologs, several eukaryotic proteins contain only a J-domain, which serves to recruit Hsp70 family members to specific cellular sites (Kelley, 1998). For instance, the ER membrane protein, Sec63, binds to the luminal Hsp70 (BiP/Kar2p) *via* a J-domain (Brodsky and Schekman, 1993). Similarly, auxilin, a mammalian protein associated with the clathrin coat of endocytic vesicles, interacts through its J-domain with cytosolic

Hsp70, which is required for the ATP-dependent uncoating of the vesicles (Ungewickell *et al*., 1995).

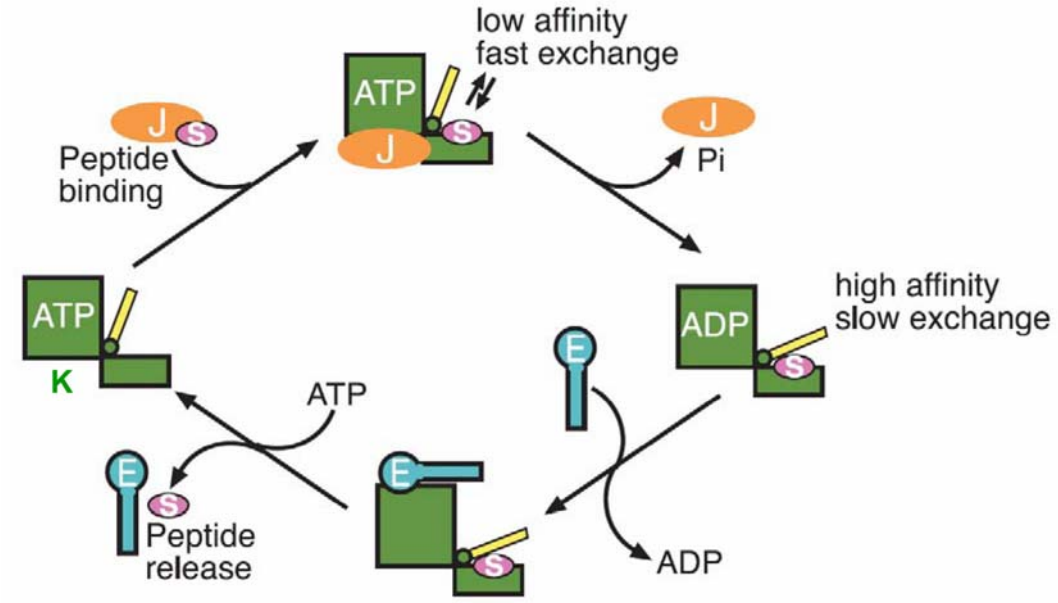

Figure 5. The reaction cycle of the DnaK system

K: DnaK, J: DnaJ, E: GrpE, S: substrate peptide. Non-native substrate polypeptides associate with either DnaJ (J) or DnaK (K) in the ATP-bound open state. DnaJ and substrate protein stimulate ATP hydrolysis of DnaK**,** leading to closure of its substrate binding pocket. GrpE is required for efficient release of ADP from the complex, and subsequent ATP binding results in opening of the substrate binding pocket and exchange of substrate polypeptides. The released substrate can either fold towards the native state or rebind to DnaJ or DnaK. Adapted from Hartl and Hayer-Hartl (2002).

The regulation of eukaryotic Hsp70 differs from DnaK in significant aspects, and is considerably more complex. Unlike DnaJ homologs, GrpE-related proteins in eukaryotes appear to be restricted to compartments of prokaryotic origin, *i.e*. mitochondria and chloroplasts. Interestingly, nucleotide exchange activities in the eukaryotic cytosol are performed by structurally unrelated proteins, such as the mammalian co-chaperone BCL2-associated athanogene-1 (Bag1), and the mammalian HSP70-binding protein (HspBP1) and its *S. cerevisiae* homolog Fes1p (Kabani *et al*., 2002b; Young *et al*., 2003). Furthermore, mammalian cells contain a 48-kDa Hsp70-interacting protein (Hip) that binds to the ATPase domain of Hsp70 and prevents ADP release from Hsp70 (Hohfeld *et al*., 1995). Similar to DnaJ in the bacterial reaction cycle, Hdj1 or Hdj2 (Hsp40s) stimulate the Hsp70 ATPase and generate the ADP-bound form with a high affinity for substrate. However, whereas GrpE promotes nucleotide exchange on DnaK, the eukaryotic Hsp70 complex appears to be the target of multiple regulatory factors. Hip binding slows dissociation of ADP from Hsp70. Through stabilization of the ADP-bound conformation of Hsp70, Hip presumably stabilizes the Hsp70 substrate complex. The action of Hip appears to antagonize that of proteins that stimulate nucleotide exchange by Hsp70, such as Bag1.

The modular domain structure of these Hsp70 regulators raises the interesting possibility that nucleotide exchange and subsequent sequence release may be coupled to downstream cellular processes. For instance, Bag1 contains an ubiquitin-homology domain in its N-terminal extension and has been proposed to direct Hsp70-bound substrates to the 26S proteasome (Luders *et al*., 2000). The N-terminal extensions of Bag1 homologs may play a role in directing the released polypeptide toward different folding or degradation pathways.

2.3.4. The chaperonins

The chaperonins are large cylindrical protein complexes consisting of two stacked rings (Hartl, 1996). In some chaperonins, the rings have seven subunits, while others contain eight or nine subunits. Chaperonins differ substantially from Hsp70 in architecture, as well as in their mechanism. However, as in the case of Hsp70 chaperones, ATP binding and hydrolysis also produce conformational changes that drive cycles of substrate binding and release. There are substantial differences between group I chaperonins, found in eubacterial cells, and the distantly related group II chaperonins found in Archaea and Eukarya. Group I chaperonins, such as GroEL of *E. coli* and Hsp60 in mitochondria and chloroplasts, function in conjunction with a ring-shaped co-factor, GroES or Hsp10, that forms the lid on a folding cage in which polypeptide substrates are enclosed during folding (Mayhew *et al*., 1996; Weissman *et al*., 1996). In contrast, such a cofactor has not been found for group II chaperonins, although helical protrusions in the apical domains of the chaperonin subunits may function as lid (Ditzel *et al*., 1998).

The group I chaperonins are perhaps the best characterized chaperone system to date, though significant questions remain about their mechanism (Hartl, 1996; Sigler *et al*., 1998). The Hsp60 chaperonins from chloroplast and mitochondria were the first complex implicated in oligomeric assembly (Ellis and van der Vies, 1991) and protein folding (Ostermann *et al*., 1989). Structurally, the *E. coli* chaperonin, GroEL, contains 14 identical subunits arranged in two stacked rings of seven subunits each (Figure 6A). The ring-shaped structure of GroEL is essential for its folding activity (Weber *et al*., 1998), which allows it to promote the folding of substrates that the Hsp70 system is unable to fold. Each subunit of GroEL consists of two discrete domains, joined by a hinge-like intermediate domain (Braig *et al*., 1994). The equatorial domain contains the ATP-binding pocket, whereas the apical domain contains a patch of hydrophobic amino acids that face the interior of the cavity and bind the unfolded substrate polypeptide through hydrophobic contacts. Unlike Hsp70s,

GroEL does not bind linear peptides, but interacts efficiently with nonnative proteins. Binding to GroEL appears to be multivalent, and bound folding intermediates presumably expose hydrophobic surfaces that allow them to interact with several GroEL subunits simultaneously (Farr *et al*., 2000; Hayer-Hartl *et al*., 1994). The substrate-binding residues in the apical domain of GroEL are also responsible for interacting with GroES, a ring shape molecule composed of seven identical subunits, which is essential for GroEL-mediated folding (Hunt *et al*., 1996). In the presence of ATP, GroES binds to GroEL and induces a conformational change in the apical domains that displaces the substrate from its binding sites. The substrate is thus released into the central cavity, which is now lined with hydrophilic side chains (Xu *et al*., 1997). GroES also promotes ATP hydrolysis in the proximal (*cis*) ring of GroEL. Enclosure of the substrate polypeptide within the chamber of this GroEL-GroES-ADP complex is essential for folding. The substrate remains enclosed in the cavity for approximately 15 seconds. It has been proposed that the enclosed cavity functions as an Anfinsen cage, *i.e.* a protected chamber that isolates the polypeptide under conditions of infinite dilution and allows it to fold according to its thermodynamic potential (Ellis, 1994). There is negative cooperativity between the two GroEL rings: the GroEL-ADP-GroES complex is dissociated by ATP binding to the *trans* ring (Figure 6B) (Rye *et al*., 1997). GroES release returns the apical domains to the conformation that exposes their hydrophobic binding sites toward the cavity, which permits a still unfolded polypeptide to rebind and undergo another cycle of folding. However, if the substrate has achieved a folded conformation, it will no longer expose sufficient hydrophobic surfaces to mediate binding and will be released.

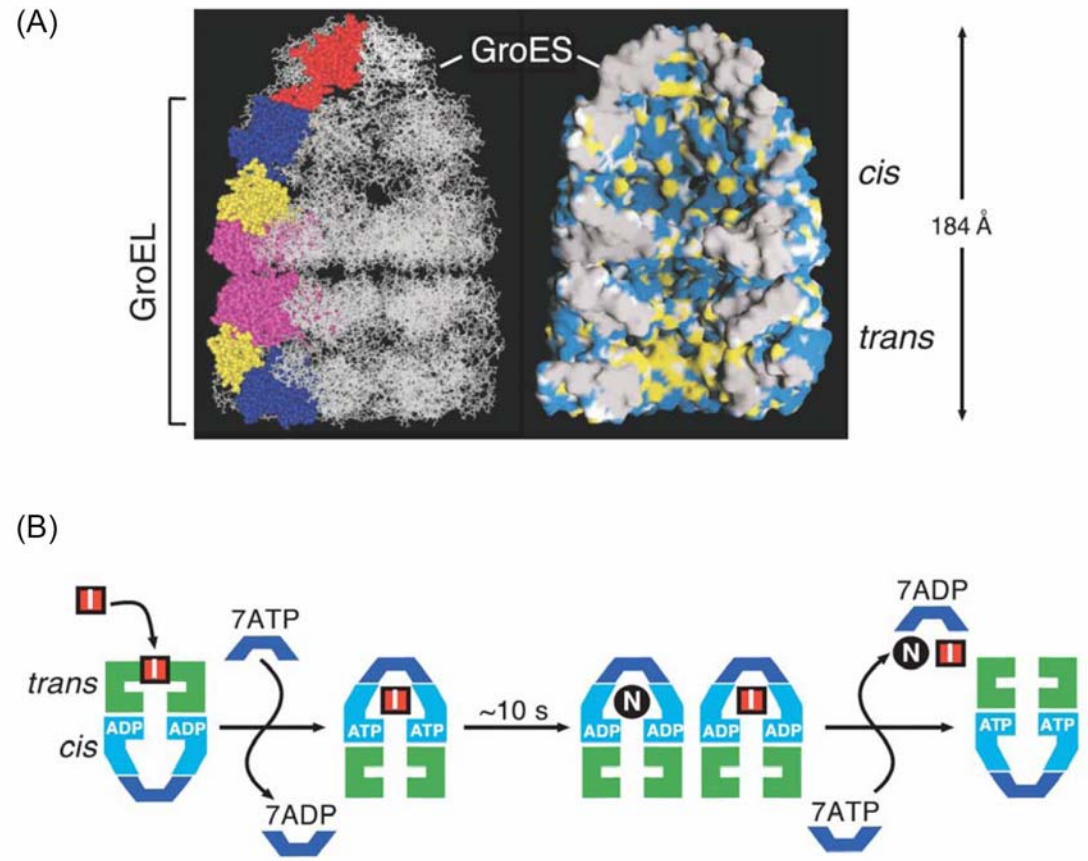

Figure 6. The GroEL-GroES chaperonin system

(A) (Left) View of the asymmetric GroEL-GroES-$(ADP)_7$ complex. The equatorial, intermediate, and apical domains of one subunit each in the *cis* and *trans* ring of GroEL are colored in pink, yellow, and dark blue, respectively, and one subunit of GroES is colored red. (Right) The accessible surface of the central cavity of the GroEL-GroES complex. Polar and charged side-chain atoms, are colored in blue; hydrophobic side-chain atoms in yellow; backbone atoms in white; and solvent-excluded surfaces at subunit interfaces in gray. Adapted from Xu *et al*., (1997). (B) Simplified reaction of protein folding in the GroEL-GroES cage. I, folding intermediate bound by the apical domains of GroEL; N, native protein folded inside the cage. For a typical GroEL substrate, multiple rounds of chaperonin action are required for folding; both I and N accumulate after a single reaction cycle and exit the cage upon GroES dissociation. I can then be rapidly re-bound by GroEL. Adapted from Hartl and Hayer-Hartl, (2002).

Group II chaperonins are more heterogeneous in sequence and structure than members of group I (Leroux and Hartl, 2000). Although group II members also have a double ring structure, they are hetero-oligomeric and the number of subunits varies between eight and nine *per* ring. Archaeal forms have two or three different subunits *per* complex (arranged in eight- or nine-fold symmetrical rings, respectively). The eukaryotic chaperonin, named TRiC (for tailless complex polypeptide-1 [TCP-1] ring complex) or CCT (for chaperonin containing TCP-1), is also ring shaped and consists of eight different, yet homologous, subunits *per* ring, ranging between 50 and 60 kDa (Frydman *et al*., 1992; Gao *et al*., 1992). The crystal structure of the archetypal group II chaperonin, the thermosome complex from *Thermoplasma acidophilum*, revealed that individual subunits have a domain arrangement similar to those in GroEL (Ditzel *et al*., 1998). Indeed, the equatorial (ATP-binding) domain is relatively conserved among all chaperonins. Most sequence divergence between subunits is found in the apical domains, which probably contain the substrate binding sites (Kim *et al*., 1994). Strikingly, the backbone trace of apical domains of the thermosome is almost identical to that of GoEL, but it has an insertion that extends as a large protrusion toward the central cavity (Klumpp *et al*., 1997). Since a major difference between group I and group II chaperonins is the lack of a GroES-like cofactor for members of group II chaperonins, this protrusion is thought to function as a built-in lid. Thus, a functional equivalent to GroES may be integral to the primary sequence of group II apical domains. Substrate binds in the central cavity of group II chaperonins (Llorca *et al*., 1999) and is folded in an ATP-dependent manner. The mechanism by which group II chaperonins mediate folding is very poorly defined. On the basis of the crystal structure of the thermosome, the complex appears to close into an Anfinsen cage (Ditzel *et al*., 1998). However, it is not clear whether the substrate is completely encapsulated during productive folding. The eukaryotic chaperonin supports the folding of proteins, such as actin, that cannot be assisted by any other chaperone system (Tian *et al*., 1995).

One important difference from GroEL is that several subunits of TRiC lack hydrophobic residues within the regions in the apical domains that correspond to the substrate binding site in GroEL. This difference led to the proposal that the substrate binding site in TRiC is located at the apical protrusions, which contain an obvious hydrophobic surface that faces the central cavity (Klumpp *et al*., 1997). This view is based on the assumption, still untested, that association of the substrate with TRiC is mediated primarily by hydrophobic interactions. The substrate binding site within the individual subunits remains to be defined. The origin of the subunit heterogeneity in group II chaperonins has not been systematically addressed. One intriguing possibility is that heterogeneity is directly linked to substrate specificity. A number of biochemical studies using endogenous model substrates suggest that each subunit contributes to the recognition of specific motifs within the substrates (Llorca *et al*., 1999; McCallum *et al*., 2000). Experiments with truncations and peptide libraries indicate that TRiC interacts with defined regions within actin, tubulin, and the tumor suppressor protein, VHL (Feldman *et al*., 1999). For instance, deletion analysis of actin suggested that chaperonin binding requires at least three discrete regions in the polypeptide (Hynes and Willison, 2000; Rommelaere *et al*., 1999). In addition, cross-linking experiments that monitor the interaction of nascent actin chains emerging from the ribosome indicate that the elongating polypeptide interacts specifically with individual subunits of the chaperonins (McCallum *et al*., 2000). In these experiments, the extent to which a nascent chain was cross-linked to multiple TRiC subunits was correlated with the stability of the TRiC-nascent chain complexes to undergo immunoprecipitation. This supports the idea that the frequency and number of different cross-links indeed reflects subunit-specific interactions with different binding sites within the nascent chains. Thus, as in GroEL, stable interactions between a folding polypeptide and the eukaryotic chaperonin may also result from a multivalent set of weak interactions between defined motifs in the substrate and individual chaperonin subunits. These experiments also demonstrate that, in contrast to the

bacterial chaperonin, group II chaperonins are capable of interacting with nascent chains in a co-translational fashion.

2.4. Aim of the study

Protein folding of newly synthesized proteins in the cell is generally dependent on the assistance of molecular chaperones. Chaperone systems throughout kingdoms have been the subject of intensive study and the working mechanisms of many chaperones are well characterized. However, little is made to compare the resulting protein folding kinetics and capacities in bacteria and in eukaryotes, until a pioneer study demonstrated that better folding of protein domain fusions indeed facilitated by co-translational folding predominantly occurring in eukaryote (Netzer and Hartl, 1997). As it would provide knowledge in understanding fundamental difference in protein folding *in vivo*, further investigation in molecular details seemed to be highly advantageous.

As many factors can be considered for the above observation, the main goal of this study is to systematically analyze, subsequently identify these factors, through genetic disruptions of nascent chain binding chaperones in both *E. coli* and *S. cerevisiae* cells. Together with biochemical analysis of purified chaperones toward the *de novo* folding of assay-applicable multi-domain proteins firefly luciferase (FL) and β-galactosidase (β-gal), as well as constructing a large pool of substrates consist of bacterial domains and artificial protein domain fusions, the results would allow us to compile a view of the cooperative network of nascent chain binding chaperones and their impact on determing the route of *de novo* protein folding. The results obtained in this study may also provide potential solution for recombinant protein misfolding problems occurring often in bacteria.

3. Materials and Methods

3.1. Materials

3.1.1. Chemicals

L-Amino acids	Sigma-Aldrich
Acetic acid	Merck
Adenosine triphosphate, disodium salt (ATP)	Sigma-Aldrich
Agarose (SeaKem LE)	Cambrex Bio Science
Ammonium persulfate (APS)	Sigma-Aldrich
Ampicillin	Merck
Arabinose	Sigma-Aldrich
Bacto agar	Difco
Bacto trypton	Difco
Bacto yeast extract	Difco
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Calcium chloride	Merck
Chloramphenicol	Sigma-Aldrich
Complete EDTA-free protease inhibitor	Roche
Coomassie brilliant blue R-250	Roth
Dimethylsulfoxide (DMSO)	Merck
Dithiothreitol (DTT)	Roche
ECL™ detection kit	Amersham Pharmacia Biotech
Ethanol	Merck
Ethidium bromide	BioRad
Ethylenediaminetetraaceticacid –sodium salt (EDTA)	Merck
Galactose	Sigma-Aldrich
Glucose	Sigma-Aldrich
Glycerol	Merck

Glycine	Roth
Guanidium hydrochloride	Sigma-Aldrich
HEPES	Sigma-Aldrich
Hydrochloric acid (37%)	Merck
Imidazol	Merck
Isopropyl-β-D-thiogalactopyranoside (IPTG)	BioMol
Kanamycin	Sigma-Aldrich
Luciferin (potassium salt)	Promega
Magnesium chloride	Merck
^{35}S-Methionine	Amersham Pharmacia Biotech
β-mercaptoethanol	Sigma-Aldrich
Methanol	Merck
Nickel-NTA agarose beads	Qiagen
o-nitrophenyl-β-D-galactopyranoside	Sigma-Aldrich
PIPES	Sigma-Aldrich
Phenyl-methyl-sulfonyl-fluoride (PMSF)	Sigma-Aldrich
2-phosphoglyceric acid	Sigma-Aldrich
Polyacrylamide/bisacrylamide solution 30 % (30 : 0.8)	Roth
Polyethylene glycol 3500 (PEG 3500)	Merck
Potassium hydroxide	Sigma-Aldrich
Rabbit reticulocyte lysate (RRL)	Promega
Raffinose	Fluka
RTS HY transcription/translation system	Roche
RNaseH	Promega
Sodium chloride	Merck
Sodium dodecylsulfate (SDS)	Sigma-Aldrich
Sodium hydroxide	Sigma-Aldrich
Sorbitol	Merck
Spectinomycin	Sigma-Aldrich
Sucrose	Merck
Tetracycline	Merck

N, N, N', N'-Tetramethylethylendiamine (TEMED)	Sigma-Aldrich
Tris-base	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Tween-20	Calbiochem
Yeast-Marker carrier DNA	Clontech

3.1.2. Enzymes

Apyrase	Sigma-Aldrich
Benzonase	Merck
Firefly luciferase	Sigma-Aldrich
Lysozyme	Sigma-Aldrich
Pfu DNA polymerase	Stratagene
Restriction enzymes	New England Biolabs
Shrimp Alkaline Phosphatase	Roche
T4 DNA ligase	New England Biolabs
Vent DNA polymerase	New England Biolabs
Zymolyase 100T	ICN Biomedicals

3.1.3. Materials

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

3.1.4. Instruments

Axiovert 200M microscope	Carl Zeiss
AIDA gel imaging software version 2.31	Raytest
Balance AG285, PB602	Mettler Toledo
Centrifuges: Avanti J-25, Avanti J20 XP, J-6B, GS-6R	Beckmann
Centrifuges 5415C and 5417R	Eppendorf
Chromatography columns (HiPrep Desalting, HiTrap metal chelating, Superdex 200, Superose 6)	Amersham Pharmacia Biotech
Deionization system MilliQ plus PF	Millipore
Electrophoresis chambers MiniProtean 3	Bio-Rad
Electrophoresis power supply Power PAC 300	Bio-Rad
Fluorescence spectrometer Fluorolog 3	HORIBA Jobin Yvon
FPLC systems	Amersham Pharmacia Biotech
EmulsiFlex high pressure homogenizer	Avestin
Gene Pulser II electroporation system	Bio-Rad
Gilson Pipetman (2, 10, 20, 100, 200, 1000 µl)	Abimed
Incubators Innova 4430	New Brunswick Scientific
Luminescent Image Analyzer LAS-3000	FUJIFILM
Luminometer (Lumat LB 9507)	BERTHOLD
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad
PCR-Thermocycler T3	Biometra
pH meter Accumet Basic	Fisher Scientific
SMART system	Amersham Pharmacia Biotech
Sonicator Ultrasonic Processor XL	Misonix Inc.
Spectrophotometer DU 640 UV/VIS	Beckmann
Synergy HT UV/VIS/fluorescence/luminescence plate reader	Bio-Tek
UV/VIS Spectrometer V-560	Jasco
Thermomixer Comfort	Eppendorf

Vortex	Ikamag
Water bath	Bioblock Scientific

3.1.5. Media

LB medium: 10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, (+ 15 g/l agar for solid medium). Adjusted to pH 7.0 with NaOH (Sambrook *et al*., 1989).

SOC medium 20 g/l tryptone, 5 g/l yeast extract, 0.5 g/l NaCl, 0.186 g/l KCl, 0.95 g/l $MgCl_2$. After autoclave, add 20 ml of filter sterilized 1M glucose (Sambrook *et al*., 1989).

M63 medium: 2 g/l $(NH_4)_2SO_4$, 13.6 g/l KH_2PO_4, 0.5 mg/l $FeSO_4$ x 7 H_2O. Before use, 1 ml $MgSO_4$ (1 M), 10 ml glucose (20 % w/v), L-amino acid mix (to 0.5 mM of each amino acid final) were added *per* 1 l medium and filter sterilized (Sambrook *et al*., 1989). 250 mM sucrose should be included for recovering bacterial spheroplasts.

SC medium 6.7 g/l yeast nitrogen base, 2 g/l Drop-out L-amino acid mix (contains 10 mg of para-aminobenzoic acid, 25 mg of adenine, 0.1 g of inositol and 0.1 g of each amino acid, except for histidine, leucine, tryptophan for selection purposes) were added *per* 1 l medium (+ 20 g/l glucose and 20 g/l agar for solid medium). (Ausubel *et al*., 2003).

YPD medium 10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose, 0.1 g/l adenine (optional), (+ 15 g/l agar for solid medium). (Ausubel *et al*., 2003).

3.1.6. Antibiotic stock solutions

Antibiotic additives to growth media were prepared as 1000x stock solutions and filter-sterilized before usage: ampicilin: 100 g/l, kanamycin: 25 g/l, spectinomycin: 50 g/l, chloramphenicol: 20 g/l (in ethanol), G418: 200 g/l.

3.2. Bacterial strains and plasmids

3.2.1. *E. coli* strains

DH5α F’	F’/ *endA1 hsdR17*(r_k^-, m_k^+) *glnV44 thi-1 recA1 gyrA* (Nal^r) *relA1*Δ*(lacIZYA-argF) U169 deoR (φ80dlac*Δ*(lacZ)M15)*
BL21(DE3) Gold (Stratagene)	B strain, F- *dcm*+ Hte *ompT hsdS*(r_B- m_B-) *gal l* (DE3) *endA* Tet^r
MC4100 (Genevaux *et al*., 2004)	F- *araD139* Δ*(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR*
MC4100Δ*tig* (Genevaux *et al*., 2004)	F- *araD139* Δ*(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR* Δ*tig*:: Cm^r
MC4100Δ*dnaK dnaJ* (Genevaux *et al*., 2004)	F- *araD139* Δ*(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR* Δ*dnaKdanJ*:: Kan^r
MC4100Δ*tig* Δ*dnaK dnaJ* (Genevaux *et al*., 2004)	F- *araD139* Δ*(argF-lac) U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 pstF25 rbsR* Δ*tig*:: Cm^r Δ*dnaKdanJ*:: Kan^r

3.2.2. *S. cerevisiae* strains

YPH499 (Sikorski and Hieter, 1989)	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1*
DS10 (Nelson *et al*., 1992)	*MAT*a, *lys1*, *leu2-3*, *112*, *ura3-52*, *his3-11*,*15 trp1*-D*1*
WY1 (Becker *et al*., 1996)	*MAT*a, *lys1*, *leu2-3*, *112*, *ura3-52*, *his3-11*,*15 trp1*-D*1,* Δ*ydj1*::*HIS3*
JY053	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*fes1::KAN*
YPH499 Δ*zuo1*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*zou1::KAN*
YPH499 Δ*ssz1*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*ssz1::KAN*
YPH499 Δ*ssa1/2*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*ssa1::KAN,* Δ*ssa2::klTRP1*
YPH499 Δ*ssb1/2*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*ssb1::KAN,* Δ*ssb2::klTRP1*
YPH499 Δ*gim1*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*gim1::KAN*
YPH499 Δ*egd2*	*MAT*a, *ura3–52, lys2–801, ade2–101, trp1-*Δ*63, his3-*Δ*200, leu2-*Δ*1* , Δ*egd2::KAN*

3.2.3. Plasmids

3.2.3.1. Construction of vectors for expression of fusion proteins in *E. coli*

A high copy number, T7 promoter-driven vector for recombinant protein expression in *E. coli* was created by removing the ROP gene from a pET-22b plasmid backbone (Novagen) by digestion with *Sap* I and *Ppu* MI restriction endonucleases, followed by end-filling with Vent DNA polymerase (New England Biolabs) and blunt-end ligation with T4 DNA ligase (New England Biolabs). For C-terminal GFP fusions, the sequence of the complete multiple cloning site of this vector was replaced by a stuffer fragment containing 5' *Nde* I and 3' *Hind*III sites flanking additional *Spe* I, *Not* I and *Nhe* I sites, as well as sequences encoding a His_6 tag and a stop codon between the *Nhe* I and *Hind* III sites (creating the pCH sereies). Synthetic oligonucleotides encoding the amino acid linkers indicated in Table 2 were introduced between the *Spe* I and *Not* I sites. A PCR-generated DNA fragment encoding GFPuv (Crameri *et al*., 1996) was inserted between the *Not* I and *Nhe* I sites, with the initial Met substituted by Ile, to exclude internal initiation of translation (creating the pCH-L-cGFP series, where L denotes the linker). DNA fragments encoding the fusion partners (NusA, MBP, Eno and MreB) were amplified by PCR and inserted between the *Nde* I and *Spe* I sites of pCH-LcGFPvectors (creating the pCH-ORF-cGFP series). For N-terminal GFP fusions, a stuffer sequence containing *Nco* I (instead of *Nde* I), *Spe* I, *Not* I, *Nde* I and *Nhe* I was introduced between the *Nde* I and *Nhe* I sites of a pCH-L-cGFP vector, by blunt- and sticky-end ligations. PCR-generated fragments encoding GFPuv and the GFPuv sequence with the L5 linker as a new N-terminus were inserted between the *Nco* I and *Spe* I sites (creating pCH-nGFP-L5 andpCH-L5-nGFP-L5). The C-terminal L5 linker was further substituted with oligonucleotides encoding sequences for the L16 and L25 linkers. DNA fragments for the enolase and MBP fusion partners were cut by digestion with *Nde* I and *Spe* I from the corresponding pCH-ORFcGFP vectors and inserted between the *Nde* I and *Nhe* I sites (creating the pCH-nGFP-L-ORF and pCH-L5-nGFP-L-ORF series).

Vectors for expression of luciferase-GFP and GFP-luciferase fusions were generated as described above, except that only vectors with L16 linkers were generated. For regulated expression in *E. coli*, Eno-GFP and GFP-Eno with L16 linkers were generated by sub-cloning the *Xba* I – *Hind* III fragments from the vectors described above into the *Xba* I – *Hind* III –cut arabinose-controlled promoter vector pBAD18 (Guzman *et al*., 1995). The sequences of all final constructs were confirmed by DNA sequencing (Sequiserve and Medigenomix GmbH).

For the *E. coli* trigger factor expression construct, the *tig* gene was PCR amplified from *E. coli* MG1655 genomic DNA and inserted into pCH vector *via Nde* I and *Nhe* I sites for C-terminally His_6-tagged protein. The resulting construct was designated as pCHTF.

For the arabinose promoter-controlled firefly luciferase expression construct, the coding region of the firefly luciferase gene (containing Myc and His_6-tags) were excised from the pET3a-Luciferase construct (from the laboratory collection) with *Nde* I and *Hind* III sites and introduced into pBAD18 vector. The resulting construct was designated as pBADLuc.

Table 2. Sequences and properties of the different amino acid linkers used in this study.

Designation	Amino acid Sequence	Properties	References
L5	TSAAA	5 aa linker that results from consecutive *Spe* I and *Not* I restriction endonuclease sites	--
L15a	TSMTATADVLAMAAA	15 aa naturally occurring interdomain linker highly conserved across kingdoms	(Nett *et al*., 2000)
L15b	TSGGSGGSGGSGAAA	15 aa uncharged flexible linker previously used to fuse multi-domain proteins	(Netzer and Hartl, 1997)
L16	TSGSAASAAGAGEAAA	16 aa flexible linker previously used in combination with GFP in *E. coli*	(Waldo *et al*., 1999)
L25	TS$(GGGGS)_4$AAA	25 aa flexible linker used extensively in recombinant antibody fragments	(Freund *et al*., 1993)
L37	TSAG$(EAAAK)_6$AAA	37 aa α-helical linker used for intramolecular FRET between fluorescent proteins	(Arai *et al*., 2001)

3.2.3.2. Construction of vectors for expression of fusion proteins in *S. cerevisiae*

For expression of GFP fusion proteins under control of the GAL1-promoter in *S. cerevisiae,* plasmids were generated by sub-cloning the *Xba* I – *Hind* III fragments encoding the corresponding fusion proteins from the *E. coli* vectors described above into *Xba* I – *Hind* III cut p415-GAL1 (Mumberg *et al*., 1994) vector.

For expression of firefly luciferase under control of the GAL1-promoter in *S. cerevisiae,* plasmids were generated by sub-cloning the *Xba* I – *Hind* III fragments encoding the corresponding gene from the *E. coli* pBDLuc vector described above into *Xba* I – *Hind* III cut p415-GAL1 vector. The resulting construct was designated as p415Gal-Luc. A copper-promoter controlled construct was generated similarly, except for the utilization of *BamH* I, *Hind* III sites for sub-cloning into the p425-CUP vector. The resulting construct was designated as p425Cup-Luc.

For expressing GroEL Class II and Class III substrates in yeast, constructs were generated by subcloning the coding region of each GroEL substrate from pET22b-based clones (Kerner *et al*., 2005) into the p415-GAL1 vector. The restriction endonuclease sites utilized were as follows: ADD, ALR2, DAPA, HEM2, METF, METK, NPL, XYLA and YAJO (*Xba* I and *Xho* I); END4, LLDD, DCEA and SYT (*Xba* I and *Hind* III); GATY (*Nde* I, *Xho* I); GATD (*Xba* I and *Sal* I).

For expressing GroEL and GroES in yeast, the GroEL encoding gene was cut out from the pDN2 vector (from the laboratory collection) *via EcoR* I digestion and inserted into a copper-inducible pSI215 vector. The GroES encoding gene was cut out from the pET11aES vector (from the laboratory collection) *via BamH* I and *Xba* I sites and inserted into an ADH promoter-regulated p426ADH vector (Mumberg *et al*., 1995), which allows constitutive expression of GroES in yeast.

3.3. Molecular cloning methods

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

For preparation of chemically-competent *E. coli* cells, a single colony was used to inoculate 500 ml LB medium (including antibiotic, if applicable) and grown to an optical density (OD_{600}) of 0.25 - 0.5 at 37 °C. The cells were then chilled on ice for 15 min and harvested at 5000 x *g* for 10 min at 4 °C. The cell pellet was washed with 80 ml ice-cold Ca/glycerol buffer (10 mM PIPES, 60 mM $CaCl_2$, 15 % glycerol; pH 7.0, adjusted with NaOH, and filter-sterilized) once and incubated with additional 80 ml Ca/glycerol buffer on ice for 30 min. Finally, the cells were pelleted and resuspended in 6 ml of Ca/glycerol buffer. 100 µl aliquots were frozen in liquid nitrogen and stored at –80 °C.

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

Alternatively, electroporation was applied to improve the transformation efficiency and avoid the heat shock process for certain bacterial strains (*e.g.* MC4100Δ*tig*Δ*dnaKdnaJ* strain). Electrocompetent cells were prepared as follows: 500 ml bacterial culture was grown to an optical density (OD_{600}) of 0.8 in LB medium at the appropriate temperature (25 °C for MC4100Δ*tig*Δ*dnaKdnaJ* strain). The cells were washed carefully with 250 ml ice-cold sterilized water for two times and finally the cells were pelleted and resuspended in 2 ml of 10% glycerol. 40 µl aliquots were frozen in liquid nitrogen and stored at –80 °C. For electroporation transformation, competent cells (40 µl) were mixed with 1-2 µl plasmid DNA (or ligation product) and transferred into a 0.2 cm Gene Pulser cuvette. The electroporation was done at 2.5 kV, 25 µFD and 200Ω settings with a Gene Puser II

elecporation device. The transformed cells were allowed to recover in 1 ml of SOC medium with 225 rpm shaking at appropriate temperature for 1 h. The cell suspension was then plated on selective plates and incubated until colonies had developed (Dower *et al*., 1988).

3.3.2. Preparation and transformation of *S. cerevisiae* competent cells

To prepare competent *S. cerevisiae*, yeast cells were inoculated in YPD medium (including antibiotic, if applicable) and grown to an optical density (OD_{600}) of 0.5 - 0.6 at 30 ^{o}C. The cells were then harvested at 2000 x *g* for 5 min at room temperature. After washing with 1 volume of water and 1/4 volume of LiSorb buffer (10 mM Tris-Cl, I mM EDTA, 100 mM LiOAc, 1 M sorbitol; pH 8.0, filter-sterilized), the cells were pelleted and resuspended in 1/150 volume of LiSorb buffer containing 1 mg/ml of Yeast-marker carrier DNA. 10 µl aliquots were frozen in liquid nitrogen and stored at –80 ^{o}C.

For transformation, 10 µl competent cells were mixed well with 0.1 - 0.2 µg (in ~ 1 µl volume) plasmid DNA and 60 µl LiPEG buffer (10 mM Tris-Cl, I mM EDTA, 100 mM LiOAc, 40% PEG3500; pH 8.0, filter-sterilized), followed by a 20 min incubation at room temperature. After adding 7 µl DMSO, the cells were heat-shocked at 42 °C for 15 min. The cells were then collected (1500 x *g*, 1 min), plated on YPD or SC selective plates and incubated at 30 °C until colonies became visible.

3.3.3. Plasmid purification

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

3.3.4. PCR amplification

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

DNA Template:	10-20 ng (plasmid DNA) 250 ng or less (bacterial genomic DNA)
Primers:	20 pmol each
dNTPs:	200 µM each
Pfu DNA Polymerase:	2.5 U
Polymerase buffer:	1 x
Additives:	4 % DMSO if GC content was >50 %,
Final volume:	50 µl

Cycling conditions (35 cycles):

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

Stored at 4 °C or -20 °C.

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

3.3.5. DNA restriction and ligation

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

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

3.3.6. DNA analytical methods

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

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

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

3.3.7. Gene disruption in *S. cerevisiae*

Yeast strains devoid of nascent chain binding chaperones Zuotin, Ssz1p, Ssb1/2p Gim1p, Egd2p (αNAC) and Ssa1/2p were constructed by direct replacement of the corresponding genes by the *kanMX4* cassette (Wach *et al*., 1994), or in combination with the *klTRP1* cassette (Knop *et al*., 1999), if for simultaneously deletion of two genes. PCR

primers were designed to amplify the selection cassettes with 40 bp of DNA homologous to the genomic target locus. PCR fragments (2 μg) were gel purified and transformed into 50 μl of YPH499 yeast competent cells as described in 3.3.2. After 2 hr recovered in 1 ml YPD at 30°C, cells were plated on YPD plates or SC-Trp plates supplemented with 200 mg/l G418. G418 resistant colonies were confirmed to have correct integration of cassettes at the target locus by PCR analysis.

3.4. Protein purification

Plasmids for expression of the following proteins were obtained from the Hartl laboratory collection: wild-type TF (wt-TF) and the FRK/AAA TF mutant carrying C-terminal His_6-tags, DnaK, DnaJ and GrpE. Proteins were overexpressed in *E. coli* and purified as described (Hesterkamp *et al*., 1997; Szabo *et al*., 1994).

E. coli enolase and GFPuv were expressed and purified from *E. coli* BL21 (DE3) Gold cells harboring the corresponding pCH-Eno and pGFPuv constructs, respectively. Cultures were grown in 1 l LB medium containing 100 mg/l ampicillin until the cell density reached A_{600} = 0.5, and followed by 2 h induction at 37°C with 1 mM IPTG. Cells were harvested and resuspended in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) with Complete EDTA-free protease inhibitor (1 tablet/ 25 ml). Lysis was achieved by homogenization of the cell suspension in an EmulsiFlex high pressure homogenizer device kept on ice. The lysate was cleared by centrifugation (40,000 x *g* for 1 h at 4 °C) and the supernatant was applied to a 5 ml HiTrap metal chelating column pre-charged with Ni^{2+}. The column was washed with a gradient of 10 to 50 mM imidazole in PBS (for over 10 column volumes) and the proteins were eluted with 250 mM imidazole in PBS, followed by desalting into PBS with a HiPrep desalting column. After the determination of protein

concentrations (described below), proteins were aliquoted, flash-frozen in liquid nitrogen and stored at -80 ºC.

Eno-GFP and GFP-Eno fusion proteins (with L16 linkers) were expressed and purified from *S. cerevisiae* YPH499 cells. Yeast cells harboring GFP fusion constructs were grown in 100 ml SC-Leucine starting medium (containing 2% raffinose) at 30 ^{o}C until cell density reached A_{600} = 3 - 4. The starting culture was further poured into 2 l SC-Leucine inducing medium (containing 2% galactose) for recombinant GFP fusion protein expression for 24 h at 30 ^{o}C. The purification process was essentially as described above.

3.5. Protein analytical methods

3.5.1. Determination of protein concentration

Protein concentrations were determined spectrophotometrically by A_{280} (in 6 M Gdm-HCl), 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).

3.5.2. SDS-PAGE (sodium-dodecylsufate polyacrylamide gel electrophoresis)

SDS-Polyacrylamide gels were prepared as follows:

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

SDS-PAGE was performed using a discontinuous buffer system (Laemmli, 1970) in BioRad Mini-Protean 3 electrophoresis chambers employing a constant current of 30 mA/gel in 50 mM Tris-Base, 380 mM glycine, 0.1 % SDS (pH 8.3). SDS loading buffer was added to the protein samples (final concentration: 60 mM Tris-HCl, pH6.8, 1% SDS, 10 % glycerol, 0,01% Bromophenol blue, 0,1 mM β-mercaptoethanol). Coomassie blue staining polyacrylamide gels were fixed and stained in 0.1 % Coomassie brilliant blue R-250, 40 % ethanol, 7 % acetic acid for 1 h or longer and de-stained in 20 % ethanol, 7 % acetic acid.

3.5.3. Western-blotting

Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes in a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) in 25 mM Tris, 192 mM glycine, 20 % methanol, pH 8.4 at constant current of 150 mA/gel for 1 h (Towbin *et al*., 1979). Nitrocellulose membranes were blocked in 5 % skim milk powder in TBST (50 mM Tris-

Cl, pH8.0, 150 mM NaCl, 0.05% Tween 20) for 1 h. The membranes were then incubated with a 1:2000 – 1:10000 dilution of primary antibody serum in TBST and extensively washed in TBST before incubation with a 1:5000 (for anti-mouse IgG) or 1:10000 (for anti-rabbit IgG) dilution of secondary antibody in TBST (anti-rabbit IgG and anti-mouse IgG, whole molecule – horseradish peroxidase conjugate, Sigma-Aldrich). After extensive washing, protein bands were detected by incubating the membranes with ECL chemiluminescence solution and exposure to X-ray film (High performance chemiluminescence film) or a Luminescent Image Analyzer LAS-3000 system.

3.6. Expression and assessment of protein solubility

3.6.1. Expression and assessment of protein solubility in *E. coli*

E. coli cells transformed with plasmids encoding the desired proteins were grown to OD_{600} = 0.5 in LB medium and induced with 0.2% arabinose (for arabinose promoter controlled constructs) or 1 mM IPTG (for T7 promoter controlled constructs) for the indicated times and at desired temperatures. Normally, 2 OD_{600} units of cells (2 ml of OD_{600} = 1) were taken for recombinant protein solubility assessment. They were harvested by centrifugation (5000 x *g*, 5 min) and spheroplasts were prepared as follows (Ausubel *et al*., 2003): the cell pellet was resuspended in 100 µl of 50 mM Tris-Cl, pH 8.0, 20% sucrose, followed by addition of 20 µl of 5 mg/ml lysozyme solution and incubating the cells on ice for 5 min. 40 ul of 20 mM of EDTA were further added and the cells were incubated on ice for 5 min. Finally, the cell suspension was mixed with 40µl of 50 mM Tris-Cl, pH 8.0 and the incubation temperature was shifted to 30 ^{o}C for 5 min. Spheroplasts were lysed by dilution into an equal volume (200 µl) of 2X lysis buffer (0.2% Triton X-100, Complete EDTA-free protease inhibitors, 100 U/ml Benzonase in PBS or Tris buffer as indicated). Aliquots were fractionated into supernatant and pellet by centrifugation (20,000 x *g* for 30

min) and analyzed by 12% SDS-PAGE and Coomassie blue staining or immunoblotting, as indicated.

3.6.2. Expression and assessment of protein solubility in *S. cerevisiae*

Wild-type YPH499 *S. cerevisiae* cells transformed with plasmids encoding the desired proteins were grown to OD_{600} = 0.8 in SC –Leucine medium (or other auxotrophic selection medium, as indicated) and induced with 2 % galactose for 2-4 h at 30 ºC. Normally, 5 OD_{600} units of cells (5 ml of OD_{600} = 1) were taken for recombinant protein solubility assessment. They were harvested by centrifugation (3000 x *g*, 5 min) and spheroplasts were prepared by resuspending the cell pellet in 250 µl of Zymolyase buffer (100 mM Tris-Cl, pH 7.6, 1.2 M sorbitol, 10 mM $CaCl_2$, 0.5 mg/ml Zymolyase 100T) at 30 ^oC for 30 min. Spheroplasts were washed with 1 ml wash buffer (100 mM Tris-Cl, pH 7.6, 1.2 M sorbitol, 10 mM $CaCl_2$) for at least two times and harvested at 1200 x *g* for 5 min. Spheroplasts were lysed with 250 µl of lysis buffer (0.1% Triton X-100, Complete EDTA-free protease inhibitors, 50 U/ml Benzonase in PBS or Tris buffer as indicated). Aliquots were fractionated into supernatant and pellet by centrifugation (20,000 x *g* for 30 min) and analyzed by 12% SDS-PAGE and Coomassie blue staining or immunoblotting, as indicated.

3.7. *In vivo* experiments

3.7.1. Quantitation of rates of accumulation of luciferase activity in *E. coli* and *S. cerevisiae*

E. coli and *S. cerevisiae* cells transformed with expression constructs for luciferase-GFP and GFP-luciferase were grown in LB and SC–Leucine media, respectively, at 30 ºC and induced at OD_{600} = 0.8 with IPTG (1 mM) or galactose (2%), respectively. Equivalent OD_{600} units of cells were lysed under native conditions as described above at the time points indicated. Luciferase activity was determined (see below, 3.7.2) and normalized by ribosome content in each system [with 1 OD_{600} unit = 8 x 10^8 bacterial cells and 1 OD_{600} unit = 2 x 10^7 yeast cells (Ausubel *et al*., 2003), and 20,000 and 200,000 ribosomes *per* cell in *E. coli* (Bremer and Dennis, 1996) and *S. cerevisiae* (Warner, 1999), respectively]. Total fusion protein production was examined by loading equivalent amounts of total lysed cytosol and analyzing them on SDS-PAGE followed by Coomassie blue staining.

3.7.2. Determination of enzyme activity and solubility *in vivo*

Wild-type, Δ*tig*, Δ*dnaKdnaJ*, or Δ*tig*Δ*dnaKdnaJ E. coli* MC4100 strains *(Genevaux et al*., 2004) transformed with arabinose-controlled expression plasmids for firefly luciferase (FL) or β-galactosidase (β-gal) with C-terminal c-Myc His_6-tags were grown in LB medium to an OD_{600} = 0.5 at 30 ^oC. Protein expression was induced with 0.2% arabinose for 15 min. Spheroplasts were produced as described above and lysed in an equal volume of 2X lysis buffer (0.2% Triton X-100, 100 U/ml Benzonase, Complete EDTA-free protease inhibitors) in 50 mM Tris-HCl (pH 7.5), 5 mM $MgSO_4$ for FL assays, or lysis buffer in 200 mM sodium phosphate (pH 7.3), 2 mM $MgCl_2$, 100 mM β-mercaptoethanol for β-gal assays. Aliquots were fractionated into supernatant and pellet by centrifugation (20,000 x *g* for 30

min). Activities were measured with the Luciferase Assay System (Promega, E1501) in a Luminometer (Lumat LB 9507) and the β-galactosidase Enzyme Assay System (Promega, E2000) in a Spectrophotometer DU640. Protein quantitation was performed by immunoblotting using an anti-c-Myc 9E10 monoclonal antibody, followed by densitometry. GroEL/ES (from plasmid pOFXtac-SL2, (Castanie *et al*., 1997) and DnaK/DnaJ/GrpE (from plasmid pOFXtac-KJE1, (Castanie *et al*., 1997)) were overexpressed in the above strains by induction with 0.5 mM IPTG for 30 min before induction of FL, which was carried out under identical conditions as above. Overexpression of FL in *S. cerevisiae* (YPH499) was carried out in cells transformed with an expression plasmid (p415GAL-Luc) for FL under galactose promoter control grown in SC-Leucine medium to an OD_{600}=0.8 at 30 °C. Protein expression was induced with 2% galactose for 4 h. Spheroplasts were prepared and analyzed as described above.

3.7.3. Determination of folding kinetics *in vivo*

Live spheroplasts from wild-type and mutant bacterial strains transformed with expression plasmids for FL or β-gal under an arabinose promoter were allowed to recover at 30 °C for 30 min with gentle shaking in M63/sucrose medium. Labeling and induction were performed by adding 60 µCi/ml ^{35}S-Methionine and 0.5% arabinose at 30 °C. Aliquots were taken at the time points indicated and lysed immediately by mixing in an equal volume of 2X lysis buffer as described above, containing 10 U/ml apyrase and 50 µg/ml chloramphenicol and placed on ice. Enzyme activities were measured as described above. The amount of full-length protein in each aliquot was determined by SDS-PAGE followed by Phosphorimager quantitation (AIDA gel imaging software version 2.31). After 40–50 min of incubation (arrows in Figure 20), the spheroplast preparation was divided into halves, one of which was treated with chloramphenicol (CAM; 200 µg/ml).

S. cerevisiae (YPH499) cells transformed with an expression plasmid for FL under copper promoter control (p425Cup-Luc) were grown in SC-Leucine-Methionine medium to an OD_{600} = 0.8 at 30 °C. Labeling and induction were performed at 30 °C by adding 100 µCi/ml ^{35}S-Methionine and 1mM $CuSO_4$. Two aliquots were taken for each time point. One was placed immediately in liquid nitrogen and used for SDS-PAGE analysis. The other aliquot was used to determine FL activity in intact cells (Greer and Szalay, 2002) by mixing with 20 volumes of 8 mM luciferin (potassium salt) in water and measuring light emission immediately. After 40 min of incubation, the culture was divided into halves, one of which was treated with cycloheximide (arrows in Figure 21) (CHX; 1.4 mg/ml).

3.7.4. *De novo* folding of firefly luciferase in *S. cerevisiae* Δ*fes1* strain

S. cerevisiae strains JY053 and YPH499 were transformed with p425Cup-Luc plasmid for firefly luciferase (FL; containing a c-Myc and a His_6 tag at the C-terminus) expressions. Cells were grown in SC-Leucine medium to an OD_{600} = 0.8 at 30 °C. Protein expression was induced by addition of 0.25 mM $CuSO_4$ for 3 h either at 30 °C or 37 °C. Spheroplasts were prepared as described above and lysed in luciferase dilution buffer (25 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM CDTA, 10% glycerol, and 0.1% Triton X-100) containing Complete EDTA-free protease inhibitors. Samples were fractionated into supernatant and pellet by centrifugation (20,000 x *g* for 30 min). FL activities were determined with the Luciferase Assay System (Promega). Protein quantitation was performed by immuno-blotting with the anti-c-Myc 9E10 monoclonal antibody followed by densitometry.

For Size exclusion chromatography, fresh lysates were made in the presence of apyrase (20 U/ml, Sigma), further separated by a Superdex 200 PC 3.2/30 column installed in a SMART system device. The column was equilibrated at room temperature in luciferase

dilution buffer. The luciferase activity assay and fraction quantitations were performed as described above.

Alternatively, apyrase-treated lysate was diluted 1:1 with 2X binding buffer (50 mM HEPES-KOH, pH 7.5, 200 mM KOAc, and 10% glycerol) containing 2% BSA and 15 µl Nickel-NTA agarose beads and incubated on a rotator at 4 ^{o}C for 2 h for the pull-down experiment. After three washes with 200 µl binding buffer containing 0.1% Triton X-100, bound protein was eluted in an equal volume of warmed SDS sample buffer and subjected to SDS-PAGE analysis. Ssa1 and Ydj1 were detected by Western blotting with specific antisera.

3.7.5. Expression of GroEL substrates in *S. cerevisiae*

Bacterial proteins of class I, II and III (Kerner *et al*., 2005) were expressed in *S. cerevisiae* YPH499 cells transformed with expression plasmids p415GAL1 under galactose promoter control grown in SC–Leucine medium at 30°C. Protein expression was induced at OD_{600}= 0.5 with 2% galactose for 4 h. Spheroplasts were prepared as described above and lysed in PBS containing 0.1% Triton X-100 and Complete EDTA-free protease inhibitors. Samples were fractionated into soluble and pellet fractions by centrifugation (20,000 x *g* for 30 min). Protein amounts were analyzed by immuno-blotting. For GroEL/GroES co-expression with class III substrates (ADD, ALR2, DAPA, END4, GATY, HEM2, LLDD, METF, METK, NPL, XYLA and YAJO), 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–Leucine–Trptophan–Uracil medium. GroEL was induced with 0.5 mM $CuSO_4$ for 3 h before induction of bacterial GroEL substrates. To examine bacterial class I (ENO) and class II (DCEA, GATD and SYT) substrate solubility

in *YDJ1*-deleted yeast, proteins were expressed in the strain wy1 (Δ*ydj1*) and its isogenic wild-type strain DS10 (Becker *et al*., 1996) and analyzed as above.

3.7.6. Fluorescence microscopy

S. cerevisiae cells expressing the indicated GFP-fusion proteins were washed with water, placed between microscope slides and cover slips and visualized under UV illumination in an Axiovert 200M microscope equipped with Filter set 38, an AxioCAM HRmdigital camera and Axiovision 3.1 software (Carl Zeiss).

3.8. *In vitro* protein assays

3.8.1. *In vitro* refolding assays

Native GFP, enolase, Eno-GFP and GFP-Eno purified as described above were denatured in 6 M Gdm-HCl, 2 mM DTT in PBS in concentrations of 50, 25 and 12.5 µM at 25 ºC for 30 min. To begin refolding, proteins were diluted 100-fold into refolding buffer (100 mM KCl; 10 mM $MgCl_2$; 20 mM MOPS-KOH, pH 7.5) at 25 ºC to final concentrations of 500, 250 and 125 nM. Refolding of the GFP domain was determined directly by monitoring green fluorescence (excitation at 398 nm and emission at 508 nm) at 25 ºC in a Fluorolog spectrofluorometer (ISA Instruments). Refolding of the enolase domain was determined by diluting aliquots of the refolding reaction at the time points indicated into assay mix (1 mM D(+) 2-phosphoglyceric acid in 100 mM KCl; 1 mM $MgSO_4$; 10µM EDTA; 50 mM Tris, pH 8.0) and measuring ΔA_{240}/time at 25 ºC (Spring and Wold, 1975) in a V-560 spectrophotometer (Jasco). Fluorescence and enolase activity were plotted relative to those of the respective native non-fused domains and fusion proteins at the same

concentrations, which were set to 100%. Equivalent molar amounts of each domain in the fusion proteins and non-fused domains displayed similar fluorescence/enolase activity.

3.8.2. Translation and determination of enzyme activity *in vitro*

Protein expression for *in vitro* transcription/translation systems was from plasmids with a T7 promoter. Bacterial S30 translations were carried out in the coupled RTS 100 HY transcription/translation system (Roche) and RRL translations in the TNT coupled system (Promega). Translation reactions were started by adding template DNA to a final concentration of 5 ng/µl and run for 1 h at 30 ^{o}C. Further centrifugation were applied (22,000 x *g* for 15 min at 4 ^{o}C) to separate soluble and insoluble fractions. For visualizing the translated product by autoradiography, 1 µl ^{35}S-methionine (Amersham; 1000 Ci/mmol, 15 mCi/ml) were included in a typical 50 µl reaction. Unless indicated otherwise, chaperones were added to S30 translations at the following concentrations: TF, 5 µM; DnaK/DnaJ/GrpE, 10/2/6µM; and GroEL/GroES, 1 /2 µM, respectively.

β-galactosidase activity was assayed as described above. Initial velocities ($\Delta A_{420}/\Delta$time) *versus* time of translation were plotted to estimate the kinetics of β-gal folding. Prior to spectrophotometric measurements, translation aliquots were diluted 5-fold in stopping buffer (20 mM HEPES-KOH, pH 7.3, 100 mM KCl, 5 mM $MgCl_2$, 100 µl/ml RNase A, 10 U/ml apyrase) at 30 ^{o}C for 1 min. The observed initial velocity data were always linear and independent of the stopping step. Thus, tetramer assembly was not rate limiting and was thus tightly coupled to folding and translation.

FL activity measurements were performed as described above. Translation aliquots were diluted 100-fold into luciferase dilution buffer (25 mM Tris-phosphate, pH 7.4, 2 mM

CDTA, 2 mM DTT, 1% Triton X-100, 1 mg/ml BSA) before measuring enzymatic activity. Relative specific activities were calculated by normalizing activity values with the relative intensities of full-length protein bands measured by phosphorimaging.

3.8.3. Post-translational folding assay

FL S30 translations were stopped after 22 min by adding RNaseA (50 µg/ml) or CAM (200 µg/ml). FL activity was measured immediately before the addition of RNaseA or CAM and at regular intervals until 60 min. The resulting activities were normalized by setting the initial value before RNaseA addition to unity. Ribosome-associated nascent chain complexes were prepared as published (Beck *et al*., 2000). An antisense oligonucleotide (21-mer) directed to the C terminus of the luciferase construct was used at a final concentration of 190 µg/ml, and the anti-ssrA oligonucleotide and RNaseH were present at 50 µg/ml and 80 U/ml, respectively.

3.8.4. Ribosome binding of TF

Translation mixes (without ^{35}S-Met and DNA) were incubated at 30 ^{o}C with increasing concentrations of purified TF-His_6. Total ribosomes were isolated by sucrose cushion centrifugation (Hesterkamp *et al*., 1996). These ribosomes were resuspended in 20 mM HEPES-KOH (pH 7.2), 10 mM $MgCl_2$, and 100 mM K-Acetate and quantitated using A_{260} (Spedding, 1990). Amounts of bound TF-His_6 were determined by quantitative immunoblotting.

3.8.5. Ribosome recruitment assay

The postribosomal supernatant (PRS) from an in vitro translation of wt-TF with ^{35}S-Met was diluted 15-fold into translation reactions (with 1.25 mM unlabeled Met) of FL or GFP in the presence of excess (6 μM) unlabeled, purified wt-TF. At different times following initiation of translation, aliquots were removed and treated with CAM (100 μg/ml) on ice and then centrifuged at 4 ^{o}C (22,000 x g for 5 min). The resulting supernatants were subjected to sucrose cushion centrifugation to isolate ribosomes. Fifteen minutes after initiation of translation, the reaction was separated into halves, one of which was treated with CAM (100 μg/ml) and the other with RNaseA (50 μg/ml) at 30 ^{o}C, and processed as above. The ^{35}S-Met-labeled PRS from a translation of the TF mutant FRK/AAA was also diluted similarly into an independent translation reaction and processed identically. Under the conditions described, there was no translation of fresh wt or mutant TF when the PRS was added to a fresh translation in the presence of ^{35}S-Met.

3.8.6. Kinetic simulation

A simple three-state model ($U_1 \rightarrow U_2 \rightarrow N$) was used to simulate FL folding kinetics in the context of translation, with U_1 representing all species preceding the complete polypeptide chains, U_2 representing full-length but nonnative chains, and N representing the folded, full-length polypeptide chains. The value of k_1 was estimated from the *in vitro* translation kinetics (Figure 19B) to be ~0.1 min^{-1} ($U_1 \rightarrow U_2$), and k_2 was set to 0.0693 min^{-1} ($U_2 \rightarrow N$), which corresponds to a $t_{1/2}$= 10 min for the KJE-assisted refolding of FL.

3.8.7. Prediction of DnaK binding sites

The protein sequences of firefly luciferase, β-galactosidase and Semliki Forest Virus Protease (SFVP) (Swiss-Prot identifiers: LUCI_PHOPY, BGAL_ECOLI and POLS_SFV) were analyzed as described (Rudiger *et al*., 1997b) using a spreadsheet template kindly provided by the authors.

4. Results

4.1. *De novo* folding of multi-domain GFP fusion proteins in *E. coli* and *S. cerevisiae*

Recent genome sequencing projects have provided vast amounts of information and molecular tools to study biological phenomena at a Systems Biology level across the kingdoms of life (Aebersold and Mann, 2003). Extensive sequencing analysis has revealed that eukaryotic genomes encode a higher abundance of multi-domain proteins than their prokaryotic counterparts. To date, 40% of proteins encoded by prokaryotes have been predicted to contain two or more domains, compared to 65% in eukaryotes, regardless of whether a domain is considered an "independently folding" or "independently evolving" unit (Ekman *et al*., 2005). This phenomenon probably occurred as a result of random gene fusion events that led to the production of modular polypeptides with novel functions as genomes became more complex during evolution (Kummerfeld and Teichmann, 2005). In order to systematically address whether recombinant multi-domain proteins fold differently in prokaryotes *versus* eukaryotes, we generated a series of fusions composed of two proteins with different folding properties and analyzed their behavior upon expression in *E. coli* and the yeast *S. cerevisiae*.

One of the proteins selected was the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. GFP is a 27 kDa, single domain protein in which covalent cyclization of three contiguous amino acid side chains (Ser65-Tyr66-Gly67) leads to the formation of an internal fluorophore (Tsien, 1998). Importantly, fluorophore formation occurs only after GFP has acquired its native β-barrel structure. Thus, GFP fluorescence is a reliable indicator of correct folding and has been used in *E. coli* to monitor folding of proteins to which it is recombinantly fused (Waldo *et al*., 1999). In order to avoid a strong dependency on the

bacterial GroEL/ES system for folding, we utilized a mutant version of GFP, called GFPuv or the Cycle3 mutant (Crameri *et al*., 1996), which does not require assistance by the chaperonin for folding. GFPuv was obtained through a DNA shuffling method and shown to result in much higher green fluorescence upon the expression in both prokaryotic (*E. coli*) and eukaryotic (Chinese Hamster Ovary, CHO) cells. The three mutation sites, F99S, M153T and V163A, are all located near the surface of the GFP molecule, allowing GFPuv to escape from a strong tendency to aggregate (Fukuda *et al*., 2000). The ability to fold efficiently rendered GFPuv a suitable candidate for a fusion partner, since it allowed us to focus mainly on the effects of the domain fusion, presumably causing an increased complexity of folding. Based on this idea, the four proteins selected as N-terminal GFP fusion partners are also characterized by robust folding and can be expressed to high levels in their native form in *E. coli*: maltose binding protein (MBP; 42 kDa), NusA (55 kDa), MreB (37 kDa) and enolase (Eno; 46 kDa). MBP is a globular protein divided into two compact domains by a maltose-binding groove (Spurlino *et al*., 1991)), while NusA is a rod-shaped molecule composed of four sequential domains (Worbs *et al*., 2001). Both are monomeric proteins that have been extensively used as N- and C- terminal carrier proteins to aid in the solubilization of aggregation-prone eukaryotic proteins (Terpe, 2003). Enolase is a single domain TIM-barrel protein that exists as a dimer, but it is enzymatically active as a monomer (Kuhnel and Luisi, 2001). MreB is a structural member of the actin superfamily fold, characterized by four sub-domains with extensive contacts surrounding a nucleotide-binding pocket. It can polymerize into filaments similar to actin (van den Ent *et al*., 2001). We utilized a protein of eukaryotic origin (GFP) with the other four bacterial proteins (MBP, NusA, MreB and Eno) for generating our artificial multi-domain proteins in order to minimize possible bias introduced by the expression of heterologous proteins in both *E. coli* and *S. cerevisiae*. This allowed us to more closely determine the true capabilities of these systems in producing correctly folded multi-domain proteins.

4.1.1. *De novo* folding of GFP fusions is inefficient in *E. coli* yet efficient in *S. cerevisiae*

Plasmids encoding individual fusion partners, *i.e.* GFPuv, Eno, MBP, MreB and NusA were generated for expression in *E. coli* at 37 °C. As expected, these four bacterial proteins displayed greater than 90% solubility even when overexpressed to >20% of the total cellular protein. GFPuv also accumulated significant amount of soluble material of greater than ~50% of total expressed protein, which is in agreement with previous observations (Figure 7) (Fukuda *et al*., 2000).

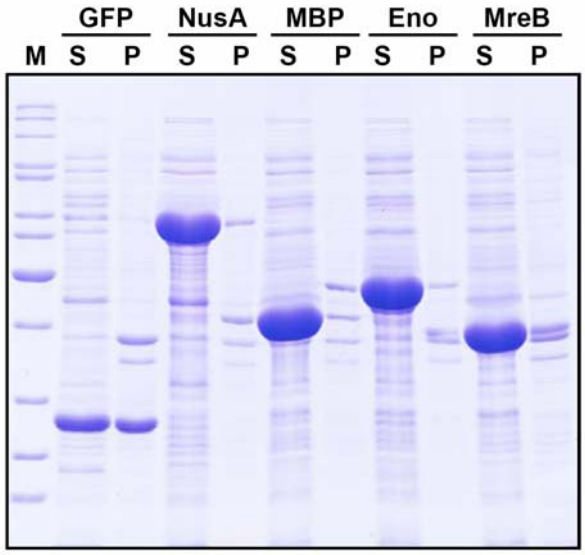

Figure 7. Solubility of GFP and nonfused protein partners upon expression in *E. coli*.

Protein samples were examined by SDS-PAGE and staining with Coomassie brilliant blue. Protein in the supernatant (S) and pellet (P) fractions is indicated. Marker proteins (M) are in descending order: 200 kDa, 150 kDa, 120 kDa, 100 kDa, 85 kDa, 70 kDa, 60 kDa, 50 kDa, 40 kDa, 30 kDa, 25 kDa and 20 kDa.

In addition to selecting appropriate fusion partners, the properties of the amino acid sequences used as domain linkers should be considered when performing protein fusion studies. To test whether the linker between GFPuv and the fusion partner had an influence on the folding of the resulting fusions, we selected six different linkers which varied in their length, predicted conformational rigidity, hydrophobicity and net charge (Table 2). The resulting 24 GFPuv fusion proteins were generated with GFPuv as the C-terminal fusion partner of Eno, MBP, MreB and NusA (Figure 8).

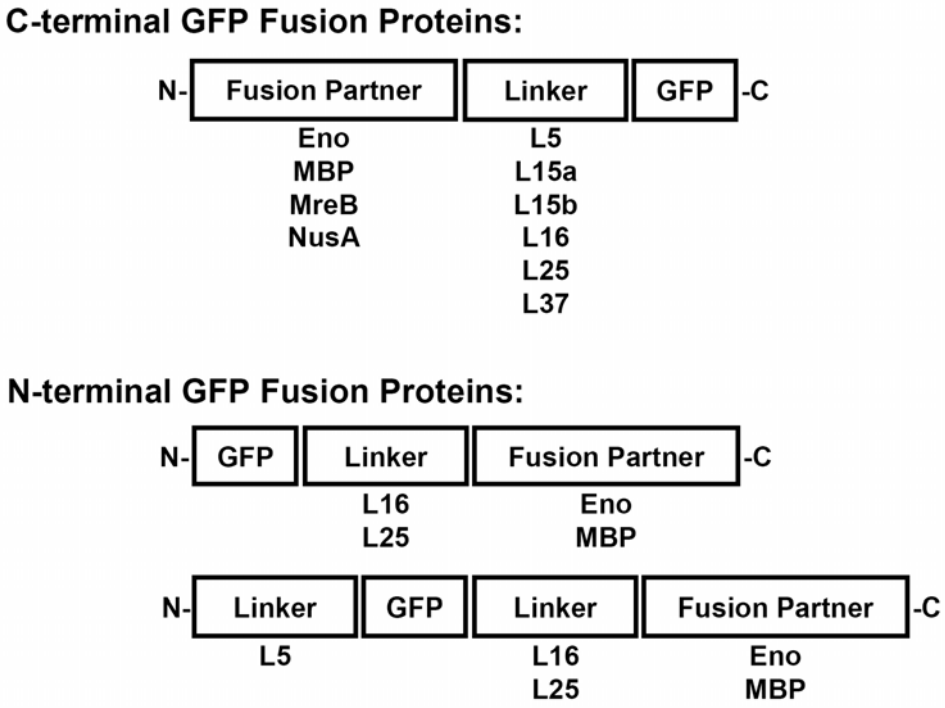

Figure 8. Constructs for expression of GFP fusions in *E. coli*

Diagram of the different phage T7 promoter-driven constructs for expression of GFP fusions in bacteria, with each component represented by a box. Fusion partners and linkers are listed underneath each box. Linker characteristics are given in Table 2.

After expression in *E. coli* of these fusions to levels similar to those of the non-fused proteins, we observed that the solubility of the fusion proteins was dramatically lower than that of their non-fused counterparts in all cases tested. As shown in Figure 9A, >90% of the Eno-, MreB-, and NusA-GFPuv fusions and >80% of the MBP-GFPuv fusions were present in the insoluble fraction, regardless of which linker was used. Substantial aggregation was observed when the robustly folding proteins were present at the N-terminus of the fusion proteins, which suggests that the C-terminal GFP moiety interfered with the folding of the more N-terminal partner. This also indicates that the N-terminal fusion partners failed to acquire their correct structure by a sequential, co-translational mechanism while the GFP moiety was still emerging from the ribosome.

To be certain that indeed fusing the domains *per se* was responsible for the observed misfolding instead of, for example, a steric problem due to the incorrect fusion order, we generated constructs where GFPuv was used as the N-terminal fusion partner to the bacterial proteins. As has been repeatedly described, robustly folding "carrier proteins" such as NusA or MBP, are fused to the C-terminus of aggregation-prone eukaryotic proteins in an attempt to aid their solubilization. Therefore it remains possible that GFPuv, a eukaryotic moiety, could fold better if fused to carrier proteins at its N-terminus. Since there was no major difference among the linkers tested, we decided to utilize either L16 or L25 for further experiments. Additionally, since the linker sequence present in front of GFP in the fusions mentioned above could potentially be responsible for the observed misfolding, we decided to include in this analysis N-terminal GFPuv fusions containing a sequence identical to linker L5 as an N-terminal segment (Figure 8).

Interestingly, the resulting eight fusions displayed an even lower solubility when compared to their C-terminal GFPuv fusion counterparts (Figure 9B). This is particularly obvious for the GFPuv-MBP fusions, which are present almost exclusively in the insoluble fraction, in contrast to the MBP-GFPuv fusions, in which about 15-20% of the material was

soluble (compare Figures 9A and 9B). This suggested that the fusion event itself, probably by increasing the complexity of the folding reaction, was responsible for misfolding. In conclusion, our results showed that domain order and the presence of additional sequences at the extreme N-terminus of GFPuv are not critical for fusion protein misfolding.

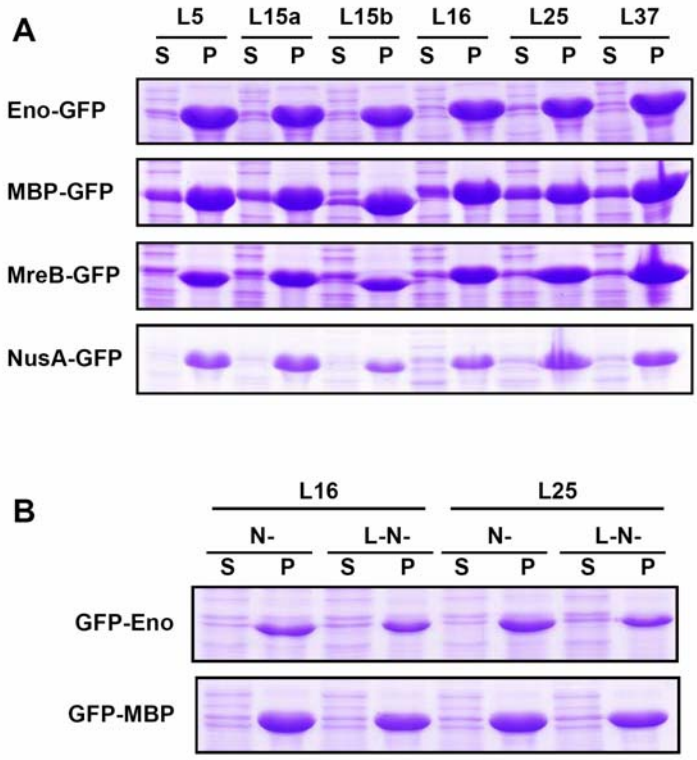

Figure 9. GFP fusion proteins aggregate upon expression in *E. coli*

(A) Solubility of C-terminal GFP fusion proteins upon expression in *E. coli* examined by SDS-PAGE and staining with Coomassie brilliant blue. Protein in the supernatant (S) and pellet (P) fractions for each fusion protein resulting from the combination of the six linkers (listed on top) and the four fusion partners (listed on the left) is indicated. (B) Solubility of N-terminal GFP fusion proteins upon expression in *E. coli* examined as above. Protein in the supernatant (S) and pellet (P) fractions of each fusion protein resulting from the combination of the two linkers (listed on top) and two fusion partners (listed on the left) is indicated. Proteins with the original GFP N-terminus are indicated as N- and those with an additional L5 linker sequence as L-N-.

In order to assess the folding efficiency of GFPuv fusions in the eukaryotic cytosol, we analyzed these proteins by recombinant expression in *S. cerevisiae* (Figure 10).

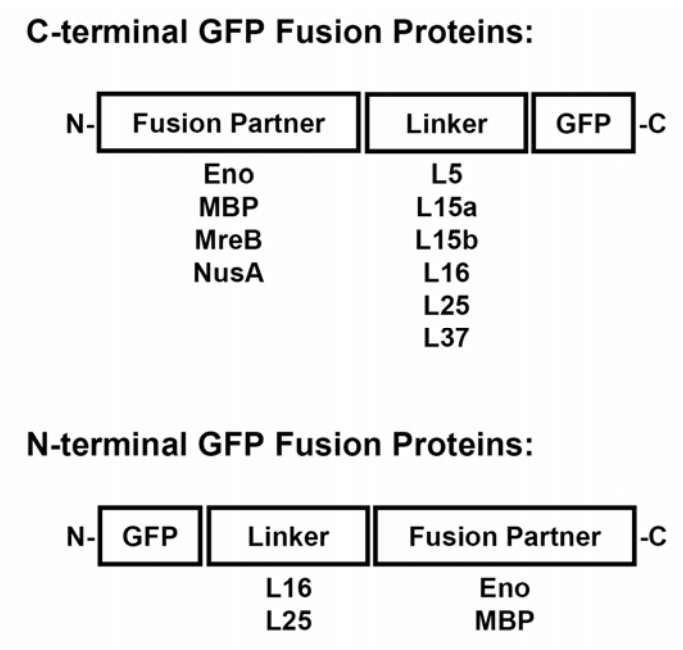

Figure 10. Constructs for expression of GFP fusions in *S. cerevisiae*

Diagram of the different GAL1 promoter-driven constructs for expression of GFP fusions in yeast, with each component represented by a box. Fusion partners and linkers are listed underneath each box. Linker characteristics are given in Table 2.

Surprisingly, we observed that a majority of total protein was soluble in each case tested, regardless of the fusion position of GFPuv (Figure 11A and 11B). *In vivo*, cellular localization of green fluorescence supported the notion that both domains of the fusion proteins were correctly folded when expressed in yeast (Figure 11C). For instance, the transcription factor NusA is a nucleic acid binding protein and its GFPuv fusions displayed clear accumulation in the nucleus, most probably as a result of the interaction between NusA and yeast chromatin. Similarly, fusions of GFPuv with the filament-forming protein MreB were present in filamentous structures reminiscent of actin cables (Doyle and Botstein, 1996). On the other hand, enolase and MBP would be expected to remain soluble

in the yeast cytosol, and these proteins showed a diffuse cytosolic distribution when fused to GFPuv.

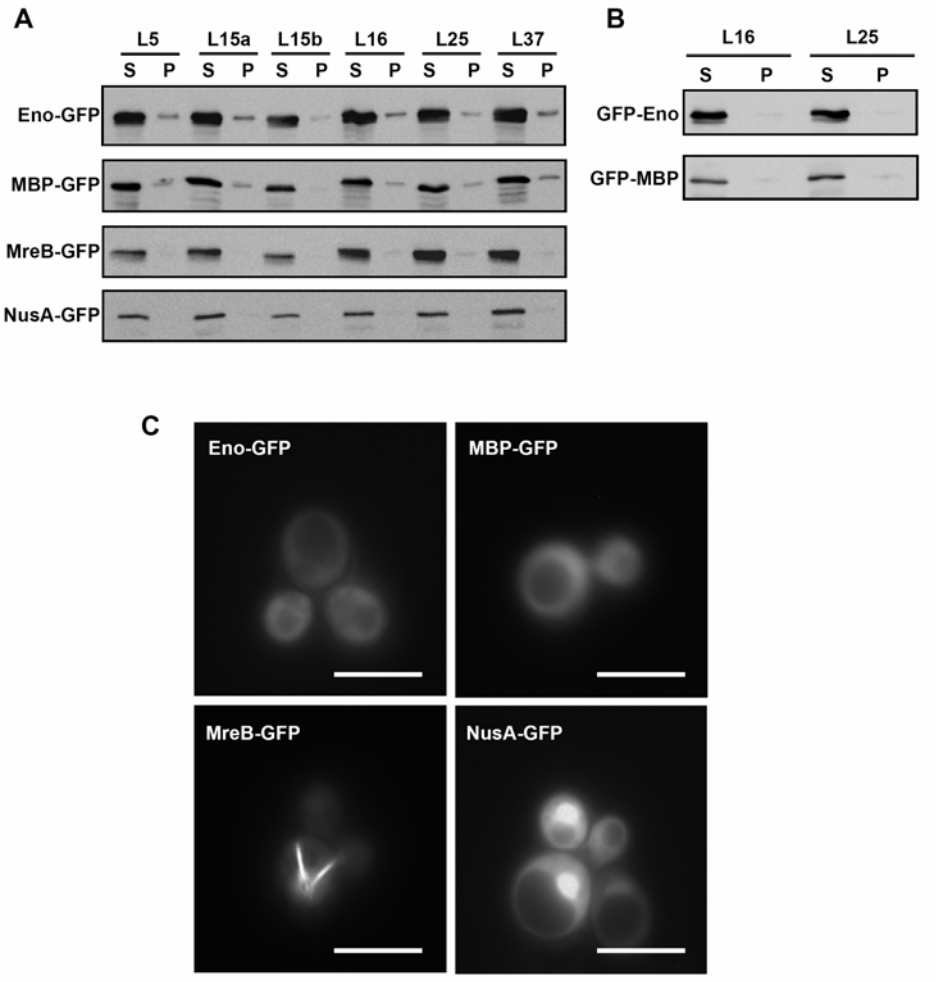

Figure 11. GFP fusion proteins are highly soluble upon expression in *S. cerevisiae*

(A) Solubility of C-terminal GFP fusion proteins upon expression in *S. cerevisiae* examined by SDS-PAGE, followed by immunoblotting with anti-GFP monoclonal antibodies. Protein in the supernatant (S) and pellet (P) fractions of each fusion protein resulting from the combination of the six linkers (listed on top) and the four fusion partners (listed on the left) is indicated. (B) Solubility of N-terminal GFP fusion proteins upon expression in *S. cerevisiae* examined as above. Protein in the supernatant (S) and pellet (P) fractions of each fusion protein resulting from the combination of the two linkers (listed on top) and two fusion partners (listed on the left) is indicated. (C) Fluorescence microscopy of GFP fusion proteins in living *S. cerevisiae* cells. The indicated fusion proteins contained linker L16. Scale bars represent 10 μm.

4.1.2. High folding efficiency in yeast is independent of expression levels

Since the expression levels of different fusions tested were much higher in *E. coli* than in yeast, it is plausible that the inefficient folding observed in bacteria could be due to the higher concentrations of recombinant proteins synthesized compared to those reached in yeast. To test this possibility, we searched for conditions that led to the production of very similar amounts of recombinant protein in bacteria and yeast at the same temperature (30 ^{o}C). We found that, at low levels of expression (~1-2% of the total cellular protein), the majority of Eno-GFPuv and GFPuv-Eno fusion proteins were still present in the insoluble fraction in *E. coli*. In contrast, at very similar expression levels in yeast, almost no insoluble fusion proteins could be detected by Coomassie blue staining (Figure 12A). Moreover, when green fluorescence was measured and compared to purified protein standards (see 4.1.3), >90% of the GFP moiety of the Eno-GFPuv and GFPuv-Eno fusion proteins was found to be native in lysates from yeast cells, while only 5-10% of each GFPuv domain was native in lysates from *E. coli* cells (Figure 12B). These results demonstrate that in *E. coli*, misfolding of fusion proteins containing a GFPuv domain (at either terminus) is not due to the high level of expression reached or to the higher temperature (37 ^{o}C) commonly utilized for expression of recombinant proteins in this organism. Rather, the misfolding of GFPuv fusions in the bacterial cytosol occurred largely independent of expression level already during the early stage of expression.

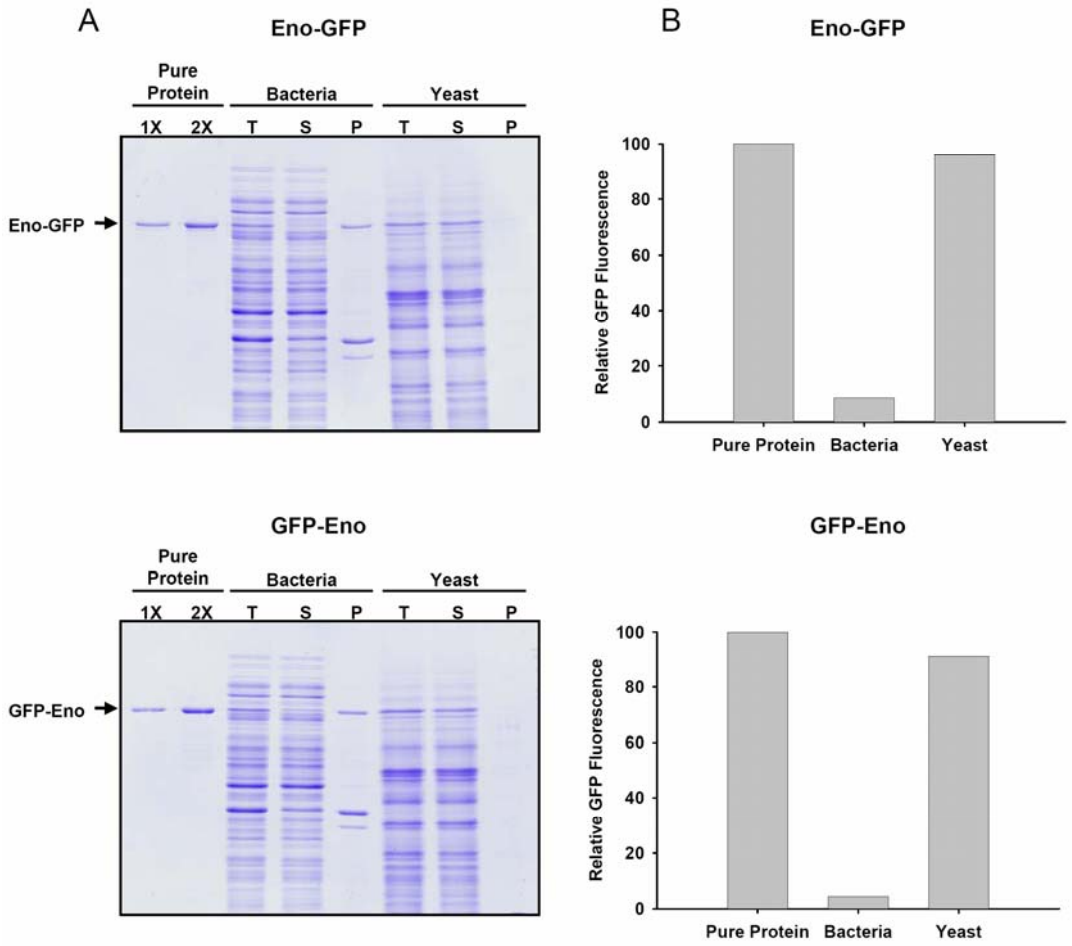

Figure 12. Different folding behaviors of GFP fusion proteins observed upon expression to similar levels and at the same temperature in bacteria and yeast

(A) Solubility of Eno-GFP (upper panel) and GFP-Eno (lower panel) upon expression to similar levels in bacteria and yeast at 30 °C, examined by SDS-PAGE and staining with Coomassie brilliant blue. Protein in the total (T), supernatant (S) and pellet (P) fractions is indicated. Lanes containing 0.2 µg (1X) and 0.4 µg (2X) of purified Eno-GFP and GFP-Eno were used as standards for the estimation of the amount of fusion proteins in the lysates. Equivalent amounts of total protein from each lysate (14 µg) were loaded in the total protein lanes (T) of each gel. Estimation by visual inspection revealed that the amount of fusion protein expressed was ~0.2 µg *per* lane of total lysate. (B) Quantification of relative green fluorescence of Eno-GFP (upper panel) and GFP-Eno (lower panel) fusion proteins expressed at similar levels in bacteria and yeast at 30 °C. Fluorescence values of identical amounts of protein used for analysis of the total lysate by SDS-PAGE were determined and normalized to the fluorescence of 0.2 µg (1X) of each purified protein, set as 100%.

4.1.3. Misfolding of GFP fusions in *E. coli* is due to an intra-molecular misfolding event

Since the misfolding of GFPuv fusions in bacteria is independent of their expression level, it could be reasoned that the folding problem is due to an intra-, rather than an inter-molecular event. We addressed this question through a biochemical approach. Typically, proteins that misfold upon *in vitro* refolding may display reduced yields, slower folding kinetics or a combination of both (Netzer and Hartl, 1997). The discriminating feature of intramolecular misfolding is that the observed effects are independent of protein concentration over a broad range. In contrast, effects resulting primarily from inter-molecular aggregation during refolding are highly sensitive to the concentration of protein utilized, with higher concentrations leading to reduced folding yields. We began by purifying recombinant native GFPuv and enolase from *E. coli*, and Eno-GFPuv and GFPuv-Eno from *S. cerevisiae* to ensure that they were in a native state. Importantly, the same molarity of purified enolase, Eno-GFPuv and GFPuv-Eno displayed identical enolase activities, and the same was observed when comparing green fluorescence activities among GFPuv, Eno-GFPuv and GFPuv-Eno proteins. This further strengthened the idea that both domains of the fusion proteins were correctly folded when expressed in yeast.

We monitored refolding of the purified proteins upon dilution from denaturant by measuring both green fluorescence and enolase activity (Figures 13A and 13B, respectively). Since the covalent GFP fluorophore is already present in these proteins, green fluorescence can be utilized to monitor refolding in real time. As can be observed, the fusion proteins reached similar refolding yields to those of the individual, non-fused proteins. Significantly, the refolding yields remained constant at increasing protein concentrations from 125 to 500 nM (Table 3), suggesting that these proteins do not undergo significant intermolecular aggregation under the conditions tested. Additionally, we observed that the individual fusion partners refolded considerably faster (3 to 4-fold) on their own than when

present as part of a fusion protein. The half-times of refolding were ~20 s and ~80 s for GFPuv alone and in the fusion proteins, respectively; and ~1.5 min and 5 min for enolase alone and in the context of the fusion proteins, respectively. These slower kinetics of refolding remained concentration independent from 125 to 500 nM (Table 3). The lack of dependence of both refolding yield and kinetics on total protein concentration for both Eno-GFPuv and GFPuv-Eno suggests that their delay during *in vitro* refolding is due to intra-molecular interactions caused by the presence of an additional domain, which result in the accumulation of kinetically trapped folding intermediates. Taken together, our results argue that in *E. coli*, GFPuv interferes with the folding of fusion proteins through an intra-molecular misfolding event.

Although the *in vitro* refolding results demonstrated that the GFPuv fusion proteins can fold, in principle, through an entirely post-translational mechanism, the *de novo* folding yields of GFPuv fusions *in vivo* is remarkably low (Figure 9A and 9B). A possible explanation for this difference could be the fact that protein concentration and macromolecular crowding in *E. coli* are very high compared to the situation *in vitro*. Thus, kinetically trapped folding intermediates, resulting mostly from an in intra-molecular misfolding event, would aggregate more easily. Alternatively, the *in vivo* misfolding event occurring during translation may lead to off-pathway intermediates that fail to fold post-translationally, eventually resulting in aggregation.

Table 3. Concentration independent *in vitro* refolding of Eno-GFP and GFP-Eno

Protein	Concentration (nM)	Fluorescence		Enolase Activity	
		Yield (%)	$t_{1/2}$ (s)	Yield (%)	$t_{1/2}$ (s)
GFP	125	92	17	n.a.	n.a.
GFP	250	88	19	n.a.	n.a.
GFP	500	95	19	n.a.	n.a.
Eno	125	n.a.	n.a.	63	210
Eno	250	n.a.	n.a.	83	90
Eno	500	n.a.	n.a.	81	96
Eno-GFP	125	82	82	53	366
Eno-GFP	250	93	80	76	306
Eno-GFP	500	107	104	84	240
GFP-Eno	125	78	107	58	456
GFP-Eno	250	86	80	71	258
GFP-Eno	500	95	109	75	244

n.a.: not applicable. The lower yields and slower $t_{1/2}$ observed at the lowest concentrations are probably the result of non-specific adsorption of the proteins to the walls of the tubes/cuvettes during refolding. Values represent typical results of experiments performed multiple times.

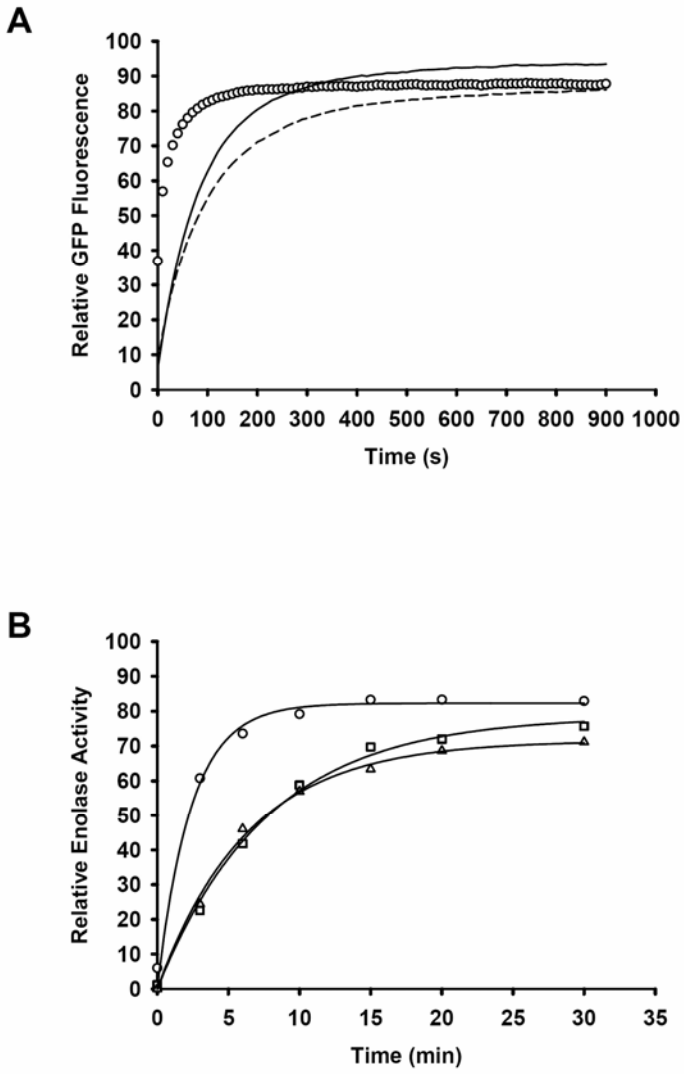

Figure 13. GFP imposes significant constraints on refolding of its fusion partner

(A) Kinetics of refolding at 25 ^{o}C upon dilution from denaturant measured by fluorescence of non-fused GFP (circles), Eno-GFP (continuous line) and GFP-Eno (broken line) at 250 nM (both fusion proteins contain linker L16). Native GFP at the same concentration was set as 100%. (B) Kinetics of refolding at 25 ^{o}C upon dilution from denaturant measured by enolase activity of non-fused enolase (circles), Eno-GFP (squares) and GFP-Eno (triangles).

4.1.4. The rate of production of folded protein *per* ribosome is higher in yeast than in bacteria

If eukaryotic cells indeed have a higher capacity for the folding of multi-domain proteins in a co-translational manner, one would expect that the fraction of correctly folded protein produced *per* ribosome in a certain period of time should be greater in yeast than in bacteria. In order to test this possibility, we took advantage of the fact that translation on *S. cerevisiae* ribosomes is slower than on *E. coli* ribosomes under standard growth conditions (translation rates in *S. cerevisiae* and *E. coli* are ~3-8 and ~12-22 aa/s, respectively)(Bremer and Dennis, 1996; Mathews *et al*., 2000). If one sets up an experiment in which equal amounts of yeast and bacterial ribosomes are used to translate identical polypeptides for the same period of time, the amount of total polypeptide synthesized by yeast cannot be higher than that produced by bacteria. Enzymatic activity of the translated protein can be utilized as a measure of the total amount of folded polypeptide. In the above experiment, a higher amount of enzymatic activity upon translation in yeast could only be explained by production of a larger fraction of correctly folded polypeptide chains in yeast than in bacteria.

To carry out such an experiment, we constructed plasmids of GFPuv fused to firefly luciferase for expression in *E. coli* and *S. cerevisiae*. Luciferase as a fusion partner has several advantages over the fusion partners used in the previous experiments. For instance, since the GFP fluorophore does not form immediately upon folding (Heim *et al*., 1994), it cannot readily be used to monitor *de novo* protein folding. MBP lacks enzymatic activity, which renders it not useful for this kind of test. Unlike enolase, there is no preexistent luciferase activity in either organism and thus there is no background enzymatic activity from endogenous sources.

When we compared the accumulation of FL activity *per* ribosome content after 1-4 h of induction at 30 ^{o}C in both organisms, we found that yeast had produced substantially more active enzyme than bacteria *per* ribosome (Figure 14). This result indicates that the fraction of correctly folded GFPuv fusion protein synthesized as native protein from each ribosome is higher in yeast than in bacteria. The above results rule out the possibility that the high fraction of correctly folded protein in yeast is due to efficient degradation of misfolded polypeptides. Regardless of whether some of the translated polypeptides (folded or misfolded) are rapidly degraded, a higher amount of active enzyme is in fact translated by equal amounts of ribosomes during the same period of time.

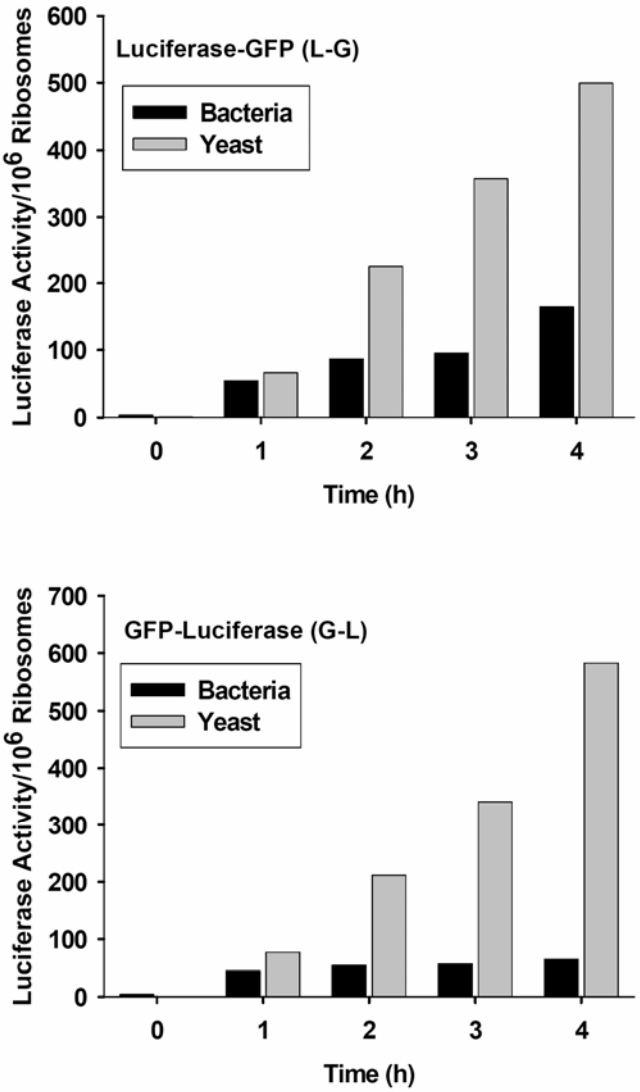

Figure 14. The rate of production of folded fusion proteins *per* ribosome is higher in *S. cerevisiae* than in *E. coli*

Quantification of the accumulation of total luciferase activity *per* 10^6 ribosomes of luciferase-GFP (upper panel) and GFP-luciferase (lower panel) fusion proteins upon induction in bacteria (black bars) or yeast (gray bars) at the time-points indicated. The linker sequence between the proteins was L16.

4.2. The influence of nascent chain binding chaperones on the folding of naturally occurring multi-domain proteins

In an attempt to understand the problem of multi-domain protein folding occurring in bacteria, we analyzed the effect of nascent chain binding chaperones on the folding of such proteins. Of the many chaperones that can be encountered by a polypeptide chain emerging from the ribosome, trigger factor (TF) and DnaK are considered the two major nascent chain binding chaperones in *E. coli* because of their high cytosolic abundance and functional significance. Additionally, the findings that TF and DnaK have overlapping functions in protein folding, and that *E. coli* does not tolerate the deletion of *dnaK* in a Δ*tig* background at temperatures above 30 °C (Deuerling *et al*., 1999; Teter *et al*., 1999), underscored the significance of these components for folding in the bacterial cytosol.

Interestingly, despite their broad function in assisting the folding of a wide spectrum of polypeptide chains, it was recently shown that the *dnaKdnaJ* operon can be deleted in a Δ*tig* strain at 20 °C (Genevaux *et al*., 2004). The resulting mutant strain (Δ*tig*Δ*dnaKdnaJ*) can be gradually adapted to growth at 30 °C (or maybe higher) and thus provides the opportunity to examine the fate of newly synthesized polypeptides *in vivo* in a genetic background with complete absence of the major nascent chain-interacting chaperones. Δ*tig*Δ*dnaKdnaJ* cells are severely defective in protein folding: they display a pronounced filamentous phenotype and accumulate substantial amounts of endogenous protein in an aggregated state, visible as inclusion bodies by phase contrast microscopy (Figure 15).

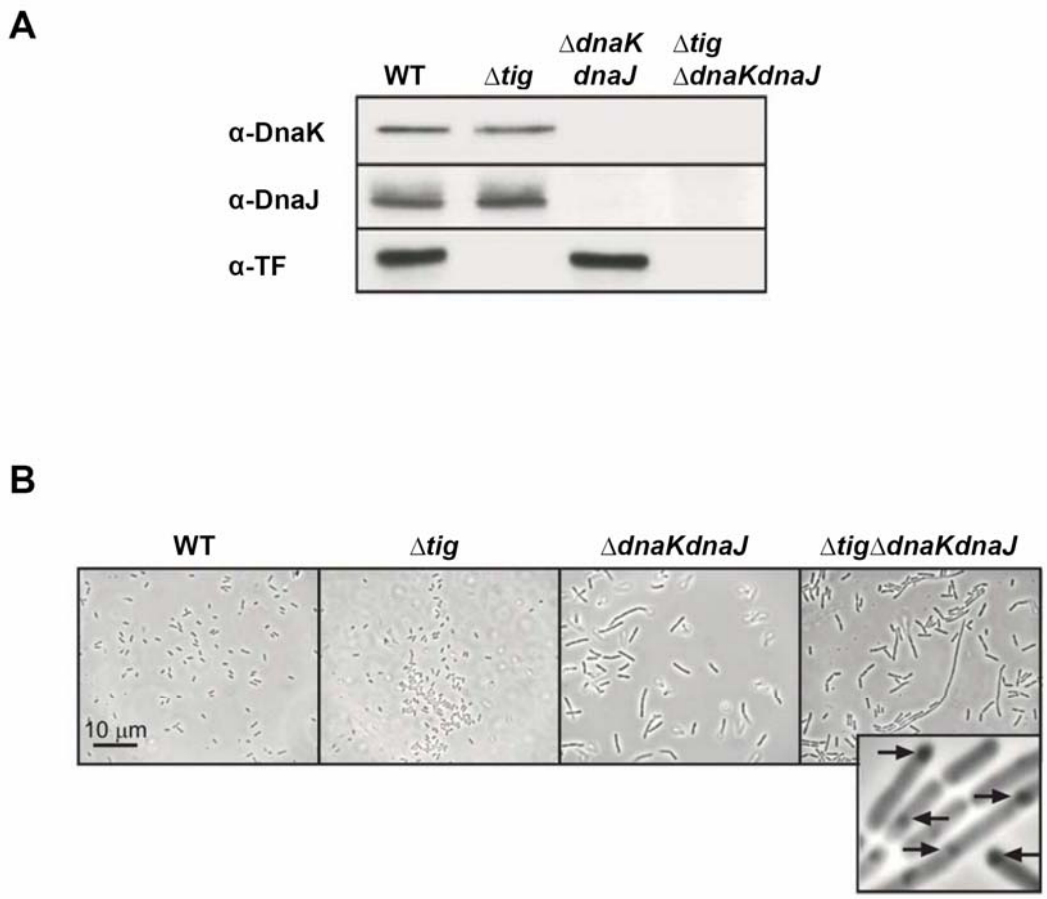

Figure 15. Characterization of *E. coli* chaperone mutant strains

(A) Immunoblots of total lysates from MC4100 wild-type (WT), Δ*tig*, Δ*dnaKdnaJ*, and Δ*tig*Δ*dnaKdnaJ* strains, probed with antibodies against DnaK, DnaJ, and TF. (B) Phase-contrast microscopy of *E. coli* cells from WT and mutant strains as above. Arrows indicate inclusion bodies in Δ*tig*Δ*dnaKdnaJ* cells (inset). *E. coli* strains were grown at 30°C in LB medium for 12 h. Samples for microscopy were prepared by fixing the cells in 80% methanol for 30 min at room temperature and then photographed under phase contrast illumination.

4.2.1. Effect of trigger factor and DnaK on the folding yields of firefly luciferase and β-galactosidase in *E. coli*

In an effort to investigate the effects of TF and the DnaK system on the folding of multi-domain nascent polypeptides, we utilized firefly luciferase (FL) and bacterial β-galactosidase (β-gal), two naturally occurring enzymes, as models. FL is a monomeric two-domain protein of 62 kDa (Conti *et al*., 1996), and β-gal is active as a tetramer of identical 116 kDa subunits composed of five compact domains (Conti *et al*., 1996; Jacobson *et al*., 1994). Since the GroEL chaperonin system is unable to mediate the folding of either of these proteins (Ayling and Baneyx, 1996; Buchberger *et al*., 1996), these two enzymes are particularly suitable models to study exclusively the folding process coupled to translation, without the possible complication of a post-translational contribution of GroEL in folding.

Upon regulated expression of FL in wild-type *E. coli* for 15 min, we detected that only about 35% of the protein was soluble (Figure 16A). FL activity in the Δ*tig* and Δ*dnaKdnaJ* strains was ~50% reduced relative to wild-type cells, whereas a ~90% reduction was measured in the Δ*tig*Δ*dnaKdnaJ* cells. This effect correlated well with a corresponding decrease in the amount of soluble FL from ~20% in Δ*tig* cells to less than 10% in the Δ*tig*Δ*dnaKdnaJ* cells. Importantly, similar results were obtained with β-gal, an endogenous *E. coli* protein, with the notable difference that this protein was mostly soluble and active in wild-type cells (Figure 16B). Δ*tig* and Δ*dnaKdnaJ* cells produced ~80% and ~50% of active β-gal, respectively, and the Δ*tig*Δ*dnaKdnaJ* strain yielded only ~20% soluble and active protein relative to wild-type cells. The above results indicated that TF and DnaK system cooperate to increase the *de novo* folding efficiency of FL and β-gal, yet in different capacities. These experiments suggest that any compensatory mechanisms allowing Δ*tig*Δ*dnaKdnaJ* cells to survive apparently do not correct the folding defects for multi-domain proteins such as FL and β-gal.

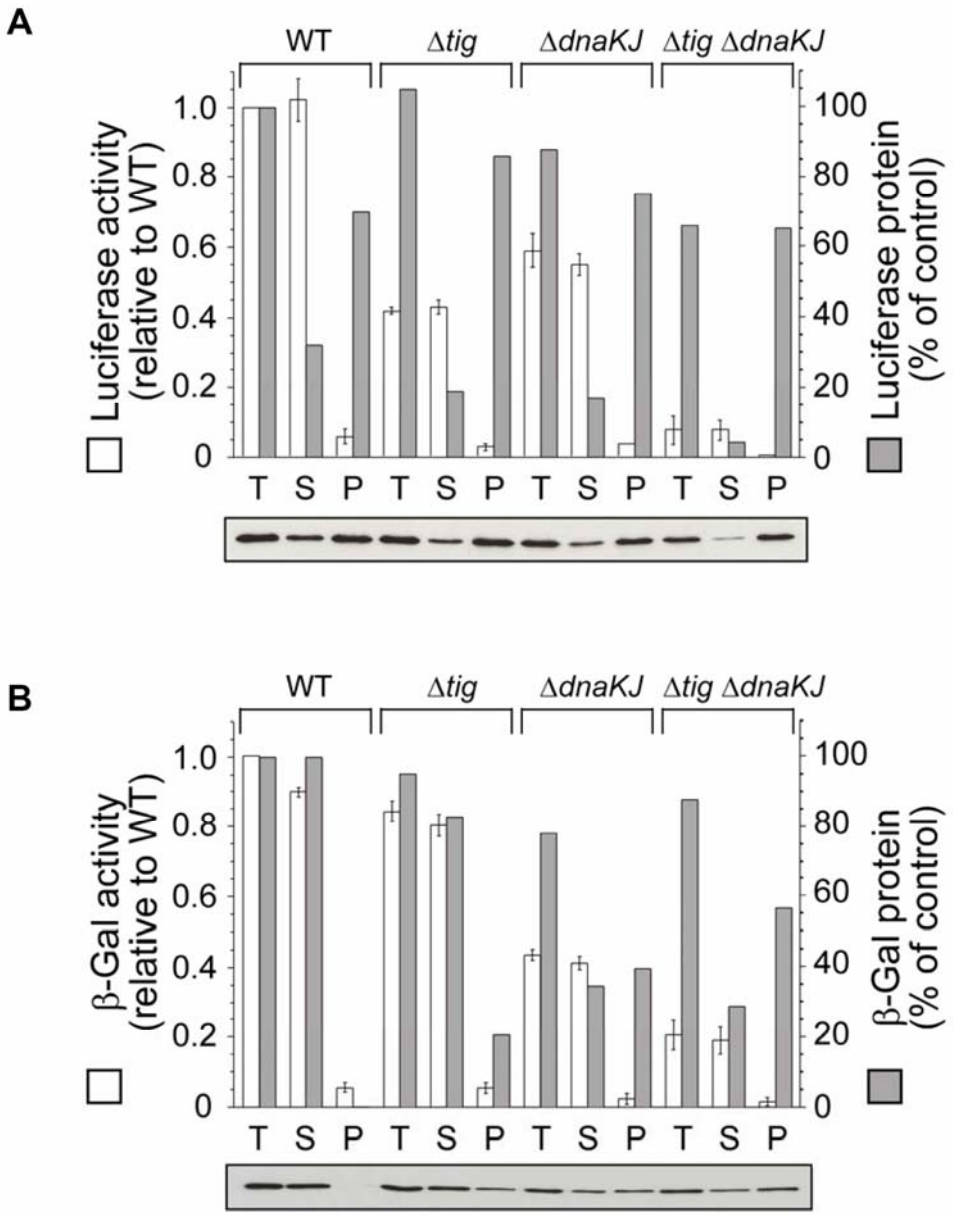

Figure 16. Effect of chaperones on folding yields of FL and β-gal

Expression of firefly luciferase (A) and β-galactosidase (B) in *E. coli* wild-type (WT) and mutant strains *in vivo*. In upper panels present enzyme activities (white bars) and protein amounts (black bars) after a 15 min induction of total (T), supernatant (S), and pellet (P) fractions. Activities are shown as values relative to WT (set to 1) and protein amounts as percent of the WT control. Immunoblots of samples used for quantitation were shown in the lower panel.

Although TF and DnaK/DnaJ/GrpE (KJE) increased the folding yield of FL to a limited extent, the folding of FL in bacteria was again found to be much less efficient than in eukaryotic cells (Figure 17), as observed above with the fusion proteins (Figure 9 and 11). In yeast, essentially all the FL was recovered in the soluble fraction, whereas only about 35% of the bacterially expressed protein was soluble. Notably, similar amounts of soluble FL from yeast gave ~12 times higher activity when compared to that from bacteria. This result suggested the presence of a non-native, soluble FL species in the isolated soluble fraction from bacteria.

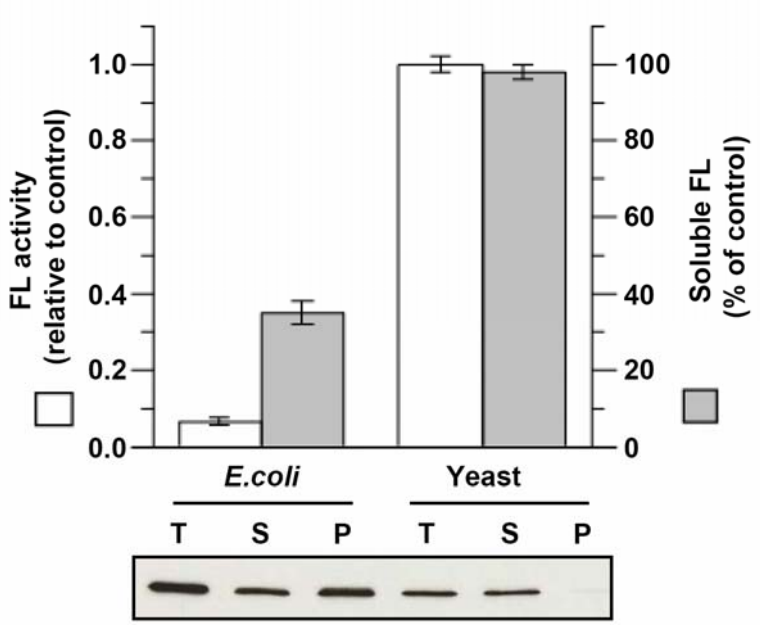

Figure 17. FL folding is inefficient in bacteria compared to eukaryotes

FL was expressed to similar levels *in vivo* in wild-type *E. coli* and *S. cerevisiae* cells. Top panel: FL-relative specific activities (white bars) and amounts of soluble FL as percent of total (gray bars). The specific activity in yeast was set to 1. Bottom panel: The corresponding immunoblot revealed distribution of FL protein in total (T), supernatant (S), and pellet (P) fractions.

To exclude the possibility that the low folding efficiency of FL in *E. coli* was caused by a limited availability of KJE or a negative interference by GroEL/GroES, we co-expressed these chaperones with FL. Over-expression of GroEL/GroES showed no effect on the FL folding yield when compared to its wild-type bacterial background. Over-expression of KJE, on the other hand, reduced the specific activity of FL further (Figure 18A and 18B). Interestingly, in both cases, the increased amount of soluble FL (~60%) observed was most present likely in a misfolded state, since the activity yield did not increase. This result suggested that the low folding yield of FL in *E. coli* was not due to the limited availability of KJE. Instead, the incompatibility between bacterial nascent-chain binding chaperones, such as TF and KJE, and nascent multi-domain eukaryotic proteins such as FL appeared to be responsible, at least in part, for the limited folding efficiency of multi-domain proteins in *E. coli*.

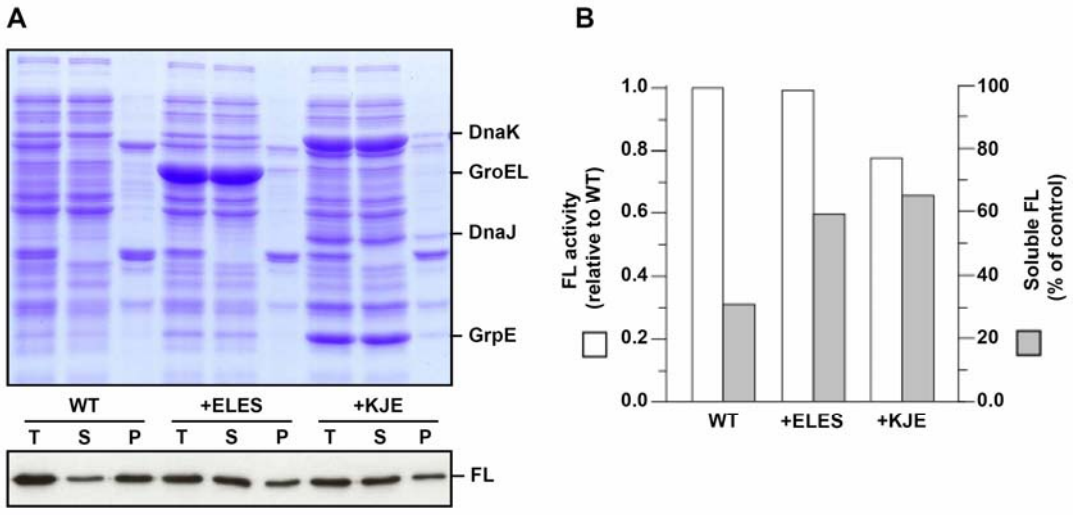

Figure 18. Solubility and folding yields of FL upon translation in *E. coli* wild-type, GroEL/ES, and KJE overproducing cells

(A) Coomassie-stained SDS-PAGE (upper panel) and the immunoblot for FL (bottom panel) analyzing cell lysates divided into total (T), supernatant (S), and pellet (P) fractions. (B) Specific activities and solubility of FL are shown with the activity in wild-type (WT) cells set to 1.

4.2.2. Distinct folding kinetics of firefly luciferase in bacteria *vs.* yeast

The contribution of TF and DnaK to FL and β-gal folding was further explored *in vitro* in S30 translation lysates from *E. coli*. These lysates support efficient protein synthesis (~200 μg/ml per hour) nevertheless due to dilute cytosol preparations, contain only low levels of endogenous chaperones. The *in vivo* concentrations of TF and DnaK under standard growth conditions are ~40 μM and ~50 μM, respectively (Hesterkamp and Bukau, 1998; Lill *et al*., 1988). In contrast, the S30 translation reaction used here contains only ~0.5 μM TF and DnaK, determined by quantitative western blotting (Table 4), and thus should mimic the Δ*tig*Δ*dnakdnaJ* deletion mutant with respect to FL and β-gal folding.

Table 4. Chaperone concentration in the bacterial S30 lysate

	Cytosol	S30 lysate
Total protein (g/L)	200	20
TF (μM)	40	0.5
DnaK (μM)	50	0.5
DnaJ (μM)	10	0.1
GrpE (μM)	30	0.3
GroEL (μM)	2	0.2
GroES (μM)	4	0.5
Ribosome (μM)	30	0.5

It has been shown that newly translated FL in a eukaryotic cell free, translation system (rabbit reticulocyte lysate, RRL) is fully active within seconds upon completion of synthesis (Kolb *et al*., 1994). This process involves the co-translational folding of the N-terminal domain of FL during its synthesis (~2 min), followed by the rapid completion of folding to

the active enzyme upon release from the ribosome (Frydman *et al*., 1999; Frydman *et al*., 1994). To test whether in the bacterial *E. coli* S30 lysate, newly synthesized FL follows a similar rapid folding mechanism, we first compared the kinetics of translation and folding in the bacterial S30 lysate with those in a rabbit reticulocyte lysate. Interestingly, regardless of a marked difference in folding yield in both systems (~5% in S30 *versus* ~60% in RRL; data not shown), FL activity appeared virtually concurrently with the production of full-length chains (Figure 19A), the hallmark of co-translational FL folding. More strikingly, upon addition of TF (5 μM) to the S30 lysate, the kinetics of FL folding showed a significant deceleration without affecting the speed of translation (Figure 19B). Increasing the amount of added TF (up to 15 μM) had no further effect (data not shown). Addition of KJE in the absence of TF did not slow the folding reaction. However, the delay in folding relative to translation was more pronounced when TF and KJE were added together (Figure 19B), reflecting the functional cooperation between TF and the DnaK system. To estimate the extent to which FL may fold post-translationally under these conditions, we simulated the kinetics of *de novo* folding based on the rate measured for the KJE-mediated refolding of denatured FL ($t_{1/2}$ ~10 min). The theoretical curve (Figure 19B) agrees well with the observed kinetics of TF/KJE- assisted *de novo* folding, suggesting that the chaperones may shift the majority of FL folding from a co-translational to a post-translational pathway. In contrast, the rapid folding observed in the unsupplemented S30 lysate (Figure 19A) apparently represents an unchaperoned co-translational default pathway, which is however inefficient.

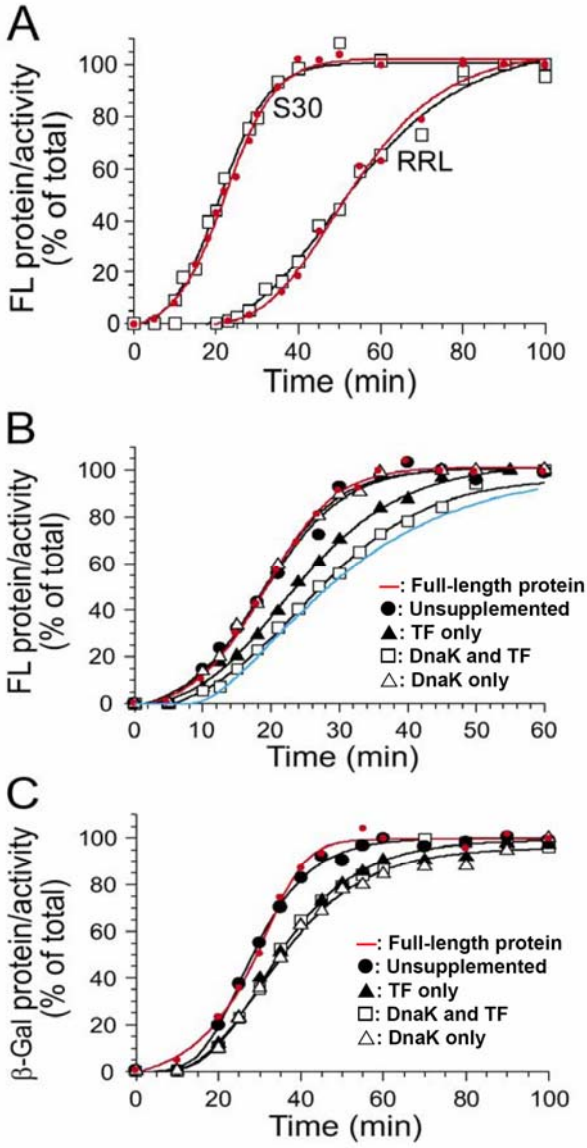

Figure 19. TF and the DnaK system delay folding relative to translation *in vitro*

(A) Folding kinetics of FL in *E. coli* S30 and rabbit reticulocyte lysate (RRL) translation reactions. Appearance of full-length protein (in red) and activity (open squares) were followed with time. Final values are set to 100%. (B) Appearance of FL activity during S30 translation in the absence of added chaperones (●) and in the presence of added KJE (Δ), TF (▲), or a combination of both (□). The appearance of full-length FL in all these translations was identical and is represented in red. The blue line represents a kinetic simulation of the evolution of FL activity, assuming that *de novo* folding follows the kinetics of KJE-mediated refolding, a post-translational folding process. (C) Appearance of β-gal activity during S30 translation in the absence of added chaperones (●), in the presence of added KJE (Δ), TF (▲), or both (□). The appearance of full-length β-gal in all these translations was identical and is represented in red. Experiments were performed in cooperation with V. Agashe and J. M. Barral.

Similar observations were made for the bacterial protein β-gal. In the bacterial S30 lysate, the appearance of full-length protein virtually coincided with that of β-gal activity (Figure 19C). This indicated that, in the default pathway, both folding and assembly of β-gal tetramers are tightly coupled to translation, in contrast to the process of refolding from denaturant (Nichtl *et al*., 1998). The addition of TF and KJE, either separately or together, caused a substantial delay in the appearance of β-gal activity relative to translation (Figure 19C). Thus, different from the combined action required for the folding delay of FL, TF and KJE have virtually overlapping roles in the folding and assembly of β-gal.

The delay in folding imposed by TF and DnaK suggested that native, active protein continues to be produced upon termination of translation. To test whether this was indeed the case *in vivo*, we utilized live *E. coli* spheroplasts expressing FL or β-gal from a tightly controlled arabinose-regulated promoter. The rate of FL synthesis was found to be maximal after ~50 min of induction. Addition of the translation inhibitor chloramphenicol (CAM) resulted in an immediate stop of protein synthesis, as shown representatively for wild-type cells (Figure 20, inset). In wild-type cells, a substantial amount of FL activity continued to be produced for more than 5 min after inhibition of translation (Figure 20A), indicating a significant post-translational phase of folding. A similar, but less pronounced post-translational phase of folding was observed in Δ*tig* and Δ*dnaKdnaJ* strains, implying that TF and the DnaK system could contribute to post-translational folding independently, but to a lesser extent than when present in combination (data not shown). Strikingly, in Δ*tig*Δ*dnaKdnaJ* cells, production of FL activity stopped instantaneously with the inhibition of translation (Figure 20A). A very similar behavior was observed for β-gal, with significant post-translational production of activity in wild-type but not in Δ*tig*Δ*dnaKdnaJ* cells (Figure 20B).

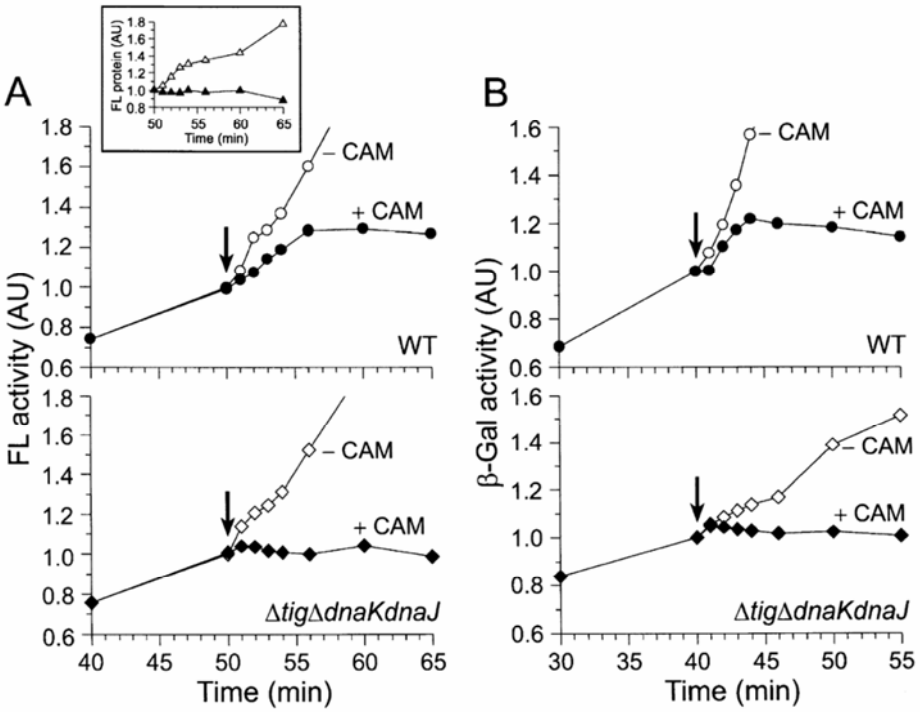

Figure 20. Kinetics of FL and β-gal folding in wild-type *E. coli* and chaperone-deleted cells *in vivo*

Accumulation of enzymatic activity in live spheroplasts of *E. coli* wild-type (WT) and Δ*tig*Δ*dnaKdnaJ* cells expressing FL (A) or β-gal (B) upon induction with arabinose at 0 min. Reactions were split at the time points indicated (arrow) and left untreated (open symbols) or treated with chloramphenicol (CAM) to stop translation (filled symbols). Inset shows the accumulation of ^{35}S-labeled full-length FL in WT cells, demonstrating the immediate stop of translation upon CAM addition. The same effect was observed for the other experiments in this figure. Enzyme activities and protein amounts are plotted relative to the point of translation inhibition (set to 1).

These results confirm the observations from the *in vitro* translation experiments (Figure 19). In the absence of TF and DnaK, folding of FL and β-gal to their enzymatically active forms *in vivo* is tightly coupled to translation, but this reaction is inefficient. The increased folding yield of these proteins in wild-type cells appears to result from a post-translational folding component introduced by the action of TF and KJE.

We considered the possibility that different folding mechanisms followed in the bacterial and eukaryotic systems may be responsible for the different folding yields. A similar *in vivo* experiment to those shown in Figure 20 was designed to investigate whether FL folding in yeast has a significant post-translational component. Expression of FL in yeast cells from a copper-regulated promoter was accompanied by the production of FL activity (Figure 21), and addition of the translation inhibitor cycloheximide (CHX) caused an immediate stop in translation (Figure 21, inset). Significantly, no post-translational accumulation of FL activity was observed, in contrast to the situation in wild-type *E. coli* (Figure 20). This result and the observation that co-translational folding of FL with ~60% efficiency was also observed in the RRL (Figure 19A and data not shown) suggests that the efficient folding of FL in the eukaryotic system is tightly coupled to translation, consistent with co-translational domain folding and rapid acquisition of enzymatic activity upon chain release from the ribosome.

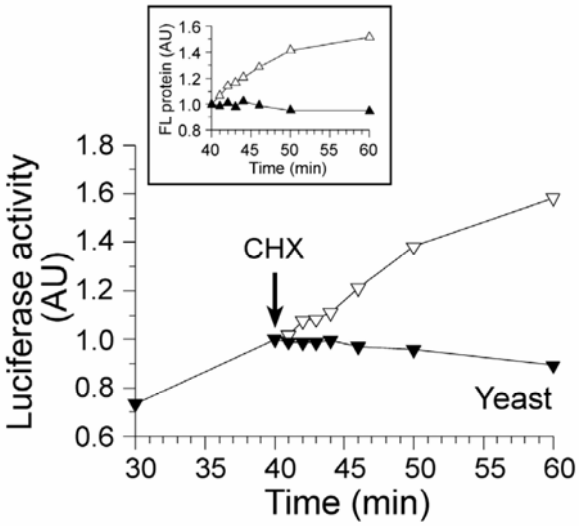

Figure 21. Kinetics of FL folding in wild-type *S. cerevisiae* cells *in vivo*

Accumulation of FL activity in *S. cerevisiae* cells upon induction of FL from a copper-regulated promoter at 0 min. Samples were split at the time point indicated (arrow) and left untreated (open symbols) or treated with cycloheximide (CHX, filled symbols) to stop translation. Inset shows the accumulation of ^{35}S-labeled full-length FL in yeast cells. Enzyme activities and protein amounts are plotted relative to the point of translation inhibition, which is set to 1.

4.2.3. TF and DnaK act co-translationally to cause a shift in folding mechanism

Next, we performed translation experiments in the S30 system to determine whether the delay in folding caused by TF and KJE requires the co-translational action of these chaperones. Production of FL activity was followed after inhibition of translation by RNaseA (50 µg/ml) or CAM (200 µg/ml) 22 min after initiating translation. No post-translational increase in FL activity was detectable in the unsupplemented S30 lysate (Figure 22), consistent with the virtually concurrent appearance of full-length protein and enzymatic activity (Figure 19A). In contrast, in the TF/KJE-supplemented lysate, a more than 2-fold increase in FL activity was observed after termination of translation with kinetics corresponding to the KJE-mediated refolding of denatured FL ($t_{1/2}$~10 min) (Figure

22). TF addition alone caused a similar post-translational folding phase but with a lower amplitude (Figure 22). This effect required the binding of TF to the ribosome, since it was not observed with a triple mutant form of TF, TF-FRK/AAA, which is deficient in ribosome binding (Kramer *et al*., 2002). Immunodepletion of DnaK abolished the post-translational folding caused by TF addition (data not shown), as did depletion of ATP, by adding apyrase together with the translation inhibitor (Figure 22), consistent with an ATP requirement for DnaK function.

Importantly, addition of TF and KJE to the translation reaction together with the translation inhibitor failed to produce any post-translational folding phase (Figure 22). Thus, the co-translational action of both TF and KJE is essential for the delayed folding of FL and the increase in folding yield produced by these chaperones.

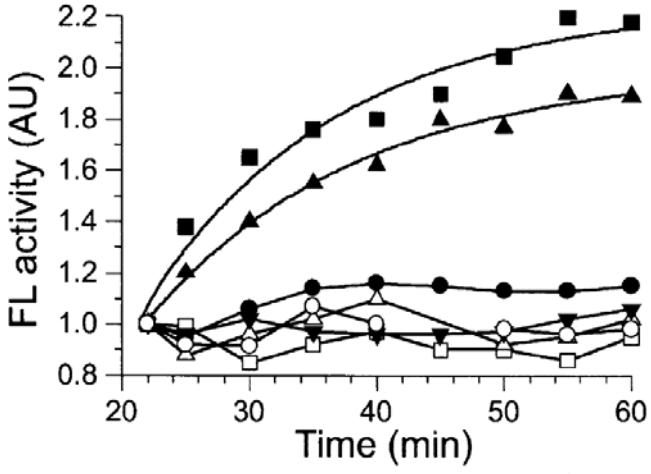

Figure 22. TF and DnaK system act co-translationally to change the mechanism of folding

The post-translational production of FL activity was followed in S30 translation reactions upon inhibiting protein synthesis with RNaseA. S30 lysate without added chaperone (○) and with chaperones added at the beginning of translation: wt-TF (▲) and FRK/AAA-TF (▼), KJE (Δ), wt-TF and KJE (■), wt-TF, KJE, and apyrase (□). In another reaction, wt-TF and KJE were added together with RNaseA at the time of stopping translation (●). Experiments were performed in cooperation with V. Agashe and J. M. Barral.

4.2.4. Additional TF molecules are recruited to tanslating ribosomes

It is surprising that TF has such a pronounced effect on the kinetics of folding while bound to the ribosome and interacting with nascent chains only in its immediate vicinity. We therefore asked whether the association of TF with the ribosome is dynamic during translation. TF binding to nontranslating ribosomes in the S30 lysate saturated at ~3 μM TF (Figure 23A), in agreement with the reported K_D value of ~1 μM for TF binding to purified ribosomes (Maier *et al*., 2003). Due to the ribosomal concentration in the bacterial S30 lysate is as low as about 0.5 μM (Table 4), a concentration of 6 μM TF was chosen to achieve effective ribosomal saturation. ^{35}S-labeled TF produced by *in vitro* translation was used as a marker for ribosomal binding of bulk TF.

In the absence of translation, addition of a small amount of ^{35}S-TF to an S30 reaction containing 6 μM unlabeled TF resulted in ^{35}S-TF binding to ribosomes that was unaffected by the addition of CAM or RNaseA (baseline in Figure 23B). Interestingly, upon initiation of transcription/ translation of FL, additional TF was recruited on to the translating ribosomes, as measured by the increased binding of ^{35}S-TF (Figure 23B). Such recruitment was not observed with the ribosome binding-deficient ^{35}S-TF-FRK/AAA protein (Figure 23B). Recruitment of TF reflected the occupancy of ribosomes with nascent FL chains. It occurred at a rate faster than the production of full-length FL (compare Figures 23B and 19) and saturated in ~10 min at a level ~3-fold higher than that at the beginning of translation. Addition of CAM, which blocks translation and stabilizes the ribosome-nascent chain complexes, resulted in retention of the recruited TF (Figure 23B). On the other hand, the release of nascent polypeptides by RNaseA caused a concomitant release of the recruited ^{35}S-TF (Figure 23B). Thus, although the starting population of ribosomes is saturated through the stoichiometric binding of TF, additional TF from the bulk solution is recruited during the process of translation. The extent of TF recruitment to translating ribosomes was found to be dependent on the properties of the nascent polypeptide chain being synthesized.

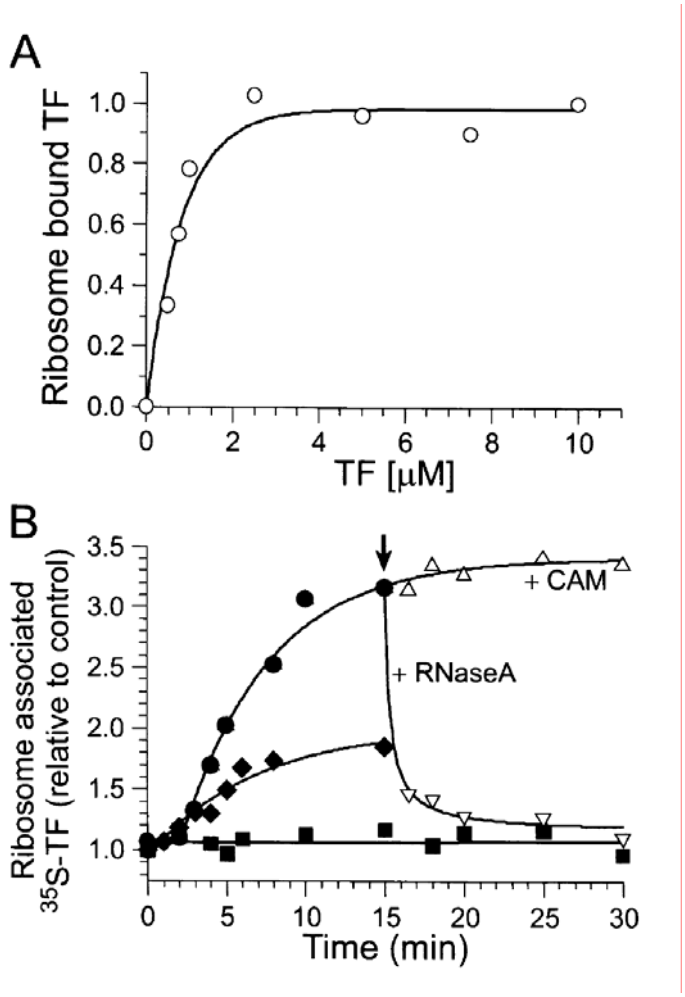

Figure 23. Recruitment of TF to translating ribosomes

(A) Increase in ribosome-associated TF upon incubation of the S30 lysate with increasing concentrations of purified wt-TF-His_6 in the absence of translation. The amount of wt-TF-His_6 bound to isolated ribosomes was determined by quantitative immunoblotting. (B) Binding of ^{35}S-labeled TF to ribosomes was followed during the translation of FL (●) or GFP (♦). The lack of recruitment of ^{35}S-TF-FRK/AAA mutant protein (■) during FL translation is also shown. Ribosomal retention of the freshly recruited ^{35}S-TF in the presence of chloramphenicol (CAM, Δ) and its release on the addition of RNaseA (▽) is shown for the FL translation. The arrow indicates the time of RNaseA or CAM addition. Binding of ^{35}S-TF is shown relative to the background binding in the absence of translation (set to 1). Experiments were performed in cooperation with V. Agashe and S. Guha.

The amount of recruitment on ribosomes translating green fluorescent protein (GFP; 25 kDa) was only about half that seen during translation of FL (62 kDa) (Figure 23B). This finding indicates a correlation of the level of TF recruitment with the size of the protein being translated and/or differences in the occurrence of hydrophobic peptide regions recognized by TF (Patzelt *et al*., 2001).

4.3. Efficient folding of multi-domain proteins in yeast is supported by chaperones

The observation of a higher capacity for multi-domain protein folding in yeast can potentially be due to nascent-chain binding chaperones in the eukaryotic cytosol actively supporting this process. Unlike the situation in bacteria where trigger factor and the DnaK system prevent nascent chain misfolding by imposing an alternate folding pathway, it is reasonable to assume that the chaperone network in eukaryotic cells mediates proficient multi-domain protein folding as exemplified by the high folding yields of GFPuv fusions and FL. To verify the existence of such chaperone network in eukaryotes, we challenged the eukaryotic cytosol in two ways. First, we created null mutations of effectors which may contribute to efficient folding of nascent chains in yeast. Second, we expressed prokaryotic proteins in yeast to see whether the cytosol of this organism is suitably equipped to fold these evolutionarily distant proteins.

4.3.1. Nucleotide exchange factor Fes1p in yeast facilitates the folding of firefly luciferase *in vivo*

It has been shown that the Hsp70 system (Ssa1p and Ydj1p) is involved in the folding of FL in the yeast cytoplasm (Lu and Cyr, 1998). Therefore, factors that regulate the Hsp70 reaction cycle, *e.g.* nucleotide exchange factors, could play a role in FL folding. Two major

classes of proteins have been described containing nucleotide exchange activities in the eukaryotic cytosol: Bag proteins and HspBP1 (Mayer and Bukau, 2005). In *S. cerevisiae*, there is only a single BAG-1 homolog, Snl1p, which is a transmembrane protein of the ER and nuclear membranes with functions apparently specialized for these organellar compartments (Sondermann *et al*., 2002). On the other hand, Fes1p, the *S. cerevisiae* ortholog of HspBP1, is the only known cytosolic nucleotide exchange factor for Hsp70 chaperones in this species. Deletion of the *FES1* gene in *S. cerevisiae* moderately impaired growth compared to wild-type cells at 37 ^{o}C as tested by serial dilution onto rich media (YPD), and the growth of this strain at 30 ^{o}C was identical to wild-type (Figure 26B and data not shown). This temperature-sensitive behavior is consistent with the previously reported phenotype (Kabani *et al*., 2002a). To further investigate the effect of the *FES1* deletion on Hsp70 function, we analyzed the efficiency of *de novo* folding in the cytosol of wild-type and Δ*fes1* cells with FL as a sensitive model protein. Upon expression in Δ*fes1* cells at 37 ^{o}C, the specific activity of FL was reduced by about 50% relative to wild-type cells without affecting the expression level and translation efficiency of the protein (Figure 24).

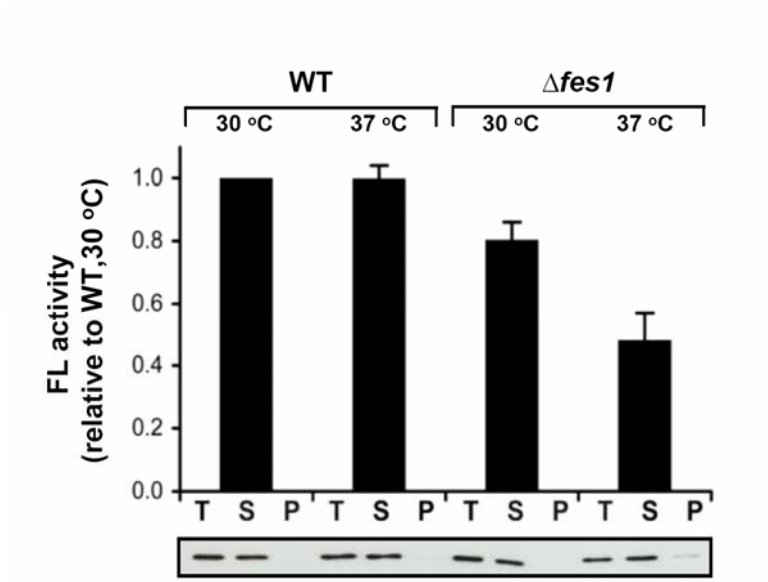

Figure 24. Effect of Fes1p on FL folding in *S. cerevisiae*

FL was expressed in the *FES1*-deleted *S. cerevisiae* strain, JY053 (Δ*fes1*), and in an isogenic wild-type strain, YPH499 (WT). Enzyme activities of FL were monitored (upper panel), and values for wild-type yeast grown at 30°C were used as an internal reference (set to 1). The lower panel shows an immunoblot used for quantitation of FL protein. The bands correspond to total (T), soluble (S), and insoluble FL (P).

Formation of insoluble aggregates of FL was not prominent in Δ*fes1* cells, but size exclusion chromatography of cell lysates revealed that FL eluted as a broad peak centered around 160 kDa. In contrast, FL protein and activity produced in wild-type cells co-eluted as a sharp symmetrical peak at ~70 kDa (Figure 25A). We suspected that the high molecular weight form of FL represented protein that remained associated with chaperones.

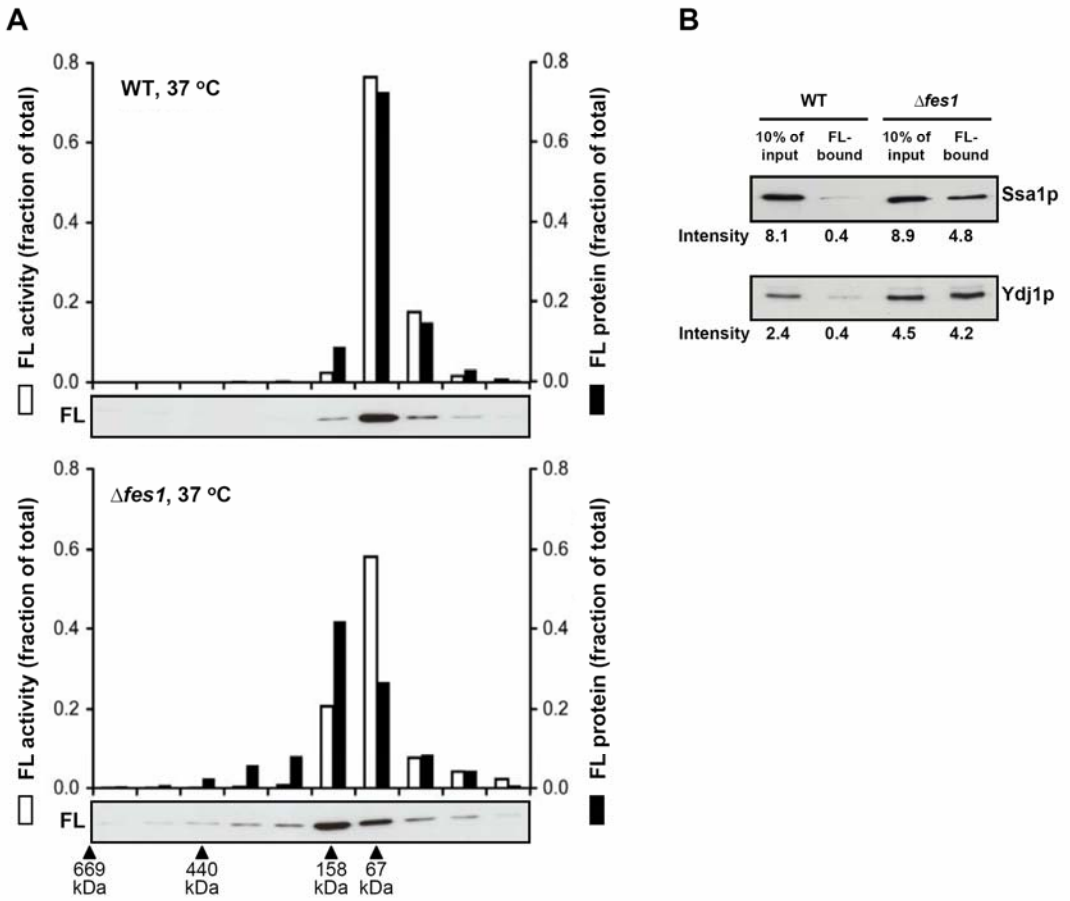

Figure 25. Nonnative FL accumulation in Δ*fes1* yeast cells

(A) Analysis of apyrase-treated yeast lysates expressing FL by Superdex-200 (Amersham) size exclusion chromatography. The upper and lower panels show wild-type (WT) and Δ*fes1* lysates, respectively. White bars denote normalized FL activity *per* fraction; black bars represent FL protein as determined by quantifying the immunoblots shown below. The positions of molecular mass markers are indicated. (B) Association of luciferase with Ssa1p and Ydj1p in Δ*fes1* cells at 37°C. Apyrase-treated yeast lysates expressing FL with a C-terminal His-tag (input) were incubated with Ni-NTA agarose beads and bound proteins (FL-bound) probed with antisera specific for Ssa1p and Ydj1p. The intensity of the bands was analyzed by densitometry.

Indeed, co-precipitation experiments from cell lysates using the C-terminal His-tag of FL showed that a substantial fraction of total Ssa1p (~5%) was associated with FL in Δ*fes1* cells at 37 °C, but not in wild-type cells (Figure 25B). Likewise, ~10% of total Ydj1p was associated with FL in Δ*fes1* cells. Although Δ*fes1* cells expressed approximately 2-fold

more Ydj1p than wild-type cells under these conditions, the amount of Ydj1p associated with FL in Δ*fes1* cells was at least 10-fold greater than in the wild-type (Figure 25B). These results indicate that Fes1p is required for fully efficient Hsp70-mediated folding of some proteins at 37 ^{o}C, consistent with a role of Fes1p as a nucleotide exchange factor in regulating substrate binding and release by Hsp70. Because eukaryotic Hsp70s are not strictly dependent on nucleotide exchange factors, unlike bacterial DnaK (Minami, 1996), it is plausible that Fes1p function becomes limiting only at elevated growth temperatures.

To correlate the physiological relevance of the above observation that Fes1p is indeed involved in regulating Ssa function, our coworkers, Yasuhito Shomura and Andreas Bracher, predicted the crucial contact residues of Fes1p to the Ssa1p ATPase domain based on their crystal structural work of the HspBP1 core domain (the mammalian ortholog of Fes1p)/Hsp70 ATPase domain complex, and designed a Ssa1p-binding deficient Fes1p (A79R, R195A) mutant. Taking advantage that *Δfes1* yeast cells failed to grow at 37 ^{o}C, we expressed the Ssa1p-binding deficient Fes1p (A79R, R195A) mutant in a *Δfes1* background and examined whether these cells displayed the same temperature sensitive phenotype. Indeed, Fes1p (A79R, R195A) does not compensate for the temperature sensitive phenotype of *Δfes1* cells (Figure 26), indicating Fes1p in the yeast cytosol is acting on the regulation of Ssa-class Hsp70 function, during the efficient folding of nascent polypeptide chains.

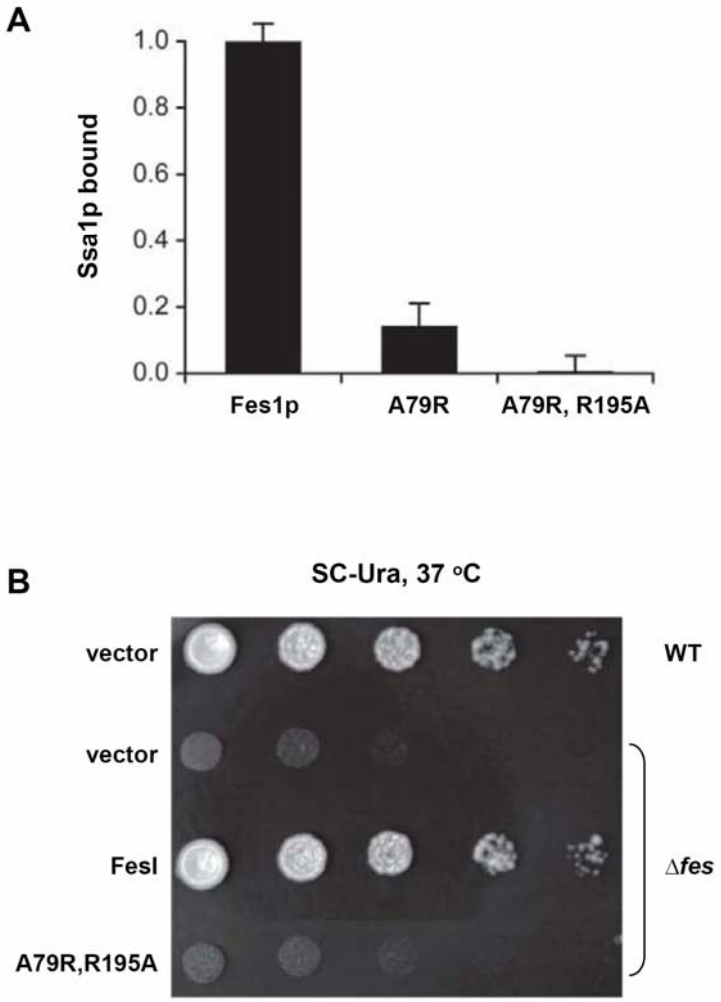

Figure 26. Molecular interactions between Fes1p and Ssa1p

(A) Ssa1p binding to Fes1p is strongly reduced by the combined mutation A79R and R195A. His-tagged Fes1p, Fes1p (A79R), and Fes1p (A79R, R195A) were incubated with purified Ssa1p at a ratio of 1:1 and applied to Ni-affinity resin. After being washed with three column volumes, bound protein was eluted and analyzed on a Coomassie-stained SDS-PAGE gel. Bound Ssa1p was quantified by densitometry and corrected for nonspecific binding in the absence of Fes1p. Amounts of Ssa1p bound relative to that observed with wild-type Fes1p are shown. (B) Fes1p (A79R, R195A) does not restore thermotolerance to Δ*fes1* cells. The *FES1*- deleted *S. cerevisiae* strain JY053 was transformed with centromeric plasmids encoding Fes1p (WT), Fes1p (A79R, R195A), or a plasmid without insert (vector). Serial dilutions of overnight cultures were inoculated on agar plates containing selective medium at 37°C for two days. No growth differences were observed at 30°C (data not shown). Experiments were performed in cooperation with Y. Shomura and Z. Dragovic.

4.3.2. FL folding in yeast strains lacking nascent chain binding chaperones

Subsequent work was designed to identify eukaryotic cytosolic factors that may contribute to efficient co-translational folding by examining a series of chaperone deletions in *S. cerevisiae*. As mentioned previously, there are at least five factors which can be regarded as potential candidates: RAC, Ssb, GimC, NAC and Ssa. All of these factors have been either demonstrated to bind nascent chains by cross-linking approaches or have been actually isolated together with ribosomes in the presence of nascent chains, and therefore could potentially assist the *de novo* folding in the yeast cytosol.

Deletions of Zuotin, Ssz1p, Ssb1/2p, Gim1p (GimC subunit 1), Egd2p (αNAC) and Ssa1/2p in the *S. cerevisiae* strain YPH499 were prepared in this work, which allowed us to analyze FL folding based on the same genetic background.

Upon induction of FL in these chaperone-deficient yeast strains at 30°C, we observed that soluble yet inactive FL accumulated in the mutant strains (Figure 27). The specific activity of FL expressed in *Δzuo1* was decreased to ~40% relative to wild-type cells, similar to the defect of FL folding in cells lacking Ssb1/2p. The lack of the other RAC component, Ssz1p, however, resulted in a diminished compromise of FL folding, with only ~35% reduction in FL activity. While essentially the solubility of FL was not altered in those three chaperone-deficient backgrounds, the *SSA1/2* deletion strain is the only strain tested in this work that resulted in insoluble FL formation (~30% of total synthesized FL), together with ~50% FL activity loss. Specific activities of FL in *GIM1* and *EGD2* deletion strains are ~75% and 90%, respectively, relative to wild-type yeast cells. Thus devoid of GimC and NAC seemed to be less significant in affecting FL folding compared to the other four chaperone-deficient strains. Perhaps their effects can be only seen at more strict conditions, *e.g.* at elevated growth temperature, as observed for Fes1p.

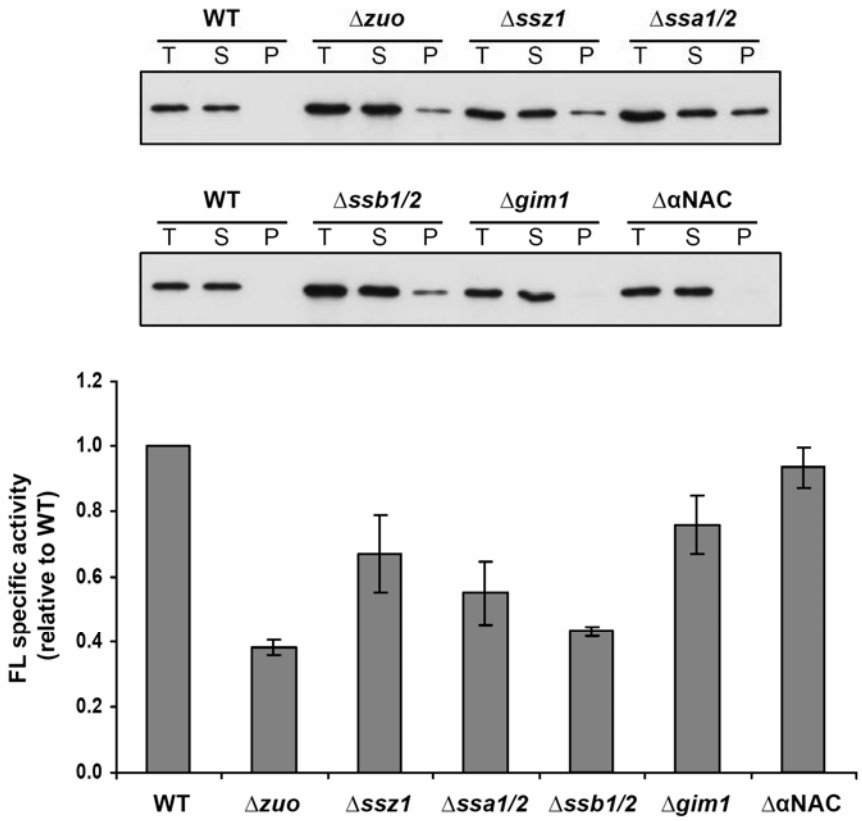

Figure 27. Effect of nascent chain binding chaperones on FL folding in *S. cerevisiae*

Accumulation of FL activity in *S. cerevisiae* cells upon induction of FL from a copper-regulated promoter. FL was expressed (200 μM $CuSO_4$ induction for 90 min at 30 ^{o}C) in the *ZUO1-*, *SSZ1-, SSA1/2-, SSB1/2-, GIM1-, and EGD2* (α-NAC)-deleted *S. cerevisiae* strains, and in an isogenic wild-type strain, YPH499 (WT). The upper panels show the immunoblots used for quantitation of FL protein. The amount of FL applied in total fraction (T) corresponds to 10^8 RLU (relative light unit) in activity. Specific activities of FL were calculated (lower panel), and values for wild-type yeast were used as an internal reference (set to 1).

This results suggested that for efficient co-translational folding of FL in yeast, the RAC/Ssb functional triad appears to play an important role for the first line of promoting the folding (or preventing the aggregation) of nascent FL, and likely with the further involvement of Ssa proteins to complete the folding process. This is reminiscent of the action of TF/DnaK in *E. coli*, yet may act in a much more productive manner.

4.3.3. The yeast cytosol can support folding of recombinantly expressed bacterial proteins unless these are strictly dependent on the specialized bacterial chaperonin GroEL

In addition to FL, a eukaryotic multi-domain protein, and the set of artificial GFPuv fusion proteins, we further utilized a variety of *E. coli* proteins to evaluate the ability of yeast cells in supporting recombinant protein folding. Bacterial proteins are evolutionarily distant from those in eukaryotes, and thus would pose a challenge to the eukaryotic cytosol and allowed us to test whether yeast cells are suitably equipped to fold a large spectrum of proteins of different complexities and origins. Our laboratory had previously identified several authentic GroEL-interacting proteins in *E. coli* (Houry *et al*., 1999; Kerner *et al*., 2005). Based on the degree of their chaperone dependence for folding, these bacterial proteins had been grouped into three classes, as listed below in Table 5.

Class I substrates, including *E. coli* enolase, are independent of chaperones for their refolding *in vitro*. Class II proteins do not refold efficiently in the absence of chaperones, but can utilize either the DnaK or the GroEL/ES systems. Class III substrates, in contrast, were found to be fully dependent on GroEL/ES for their folding, and 12 of these substrates were further analyzed upon expressing in *S. cerevisiae*.

Table 5. *E. coli* GroEL interacting proteins utilized for heterologous expression in yeast

	Swiss Port ID	Description	Molecular Mass (monomer)	Quaternary structure
Class I	ENO	Enolase (2-phosphoglycerate dehydratase)	45.5 kDa	homodimer
Class II	DCEA	Glutamate decarboxylase alpha	52.7 kDa	homohexamer
	GATD	Galactitol-1-phosphate 5-dehydrogenase	37.4 kDa	
	SYT	Threonyl-tRNA synthetase (ThrRS)	74.0 kDa	homodimer
Class III	ADD	Adenosine deaminase	36.4 kDa	
	ALR2	Alanine racemase	38.8 kDa	
	DAPA	Dihydrodipicolinate synthase (DHDPS)	31.3 kDa	homotetermer
	END4	Endonuclease IV	31.5 kDa	monomer
	GATY	Tagatose-1,6-bisphosphate aldolase (TBPA)	30.8 kDa	
	HEM2	Delta-aminolevulinic acid dehydratase	35.5 kDa	homooctamer
	LLDD	L-lactate dehydrogenase (Cytochrome)	42.7 kDa	
	METF	5,10-Methylene-tetrahydrofolate reductase	33.1 kDa	homotetermer
	METK	S-Adenosylmethionine synthetase	41.8 kDa	homotetermer
	NPL	N-acetylneuraminate lyase	32.5 kDa	homotetermer
	XYLA	Xylose isomerase	49.7 kDa	homotetermer
	YAJO	Hypothetical oxidoreductase	36.4 kDa	

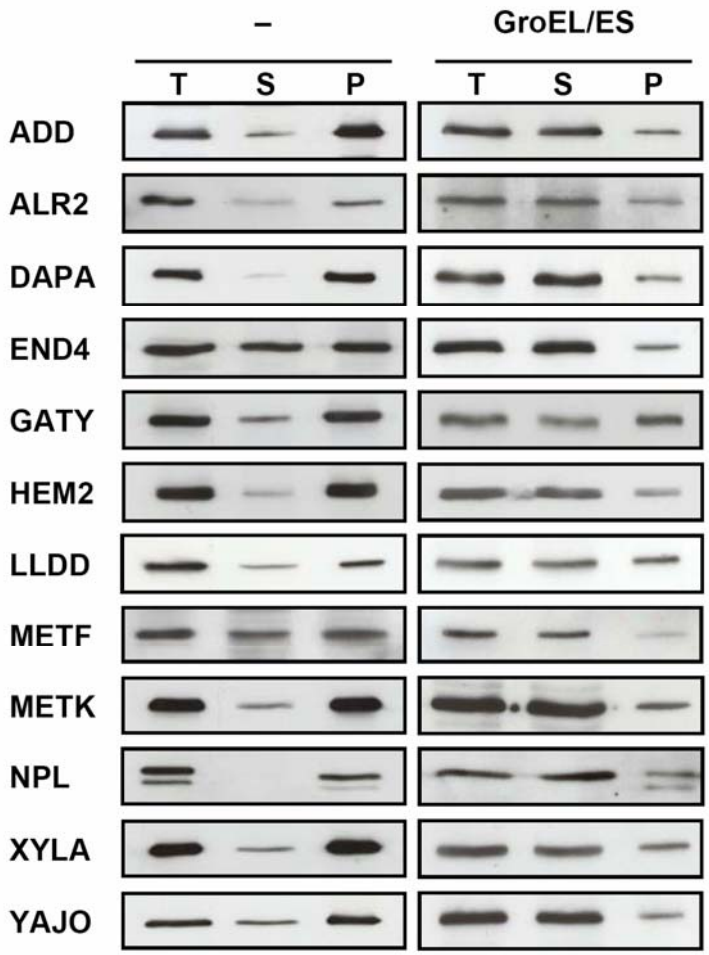

Figure 28. Solubility of *E. coli* GroEL substrates upon heterologous expression in *S. cerevisiae*

Solubility of class III proteins (ADD, ALR2, DAPA, END4, GATY, HEM2, LLDD, METF, METK, NPL, XYLA, andYAJO) that were expressed at 30°C in *S. cerevisiae* YPH499 wild-type cells either without (−) or with co-expression of GroEL/GroES. Total (T), soluble (S), and insoluble pellet fractions (P) were analyzed by immunoblotting.

Surprisingly, when moderately expressed in wild-type yeast, all class III substrates accumulated largely in the insoluble fraction (Figure 28). Class III proteins could only be isolated in the soluble fraction when bacterial GroEL and GroES were both co-expressed in the yeast cytosol. In a control experiment where GroEL and class III substrates were expressed without GroES, the substrates failed to show any improvement in solubility, which indicates that the complete chaperonin system is required for obligate GroEL substrates to fold (Figure 29A). This experiment also showed that the expression of GroEL in itself did not cause the increased solubility of the co-expressed substrates. Taken together, we can conclude that the requirement of the *E. coli* class III proteins for GroEL/GroES is specific and independent of the bacterial machinery of protein biosynthesis.

In contrast, *E. coli* class I and class II proteins were soluble upon expression in wild-type yeast (Figure 29B), confirming that the yeast cytosol indeed has a broad capacity for folding proteins correctly regardless of their origin. Substantial aggregation of class II proteins was however observed in the mutant strain Δ*ydj1* that lacks the Hsp70 cofactor Ydj1p, supporting the conclusion that class II proteins are chaperone-dependent but can utilize either the Hsp70 system or GroEL/GroES for folding.

Together these findings suggest that the chaperone machineries of cells and their substrates have co-evolved. While the bacterial chaperone system is adapted to trigger post-translational folding, the eukaryotic chaperone machinery is optimized for the co-translational folding of multi-domain protein but fails to supply for the post-translational folding of heterologous proteins that are strongly dependent on the bacterial GroEL chaperonin.

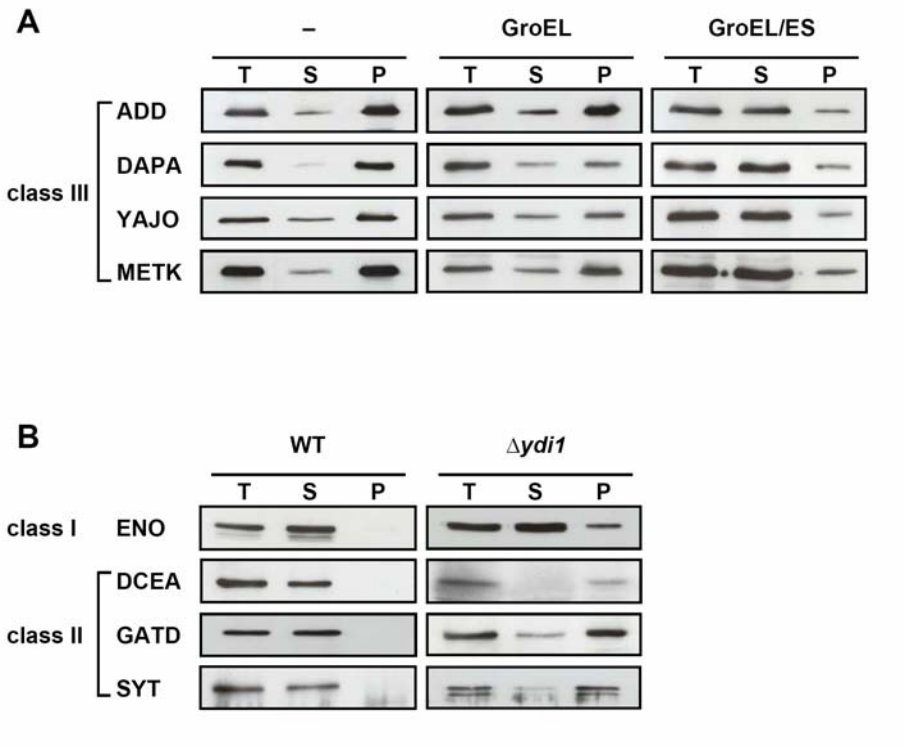

Figure 29. Solubility of bacterial class I, II and III substrates upon heterologous expression in *S. cerevisiae.*

(A) Solubility of class III proteins ADD, DAPA, YAJO, and METK, that were expressed at 30°C in *S. cerevisiae* YPH499 wild-type cells either without (−) or with co-expression of GroEL or GroEL/GroES. Total (T), soluble (S), and insoluble pellet fractions (P) were analyzed by immunoblotting. (B) Solubility of class I and II proteins in wild-type and *YDJ1*-deleted yeast cells. ENO (class I) and DCEA, GATD, and SYT (class II) were expressed in a *YDJ1*-deleted (Δ*ydj1*) strain and in its isogenic wild-type DS10 (WT) strain. Cell fractions were analyzed as above.

5. Discussion

In this study, a large number of GFP fusion proteins were constructed and compared with regard to their folding efficiencies in *E. coli* and *S. cerevisiae.* By varying the linker sequences between domains and the expression conditions for these GFP fusions, it became clear that these artificial multi-domain proteins fold remarkably inefficient in bacteria.

Further *in vitro* and *in vivo* experiments provided insight into the role of bacterial chaperones TF and the DnaK system in the *de novo* folding of multi-domain proteins FL and β-gal. Significantly, TF and DnaK working in concert were shown to cause a delay in folding relative to translation for both FL and β-gal. While the bacterial protein β-gal is adapted to a post-translational folding regime, such a mechanism is incompatible with efficient folding of FL. In sharp contrast to the situation in bacteria, in yeast cells FL folding follows an efficient co-translational mechanism, suggesting that the eukaryotic chaperone system specifically supports the sequential productive folding of nascent chains.

Indeed, analysis of the defect of FL folding in yeast cells devoid of six different nascent chain binding chaperones, and a nucleotide exchange factor suggested that the superior folding capacity of the yeast cytosol is dependent on a cooperative network of chaperone systems. These results may prove valuable in engineering bacterial strains expressing eukaryotic chaperones for the improved expression of complex heterologous protein.

5.1. Folding complexity arising from domain fusion impeded the folding of GFP fusions in bacteria

Protein fusion strategies have been broadly applied in biological research, including the visualization of subcellular localization of target protein (Huh *et al*., 2003), high-affinity epitope tagging for global analysis of protein expression (Ghaemmaghami *et al*., 2003),

protein-protein interaction mapping (two hybrid techniques), as well as protein purification with ligand affinity chromatography. While certainly providing an informative resource for studies in eukaryotic systems, the present study indicated that protein fusion approaches in bacteria must be interpreted with care due to the frequent occurrence of protein misfolding.

To be able to evaluate the severity of the problem of fusion protein folding in bacteria, we chose individual domains for the generation of artificial multi-domain fusions based on the following criteria: i) Absence of obligate chaperone dependency; ii) High solubility when expressed individually in bacteria and eukaryotic cells; iii) Absence of post-translational modifications or disulfide bond formation as a requirement for domain folding. In addition, we selected a set of representative linkers to connect the two fusion partners, varying characteristics of length, predicted secondary structure, amino acid composition and natural-occurrence. Model proteins that fulfilled these criteria were the endogenous *E. coli* proteins, Eno, MBP, NusA, and MreB, as well as the GFP variant, GFPuv.

Intriguingly, the resulting fusions of GFPuv with any of the four robust folders and with any of the linkers failed to fold efficiently in bacteria. An immediate explanation for this result was that the fusion event increased the folding complexity of these longer polypeptides. As has been shown in numerous *in vitro* refolding studies, the number of conformations accessible to a polypeptide chain grows exponentially with chain length (Dobson, 2003). It is well accepted that for small proteins (fewer than 100 residues), the folding reaction can be simply modeled as a fast, two-state transition process between a disordered denatured state and the ordered native state, without partially folded intermediates being significantly populated. In contrast, the folding kinetics of larger proteins may in some cases be dominated by the formation of multiple intermediates. The misfolding of GFPuv fusions observed in bacteria strongly suggested that folding of these multi-domain proteins in the bacterial cytosol is post-translational, since a mechanism of co-translational and sequential domain folding would avoid the increased folding complexity

from domain fusion. Co-translational folding of one domain well before another can prevent the formation of non-productive intermediates, as observed upon expression of the GFPuv fusions in *S. cerevisiae*.

Since GFPuv was able to affect the overall folding of the fusion protein even when present as the C-terminal partner, the N-terminal domain could not have acquired its structure by the time the GFPuv portion started to emerge from the ribosome. Similarly, folding interference due to lack of (or delayed) co-translational folding of GFPuv as the N-terminal domain would explain the remarkably low solubility of fusion proteins in which the robustly folding domains (Eno, MBP) were placed at the C-terminus. Our findings are consistent with the view that bacteria have a notably reduced capacity for co-translational domain folding compared to eukaryotes, at least for certain domain structures (Netzer and Hartl, 1997).

5.2. Co- *vs.* post-translational folding: effect of nascent chain binding chaperones on multi-domain protein folding

By acting on translating polypeptide chains, TF and the DnaK chaperone system improve the folding yield for the multi-domain protein β-gal and to a lesser extent of FL in the *E. coli* cytosol. Remarkably, this increase in yield is coupled to a substantial deceleration of the folding process, compared to folding in chaperone-depleted systems. Thus, the bacterial chaperone machinery does not support the kinetically most efficient folding route available in the context of translation, but instead favors a folding mechanism with a pronounced post-translational component. As a consequence, proteins such as FL fold less efficiently than in the eukaryotic system.

5.2.1. Default folding *versus* chaperone-assisted folding

In the absence of TF and the DnaK system, folding of the multi-domain proteins FL and β-gal occurs with kinetics tightly coupled to translation, albeit in a much less efficient fashion (Figure 19 and 20). We propose that this process represents an unassisted default pathway that involves the co-translational formation of native or a native-like domain structure, followed by the rapid completion of folding upon chain release from the ribosome (Figure 30, pathway 1). However, while kinetically efficient, this reaction is characterized by a low folding yield, either as a result of intramolecular misfolding or interchain aggregation (Figure 30, pathway 2) due to the absence of chaperones that would prevent aggregation. However, surprisingly, TF and DnaK do not increase the folding yield by improving the efficiency of the default pathway. In the case of FL, folding (and misfolding) is delayed by the chaperones until the nascent chain is released from the ribosome (Figure 30, pathway 3). As in refolding, the native conformation is then reached by multiple cycles of chaperone binding and release to nonnative states (Figure 30, pathway 7), which is inefficient in yield and kinetically slow. In this reaction, TF and DnaK not only slow the completion of folding of a co-translationally prefolded intermediate, but also essentially shift the folding mechanism toward a post-translational pathway.

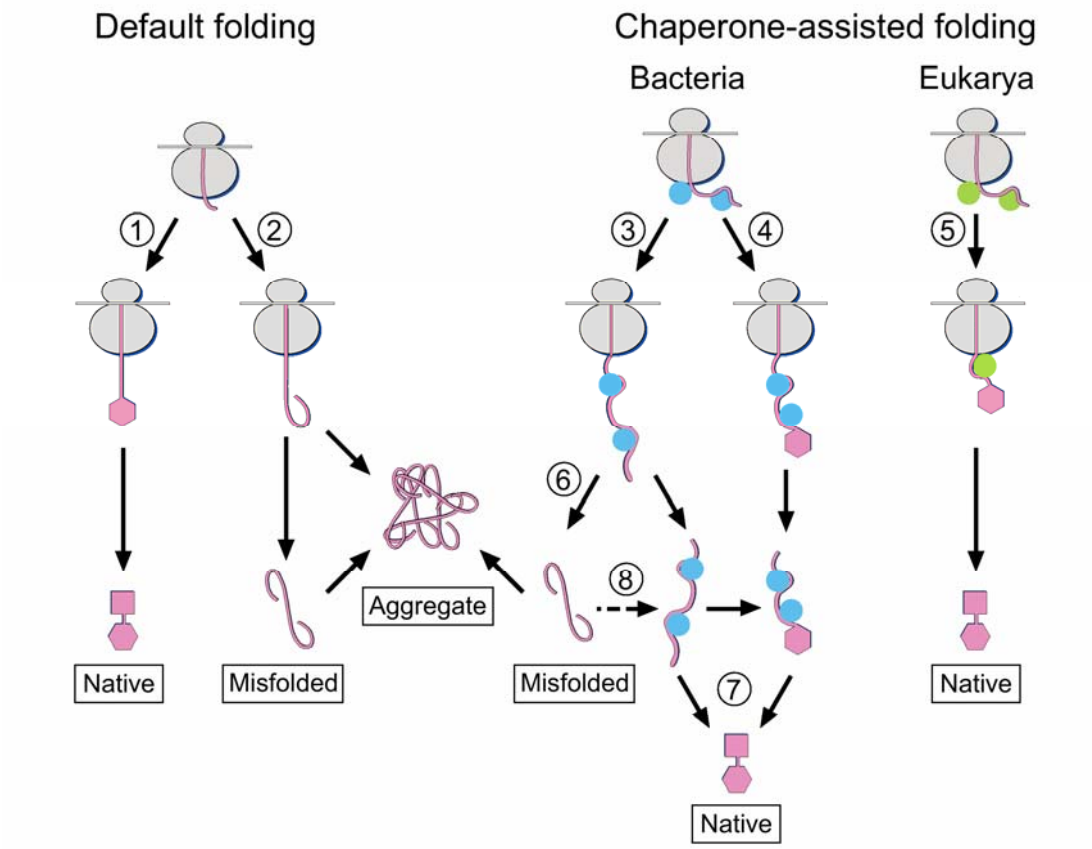

Figure 30. Effects of nascent chain binding chaperones on the folding of multi-domain proteins, a working model

The translating polypeptide chain of a hypothetical two-domain protein is shown in pink with folded domains represented by hexagons and squares. Bacterial chaperones are in blue, and eukaryotic chaperones are in green.

However, the capacity of TF and DnaK to retard folding during translation is likely to be insufficient for very long nascent chains, and, consequently, large bacterial proteins such as β-gal could potentially initiate productive domain folding co-translationally (Figure 30, pathway 4). Thus, the mechanism of chaperone-assisted *de novo* folding of FL in bacteria differs from that in the eukaryotic cytosol (Figure 30, pathway 5), where rapid and highly efficient co-translational folding is presumably supported by the Hsp70/Hsp40 chaperone system (Frydman *et al*., 1994). In contrast, the bacterial chaperones fail to efficiently prevent the misfolding of newly synthesized FL chains (Figure 30, pathway 6) and are able

to shift only a fraction of molecules to a productive post-translational folding regime (Figure 30, pathway 7).

5.2.2. Mechanism of delayed folding

An important question regarding the above model is: How do TF and DnaK delay the folding relative to translation? It has been demonstrated that both chaperones recognize similar hydrophobic regions in nascent polypeptides (Patzelt *et al*., 2001; Rudiger *et al*., 1997a). Our results indicate that TF acts to delay the folding and misfolding of nascent chains by a dynamic interaction cycle with translating ribosomes. Upon initiation of translation, ribosome-associated TF binds to the emerging chain, and additional TF molecules are recruited to the translating ribosome. Recruitment depends on the ability of TF to bind to the large ribosomal subunit, suggesting that the initially bound TF leaves the ribosome docking site but maintains contact with the elongating chain, thereby inhibiting folding/misfolding (Figure 31). Since TF does not form long-lived complexes with substrate proteins after the completion of synthesis (Hesterkamp *et al*., 1996; Maier *et al*., 2003), maintenance of folding competence in regions of the nascent protein far removed from the ribosome may require the engagement of DnaK, which acts independently of the ribosome.

DnaK binds and releases nonnative polypeptides in an ATP-dependent manner regulated by DnaJ and GrpE (Bukau and Horwich, 1998) and could in principle facilitate co-translational folding. However, following release, DnaK rebinds the nonnative protein within seconds (Pierpaoli *et al*., 1997), limiting the time available for the folding of an average domain. Moreover, both FL and β-gal contain numerous predicted high affinity binding regions for DnaK (22 in FL and 25 in β-gal; Agashe *et al*., 2004), *i.e.*, several sites in each structural domain. This suggests that rapid domain folding would require a mechanism of coordinated nascent chain release from multiple DnaK molecules. Given the

fast speed of bacterial translation (~12-22 aa/s) ((Bremer and Dennis, 1996; Mathews *et al*., 2000), the TF/KJE system would be geared toward stabilizing nascent chains of average size in an unfolded state, except for larger proteins with relatively long translation times (~50 s for β-gal, for example).

Based on these considerations, efficient co-translational folding in *E. coli* would occur for proteins consisting of relatively small, fast folding domains with few chaperone recognition motifs. Such a reaction has been observed with the 149 residue Semliki Forest Virus Protease (SFVP) as a model protein (Nicola *et al*., 1999), which contains only three predicted high affinity sites for DnaK. In mammalian host cells, this module has been under strong selective pressure to fold co-translationally. It must cleave itself from the growing nascent chain to expose a signal sequence that co-translationally targets the remainder of the viral polyprotein to the endoplasmic reticulum (ER). As shown recently, SFVP is the fastest refolding two-domain protein known ($t_{1/2}$ ~50 ms) (Sanchez *et al*., 2004), and its folding upon synthesis is apparently Hsp70-independent (Nicola *et al*., 1999). More generally, however, the efficient co-translational folding of multi-domain proteins in eukaryotes may require chaperone assistance, as shown for FL (Frydman *et al*., 1994). We speculate that this could be accomplished through a functional regulation and cooperation of the eukaryotic Hsp70/Hsp40 system with additional chaperone components and may be facilitated by the reduced translation speed in eukaryotes.

5.2.3. Structure of trigger factor and current model of its function

Trigger factor is known to be composed of a ribosome binding domain (BD), a peptidyl-prolyl-cis/trans isomerase (PPIase) domain and a C-terminal portion without sequence homology to any other proteins in the database. The recent crystal structure of TF showed how these domains are arranged (Ferbitz *et al*., 2004) (Figure 31). Interestingly, the TF

ribosome-binding domain (TF-BD) is connected *via* a long linker to the PPIase domain at the opposite end of the elongated molecule and the C-terminal domain is positioned in between.

The crystal structure of full-length TF in complex with the ribosome has not been solved to date. However, crystal structures of the amino-terminal TF-BD with the large ribosomal subunit from bacteria and archaea have allowed to model full-length TF onto the ribosome (Baram *et al*., 2005; Ferbitz *et al*., 2004; Schlunzen *et al*., 2005), suggesting its probable position. All of these models predict that the hydrophobic crevice of TF is oriented towards the exit tunnel and thus would be accessible to the emerging nascent chain (see Figure 31 and 32).

In the earliest of these models, built from a co-crystal structure of the amino terminal 144 residues of *E. coli* TF in a heterologous complex with the 50S subunit from the archaeon *H. marismortui*, TF was proposed to crouch over the ribosomal exit tunnel. The resulting "cradle"-shaped space was suggested to provide a confined space large enough to accommodate compactly folded domains of up to ~15 kDa in size, allowing their co-translational folding in a protected environment (Ferbitz *et al.,* 2004). This model, however, was afflicted with major uncertainty, since only 35 out of 144 residues of the TF-BD could be visualized, in the crystal structure whereas the other residues could not be assigned to the experimentally determined electron density. In addition, due to the absence of TF in all archaeal genomes sequenced to date, the heterologous complex provided incomplete insight into TF ribosome interactions.

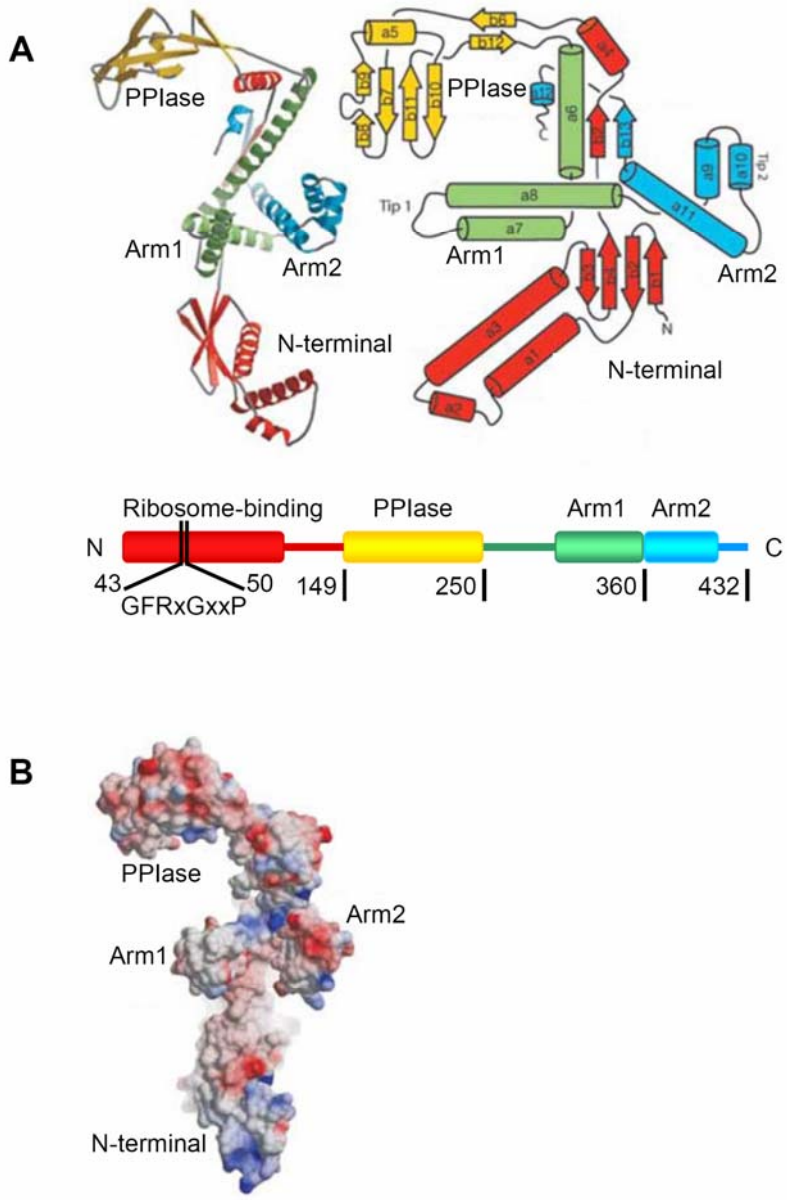

Figure 31. Structure of *E. coli* trigger factor

(A) Trigger factor adopts an extended fold. Left: ribbon diagram of the trigger factor fold. Right: schematic representation of the domain organization. Bottom: domain arrangement in sequence space. Positions of the ribosome-binding trigger factor signature (residues 43–50) and domain borders are indicated. In all parts the ribosome binding (N-terminal) domain is shown in red, the PPIase domain in yellow and 'arm' 1 and 'arm' 2 in green and blue, respectively. (B) Solvent accessible surface of trigger factor (in stereo view), colored by electrostatic potential (blue, positive; red, negative). Adapted from Ferbitz *et al*. (2004)

Subsequently, two structures of a homologous complex of the TF-BD and the 50S subunit from the eubacterium *Deinococcus radiodurans* were solved, revealing the molecular details of the interaction between the ribosome and this TF domain (Baram *et al.,* 2005; Schlunzen *et al.,* 2005). The predominant contacts that mediate ribosome binding of TF are established by the flexible loop region of TF, which is located between helices α1 and α2 and contains the TF signature motif. This region contacts the ribosomal protein L23 and ribosomal RNA. Residues at the side of the loop also contact the ribosomal protein L29, although to a minor extent (Schlunzen *et al.*, 2005). This is consistent with L29 not being essential for TF-ribosome interactions (Kramer *et al.,* 2002). In addition, helix α2 contacts an extended loop structure of L24 (Schlunzen *et al.*, 2005) (see Figure 32).

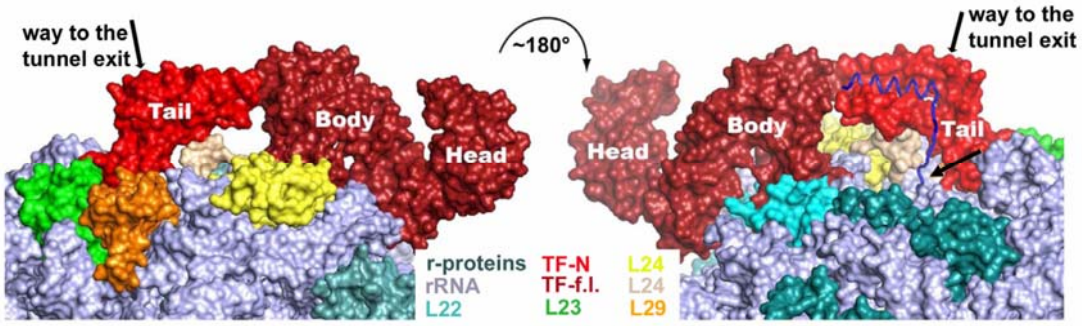

Figure 32. Modeling of full-length *E. coli* TF onto the ribosome

Full-length *E. coli* TF were modeled (Body, Head; dark red) onto the ribosome based on the co-crystal structure of the amino-terminal TF-domain of *D. radiodurans* TF (Tail; red) in complex with its 50S ribosomal subunit. Ribosomal proteins (r-proteins) and rRNA are colored as indicated. A modeled nascent chain emerging from the exit tunnel (indicated by black arrows) is shown in blue. The loop-extension of L24 (shown in gold; with the globular domain in yellow) encroaches on the hydrophobic crevice. Adapted from Schlunzen *et al.,* 2005.

A reorientation of helix α2 in the ribosome-bound state (see Figure 33) enables the interaction of TF with L24. In the ribosome-bound state, the carboxyl-terminal end of helix α2 is shifted away from the β-sheet structure of the TF-BD, as compared to the situation with unbound TF. This conformational rearrangement causes the binding loop to adopt a more open conformation. Thus subsequent rearrangement of the amino-terminal TF-BD opens up a tunnel-like structure with highly hydrophobic walls, that is aligned with the ribosomal exit tunnel on one side and with the hydrophobic crevice of the TF body on the other (Schlunzen *et al.*, 2005) (see Figure 33). These observations suggest that ribosome binding induces the formation of a channel in the TF amino-terminal domain that could guide the nascent chain towards the hydrophobic crevice.

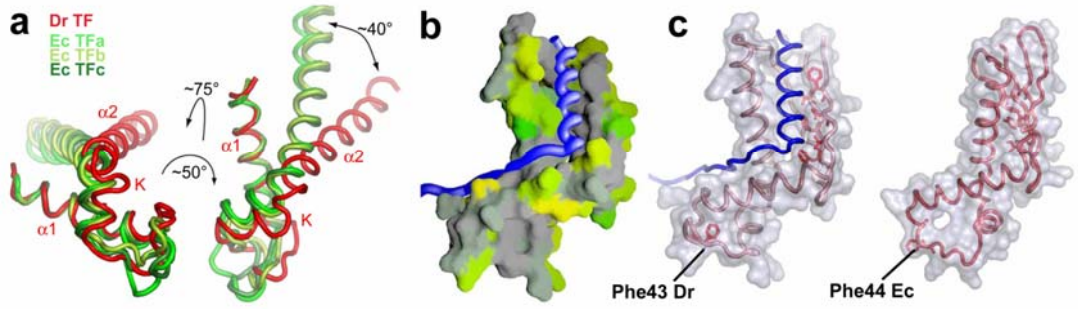

Figure 33. Conformational rearrangements in the TF ribosome binding domain.

Dr: *D. radiodurans*, Ec: *E. coli*, K: kink in helix α2. **a.** Superposition of the loop region of ribosome-bound TF from *D. radiodurans* (red) with each of the three molecules (a, b and c, in different shades of green) found in the asymmetric unit of the unbound amino-terminal TF-BD from *E. coli* (Kristensen & Gajhede 2003). The two orientations shown are related by rotation as indicated. Helix α2 of the ribosome-bound *D. radiodurans* TF-BD is shifted by ~40° when compared to the free *E. coli* TF. **b.** Surface representation of the ribosome bound amino-terminal domain of *D. radiodurans* TF. An α-helical nascent peptide is modeled into the conduit. Hydrophobic regions on the surface are shown in gray, polar regions in green/orange. **c.** Left: same as **b.**, but with a transparent surface and a backbone representation in red. Right: same representation of the free *E. coli* TF-BD. The hydrophobic channel is present only in the ribosome-bound structure, where the universally conserved Phe43-residue and the neighboring residues are flipped into the hydrophobic loop-region compared to the structure of the free *E. coli* TF. From Schlunzen *et al.,* 2005.

The docking of the full-length TF onto the ribosome-bound TF-BD from *D. radiodurans* TF resulted in a model in which the overall orientation of TF was similar to the one described by Ferbitz *et al.* The interaction between helix α2 and L24, however, was not visible in the heterologous complex, because the L24 from *H. marismortui* lacks the loop extension that contacts α2 (Ferbitz *et al.*, 2004; Schlunzen *et al.*, 2005). The L24 loop is relatively disordered in the native *D. radiodurans* 50S structure, but appears to be stabilized in the complex with the TF amino-terminal domain. In the resulting model, the space confined by the molecular cradle is substantially reduced by the L24 extension, such that a globular domain of ~15 kDa cannot be accommodated (see Figure 32). The loop extension of L24 from *E. coli* is slightly larger than in *D. radiodurans*. The crevice in the homologous complex of *E. coli* TF with its 50S ribosomal subunit is thus expected to exhibit at least the same restriction in space, rendering the occurrence of co-translational folding within the proposed cradle very unlikely (Schlunzen *et al.*, 2005).

A fluorescence-based approach was recently employed in our laboratory to allow the detailed observation of TF activity in real time during translation of nascent chains (Kaiser *et al*., accepted for publication), which provided detailed insight in clarifying the working models of TF mainly based on biochemical and structural studies. By measuring intra-molecular fluorescence resonance energy transfer (FRET) between exogenous fluorophores attached site-specifically to TF, it was shown that binding to the ribosome causes a conformational opening within TF. This structural rearrangement may position the substrate-binding region in a state competent to receive the polypeptide nascent chain. Ribosome binding of TF can also be followed by an environmentally sensitive fluorophore covalently linked to the TF molecule. Conformational compaction and release of TF from non-translating ribosomes were found to occur concomitantly, with a half-time of $t_{1/2} \approx 10$ s. During translation of firefly luciferase (FL), a previously characterized TF substrate, the kinetics of ribosome release were found to be unaltered, whereas the conformational

compaction event took place with $t_{1/2} \approx 36$ s. The delay in conformational compaction kinetics over ribosome release that was observed during FL translation suggests that TF stays associated with the nascent chain after vacating the ribosomal binding site, thus allowing the loading of another TF molecule onto the same nascent chain. Translation of substrates other than FL revealed that the half-time of TF molecular compaction correlates well with the occurrence of motifs of high hydrophobicity on the nascent polypeptide chain. TF-substrate interaction is prolonged for nascent chains exposing sequences of high mean hydrophobicity (Kaiser *et al*., data not shown). The work by Kaiser *et al.* work supported the conclusion of this study that TF delays folding and misfolding events relative to translation *in vivo*.

5.3. Factors that contribute to efficient co-translational folding in yeast (and their limitations)

As in bacteria, *de novo* folding of nascent chains in the eukaryotic cytosol is assisted by nascent chain binding chaperones. However, unlike the situation in bacteria, where TF and the DnaK system are the main players of nascent chain binding chaperones, a large variety of different chaperones in the eukaryotic cytosol are present and may contribute to efficient co-translational folding of nascent chains. Importantly, there is no TF homolog in eukaryotes. However, several cytosolic chaperones have been shown to interact with ribosome-bound nascent chains. In *S. cerevisiae*, these factors include the nascent chain-associated complex (NAC) (Wang *et al*., 1995), the ribosome-associated complex (RAC) (Gautschi *et al*., 2001; (Pfund et al., 1998), the gene involved in microtubule biosynthesis (GimC/prefoldin) complex (Hansen *et al*., 1999), and the eukaryotic chaperonin, TRiC (Frydman and Hartl, 1996).

NAC was identified by cross-linking with nascent chains and appears to interact broadly with translating polypeptides at a distance of 17-100 amino acids from the ribosomal peptidyl transferase center (Wang *et al*., 1995). The heterodimeric complex of NAC can be separated into α and β subunits. In *S. cerevisiae*, these proteins are encoded by the *EGD2* (α), *EGD1* (β_1) and *BTT1* (β_3) genes. The concentration of the yeast αNAC/β_1NAC complex was estimated to be roughly equimolar to ribosomes, while the αNAC/β_3NAC complex is 100-fold less abundant (George *et al*., 1998). The role of NAC in folding is still unclear.

Binding of Hsp70 family members to ribosome-nascent chain complexes has been demonstrated in mammalian cells and in yeast (Hartl and Hayer-Hartl, 2002). In *S. cerevisiae*, two families of Hsp70 homologs bind to ribosomes: the Ssb proteins (Pfund *et al*., 1998) and the Ssz1p (Gautschi *et al*., 2001). The association of Ssb with ribosomes is salt-sensitive in the absence of translation, but becomes resistant to 2 M NaCl in the presence of nascent chains. Moreover, Ssb can be cross-linked to ribosome-bound nascent chains as short as 54 amino acids, suggesting that Ssb is located very close to the polypeptide exit site of the eukaryotic ribosome (Hundley *et al*., 2002).

TRiC, the group II chaperonin of the eukaryotic cytosol, has been found to be photo-crosslinked with actin nascent chains as short as 133 amino acids (McCallum *et al*., 2000). Since approximately 40 amino acids at the C-terminus of the nascent chain are sequestered inside the exit tunnel of the ribosome, this length would leave only about 90 amino acids outside the exit tunnel available for TRiC binding. Given that the prefoldin/GimC complex binds first to nascent actin and tubulin chains and then delivers these substrates to TRiC, thus facilitating efficient folding of actin and tubulin (Hansen *et al*., 1999; Siegers *et al*., 1999), GimC acting together with TRiC may have an important role in general nascent chain folding in the eukaryotic cytosol.

Although no cross-linking of Ssa to ribosome-bound nascent chains has been observed in yeast (Pfund *et al*., 1998), there is indeed evidence for the involvement of Ssa during the

folding of newly synthesized polypeptides. A temperature-sensitive *ssa*-deficient strain has been shown to have reduced specific activity of the cytosolic enzyme ornithine transcarbamylase (Kim *et al*., 1998). Moreover, the Ssa subfamily is the only essential of four Hsp70 subfamilies (Ssa, Ssb, Sse, and Ssz) in the yeast cytosol. The expression of at least one of the four Ssa (Ssa1-4) proteins is essential for yeast cells to survive (Werner-Washburne *et al*., 1987). Since it has been shown that Ssa proteins are indeed responsible for the folding of newly synthesized polypeptides, it is not unreasonable to suggest that Hsp70 cofactors also participate in this process, such as the recently identified nucleotide exchange factor Fes1p, whose structure has recently been determined by our laboratory. We proceeded in studying the involvement of this cofactor by examining the folding of our well characterized model substrate FL in a series of chaperone deletion strains in *S. cerevisiae*.

5.3.1. Inefficient FL folding in yeast strains devoid of nascent chain binding chaperones and the nucleotide exchange factor Fes1p

The observation of inefficient FL folding in yeast strains lacking nascent chain binding chaperones, as well as similar results obtained in a nucleotide exchange factor (NEF) *FES1* deletion strain, supports the idea that *de novo* protein folding is largely dependent on nascent chain binding chaperones. Interestingly, whereas the *FES1* deletion showed a conditional defect only at elevated temperatures (37^{o}C), FL folding was significantly compromised in yeast cells lacking Ssa1/2p at 30^{o}C and was accompanied by aggregate formation (Figure 32). As Fes1p serves as a NEF in regulating Ssa-class Hsp70 function, this result implied that there is functional redundancy with regard to NEF function in the yeast cytosol. Indeed, the recent finding of a Hsp110 chaperone in yeast, Sse1p, acting as a potent NEF for both Ssa1p and Ssb1p provided evidence in explaining why Fes1p is not essential at physiological growth temperatures. The fact that Sse1p and Fes1p exhibit partial

functional redundancy suggests a versatile regulation mechanism for Hsp70s in the yeast cytosol (Dragovic *et al*., 2006; Raviol *et al*., 2006).

Strong impairment of the folding of nascent FL in cells lacking zuotin or Ssb1/2p, but to a lesser extent cells lacking Ssz1p, points to Ssz1p as a unique player in this functional triad. In fact, Ssz1p is an unusual Hsp70 as it associates with J-domain protein zuotion in a stable manner. In addition, Ssz1p does not associate with nascent polypeptide chains, by contrast with Ssb, and its C-terminal substrate-binding domain can be deleted without compromising its activity *in vivo* (Hundley *et al*., 2002). Although Ssz1p can bind nucleotide, it lacks ATP hydrolysis activity almost completely, which is consistent with the report that zuotin can only stimulate the ATPase activity of Ssb but not Ssz1p (Huang *et al*., 2005). Thus, it seems that Ssz1p is an Hsp70 stripped of the usual functions except for its binding to zuotin. On the other hand, even in the absence of Ssz1, zuotin may be able to stimulate the ATPase activity of Ssb for substrate protein folding, consistent with the limited effect of Ssz1 deletion on FL folding.

The accumulation of soluble nonnative FL species in yeast cells is intriguing when considering the high degree of cytoplasmic protein quality control in eukaryotic cells. It suggests that the overall concentration of chaperones participating in FL folding is high thus allowing folding intermediates to be maintained in a soluble state. The misfolded species are apparently protected against immediate degradation by the proteasome machinery, supporting the view that molecular chaperones represent a first line of protein quality control in cells. Since all our observations are based on the folding of one substrate protein, FL, further experiments with additional model substrates must be performed to ascertain the generality of our findings.

5.3.2. Inability of the yeast cytosol to support the folding of bacterial proteins that are GroEL dependent

Based on the above observations, the yeast cytosol was shown to have a robust capacity in supporting the folding of recombinant multi-domain proteins including the GFPuv fusion proteins and FL. However, a major limitation of this ability was observed when expressing GroEL-dependent bacterial proteins. None of the 12 stringently GroEL dependent (class III) proteins tested accumulated in a soluble, native state upon expression in wild-type yeast cells. Surprisingly, unlike FL, these class III proteins aggregated upon synthesis, indicating that the yeast cytosolic chaperones are not able to mediate their productive folding. In some cases, aggregation was so pronounced that the insoluble protein could be observed in Coomassie blue-stained gels (ADD, XYLA; ~1% of total protein; data not shown). Thus, aggregation of these proteins was very rapid and escaped the yeast quality control system for adequate recognition and degradation of misfolded polypeptides.

On the other hand, the class I and class II substrates (ENO, DCEA, GATD and SYT) expressed in wild-type yeast cells were efficiently folded and accumulated in the soluble fraction. This suggested that these proteins were able to utilize the endogenous yeast chaperones productively, similar to the situation with FL. Only upon expression in *Δydj1* cells, which have a functionally deficient Hsp70 system, did the class II substrates aggregate, whereas the class I substrate, enolase, remained largely soluble (Figure 25B). This is in good agreement with the initial definition of class I and class II substrates (Kerner *et al*., 2005), according to which the class I proteins are largely independent of chaperone assistance, and class II proteins can utilize either the Hsp70 system or the bacterial chaperonin for folding. Importantly, these results suggest that the functional capacity of eukaryotic Hsp70s overlap with that of the DnaK system, whereas the eukaryotic chaperonin is unable to functionally replace its bacterial counterpart.

The general ability of the eukaryotic cytosol to support the folding of 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 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. It can be further concluded from these findings that the sequestered folding environment provided by TRiC differs from that of bacterial GroEL. Notably, except for METK, a two layer α+β domain protein, all the class III substrates tested in this work are TIM barrel structured protein, indicating a remarkable bias of GroEL towards this class of proteins. How class III substrates, in particularly TIM barrel proteins, are specifically recognized and folded by the GroEL/ES system requires further investigation.

5.4. Perspectives

Highly modular proteins of eukaryotic origin are often characterized by low folding yields upon expression in bacterial hosts (Baneyx, 1999; Baneyx and Mujacic, 2004). The results presented in this study suggest that many of these proteins rely on a mechanism of co-translational domain folding which is not efficiently supported by the bacterial folding machinery.

This insight suggests new strategies in developing bacterial strains for the production of native eukaryotic multi-domain proteins with high yield, including proteins of therapeutic interest. Specifically, a constructive biology approach would be applied to the Δ*tig*Δ*dnaKdnaJ E. coli* strain, which lacks the inhibitory components of co-translational folding, by introducing components of the eukaryotic chaperone machinery from yeast, such as RAC, Ssb and Ssa, those were shown to support co-translational folding of FL. In order for the yeast chaperones to productively interact with bacterial ribosomes, they could be engineered to contain the ribosome-binding domain of TF and thus be efficiently positioned next to the polypeptide exit site on the ribosome.

An alternative strategy that holds considerable promise in advancing the bacterial chaperone machinery to exert efficient co-translational folding is to subject the bacterial Hsp70 system to *in vitro* evolution. Evidence that this may work in principle was recently provided by an intriguing study by Weissman and coworkers (Wang *et al*., 2002) who showed that successive rounds of *in vivo* screening and DNA shuffling generated GroEL variants with a greatly enhanced capacity in folding certain proteins. A contribution of these strategies may prove useful in improving recombinant protein productions, thereby resolving a major bottleneck of large-scale synthetic genome programs.

6. References

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

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

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng *14*, 529-532.

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

Ayling, A., and Baneyx, F. (1996). Influence of the GroE molecular chaperone machine on the in vitro refolding of Escherichia coli beta-galactosidase. Protein Sci *5*, 478-487.

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

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

Baldwin, R. L., and Rose, G. D. (1999a). Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci *24*, 26-33.

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

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol *10*, 411-421.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol *22*, 1399-1408.

Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A., and Yonath, A. (2005). Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. Proc Natl Acad Sci U S A *102*, 12017-12022.

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.

Beatrix, B., Sakai, H., and Wiedmann, M. (2000). The alpha and beta subunit of the nascent polypeptide-associated complex have distinct functions. J Biol Chem *275*, 37838-37845.

Beck, K., Wu, L. F., Brunner, J., and Muller, M. (2000). Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. Embo J *19*, 134-143.

Becker, J., Walter, W., Yan, W., and Craig, E. A. (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). Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. J Struct Biol *135*, 84-93.

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.

Bremer, H., and Dennis, P. P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhart, F. C., ed.), p.p.1553-1569, ASM Press, Washington, DC.

Brodsky, J. L., and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. J Cell Biol *123*, 1355-1363.

Buchberger, A., Schroder, H., Hesterkamp, T., Schonfeld, H. J., and Bukau, B. (1996). Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. J Mol Biol *261*, 328-333.

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

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

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

Conti, E., Franks, N. P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure *4*, 287-298.

Craig, E. A., Weissman, J. S., and Horwich, A. L. (1994). Heat shock proteins and molecular chaperones: mediators of protein conformation and turnover in the cell. Cell *78*, 365-372.

Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol *14*, 315-319.

Cummings, J. L. (1998). Alzheimer's disease management. J Clin Psychiatry *59 Suppl 13*, 4-5.

Cyr, D. M., Langer, T., and Douglas, M. G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends Biochem Sci *19*, 176-181.

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

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.

Dill, K. A., and Chan, H. S. (1997). From Levinthal to pathways to funnels. Nat Struct Biol *4*, 10-19.

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

Dobson, C. M. (2003). Protein folding and misfolding. Nature *426*, 884-890.

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

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

Doyle, T., and Botstein, D. (1996). Movement of yeast cortical actin cytoskeleton visualized in vivo. Proc Natl Acad Sci U S A *93*, 3886-3891.

Ekman, D., Bjorklund, A. K., Frey-Skott, J., and Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. J Mol Biol *348*, 231-243.

Ellis, R. J. (1994). Molecular chaperones. Opening and closing the Anfinsen cage. Curr Biol *4*, 633-635.

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

Ellis, R. J., and van der Vies, S. M. (1991). Molecular chaperones. Annu Rev Biochem *60*, 321-347.

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.

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.

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.

Feldman, D. E., Thulasiraman, V., Ferreyra, R. G., and Frydman, J. (1999). Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. Mol Cell *4*, 1051-1061.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature *431*, 590-596.

Flaherty, K. M., DeLuca-Flaherty, C., and McKay, D. B. (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature *346*, 623-628.

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

Fonte, V., Kapulkin, V., Taft, A., Fluet, A., Friedman, D., and Link, C. D. (2002). Interaction of intracellular beta amyloid peptide with chaperone proteins. Proc Natl Acad Sci U S A *99*, 9439-9444.

Freund, C., Ross, A., Guth, B., Pluckthun, A., and Holak, T. A. (1993). Characterization of the linker peptide of the single-chain Fv fragment of an antibody by NMR spectroscopy. FEBS Lett *320*, 97-100.

Frydman, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1999). Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. Nat Struct Biol *6*, 697-705.

Frydman, J., and Hartl, F. U. (1996). Principles of chaperone-assisted protein folding: differences between in vitro and in vivo mechanisms. Science *272*, 1497-1502.

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

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

Fukuda, H., Arai, M., and Kuwajima, K. (2000). Folding of green fluorescent protein and the cycle3 mutant. Biochemistry *39*, 12025-12032.

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

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

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

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

George, R., Beddoe, T., Landl, K., and Lithgow, T. (1998). The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria in vivo. Proc Natl Acad Sci U S A *95*, 2296-2301.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

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

Greer, L. F., 3rd, and Szalay, A. A. (2002). Imaging of light emission from the expression of luciferases in living cells and organisms: a review. Luminescence *17*, 43-74.

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.

Hansen, W. J., Cowan, N. J., and Welch, W. J. (1999). Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. J Cell Biol *145*, 265-277.

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.

Heim, R., Prasher, D. C., and Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A *91*, 12501-12504.

Hesterkamp, T., and Bukau, B. (1998). Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in E.coli. Embo J *17*, 4818-4828.

Hesterkamp, T., Deuerling, E., and Bukau, B. (1997). The amino-terminal 118 amino acids of Escherichia coli trigger factor constitute a domain that is necessary and sufficient for binding to ribosomes. J Biol Chem *272*, 21865-21871.

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

Hohfeld, J., Minami, Y., and Hartl, F. U. (1995). Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle. Cell *83*, 589-598.

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.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

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

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

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

Hynes, G. M., and Willison, K. R. (2000). Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of beta-actin. J Biol Chem *275*, 18985-18994.

Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994). Three-dimensional structure of beta-galactosidase from E. coli. Nature *369*, 761-766.

Jaenicke, R. (1991). Protein folding: local structures, domains, subunits, and assemblies. Biochemistry *30*, 3147-3161.

Jana, N. R., Tanaka, M., Wang, G., and Nukina, N. (2000). Polyglutamine length-dependent interaction of Hsp40 and Hsp70 family chaperones with truncated N-terminal huntingtin: their role in suppression of aggregation and cellular toxicity. Hum Mol Genet *9*, 2009-2018.

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

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

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

Kelley, W. L. (1998). The J-domain family and the recruitment of chaperone power. Trends Biochem Sci *23*, 222-227.

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

Kim, S., Schilke, B., Craig, E. A., and Horwich, A. L. (1998). Folding in vivo of a newly translated yeast cytosolic enzyme is mediated by the SSA class of cytosolic yeast Hsp70 proteins. Proc Natl Acad Sci U S A *95*, 12860-12865.

Kim, S., Willison, K. R., and Horwich, A. L. (1994). Cystosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. Trends Biochem Sci *19*, 543-548.

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

Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. Yeast *15*, 963-972.

Kolb, V. A., Makeyev, E. V., and Spirin, A. S. (1994). Folding of firefly luciferase during translation in a cell-free system. Embo J *13*, 3631-3637.

Kuhnel, K., and Luisi, B. F. (2001). Crystal structure of the Escherichia coli RNA degradosome component enolase. J Mol Biol *313*, 583-592.

Kummerfeld, S. K., and Teichmann, S. A. (2005). Relative rates of gene fusion and fission in multi-domain proteins. Trends Genet *21*, 25-30.

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

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

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

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

Lill, R., Crooke, E., Guthrie, B., and Wickner, W. (1988). The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. Cell *54*, 1013-1018.

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

Lu, Z., and Cyr, D. M. (1998). Protein folding activity of Hsp70 is modified differentially by the hsp40 co-chaperones Sis1 and Ydj1. J Biol Chem *273*, 27824-27830.

Luders, J., Demand, J., and Hohfeld, J. (2000). The ubiquitin-related BAG-1 provides a link between the molecular chaperones Hsc70/Hsp70 and the proteasome. J Biol Chem *275*, 4613-4617.

Maier, R., Eckert, B., Scholz, C., Lilie, H., and Schmid, F. X. (2003). Interaction of trigger factor with the ribosome. J Mol Biol *326*, 585-592.

Malkin, L. I., and Rich, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. J Mol Biol *26*, 329-346.

Mathews, M. B., Sonenberg, N., and Hershey, J. W. B. (2000). Origins and priciples of translational control. Translational Control of Gene Expression (Sonenberg, N., Hershey, J.W.B. and Mathews, M.B., eds.), pp.1-31, Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mayer, M. P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci *62*, 670-684.

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

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

McCarty, J. S., Buchberger, A., Reinstein, J., and Bukau, B. (1995). The role of ATP in the functional cycle of the DnaK chaperone system. J Mol Biol *249*, 126-137.

McNaught, K. S., Shashidharan, P., Perl, D. P., Jenner, P., and Olanow, C. W. (2002). Aggresome-related biogenesis of Lewy bodies. Eur J Neurosci *16*, 2136-2148.

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

Minami, Y. (1996). [Structure and function of the DnaJ/Hsp40 family]. Tanpakushitsu Kakusan Koso *41*, 875-882.

Minton, A. P. (2000). Implications of macromolecular crowding for protein assembly. Curr Opin Struct Biol *10*, 34-39.

Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K., and Hartl, F. U. (2000). Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. Proc Natl Acad Sci U S A *97*, 7841-7846.

Mumberg, D., Muller, R., and Funk, M. (1994). Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res *22*, 5767-5768.

Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119-122.

Nelson, R. J., Ziegelhoffer, T., Nicolet, C., Werner-Washburne, M., and Craig, E. A. (1992). The translation machinery and 70 kd heat shock protein cooperate in protein synthesis. Cell *71*, 97-105.

Nett, J. H., Hunte, C., and Trumpower, B. L. (2000). Changes to the length of the flexible linker region of the Rieske protein impair the interaction of ubiquinol with the cytochrome bc1 complex. Eur J Biochem *267*, 5777-5782.

Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature *388*, 343-349.

Nichtl, A., Buchner, J., Jaenicke, R., Rudolph, R., and Scheibel, T. (1998). Folding and association of beta-Galactosidase. J Mol Biol *282*, 1083-1091.

Nicola, A. V., Chen, W., and Helenius, A. (1999). Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. Nat Cell Biol *1*, 341-345.

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.

Otto, H., Conz, C., Maier, P., Wolfle, T., Suzuki, C. K., Jeno, P., Rucknagel, P., Stahl, J., and Rospert, S. (2005). The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. Proc Natl Acad Sci U S A *102*, 10064-10069.

Patzelt, H., Rudiger, S., Brehmer, D., Kramer, G., Vorderwulbecke, S., Schaffitzel, E., Waitz, A., Hesterkamp, T., Dong, L., Schneider-Mergener, J.*, et al.* (2001). Binding specificity of Escherichia coli trigger factor. Proc Natl Acad Sci U S A *98*, 14244-14249.

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

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

Pellecchia, M., Szyperski, T., Wall, D., Georgopoulos, C., and Wuthrich, K. (1996). NMR structure of the J-domain and the Gly/Phe-rich region of the Escherichia coli DnaJ chaperone. J Mol Biol *260*, 236-250.

Pfund, C., Lopez-Hoyo, N., Ziegelhoffer, T., Schilke, B. A., Lopez-Buesa, P., Walter, W. A., Wiedmann, M., and Craig, E. A. (1998). The molecular chaperone Ssb from Saccharomyces cerevisiae is a component of the ribosome-nascent chain complex. Embo J *17*, 3981-3989.

Pierpaoli, E. V., Sandmeier, E., Baici, A., Schonfeld, H. J., Gisler, S., and Christen, P. (1997). The power stroke of the DnaK/DnaJ/GrpE molecular chaperone system. J Mol Biol *269*, 757-768.

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.

Rommelaere, H., De Neve, M., Melki, R., Vandekerckhove, J., and Ampe, C. (1999). The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. Biochemistry *38*, 3246-3257.

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

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

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

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

Sanchez, I. E., Morillas, M., Zobeley, E., Kiefhaber, T., and Glockshuber, R. (2004). Fast folding of the two-domain semliki forest virus capsid protein explains co-translational proteolytic activity. J Mol Biol *338*, 159-167.

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

Schlunzen, F., Wilson, D. N., Tian, P., Harms, J. M., McInnes, S. J., Hansen, H. A., Albrecht, R., Buerger, J., Wilbanks, S. M., and Fucini, P. (2005). The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. Structure (Camb) *13*, 1685-1694.

Sha, B., Lee, S., and Cyr, D. M. (2000). The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1. Structure *8*, 799-807.

Sherman, M. Y., and Goldberg, A. L. (2001). Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron *29*, 15-32.

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

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

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

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

Sikorski, R. S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Sondermann, H., Ho, A. K., Listenberger, L. L., Siegers, K., Moarefi, I., Wente, S. R., Hartl, F. U., and Young, J. C. (2002). Prediction of novel Bag-1 homologs based on structure/function analysis identifies Snl1p as an Hsp70 co-chaperone in Saccharomyces cerevisiae. J Biol Chem *277*, 33220-33227.

Spedding, G. (1990). Isolation and analysis of ribosomes from prokaryotes, eukaryotes, and organelles. In Ribosomes and protein synthesis: a practical approach, G. Spedding, ed. (Oxford, Oxford University Press), pp. 1-29.

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

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

Suh, W. C., Lu, C. Z., and Gross, C. A. (1999). Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. J Biol Chem *274*, 30534-30539.

Szabo, A., Korszun, R., Hartl, F. U., and Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. Embo J *15*, 408-417.

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

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol *60*, 523-533.

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.

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

Tian, G., Vainberg, I. E., Tap, W. D., Lewis, S. A., and Cowan, N. J. (1995). Specificity in chaperonin-mediated protein folding. Nature *375*, 250-253.

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.

Tsien, R. Y. (1998). The green fluorescent protein. Annu Rev Biochem *67*, 509-544.

Ungewickell, E., Ungewickell, H., Holstein, S. E., Lindner, R., Prasad, K., Barouch, W., Martin, B., Greene, L. E., and Eisenberg, E. (1995). Role of auxilin in uncoating clathrin-coated vesicles. Nature *378*, 632-635.

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

van den Ent, F., Amos, L. A., and Lowe, J. (2001). Prokaryotic origin of the actin cytoskeleton. Nature *413*, 39-44.

Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *10*, 1793-1808.

Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol *17*, 691-695.

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, S., Sakai, H., and Wiedmann, M. (1995). NAC covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. J Cell Biol *130*, 519-528.

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem Sci *24*, 437-440.

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

Weissman, J. S., 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.

Werner-Washburne, M., Stone, D. E., and Craig, E. A. (1987). Complex interactions among members of an essential subfamily of hsp70 genes in Saccharomyces cerevisiae. Mol Cell Biol *7*, 2568-2577.

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

Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R., and Wahl, M. C. (2001). An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol Cell *7*, 1177-1189.

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

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

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

Young, J. C., Hoogenraad, N. J., and Hartl, F. U. (2003). Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. Cell *112*, 41-50.

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.

7. Appendices

7.1. Abbreviations

ADD	adenosine deaminase
ADP	adenosine 5'-diphosphate
ALR2	alanine racemase, catabolic
Amp	ampicillin
APS	ammonium peroxodisulfate
ATP	adenosine 5'-triphosphate
BSA	albumin bovine serum
β-gal	β-galactosidase
CAM	chloramphenicol
CDTA	*trans*-1,2-diaminocyclohexane-*N,N,N',N'*-tetraacetic acid
DAPA	dihydrodipicolinate synthase (DHDPS)
DCEA	glutamate decarboxylase alpha
DHFR	Dihydrofolate reductase
DNA	deoxyribonucleic acid
DnaJ	bacterial Hsp40 chaperone
DnaK	bacterial Hsp70 chaperone
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
END4	endonuclease IV
ENO	enolase (2-phosphoglycerate dehydratase)
FL	Firefly luciferase
g	acceleration of gravity, 9.81 m/s^2
GATD	galactitol-1-phosphate 5-dehydrogenase
GATY	tagatose-1,6-bisphosphate aldolase gatY (TBPA)
GdnHCl	guanidinium hydrochloride
GFP	Green fluorescent protein

GroEL	bacterial Hsp60 chaperonin
GroES	bacterial Hsp10 cochaperonin
GrpE	bacterial nucleotide exchange factor of DnaK
HEM2	delta-aminolevulinic acid dehydratase
h	hour
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	Luria Bertani
LLDD	L-lactate dehydrogenase
MBP	Maltose binding protein
METF	5,10-methylenetetrahydrofolate reductase
METK	S-adenosylmethionine synthetase
MOPS	3-(N-morpholino)propanesulfonic acid
NAC	nascent chain-associate complex
NADPH	β-nicotinamide adenine dinucleotide 2'-phosphate
NANA (NPL)	N-acetylneuraminate lyase
NTA	nitrilo-triacetic acid
OAc	acetate
OD	optical density
OmpA	outer membrane protein A
PAGE	PolyAcrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank. Repository for processing and distribution of 3-D structure data of proteins and nucleic acids. http://www.rcsb.org/pdb/
PPIase	prolyl- *cis/trans* isomerase
RAC	ribosome-associate complex
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodiumdodecylsulfate
SFVP	Semliki Forest virus protease
SYT	threonyl-tRNA synthetase (ThrRS)

TEMED	*N*,*N*,*N'*,*N'*-tetramethylethylenediamine
TF	trigger factor
TRiC	Tailless complex polypeptide ring complex
Tris HCl	tris(hydroxymethyl)aminomethane hydrochloride
XYLA	xylose isomerase
YAJO	hypothetical oxidoreductase yajO
Ydj1p	Yeast DnaJ like protein 1

7.2. Publications

Journal articles

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.

*: Equal contribution

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

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

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

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

Kaiser, C. M., **Chang, H. C**., Agashe, V. R., Lakshmipathy, S. K., Etchells, S. A., Hayer-Hartl, M., Hartl, F. U., and Barral, M. (2006). Real-time observation of trigger factor function on translating ribosomes. Nature *444*, 455-460.

Poster

Chang, Hung-Chun., Barral, J.M., and Hartl, F-U. 2003. Analysis of protein folding in Eukaryotes vs. Prokaryotes. [Poster Presentation in EURESCO and FEBS conference "Biology of Molecular Chaperones" in Tomar/ Portugal (August 2003)].

7.3. *Curriculum vitae*

Name	Hung-Chun Chang
Birth date	31, October, 1972
Birth place	Kaohsiung, Taiwan
Nationality	Taiwan, Republic of China

Dissertation

2002-present	Ph.D. dissertation directed by Prof. Dr. F. Ulrich Hartl in Department of Cellular Biochemistry, Max-Planck Institute of Biochemistry, Martinsried: *Mechanisms of de novo muti-domain protein folding in bacteria and eukaryotes*.

Education

1994-1997	Master of Science. Master thesis directed by Prof. Dr. Jenn Tu in Institute of Microbiology, Soochow University, Taipei, Taiwan: *Resolvase mediated recombination in Xanthomonas transposable element ISXc5 in vivo.*
1990-1994	Bachelor of Science. Department of Microbiology, Soochow University, Taipei, Taiwan.
1987-1990	School Leaving Certificate. Affiliate Senior High School of National Taiwan Normal University, Taipei, Taiwan.
1984-1987	School Leaving Certificate. Min-Seng Junior High School, Taipei, Taiwan.
1978-1984	Affiliate Elementary School of National Kaohsiung Normal University, Kaohsiung, Taiwan.

Experience

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

Honors

1994-1996	Full scholarship awarded by the Ministry of Education, Taiwan.