

Mechanisms of Protein Folding: Molecular Chaperones and Their Application in Biotechnology**

Axel Mogk,^[a,b] Matthias P. Mayer,^[a] and Elke Deuerling^{*[a]}

KEYWORDS:

aggregation · molecular chaperones · protein folding · recombinant protein production · trigger factor

1. Introduction

Protein folding is the process that leads from the linear amino acid sequence of a polypeptide chain to a defined spatial structure characteristic for the native protein. In 1961 Anfinsen showed that all the information needed for reaching the active 3D structure is encoded in the protein's amino acid sequence.^[1] Protein folding is initiated by collapse of the polypeptide chain, which is driven by the "desire" of hydrophobic amino acids to escape the polar solvent water. The number of theoretical possible but wrong conformations of a protein is incredibly high: even a small protein consisting of 100 amino acids can adopt 10^{30} different structures. Reaching the correct conformation by accidental scanning of all possible conformations would take 10^{11} years (assuming 1 ps per conformation). Protein folding is however a very fast process within the range of seconds or even milliseconds. How does a protein find the needle in the haystack—the native structure among all the theoretical possible conformations?

2. Energy Landscapes and Folding Funnels

Initially it was proposed that proteins have to pass along specific folding pathways to find the native structure within a very short time. Meanwhile proteins are known to reach their correct conformation from many different unfolded states. The mechanism of protein folding is therefore currently described by folding funnels and energy landscapes (Figure 1). The native structure represents the thermodynamically most favored status at the bottom of the funnel. All nonnative conformations of a protein possess a higher energy, and protein folding can be imagined as a ball rolling down the slope of the energy landscape. A folding funnel represents many different folding pathways that can be used by the unfolded protein to reach the energy minimum (native structure) quickly or slowly, depending on the shape and the gradient of the funnel.^[2]

The folding of proteins is an error-prone process, especially in the case of large proteins built up of several domains. In the course of the folding process intermediates may accumulate that have hydrophobic amino acid side chains still exposed; these can serve as sticky surfaces and promote aggregation. The

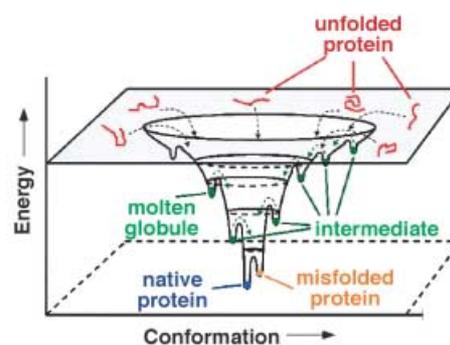


Figure 1. Folding funnels. A folding funnel represents the free energy of all potential protein structures as a function of the possible conformations. Different unfolded species of a protein "roll down" the surface of the folding funnel. A folding funnel contains many local minima, in which a protein can fall during the folding process. Some local minima represent productive folding intermediates, with a stable and nativelike structure (molten globules), while others act like a trap and keep the proteins in a nonnative state. Reproduced from ref. [2] with permission. Copyright (2000) Nature America Inc. (<http://www.nature.com/nsb>).

higher the concentration of such an aggregation-prone intermediate, the more likely it is that protein aggregation occurs.

Some proteins additionally require the correct isomeric state of a peptide bond between a proline and a neighboring amino acid. Moreover, the correct formation of disulfide bridges is especially important for secretory proteins containing cysteines. Folding intermediates with wrong conformations of peptidyl–prolyl bonds or wrong disulfide bonds are also in danger of aggregating. Inside the cell, both of these folding reactions are therefore catalyzed by specific enzymes: PPLases and protein disulfide isomerases. More precise information on these topics can be found in comprehensive reviews.^[3, 4]

[a] Dr. E. Deuerling, Dr. A. Mogk, Dr. M. P. Mayer
Zentrum für Molekulare Biologie (ZMBH)
Im Neuenheimer Feld 282, 69120 Heidelberg (Germany)
Fax: (+49) 6221-54-5894
E-mail: e.deuerling@zmbh.uni-heidelberg.de

[b] Dr. A. Mogk
Institut für Biochemie und Molekularbiologie
Hermann-Herder-Strasse 7, 79104 Freiburg (Germany)

[**] This Minireview is a translation of an article that previously appeared in German in *Unserer Zeit*, 2001, 31, 182–192.

3. Folding inside the Cell

Folding processes inside the cell are much more complex than the refolding of denatured model proteins in the test tube. Proteins are synthesized *in vivo* at the ribosomes in a vectorial manner from the N to the C terminus (Figure 2). In contrast to in

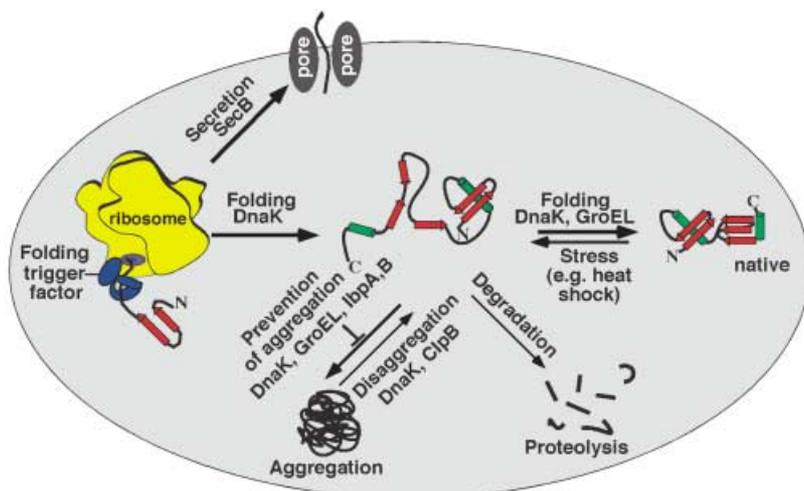


Figure 2. Life cycle of proteins inside the cell. Proteins are synthesized by the ribosomes and subsequently they are either secreted or they fold into their native structure. During these stages proteins are in danger of undergoing misfolding and aggregation. Even when proteins have already reached their native conformation they are vulnerable to misfolding due to Brownian motion, especially under heat stress. Molecular chaperones (the prokaryotic homologues are shown here) control the folding status throughout the whole lifespan of the proteins. Possible secondary structure elements in proteins are shown in red (β strands) and green (α helices).

in vitro experiments, which were performed with diluted ($< 1 \text{ g L}^{-1}$ protein) proteins and complete polypeptide chains, nascent polypeptides emerging from the ribosome do not contain the complete information necessary for folding. In addition, the cytosolic concentration of macromolecules including ribosomes, nucleic acids and proteins, is enormously high (340 g L^{-1} ; Figure 3).^[5] In this crowded macromolecular environment exposed hydrophobic amino acids of nascent polypeptides and folding intermediates may interact inappropriately leading to misfolding and aggregation (Figure 2). While, as a consequence

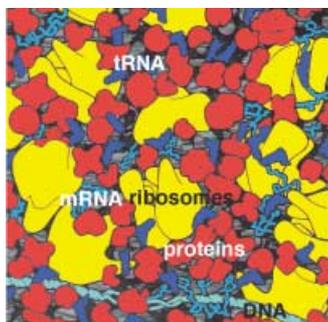


Figure 3. Molecular crowding. The cartoon illustrates the high density of macromolecules (ribosomes, proteins, RNA and DNA) in the cytosol. In this environment proteins have to fold and maintain their native structures. Derived from ref. [5] with permission. Copyright (1991) Elsevier Science.

of this problem, cytosolic proteins should fold as fast as possible, the folding of proteins destined for another compartment has to be delayed in order to allow translocation through the membrane in the unfolded state.

Due to Brownian motion and thermal vibrations even native proteins are always in danger of spontaneously unfolding and losing their active structure. This feature of proteins is probably the evolutionary price for conformational flexibility, which is essential for protein function. For most proteins there are only small energy barriers between the native and the misfolded state. A number of proteins are specifically thermolabile and their folding status is even more susceptible to changes in the cellular environment. Stress conditions, like a sudden increase in temperature, can therefore lead to unfolding, aggregation, or degradation of many proteins (Figure 2).

4. Molecular Chaperones Control Protein Folding

To optimize cellular protein folding, protective systems have developed in the course of evolution. These systems consist of families of highly conserved proteins, the so-called molecular chaperones. Chaperones are found in high concentrations in all cells, from bacteria to humans. They guide a large variety of folding processes throughout the life cycle of proteins, from synthesis to degradation (Figure 2).^[6, 7] For example,

they assist the *de novo* folding of proteins or they form repair machines for misfolded or even aggregated proteins, and they are therefore especially important for the survival of cells during stress situations. Since heat shock can induce the synthesis of many chaperones, those are also called heat shock proteins (Hsps). The name of each of the chaperone families is derived from the molecular weight of the corresponding main representative (for example, Hsp70—a protein with a molecular weight of 70 kDa). Table 1 summarizes the structure and function of the most important chaperone families and lists some of their prokaryotic and eukaryotic members, as well as their co-chaperones (regulatory and cooperating proteins). In the following sections the functions of chaperones from *Escherichia coli*, a well-understood model organism, are described.

De novo protein folding

Newly synthesized proteins in bacteria associate with the chaperone trigger factor, as soon as they leave the exit tunnel of the ribosome.^[8, 9] The trigger factor's high degree of conservation within the bacterial kingdom underlines its functional importance. So far the trigger factor has not been found in the cytosol of eukaryotic cells, but it is assumed that other proteins, like the ribosome-associated protein complex NAC, can take over trigger factor's function. Exposed hydrophobic side chains of the newly synthesized polypeptides are most likely protected from wrong

Table 1. Chaperone families: Structure and function					
Chaperone family	Structure	ATP	Examples prokaryote eukaryote	Co-chaperone	Function
Hsp100		+	ClpB		disaggregation together with Hsp70
			ClpA		proteolysis together with the ClpP protease
				Hsp104	thermotolerance disaggregation together with Hsp70
Hsp90		+	HtpG		tolerance to extreme heat shock
				Hsp90	Hop, p23, CDC37
Hsp70		+	DnaK		de novo protein folding prevention of aggregation of heat-denatured proteins solubilization of protein aggregates together with ClpB regulation of the heat shock response
				Hsp70, Hsc70	Hsp40, Bag1, Hip, Chip, Hop, HspBP1
Hsp60		+	GroEL		de novo protein folding prevention of aggregation of heat-denatured proteins
				CCT/TRiC	prefoldin
sHsp			lbpA, lbpB		prevention of aggregation of heat-denatured proteins binding to inclusion bodies
				Hsp25 (crystalline)	
trigger factor	monomer 		trigger factor		ribosome-associated potential function in de novo protein folding
NAC	heterodimer 		NAC		ribosome-associated potential function in de novo protein folding
SecB	tetramer 		SecB		protein secretion

interactions by association with the trigger factor. In addition, the trigger factor may prevent premature folding until a complete domain has emerged from the ribosomal exit site.^[10]

It is assumed that after release from the trigger factor the newly synthesized polypeptide chain can start or continue its folding into the native state. The speed of the folding process and the yield of correctly folded proteins differ from protein to protein. Small proteins can, in general, fold quickly and without guidance of any additional chaperone into their native structure. Large, multidomain proteins usually fold slowly and are often trapped in nonnative and aggregation-prone conformations. Such proteins are substrates for the chaperones of the Hsp70 (DnaK) or Hsp60 (GroEL) protein families. These chaperones support folding into the native state by cycles of binding and release of the protein substrates.^[11–13] In contrast to the GroEL “machine”, which is supposed to associate exclusively with newly synthesized proteins after their release from the ribosome, DnaK associates with both nascent polypeptides and polypeptides released from the ribosome.^[8, 9, 14, 15]

E. coli proteins destined for secretion are often bound by the specialized chaperone SecB and transported to the export machinery at the plasma membrane. Since native proteins cannot pass the translocation channel due to their compact structures, it is assumed that binding to SecB keeps the secretory proteins in an export-competent and partially unfolded state suitable for translocation. Integral membrane proteins are specifically bound by the signal recognition particle (SRP) during their synthesis and are guided to the SRP receptor at the plasma membrane.^[16]

Chaperones repair misfolded proteins

Stress conditions like a sudden increase in temperature can cause proteins to unfold and aggregate. Different chaperone systems function in the cell as a protective system to prevent protein aggregation by binding to the misfolded proteins. Examples of this are the DnaK (Hsp70) and GroEL (Hsp60) systems, both working together with regulatory co-chaperones, and the small heat shock proteins lbpAB (sHsps). Interestingly,

small heat shock proteins are also found within large protein aggregates (inclusion bodies), which are often generated during the overproduction of heterologous proteins in *E. coli* (see Section 6). Since sHsps cannot promote the refolding of misfolded proteins into the native state by themselves, it was proposed that they serve as a buffering system binding to the unfolded proteins during stress situations and transferring them after the return to optimal growth conditions to the Hsp60 or Hsp70 chaperones for refolding.^[17, 18] Circumstantial evidence from a number of different laboratories indicates that the diverse chaperone classes form a functional network in which they either compete for binding to unfolded proteins and/or cooperate during protein refolding.

The importance of molecular chaperones can be demonstrated by analyzing the phenotypes of *E. coli* cells missing a specific chaperone system (for example, Hsp70). *E. coli* cells lacking the Hsp70 chaperone DnaK accumulate large amounts of aggregated proteins at high temperatures.^[19] Figure 4 illustrates that more than 200 different proteins aggregate in these cells at 42 °C. This pronounced protein aggregation is linked to the inability of the mutant strain to grow at elevated temperatures.

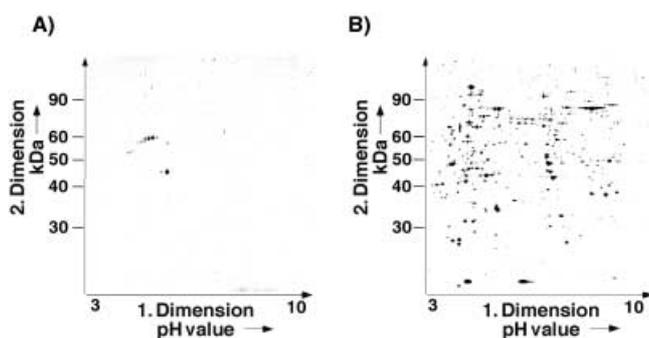


Figure 4. Cellular protein aggregation. Aggregated proteins isolated from wild type (A) and $\Delta dnaK$ cells (B) after heat shock (42 °C) were separated by two dimensional gel electrophoresis. Each spot represents an aggregated protein species and more than 200 spots are detectable in cells lacking the DnaK chaperone.

Refolding of aggregated proteins

For a long time protein aggregation was thought to be a dead end in the life of proteins, comparable with a boiled egg that cannot be unboiled again. New studies with the model systems baker's yeast (*Saccharomyces cerevisiae*), the bacterium *E. coli*, and the plant *Arabidopsis thaliana*, however, demonstrated that preexisting protein aggregates can be successfully disaggregated and even refolded to the active conformation.^[19–22] Disaggregation of protein aggregates was demonstrated in vivo and in the test tube by using thermolabile malate dehydrogenase (MDH) as a reporter enzyme.^[23] Large aggregates of MDH could be resolubilized in vitro and MDH was refolded afterwards into its native structure. So in theory, a boiled egg, which consists of denatured proteins, can be unboiled again. Protein disaggregation is achieved by a bichaperone system consisting of ClpB (Hsp104) and the DnaK (Hsp70) machine. Importantly, only the

combination of chaperone systems is active in resolubilization and refolding of aggregated proteins. This activity of the bichaperone system is directly linked to the survival of the mentioned organisms at very high temperatures.^[22, 24]

5. Molecular Mechanism of Chaperones

According to our current knowledge most chaperones bind to one or several short peptide segments that are enriched in hydrophobic amino acids. In native proteins such peptide segments are found in the interior, the hydrophobic core. Nascent or misfolded polypeptides expose these segments and are therefore bound by chaperones. Thereby the association of several segments of different polypeptides, which would otherwise lead to aggregation, is prevented. This activity of chaperones is called "holder" function and can be exerted independently of adenosine triphosphate (ATP). Refolding into the active 3D structure requires the "folder" function, which is ATP-dependent. Two chaperone machines acting as "folders" have already been well studied: Hsp60 and Hsp70. Still relatively little is known about the structure and function of the other chaperone families listed in Table 1. In the following sections the molecular mechanism for a member of each of the Hsp60 and Hsp70 chaperone families, GroEL and DnaK of *E. coli*, will be described.

The Hsp60 chaperone

The barrel shaped GroEL chaperone that is found in all prokaryotes, mitochondria, and chloroplasts consists of two stacked rings of seven subunits each.^[12, 13] Figure 5 shows the reconstruction of the GroEL structure from cryoelectron microscopic pictures and the chaperone cycle that is deduced from biochemical and structural investigations.^[25, 26] Each of the two GroEL rings forms a cavity in which misfolded polypeptides bind. Both rings fold substrates with a phase shift (Figure 5B, green substrate, I→VI; blue substrate, IV→III) and each cycle takes about 15 seconds. This mechanism can be compared to a two-stroke engine.

The folding cycle for the green substrate starts with the association to the open ring (cis ring, I). After binding of ATP (II) the substrate is enclosed in a "folding chamber" by the association of the GroES co-chaperone, which also consists of seven subunits and functions as a lid (III). The substrate is thereby prevented from interacting with proteins in solution, analogously to an infinite dilution. ATP and GroES binding and subsequent cooperative ATP hydrolysis in all GroEL subunits of the cis ring induce conformational changes in the cis ring leading to an enlargement of the folding chamber whereby the substrate binding sites of GroEL (red circles in Figure 5B) move away from the substrate (II, III, IV). This movement was shown to cause a global unfolding of the bound polypeptide.^[27] It is assumed that unproductive hydrophobic interactions are thereby broken and the protein gets a new chance to fold correctly. Consequently, the protein substrate is lifted out of a local energy minimum of the folding funnel (see Figure 1). Binding of ATP, GroES, and substrate to the subunits of the idle ring (trans ring) induces the dissociation of the GroES co-chaperone of the cis

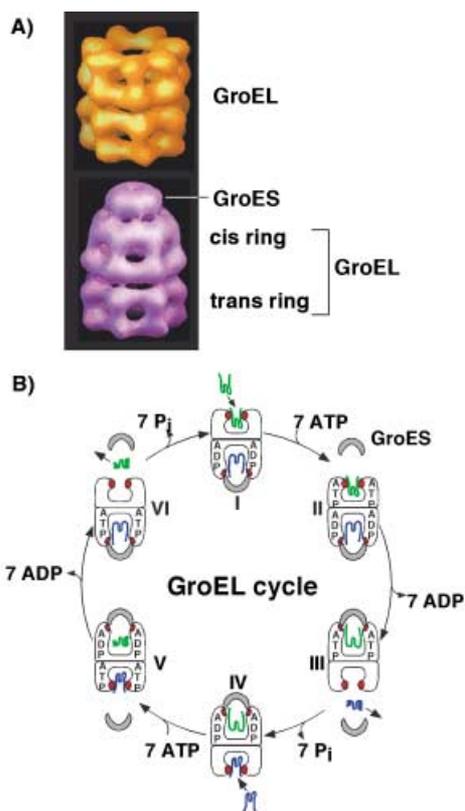


Figure 5. A) Reconstruction of the GroEL structure with and without the GroES "lid" from cryoelectron microscopy pictures. Reproduced from ref. [25] with permission. Copyright (1996) Cell Press. A video of GroEL – GroES can be found on Dr. H. Saibil's internet page (<http://www.cryst.bbk.ac.uk/~ubcg16z/chaperone.html>). B) Model of the GroEL chaperone cycle. Two misfolded proteins (green and blue) are simultaneously folded in a phase-shifted manner. The red circles symbolize the hydrophobic substrate binding sites of GroEL.

ring and the release of the substrate (IV, V, VI). Some proteins fold while still enclosed in the GroEL folding chamber (Figure 5B, green substrate), others only after their release from GroEL (Figure 5B, blue substrate). Many proteins need several binding and release cycles before they reach their native structure.

One obvious problem is the size limitation of the folding chamber of Hsp60 chaperones, which accommodates only polypeptides smaller than 60 kDa. It is difficult to imagine how Hsp60 could assist the folding of larger polypeptides, which mostly consist of several domains. However, exceptions might be possible since a recent paper reports that GroEL assists the folding of yeast aconitase, a 82 kDa protein.^[28]

The Hsp70 chaperone

In contrast to GroEL, DnaK does not enclose its substrate completely but only binds to a single short peptide segment of about five amino acids. DnaK consists of an N-terminal ATPase domain and a C-terminal substrate binding domain (Figure 6A, B). Substrate binding and release is controlled by ATP hydrolysis to adenosine diphosphate (ADP) and ADP/ATP exchange processes, which are regulated by the co-chaperones DnaJ and

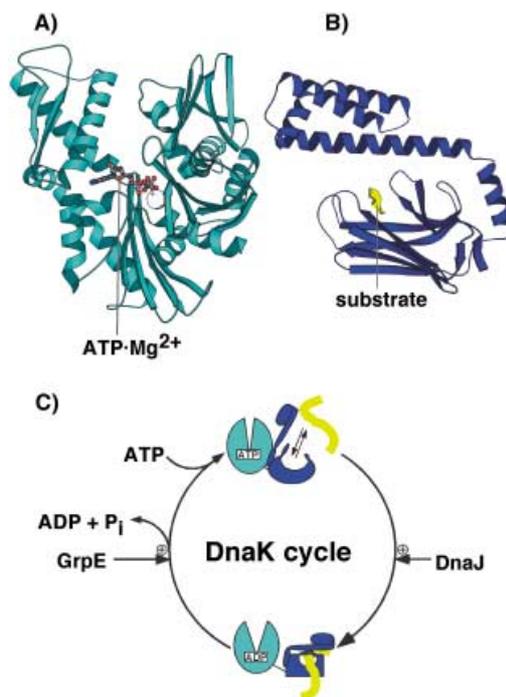


Figure 6. A) Secondary structure representation of the ATPase domain of DnaK in complex with ATP and Mg²⁺ (homology model of DnaK on Hsc70; Swiss-Prot AC code PO4475_C00001).^[37] B) Secondary structure representation of the substrate binding domain of DnaK in complex with a substrate peptide (PDB file 1DKX).^[38] C) Schematic representation of the ATPase cycle of DnaK. In the ATP state substrate is bound and released at high rates. In the ADP state the substrate is locked into the substrate binding domain. The transition from the low affinity to the high affinity conformation is effected by ATP hydrolysis, which is triggered by the simultaneous interaction with the substrate and the DnaJ co-chaperone. ADP/ATP exchange, which controls the lifetime of the DnaK substrate complex, is regulated by the GrpE co-chaperone.

GrpE. The chaperone cycle of DnaK is shown in Figure 6C. In the ATP state substrates associate to and dissociate from DnaK with high rates but affinity of DnaK to substrates is low. In the ADP state DnaK binds substrates with high affinity but substrate exchange rates are low. The DnaJ co-chaperone, which by itself also interacts with substrates through hydrophobic interactions, stimulates the ATPase activity of DnaK. Experimental evidence suggests that DnaJ "recognizes" the substrate and passes it over to DnaK whereby ATP is hydrolyzed and a segment of the substrate is tightly enclosed.^[13]

Under physiological conditions substrate dissociation is controlled by ADP/ATP exchange that, in turn, is stimulated by the GrpE co-chaperone. The cycles of substrate binding and release are, according to recent calculations, very short (about 1 second per cycle) and most likely have to be repeated many times until a protein is refolded and not any longer recognized as a substrate. However, it is not yet clear how these binding cycles promote the refolding of misfolded proteins. According to one hypothesis DnaK unfolds the substrate locally in contrast to the global unfolding by GroEL.^[29] This mechanism evidently has the big advantage that it is independent of the size of the misfolded protein substrate. It was shown that proteins of all sizes are found among the Hsp70 substrates and that proteins larger than 60 kDa are especially dependent on Hsp70 assistance during their folding.^[8, 19]

6. New Prospects in Biotechnology

Living cells contain a defined and regulatable arsenal of chaperones to prevent unproductive folding reactions and to increase the yield of properly folded native proteins. Misfolding and aggregation, however, happen frequently when recombinant genes are over-expressed in artificial cellular systems which are genetically engineered for large-scale protein production. The cellular folding capacity is overloaded; this results in the formation of biologically inactive, incorrectly folded proteins. The fate of the unfolded proteins differs: unfolded recombinant proteins can be subsequently degraded by the proteolytic machinery of the cell or are deposited into biologically inactive large aggregates known as inclusion bodies (Figure 7 and 8A). One possible strategy to solve these problems is to elevate the levels of molecular chaperones.

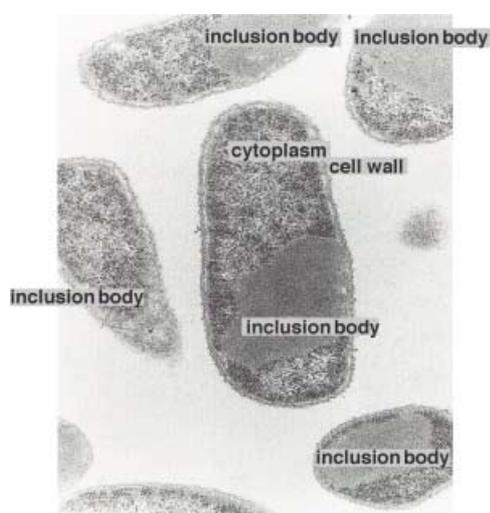


Figure 7. Bacterial cells containing inclusion bodies as visualized by electron microscopy. The recombinant protein overproduced in bacterial cells cannot fold correctly and forms large insoluble aggregates.

Protein production in biotechnology

The large-scale production of recombinant proteins in biotechnology has received fresh impetus over the past few years. Therapeutics like vaccines, growth factors, and hormones as well as diagnostic agents and proteins used in scientific research (for example, antibodies and DNA- or RNA-modifying enzymes) are of particular interest in biotechnology. The goal is to obtain the maximum yield of the properly folded active protein. Why are the majority of proteins synthesized in heterologous systems? Many proteins could not be isolated in large quantities from the organism they originate from, for example, human insulin. Very often recombinant production is the only way to achieve sufficient amounts. For that purpose, the genetic information encoding the protein is transferred into a host organism. The host cell is genetically engineered in a way that allows the large-scale production of the heterologous

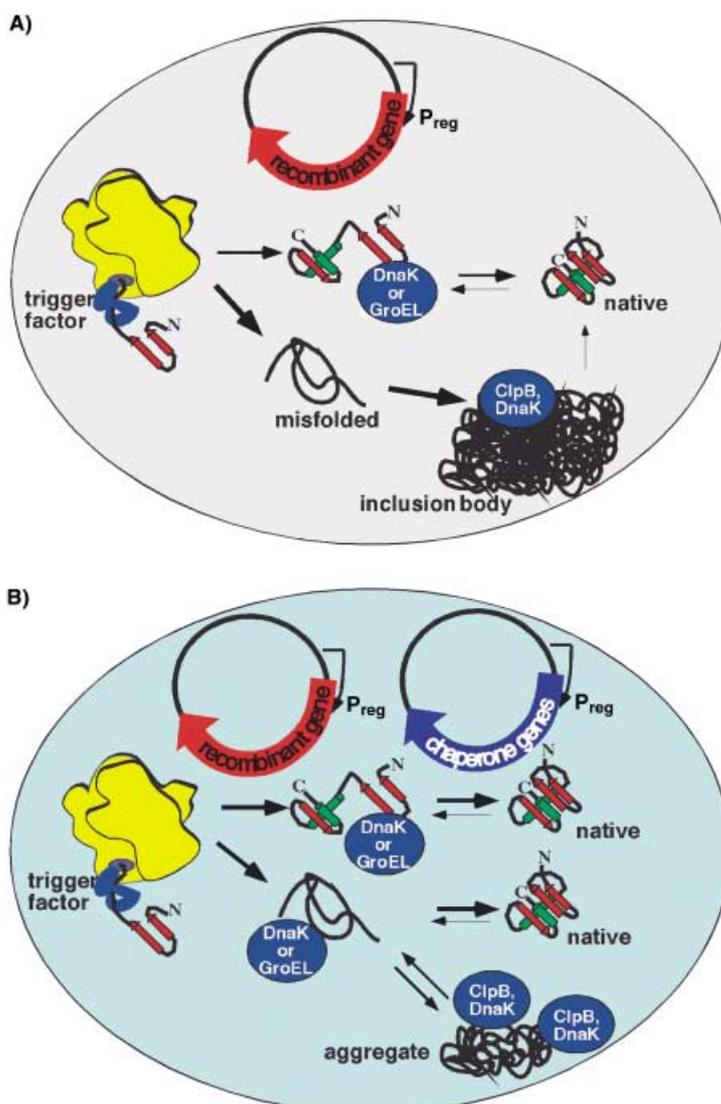


Figure 8. Biotechnological production of recombinant protein without (A) and with (B) optimized amounts of chaperones. A) Without sufficient amounts of chaperones the recombinant protein is highly prone to aggregation and forms inclusion bodies. B) The controlled co-overproduction of molecular chaperones together with the target protein leads to increased levels of the properly folded recombinant protein. P_{reg} = a regulatable promoter, for example, isopropyl-β-D-thiogalactopyranoside (IPTG) controlled. Possible secondary structure elements in proteins are shown in red (β strands) and green (α helices).

protein. Different host systems are established: plant or mammalian cell lines, insect and yeast cells, or bacteria. The decision of which host system is favored depends on the properties of the desired protein and the yield that can be obtained in the particular system. Eukaryotic systems are necessary when the recombinant protein has to be modified posttranslationally, for example, when glycosylation is required. Bacterial systems, like *E. coli*, have the advantage that they can be genetically engineered easily and produce large quantities of recombinant proteins in rapid, often inexpensive, fermentation processes. Among the many systems available for heterologous protein production, bacteria still remain the most attractive.

Improving protein folding

The application of molecular chaperones in biotechnology to increase the yield of properly folded recombinant target proteins is an attractive strategy. Nevertheless, it should be mentioned that there is no guarantee that chaperones are always capable of improving the solubility of recombinant proteins. No effect on solubility was found for chloramphenicol acetyltransferase and yeast *N*-myristoyl transferase when GroEL/GroES or DnaK were overexpressed. Moreover, GroEL/ES overproduction did not enhance the amount of soluble Adenovirus oncogene product E1A, myc protooncogene product, and *skr*-related gene product SnO.^[30–32] The beneficial effect of elevated chaperones levels may depend on the target and, even more importantly, on the levels and mixture of chaperones provided. None of the individual chaperones tested can be considered as the general folding chaperone that could prevent misfolding of all proteins. Rather, it was shown for a variety of chaperones and different heterologous protein substrates that individual chaperone systems work well for one substrate but are not beneficial for others. Different parameters have to be taken into consideration, such as the relative affinities of the chaperone to the folding intermediates of the target protein and the folding and aggregation kinetics of the recombinant protein. These parameters are difficult to predict, and therefore the application of molecular chaperones relies on trial and error experiments. With respect to recent findings on the *in vivo* role and cooperation of molecular chaperones, two alternative strategies appear to be best suited to optimize the folding yields of overproduced proteins in *E. coli*:

- 1) The simultaneous overexpression of several chaperone systems together with the recombinant target protein in the host system. This strategy would allow the provision of sufficient amounts of chaperones inside the cell to ensure productive *de novo* folding as well as to refold misfolded and aggregated proteins into native species.
- 2) Chaperone mixtures could be added during or after the purification procedure, for example, to resolubilize and refold proteins from inclusion bodies *in vitro*.

Both strategies are explained in detail below.

Overproduction of molecular chaperones inside the host cell

Introducing extra copies of chaperone-encoding genes into the host cells together with the target gene would ensure the simultaneous overexpression of the target and chaperone genes. In principle all host cells could be engineered like this, but it is still easiest to do this in bacteria. Bacterial cells can be transformed with several plasmids. These plasmids carry either the chaperone gene(s) or the target gene, both controlled by a regulatable promoter to ensure a coordinated coexpression (Figure 8B).

The most promising candidates for this approach in biotechnology are the prokaryotic ATP-controlled chaperone systems GroEL, DnaK, and ClpB and the ATP-independent ribosome-associated trigger factor. These chaperone systems are well-characterized and supposed to build a cellular folding network.

They can prevent the aggregation of misfolded proteins and assist refolding into the native conformation. The co-overproduction of the ribosome-associated chaperone trigger factor may enhance the folding of newly synthesized polypeptides and thereby increase the amount of native target protein. Enhanced levels of DnaK and GroEL systems keep the recombinant protein in a soluble state. Furthermore, the simultaneous presence of elevated levels of the bichaperone system, DnaK together with ClpB, could promote the resolubilization of any aggregated target protein.

It was shown for a variety of heterologous proteins that the coexpression of individual chaperones increases the solubility and decreases the aggregation of recombinant proteins when they are coexpressed in *E. coli*. For example the overproduction of trigger factor enhances the solubility of human lysozyme.^[33] The solubility and yield of two recombinant enzymes, dihydrofolate reductase (DHFR) from *Staphylococcus* and human protein tyrosine kinase (p50^{csk}), are improved when the GroEL system is co-overproduced in *E. coli*.^[34, 35] The overproduction of DnaJ elevates fourfold the amount of soluble recombinant transglutaminase.^[36] Some more examples are given in a comprehensive review.^[30]

Since we now know that these chaperones cooperate with each other, the new and most favorable strategy to go for in biotechnology would be to express combinations of chaperones rather than a single chaperone system. In addition, one should consider as well that there are highly specific chaperone systems known that assist the folding of defined proteins only. For example, in bacteria, the Hsp70 homologue HscA is supposed to assist specifically the folding of proteins containing iron–sulfur clusters.

Refolding of proteins trapped in inclusion bodies

An alternative strategy for large-scale protein purification is the isolation of recombinant proteins from inclusion bodies. This method is widely used since many recombinant proteins are deposited in such large insoluble aggregates. There is also an advantage in such an overexpression “accident” because the cell’s capacity for soluble proteins is limited, whereas proteins stored in large inclusion bodies do not interfere significantly with the cell’s metabolism. Moreover, proteins stored in inclusion bodies are already relatively pure. These inclusion bodies can be isolated easily by simple centrifugation steps after the disruption of the host cells. The disadvantage of this method is that it is complicated and inefficient. Treatment with strong denaturing agents is the first step to regain the functional target protein from inclusion bodies. This leads to soluble proteins but at the same time prevents the refolding of the soluble proteins into their native structure. Protein refolding is achieved either by rapidly diluting the proteinaceous solution in a large volume of buffer or by dialysis against buffer without denaturing agents. Both steps of this procedure are complicated and might be simplified by the use of molecular chaperones. Chaperones may be used to either solubilize the proteins from inclusion bodies or to promote the refolding of denatured proteins after their solubilization. It had been demonstrated that the bichaperone

system, DnaK and ClpB of *E. coli*, could do both; this system resolubilizes aggregates of malate dehydrogenase and promotes the refolding of this enzyme into the active conformation.^[23] The finding that the disaggregation capacity of this system works promiscuously for many different aggregated proteins was very important.^[19] Use of the bichaperone system may avoid the treatment of inclusion bodies with denaturing agents and may promote folding into the biologically active protein conformation. The GroEL system may further facilitate refolding of proteins. To date there is no report on the applicability of the bichaperone and GroEL systems to solubilize recombinant proteins deposited in inclusion bodies in vitro or in vivo. Moreover, two major drawbacks are the costs and the need to separate the refolded target protein from the chaperones after the refolding process. However, the high disaggregation and refolding capacity and the broad substrate specificity of these chaperones hold great promise.

- [1] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.
- [2] C. P. Schultz, *Nat. Struct. Biol.* **2000**, *7*, 7–10.
- [3] C. Schiene, G. Fischer, *Curr. Opin. Struct. Biol.* **2000**, *10*, 40–45.
- [4] S. Raina, D. Missiakas, *Annu. Rev. Microbiol.* **1997**, *51*, 179–202.
- [5] D. S. Goodsell, *Trends Biochem. Sci.* **1991**, *16*, 203–206.
- [6] B. Bukau, F. X. Schmid, J. Buchner, in *Molecular chaperones and folding catalysts—regulation, cellular function and mechanism* (Ed.: B. Bukau), Harwood Academic Publishers, Amsterdam, **1999**, pp. 3–10.
- [7] F. U. Hartl, M. Hayer-Hartl, *Science* **2002**, *295*, 1852–1858.
- [8] E. Deuerling, A. Schulze-Specking, T. Tomoyasu, A. Mogk, B. Bukau, *Nature* **1999**, *400*, 693–696.
- [9] S. A. Teter, W. A. Houry, D. Ang, T. Tradler, D. Rockabrand, G. Fischer, P. Blum, C. Georgopoulos, F. U. Hartl, *Cell* **1999**, *97*, 755–765.
- [10] B. Bukau, E. Deuerling, C. Pfund, E. A. Craig, *Cell* **2000**, *101*, 119–122.
- [11] A. L. Horwich, K. Brooks Low, W. A. Fenton, I. N. Hirshfield, K. Furtak, *Cell* **1993**, *74*, 909–917.
- [12] F. U. Hartl, *Nature* **1996**, *381*, 571–580.
- [13] B. Bukau, A. L. Horwich, *Cell* **1998**, *92*, 351–366.
- [14] E. Schaffitzel, S. Rüdiger, B. Bukau, E. Deuerling, *Biol. Chem.* **2001**, *382*, 1235–1243.
- [15] K. L. Ewalt, J. P. Hendrick, W. A. Houry, F. U. Hartl, *Cell* **1997**, *90*, 491–500.
- [16] G. Schatz, B. Dobberstein, *Science* **1996**, *271*, 1519–1526.
- [17] M. Ehrnsperger, S. Gräber, M. Gaestel, J. Buchner, *EMBO J.* **1997**, *16*, 221–229.
- [18] L. Veinger, S. Diamant, J. Buchner, P. Goloubinoff, *J. Biol. Chem.* **1998**, *273*, 11 032–11 037.
- [19] A. Mogk, T. Tomoyasu, P. Goloubinoff, S. Rüdiger, D. Röder, H. Langen, B. Bukau, *EMBO J.* **1999**, *18*, 6934–6949.
- [20] G. J. Lee, A. M. Roseman, H. R. Saibil, E. Vierling, *EMBO J.* **1997**, *16*, 659–671.
- [21] J. R. Glover, S. Lindquist, *Cell* **1998**, *94*, 73–82.
- [22] S. W. Hong, E. Vierling, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4392–4397.
- [23] P. Goloubinoff, A. Mogk, A. Peres Ben Zvi, T. Tomoyasu, B. Bukau, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13732–13737.
- [24] Y. Sanchez, S. L. Lindquist, *Science* **1990**, *248*, 1112–1115.
- [25] A. M. Roseman, S. Chen, H. White, K. Braig, H. R. Saibil, *Cell* **1996**, *87*, 241–251.
- [26] H. S. Rye, A. M. Roseman, S. Chen, K. Furtak, W. A. Fenton, H. R. Saibil, A. L. Horwich, *Cell* **1999**, *97*, 325–338.
- [27] M. Shtilerman, G. H. Lorimer, S. W. Englander, *Science* **1999**, *284*, 822–825.
- [28] T. K. Chaudhuri, G. W. Farr, W. A. Fenton, S. Rospert, A. L. Horwich, *Cell* **2001**, *107*, 235–246.
- [29] M. P. Mayer, S. Rüdiger, B. Bukau, *Biol. Chem.* **2000**, *381*, 877–885.
- [30] J. G. Thomas, A. Ayling, F. Baneyx, *Appl. Biochem. Biotechnol.* **1997**, *66*, 197–238.
- [31] S. C. Lee, P. O. Olins, *J. Biol. Chem.* **1992**, *267*, 2849–2852.
- [32] T. Yasukawa, C. Kanei-Ishii, T. Maekawa, J. Fujimoto, T. Yamamoto, S. Ishii, *J. Biol. Chem.* **1995**, *270*, 25 328–25 331.
- [33] K. Nishihara, M. Kanemori, H. Yanagi, T. Yura, *Appl. Environ. Microbiol.* **2000**, *66*, 884–889.
- [34] K. E. Amrein, B. Takacs, M. Stieger, J. Molnos, N. A. Flint, P. Burn, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1048–1052.
- [35] G. E. Dale, H. J. Schönfeld, H. Langen, M. Stieger, *Protein Eng.* **1994**, *7*, 925–931.
- [36] K. Yokoyama, Y. Kikuchi, H. Yasueda, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 1205–1210.
- [37] K. M. Flaherty, C. Deluca-Flaherty, D. B. McKay, *Nature* **1990**, *346*, 623–628.
- [38] X. Zhu, X. Zhao, W. F. Burkholder, A. Gragerov, C. M. Ogata, M. Gottesman, W. A. Hendrickson, *Science* **1996**, *272*, 1606–1614.

Received: October 24, 2001

Revised version: April 12, 2002 [M311]