



REVIEW: PROTEIN FOLDING

Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein

F. Ulrich Hartl* and Manajit Hayer-Hartl

Efficient folding of many newly synthesized proteins depends on assistance from molecular chaperones, which serve to prevent protein misfolding and aggregation in the crowded environment of the cell. Nascent chain-binding chaperones, including trigger factor, Hsp70, and prefoldin, stabilize elongating chains on ribosomes in a nonaggregated state. Folding in the cytosol is achieved either on controlled chain release from these factors or after transfer of newly synthesized proteins to downstream chaperones, such as the chaperonins. These are large, cylindrical complexes that provide a central compartment for a single protein chain to fold unimpeded by aggregation. Understanding how the thousands of different proteins synthesized in a cell use this chaperone machinery has profound implications for biotechnology and medicine.

To become functionally active, newly synthesized protein chains must fold to unique three-dimensional structures. How this is accomplished remains a fundamental problem in biology. Although it is firmly established from refolding experiments *in vitro* that the native fold of a protein is encoded in its amino acid sequence (1), protein folding inside cells is not generally a spontaneous process. Evidence accumulated over the last decade indicates that many newly synthesized proteins require a complex cellular machinery of molecular chaperones and the input of metabolic energy to reach their native states efficiently (2–5). The various chaperone factors protect nonnative protein chains from misfolding and aggregation, but do not contribute conformational information to the folding process. Here we focus on recent advances in our mechanistic understanding of *de novo* protein folding in the cytosol and seek to provide a coherent view of the overall flux of newly synthesized proteins through the chaperone system.

Protein Aggregation

Spontaneous refolding *in vitro* is generally efficient for small, single-domain proteins that bury exposed hydrophobic amino acid residues rapidly (within milliseconds) upon initiation of folding (1). In contrast, larger proteins composed of multiple domains often refold inefficiently, owing to the formation of

partially folded intermediates, including misfolded states, that tend to aggregate (Fig. 1). Misfolding originates from interactions between regions of the folding polypeptide chain that are separate in the native protein and that may be stable enough to prevent folding from proceeding at a biologically rele-

phobic forces and interchain hydrogen bonding (1, 6). This aggregation process irreversibly removes proteins from their productive folding pathways, and must be prevented *in vivo* by molecular chaperones. A certain level of protein aggregation does occur in cells despite the presence of an exclusive chaperone machinery and, in special cases, can lead to the formation of structured, fibrillar aggregates, known as amyloid, that are associated with diseases such as Alzheimer's or Huntington's disease (6, 7) (Fig. 1). Compared to refolding in dilute solution, the tendency of nonnative states to aggregate in the cell is expected to be sharply increased as a result of the high local concentration of nascent chains in polyribosomes and the added effect of macromolecular crowding.

Nascent chains. During translation, the folding information encoded in the amino acid sequence becomes available in a vectorial fashion.

The polypeptide exit channel in the large ribosomal subunit is 100 Å long, a distance spanned by an extended chain of ~30 amino acid residues or an α helix of 65 residues (8). The channel is on average only 15 Å wide and is expected to prohibit folding beyond helix formation inside the ribosome, unless the tunnel is conformationally dynamic. Because the formation of stable tertiary structure is a cooperative process at the level of protein domains (50 to 300 amino acid residues), an average domain can complete folding only when its entire sequence has emerged from the ribosome. It takes more than a minute to synthesize a 300-residue protein in eu-

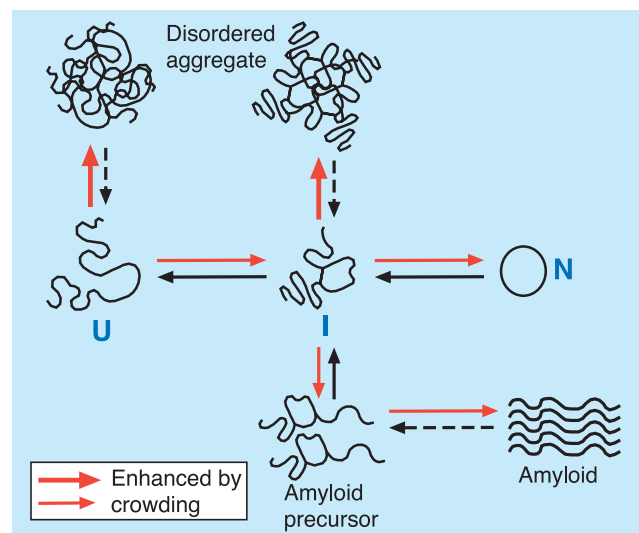


Fig. 1. Aggregation of nonnative protein chains as a side-reaction of productive folding in the crowded environment of the cell. Enhancement of aggregation and chain compaction by macromolecular crowding (red arrows). U, unfolded protein chain released from ribosome; I, partially folded intermediate; N, native, folded protein. Crowding is predicted to enhance the formation of amyloid fibrils, but this effect has not yet been demonstrated experimentally. [Adapted from (7)]

vant time scale. These nonnative states, though compact in shape, often expose hydrophobic amino acid residues and segments of unstructured polypeptide backbone to the solvent. They readily self-associate into disordered complexes (Fig. 1), driven by hydro-

karyotes. As a consequence, many nascent chains expose non-native features for a considerable length of time and are prone to aggregation. This tendency to aggregate is thought to be greatly increased by the close proximity of nascent chains of the same type

Department of Cellular Biochemistry, Max-Planck-Institut für Biochemie, Am Klopferspitz 18A, D-82152 Martinsried, Germany.

*To whom correspondence should be addressed. E-mail: uhartl@biochem.mpg.de

in polyribosome complexes (5), thus leading to the requirement for chaperones to maintain nascent chains in a nonaggregated, folding-competent conformation.

Macromolecular crowding. 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*) (9) are predicted to enhance the aggregation of nonnative protein chains substantially by increasing their effective concentrations (10) (Fig. 1). Crowding generally provides a nonspecific force for macromolecular compaction and association (11), including the collapse of protein chains during folding (9) and the interaction of nonnative proteins with molecular chaperones (12).

How Chaperones Prevent Aggregation

The cellular chaperone machinery counteracts the aggregation of nonnative proteins, both during de novo folding and under conditions of stress, such as high temperature, when some native proteins unfold. Many chaperones, though constitutively expressed, are synthesized at greatly increased levels under stress conditions and are classified as stress proteins or heat-shock proteins (Hsps) (3). In general, all these chaperones recognize hydrophobic residues and/or unstructured backbone regions in their substrates, i.e., structural features typically exposed by nonnative proteins but normally buried upon completion of folding. Chaperones that participate broadly in de novo protein folding, such as the Hsp70s and the chaperonins, promote the folding process through cycles of substrate binding and release regulated by their adenosine triphosphatase (ATPase) activity and by cofactor proteins. Chaperone binding may not only block intermolecular aggregation directly by shielding the interactive surfaces of non-native polypeptides, including unassembled protein subunits, but may also prevent or reverse intramolecular misfolding. Certain chaperones of the Hsp100 or Clp family even have the ability to unfold proteins or to disrupt small-protein aggregates by an adenosine 5'-triphosphate (ATP)-dependent mechanism (13). For a growing number of proteins, chap-

erone function is combined with an additional activity, as is the case for certain protein disulfide isomerases and peptidyl-prolyl isomerases, enzymes that catalyze rate-limiting steps in the folding of some proteins (14).

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 (Fig. 2). The majority of small pro-

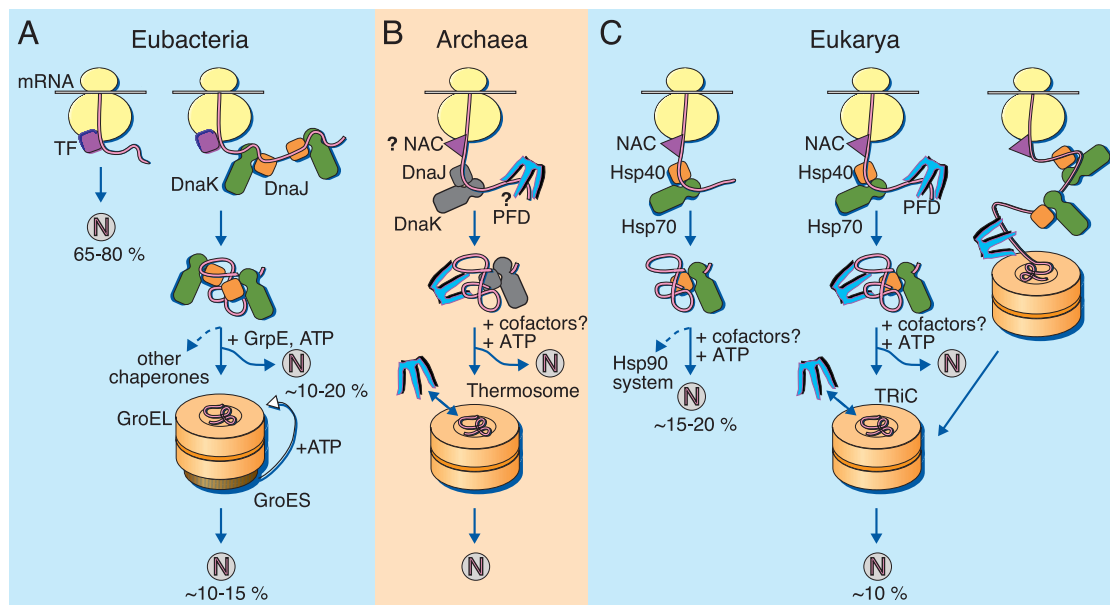


Fig. 2. Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol. (A) Eubacteria. TF, trigger factor; N, native protein. Nascent chains probably interact generally with TF, and most small proteins (~65 to 80% of total) fold rapidly upon synthesis without further assistance. Longer chains (10 to 20% of total) 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. GroEL does not bind to nascent chains and is thus likely to receive an appreciable fraction of its substrates after their interaction with DnaK. (B) Archaea. PFD, prefoldin; NAC, nascent chain-associated complex. Only some archaeal species contain DnaK/DnaJ. The existence of a ribosome-bound NAC homolog, as well as the interaction of PFD with nascent chains, has not yet been confirmed experimentally. (C) Eukarya—the example of the mammalian cytosol. Like TF, NAC probably interacts generally with nascent chains. 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, and a fraction of these must be transferred to Hsp90 for folding. About 10% of chains are co- or posttranslationally passed on to the chaperonin TRiC in a reaction mediated by PFD.

Protein Flux Through the Chaperone System

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 chaperone are conserved in all three domains of life and can cooperate in a topologically and timely ordered manner (15–17) (Fig. 2, A to C).

Although the essential nature of the chaperonins has long been recognized (18, 19), it has proved more difficult to establish the essential role of nascent chain-binding chaperones in protein folding, because of considerable functional redundancy among components (20, 21). Some of these chaperones,

teins are thought to fold rapidly and without further assistance upon completion of synthesis and release from this first set of components (Fig. 2A). 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 (20–22). In addition to stabilizing elongating chains, these chaperones also assist in co- or posttranslational folding, or facilitate chain transfer to downstream chaperones (Fig. 2, A and C) (17, 20, 21). A subset of slow-folding and aggregation-sensitive proteins (10 to 15% of total) interact with a chaperonin for folding in both prokaryotes and eukaryotes (22–24). Many eukaryotic kinases and other signal-transduction proteins use an additional chaperone pathway from Hsp70 to Hsp90 (Fig. 2C), a specialized ATP-dependent chaperone that cooperates with ancillary factors in protein folding and regulation. [For a detailed discussion of the Hsp90 system, see (25, 26).]

Ribosome-Binding Chaperones

Trigger factor (TF), a eubacterial protein of 48 kD, binds to ribosomes at a 1:1 stoichiometry and interacts with nascent chains as short as 57 residues (27). The nascent chain–TF complex dissociates, in an ATP-independent manner, after chain release from the ribosome (27). Although TF exhibits peptidyl-prolyl cis/trans isomerase (PPIase) activity in vitro, recognition of target polypeptides by TF is independent of proline residues (28) and is mediated by short sequences enriched in hydrophobic (aromatic) amino acids (28). 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 (20, 21). *E. coli* cells lacking TF (Δ tig) or DnaK (Δ dnaK) exhibit no apparent folding defects at 37°C; however, deletion of *dnaK* in a Δ tig strain is lethal. In light of this functional redundancy, the biological significance of the PPIase activity of TF remained unclear, but a recent study suggests that DnaK has a related activity in accelerating the cis/trans isomerization of nonprolyl peptide bonds (29). These isomerase activities may allow TF and the Hsp70s to maintain nascent and newly synthesized chains in a flexible state, poised for rapid folding upon release. In contrast to DnaK, a role of TF in mediating folding posttranslationally has not yet been demonstrated, but would be consistent with the finding that only half of total TF is ribosome bound (30).

The eukaryotic cytosol lacks TF but contains a ribosome-associated heterodimeric complex of α (33 kD) and β (22 kD) subunits, termed NAC (nascent chain–associated complex) (Fig. 2C) (31). A homolog of α -NAC appears to be present in some archaea (32). Although NAC lacks a PPIase domain, it has properties that suggest a functional similarity to TF. NAC associates with short nascent chains and dissociates upon chain

release from the ribosome (4, 33). However, a direct role for NAC in protein folding remains to be established.

Whereas the Hsp70 proteins in bacteria and higher eukaryotes act both co- and posttranslationally (see below), yeast and other fungi have cytosolic Hsp70 homologs that are specialized in nascent chain binding. The Ssb1 and Ssb2 proteins in the yeast *Saccharomyces cerevisiae* interact with the ribosome and with short nascent chains (34). Interestingly, this function of the Ssb proteins appears to be mediated by yet another Hsp70, Ssz1, which forms a stable ribosome-associated complex (RAC) with zototin (35, 36), the Hsp40 partner of Ssb1 and Ssb2 (30). RAC and the Ssb proteins are thought to act in concert in stabilizing nascent chains.

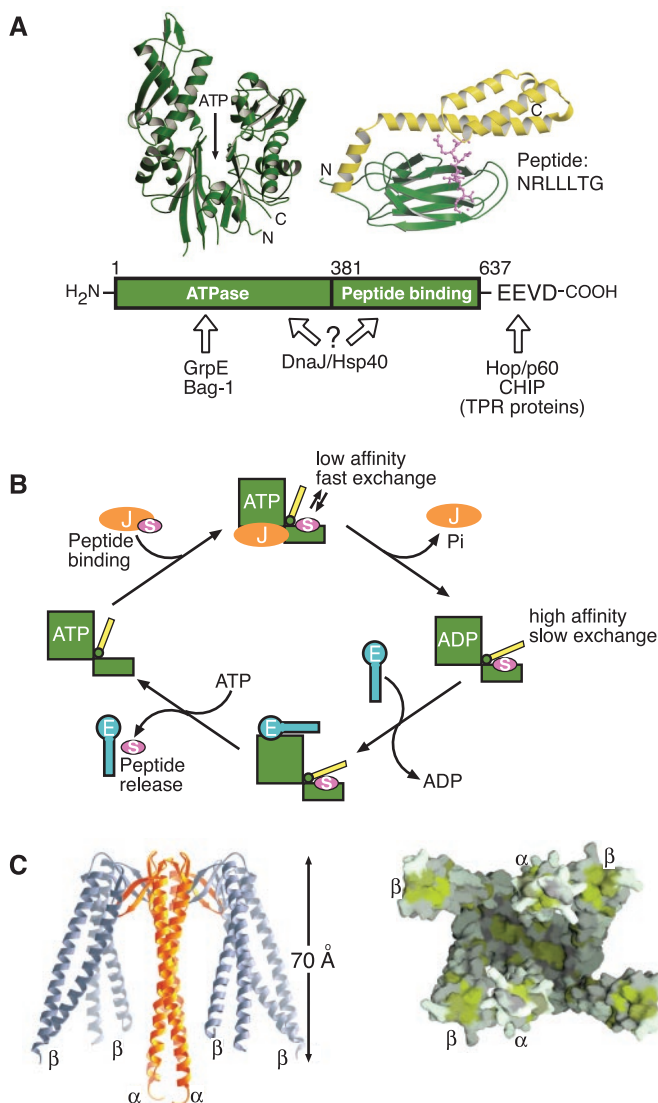
The Hsp70 System

The classic, nonribosome-binding members of the Hsp70 family exist in the cytosol of eubacteria, eukarya, and some archaea, as well as within eukaryotic organelles, such as mitochondria and endoplasmic reticulum. *S. cerevisiae* has four nonribosome-binding Hsp70 proteins in the cytosol, namely, Ssa1 to Ssa4. The cytosol of higher eukaryotes contains both constitutively expressed Hsp70 homologs (Hsc70) and stress-inducible forms (Hsp70). Together with cochaperones of the Hsp40 (DnaJ) family, these Hsp70s function by binding and releasing, in an

ATP-dependent manner, extended polypeptide segments that are exposed by proteins in their non-native states.

Structure and reaction cycle. The structural and mechanistic aspects of the Hsp70 system are best understood for the eubacterial Hsp70, termed DnaK, its Hsp40 cochaperone, DnaJ, and the nucleotide exchange factor GrpE. DnaK consists of a \sim 44-kD NH_2 -terminal ATPase domain and a \sim 27-kD COOH-terminal peptide-binding domain (37) (Fig. 3A). The latter is divided into a β -sandwich subdomain with a peptide-binding cleft and an α -helical latchlike segment (38). Target peptides are \sim seven residues long and are typically hydrophobic in their central region, with leucine and isoleucine residues being preferred by DnaK (4, 39) (Fig. 3A). These binding sites occur statistically every \sim 40 residues in proteins and are recognized with affinities of 5 nM to 5 μ M (37). The peptides are bound to DnaK in an extended state through hydrophobic side-chain interactions and hydrogen bonds with

Fig. 3. Structure and function of chaperones with the ability to bind nascent chains. (A) (Top) Structures of the ATPase domain (40) and the peptide-binding domain (38) of Hsp70 shown representatively for *E. coli* DnaK, generated with the program MOLSCRIPT (87). The α -helical latch of the peptide binding domain is shown in yellow and a ball-and-stick model of the extended peptide substrate in pink. ATP indicates the position of the nucleotide binding site. The amino acid sequence of the peptide is indicated in single-letter code (D, Asp; E, Glu; G, Gly; L, Leu; N, Asn; R, Arg; T, Thr; and V, Val). (Bottom) The interaction of prokaryotic and eukaryotic cofactors with Hsp70 is shown schematically. Residue numbers refer to human Hsp70. Only the Hsp70 proteins of the eukaryotic cytosol have the COOH-terminal sequence EEVD that is involved in binding of tetratricopeptide repeat (TPR) cofactors. (B) Simplified reaction cycle of the DnaK system with DnaJ colored as in (A). J, DnaJ; E, GrpE; S, substrate peptide. GrpE is drawn to reflect the extended shape of the protein. Not all substrates are presented to DnaK by DnaJ. The intermediate DnaK–DnaJ–substrate–ATP is probably very transient, as this is the fastest step of the cycle. (C) (Left) Side view and dimension of the structure of archaeal PFD with the two α subunits shown in gold and the four β subunits in gray. (Right) Bottom view of the PFD complex showing the central space enclosed by the six coiled-coil segments. Surface representation is shown with hydrophobic patches in yellow and hydrophilic regions in gray [reproduced from (54) with permission].



the peptide backbone (38). Thus, Hsp70 recognizes structural features common to most nascent chains: exposed hydrophobic amino acid side chains, in conjunction with an accessible polypeptide backbone.

Rapid peptide binding occurs in the ATP-bound state of DnaK in which the α -helical latch over the peptide-binding cleft is in an open conformation (Fig. 3B). Stable holding of peptide involves closing of the latch, a conformational change that is achieved by hydrolysis of bound ATP to adenosine 5'-diphosphate (ADP). The cycling of DnaK between these states is regulated by DnaJ (41 kD) and by GrpE, a homodimer of 20-kD subunits (37, 40). The NH₂-terminal J domain of DnaJ binds to DnaK and accelerates hydrolysis of ATP by DnaK, thus facilitating peptide capture (41, 42). The COOH-terminal domain of DnaJ (and of other Hsp40s) functions as a chaperone in recognizing hydrophobic peptides and can thus recruit DnaK to nascent chains (15, 43, 44). GrpE induces the release of ADP from DnaK (40), and upon rebinding of ATP the DnaK-peptide complex dissociates, completing the reaction cycle (Fig. 3B).

Some eukaryotic Hsp70 homologs, such as BiP in the endoplasmic reticulum, cooperate with J-domain proteins that lack a separate affinity for hydrophobic sequences. These Hsp70s may be able to bind extended polypeptide chains more generally, independent of exposed hydrophobic features (45). Whereas all Hsp70s seem to cooperate with J proteins, most eukaryotic Hsp70 proteins may be independent, for their general function, of a GrpE-like nucleotide exchange factor. Such a factor appears to be dispensable because the rate-limiting step in the ATPase cycle of eukaryotic Hsp70 is normally not the dissociation of bound ADP, but ATP hydrolysis itself. On the other hand, a small protein, Bag-1, acts as a nucleotide exchange factor and specific regulator of Hsp70 in the eukaryotic cytosol (46). The Hsp70-interacting Bag domain is structurally unrelated to GrpE (40, 47) and in Bag-1 is linked to an NH₂-terminal ubiquitin-homology domain (see below).

Substrates and mechanism of folding. The cellular concentration of DnaK (~50 μ M) exceeds that of ribosomes (~30 μ M) (4), assuming an even cytosolic distribution. DnaK preferentially associates with elongating polypeptides larger than 20 to 30 kD and thus acts on nascent chains subsequent to TF (21) (Fig. 2A). Upon deletion of TF, the fraction of nascent and newly synthesized polypeptides interacting with DnaK increases from ~15 to ~40% (20, 21). Whereas some chains transit DnaK with half-lives of less than 1 min, consistent with rapid folding upon completion of synthesis, other newly synthesized proteins are released from DnaK slowly, with half-lives of 10 min or more.

Large proteins >60 kD, which do not fit into the central cavity of the chaperonin GroEL (see below), constitute an appreciable fraction of these substrates, suggesting that DnaK facilitates the posttranslational folding of multidomain proteins through cycles of binding and release (21). Consistent with this conclusion, depletion of DnaK in TF-deleted cells causes the aggregation of many large, newly synthesized polypeptides (20). Similar to DnaK, mammalian Hsc70 also binds a wide range of nascent and newly synthesized chains (>15 to 20% of total) (Fig. 2C), including many multidomain proteins >50 kD (22).

How do cycles of Hsp70 binding and release promote protein folding? Generally, on release from Hsp70, an unfolded chain is free to partition to its native state. Rebinding of slow-folding intermediates to Hsp70 follows this release and prevents aggregation. Assuming that for long protein chains cycling is mediated by multiple Hsp70 molecules at the level of individual domains, the Hsp70 system could promote the folding of multidomain proteins by preventing (and perhaps reversing) intramolecular misfolding. The recently discovered isomerase activity of Hsp70 for nonprolyl peptide bonds may support this function (29). Consistent with this model, the Hsp70 system strongly accelerates the slow, spontaneous refolding of chemically denatured firefly luciferase (~60 kD) in vitro (48, 49). The enzyme ornithine transcarbamylase accumulates in a misfolded but soluble form in vivo when expressed in Hsp70 (Ssa)-deficient yeast (50).

Surprisingly, the components of the Hsp70 system are missing in certain species of archaea (32). How these cells protect nascent and newly synthesized polypeptides from aggregating is not yet clear, but a candidate chaperone for nascent chains in archaea is prefoldin.

Prefoldin

Prefoldin (PFD) (51), also known as the Gim complex (genes involved in microtubule biogenesis) (52), is a ~90-kD complex of two α and four β subunits in the archaeal and eukaryotic cytosol. The eukaryotic α and β subunits are not identical but orthologous (53). The structure of PFD resembles that of a jellyfish, with six α -helical coiled-coil tentacles emanating from a β -barrel body (Fig. 3C). At the tips these ~65 Å long coiled coils are partially unwound, exposing hydrophobic amino acid residues for the binding of nonnative protein (54) (Fig. 3C). Substrate binding and release by PFD is ATP independent, and in vitro, mammalian and archaeal PFD can stabilize nonnative proteins for subsequent transfer to a chaperonin (51, 53). PFD binds to nascent chains (55, 56) and cooperates in the folding of actin and

tubulin with the eukaryotic chaperonin (17). Interestingly, a combined deletion of the Ssb-class Hsp70s and of PFD in yeast results in a pronounced synthetic growth defect (56), resembling the synthetically lethal phenotype of the TF and DnaK deletions in *E. coli* (20, 21). These findings underscore the functional redundancy among nascent chain-binding chaperones and suggest that PFD may have a DnaK or TF-like role in the archaeal cytosol.

The Chaperonins

The chaperonins are a conserved class of large double-ring complexes of ~800 kD enclosing a central cavity. They occur in two subgroups that are similar in architecture but distantly related in sequence. Group I chaperonins, also known as Hsp60s, are generally found only in eubacteria and in organelles of endosymbiotic origin—mitochondria and chloroplasts. They cooperate with cofactors of the GroES or Hsp10 family. Group II chaperonins exist in the archaeal and the eukaryotic cytosol and are GroES independent. The chaperonin mechanism differs fundamentally from that of the Hsp70 system, although in both cases protein binding and release is ATP regulated. Nonnative substrate protein is first captured through hydrophobic contacts with multiple chaperonin subunits and is then displaced into the central ring cavity where it folds, protected from aggregating with other nonnative proteins.

Group I chaperonins—structure and reaction cycle. *E. coli* GroEL and its cofactor GroES represent the paradigmatic Group I chaperonin system. In GroEL, two heptameric rings of identical 57-kD subunits are stacked back-to-back. Each subunit consists of three domains: The equatorial domain harbors the ATP binding site and is connected through an intermediate, hingelike domain to the apical domain (Fig. 4A). The latter makes up the opening of the cylinder and exposes a number of hydrophobic residues toward the ring cavity for substrate binding. GroES is a homoheptameric ring of ~10-kD subunits that cycles on and off the ends of the GroEL cylinder, in a manner regulated by the GroEL ATPase (4, 37, 57) (Fig. 4A).

The hydrophobic surfaces exposed by the apical domains (Fig. 4A) interact with hydrophobic amino acid residues on compact folding intermediates (4, 37, 57). Hydrophobic sequences bind to a flexible groove between two amphiphilic helices in the apical domain. This region can accommodate a peptide either as a β hairpin or an amphiphilic α -helical conformation (58, 59). Stable substrate binding with nanomolar affinity relies on the interaction of a nonnative polypeptide with multiple apical domains (60). The GroES subunits have mobile sequence loops that contact the substrate-binding regions in the apical domains of GroEL and

mediate substrate dissociation (37, 57, 61, 62).

GroEL is functionally asymmetrical; the two rings are coupled by negative allostery and do not occur in the same nucleotide-bound state. The chaperonin reaction begins by the binding of substrate polypeptide to the free end (i.e., the trans ring) of a GroEL-GroES complex (Fig. 4B). This step is closely followed by the binding of seven ATP molecules and GroES, resulting in the displacement of substrate into a GroES-capped cavity and causing the dissociation of the seven ADP molecules and GroES from the former cis complex. Upon binding to GroES, the apical domains undergo a massive rotation and upward movement (61, 63), resulting in an enlargement of the cavity and a shift in its surface properties from hydrophobic to hydrophilic (Fig. 4A). Non-native proteins up to ~60 kD can be encapsulated and are free to fold in the resulting GroEL-GroES cage (also termed "Anfinsen cage") (64–67). Folding is allowed to proceed for ~10 s, timed by the hydrolysis of the seven ATP molecules in the cis ring. Upon completion of hydrolysis, binding of seven ATP molecules to the trans ring triggers the opening of the cage. Both folded and nonnative protein exit at this point (Fig. 4B), but folding intermediates that still expose extensive hydrophobic surfaces are rapidly recaptured and folding cycles are repeated until the protein reaches its native state. Oligomeric assembly occurs in solution after subunit folding inside the cage.

Substrates and folding mechanisms. About 10% of newly synthesized polypeptides normally transit GroEL posttranslationally (23, 24), consistent with the cytosolic concentration of GroEL (~3 μ M) relative to that of ribosomes

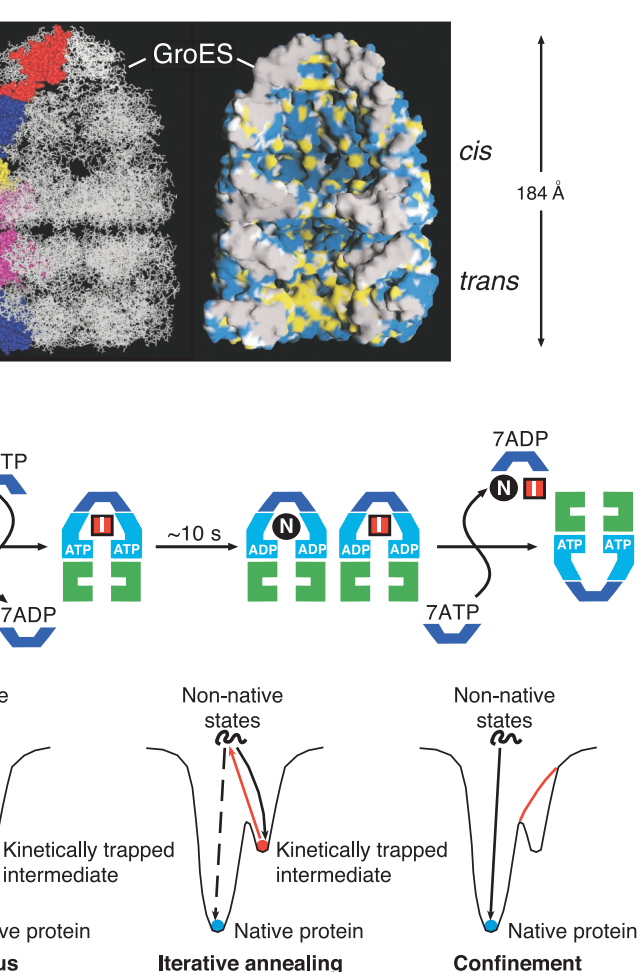


Fig. 4. The GroEL-GroES chaperonin system. (A) (Left) View of the asymmetric GroEL-GroES-(ADP)₇ complex generated with the coordinates 1AON (61) and program Weblab ViewLite 4.0 (Molecular Simulations). The equatorial, intermediate, and apical domains of one subunit each in the cis and trans ring of GroEL 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, blue; hydrophobic side-chain atoms, yellow; backbone atoms, white; and solvent-excluded surfaces at subunit interfaces, gray. [Reprinted from (61) with permission] (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 is then rapidly re-bound by GroEL. (C) Mechanisms of accelerated folding. Simple energy diagrams are shown for a protein that forms a kinetically trapped intermediate during spontaneous folding (left). In the iterative annealing model, this intermediate is thought to be actively unfolded by GroEL/GroES (69) and allowed to repartition (middle), whereas confinement of nonnative protein in the narrow, hydrophilic environment of the GroEL-GroES cage is suggested to result in a smoothing of the energy landscape (right), such that formation of certain trapped intermediates is avoided (67). Both proposed mechanisms would result in accelerated folding.

(~30 μ M). Most of these proteins are between 20 and 60 kD in size and leave GroEL with half-lives between 15 s and several minutes (23). The essential nature of GroEL and GroES (18) may be explained in principle by the existence of at least one essential *E. coli* protein with an absolute chaperonin dependence. Alternatively, loss of chaperonin could be lethal because it results in a reduced efficiency, rather than a complete loss, of folding for many proteins. Analysis of the highly streamlined pro-

teome of *Ureaplasma urealyticum* (68), the first eubacterium that lacks GroEL and GroES, may now offer an opportunity to test these hypotheses.

About 50 proteins interacting with GroEL in the *E. coli* cytosol have been identified, and many of them contain two or more domains with α/β folds (24). Proteins with such complex topologies often fold slowly and are aggregation prone, owing to the exposure of extensive hydrophobic surfaces in their non-native states. Stringent model substrates of GroEL, such as bacterial RuBisCo (ribulose-1,5-bisphosphate carboxylase-oxygenase), share this domain topology and fold efficiently only when in the GroEL-GroES cage (67).

In addition to preventing aggregation during folding, encapsulation of nonnative RuBisCo (50 kD) in the hydrophilic cage speeds up the folding reaction substantially (67). Confinement in the cage may smooth the energy landscape of folding for some larger proteins, either by preventing the formation of certain kinetically trapped intermediates or by facilitating their progression toward the compact, native state (Fig. 4C). This acceleration of RuBisCo folding had previously been attributed to a mechanism of "iterative annealing" (57, 69), rather than to an effect of confinement. In this alternative model the chaperonin is suggested to facilitate folding by cycles of unfolding kinetically trapped states, followed by repartitioning of the unfolded protein between productive and nonproductive folding pathways (57) (Fig. 4C). Active unfolding of RuBisCo was suggested to result from GroES-induced movements of the apical GroEL domains, exerting a stretching force on the bound polypeptide (69), but this effect has not yet been confirmed with any other GroEL-dependent protein (69, 70).

As shown recently, GroEL also interacts

with, and assists in the folding of, certain proteins too large to be encapsulated by GroES (23, 24, 71). Mitochondrial aconitase (82 kD), for example, can fold through ATP-regulated cycles of GroEL binding and release of nonnative states, with protein release being triggered by the binding of GroES to the opposite (trans) ring of GroEL (71).

Group II chaperonins. The Group II chaperonin of the eukaryotic cytosol, TRiC (TCP-1 ring complex, also called CCT for chaperonin-containing TCP-1), contains eight orthologous subunits per ring that differ primarily in their apical domains. The simpler archaeal chaperonin, referred to as thermosome, consists of up to three different subunits, which are arranged in eight- or nine-membered rings. The backbone trace of the chaperonin II apical domain is virtually identical to that of GroEL, with the exception of an α -helical insertion that protrudes from the ring opening (72, 73) and, in the absence of a separate GroES-like cofactor, is thought to function as a built-in lid of the central cavity.

The mechanism of Group II chaperonins is not yet well understood, and the nature and exact location of the substrate binding site(s) on the apical domains are still undefined. The most abundant substrates of TRiC are the cytoskeletal proteins actin and tubulin. Strikingly, folding of these proteins cannot be mediated by GroEL and GroES, suggesting a more specific role for TRiC in folding beyond prevention of aggregation. Actin binds to TRiC through at least two distinct regions and interacts with specific TRiC subunits (74, 75). ATP binding induces encapsulation of the protein by the apical-domain protrusions and initiates folding (75, 76). Through its built-in lid mechanism, TRiC may act cotranslationally in the folding of discrete domains of proteins that are too large to be encapsulated as a whole (16) (Fig. 2C).

The subunit heterogeneity of TRiC suggested that the cytosolic chaperonin may be adapted to assisting the folding of a small set of specific proteins, including actins and tubulins. However, as determined by a recent pulse-chase analysis in mammalian cells, TRiC interacts transiently with a wider range of newly synthesized proteins of 30 to 120 kD in size, constituting ~12% of total synthesized chains (22). The list of model substrates includes, among others, firefly luciferase, α -transducin, and the von Hippel-Lindau tumor suppressor protein (5).

Coordination of Translation and Chaperone Activities

To ensure an efficient use of the cytosolic folding machinery, protein synthesis on ribosomes must be coordinated with the activities of the various chaperone systems in stabilizing nascent chains and in promoting folding.

The mechanistic principles underlying this functional cooperation are not yet well understood, but plausible models have been developed based on a combination of *in vitro* and *in vivo* studies.

Cotranslational domain folding. It has been suggested that translation itself can have a "chaperone-like" role in the folding of larger proteins composed of multiple domains (77). Attempts to refold such proteins *in vitro* often result in intramolecular misfolding and aggregation (5). Cotranslational and sequential domain folding, i.e., the folding of one domain well before another is synthesized, avoids this problem, as demonstrated with artificial two-domain fusion proteins combining the ~20-kD proteins H-ras and dihydrofolate reductase (DHFR) (77).

Domain folding during translation occurs in the prokaryotic and eukaryotic cytosol (78), but a difference in the efficiencies of folding was noted for certain multidomain proteins when comparing both systems (77). The bacterial two-domain protein OmpR and the ras-DHFR fusion proteins fold cotranslationally in a mammalian cell lysate or in intact cells but posttranslationally upon synthesis in the *E. coli* system. Although the individual domains fold efficiently in *E. coli*, bacterial expression of ras-DHFR does not result in an active protein. Eukaryotic cells contain a much greater number of proteins with multiple domains than bacteria (77). Although the sizes of protein domains are uniformly distributed in all domains of life, in *E. coli* only 13% of all 4300 proteins exceed a length of 500 residues (~55 kD), compared with 38% of the 5800 proteins in *S. cerevisiae*. Thus, the eukaryotic translation and folding machineries may have been optimized in evolution to facilitate cotranslational domain folding. This optimization may be reflected in the 5- to 10-times slower speed of translation in eukaryotes compared with bacteria and in a functional adaptation of the eukaryotic chaperone machinery. Eukaryotic TRiC, for example, may mediate cotranslational domain folding for some proteins (16, 56, 79), whereas folding in the bacterial GroEL-GroES cage is strictly posttranslational (23) (Fig. 2).

Processivity of chaperone action. The notion of cooperation between mechanistically distinct chaperones in protein folding is now firmly established (15, 16, 20, 21, 26), but how the different components of the folding machinery are functionally integrated is not yet well understood. In principle, substrate transfer between chaperones could be accomplished by free partitioning of nonnative states through the solution. However, considering the highly crowded nature of the cytosol, it is difficult to envisage how aggregation is avoided in this model. Alternatively, ordered pathways of cellular folding may ex-

ist in which different chaperones function in a processive manner to minimize the exposure of nonnative proteins to the bulk cytosol.

Whereas in both prokaryotes and eukaryotes specific chaperones are recruited to nascent chains by virtue of their affinity for the ribosome (Fig. 2), the existence of processive chaperone pathways has so far been demonstrated only in the eukaryotic system. As shown in yeast and mammalian cells, folding intermediates generated during biosynthesis are not freely exposed to the bulk cytosol, but rather are functionally compartmentalized (17, 22). In these experiments, certain nascent and newly synthesized protein chains do not bind to a heterologously expressed, noncycling mutant of GroEL (Trap-GroEL), but instead interact productively with the endogenous eukaryotic chaperones. In the specific case of actin, an obligatory substrate of TRiC, protection from exposure to the bulk cytosol during folding is mediated by PFD. The speed and efficiency of actin folding is markedly reduced in PFD-deficient yeast, with nonnative chains being released into the cytosol (17). PFD may deliver substrate proteins to TRiC by binding both to nascent chains (55, 56) and to TRiC itself (51). In addition, PFD and TRiC seem to cooperate functionally in actin folding, such that nonnative chains are not released into the cytosol during folding cycles (17) (Fig. 2C).

Another example of chaperone coupling in the eukaryotic cytosol is the cooperation between Hsc70 and Hsp90 in the folding of signal-transduction proteins (26). Substrate transfer from Hsc70 to Hsp90 is mediated by Hop (Hsp organizing protein; also known as p60), an adaptor protein that physically links both chaperones. Hop contains two tetratricopeptide repeat (TPR) domains, which bind the extended COOH-terminal sequences of Hsc70 and Hsp90, respectively (80) (Fig. 3A). As shown recently, similar mechanisms are involved in regulating the transfer of nonnative or irreversibly misfolded proteins from these chaperones to the ubiquitin-proteasome machinery. The protein CHIP associates with Hsp90 through an NH₂-terminal TPR domain and targets certain Hsp90 substrates for degradation through a COOH-terminal ubiquitin ligase domain (81, 82). CHIP cooperates functionally with Bag-1 (see above), which binds to Hsc70 and to the proteasome (83). These findings provide the first insight into the mechanisms that integrate chaperone-assisted folding and proteolytic degradation, the two main components of protein quality control in the cytosol.

Perspectives

Recent years have seen major advances in our understanding of the basic mechanisms of chaperone-assisted protein folding. Future efforts will define more comprehensively the rules for how the thousands of different pro-

teins in a cell use the chaperone machinery. Global analyses of chaperone usage and folding properties in eukaryotes and prokaryotes will address these questions through a combination of proteomics and high-throughput protein expression (84). These studies may eventually offer a rational basis to optimize recombinant protein production in organisms that have been genetically modified to provide the appropriate folding machineries.

In addition to its biotechnological interest, understanding the complex functions of the chaperone arsenal will likely prove useful in dissecting the mechanisms by which protein misfolding and aggregation cause disease. Are chaperones capable of preventing the deposition of amyloid aggregates, and if so, why do these defense mechanisms fail in the millions of patients suffering from neurodegenerative maladies such as Alzheimer's or Huntington's disease? Recent reports that an up-regulation of the Hsp70 system can suppress the neurotoxicity of certain amyloidogenic proteins (85, 86) point toward molecular chaperones as promising targets in the quest for treatment of protein-misfolding diseases.

References and Notes

1. C. Dobson, M. Karplus, *Curr. Opin. Struct. Biol.* **9**, 92 (1999).
2. R. J. Ellis, S. M. Hemmingsen, *Trends Biochem. Sci.* **14**, 339 (1989).
3. M.-J. Gething, J. Sambrook, *Nature* **355**, 33 (1992).
4. F. U. Hartl, *Nature* **381**, 571 (1996).
5. J. Frydman, *Annu. Rev. Biochem.* **70**, 603 (2001).
6. S. E. Radford, *Trends Biochem. Sci.* **25**, 611 (2000).
7. C. M. Dobson, *Trends Biochem. Sci.* **24**, 329 (1999).
8. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **289**, 920 (2000).
9. R. J. Ellis, *Curr. Opin. Struct. Biol.* **11**, 114 (2001).
10. B. van den Berg, R. J. Ellis, C. M. Dobson, *EMBO J.* **18**, 6927 (1999).
11. A. P. Minton, *Curr. Opin. Struct. Biol.* **10**, 34 (2000).
12. J. Martin, F. U. Hartl, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1107 (1997).
13. A. P. Ben-Zvi, P. Goloubinoff, *J. Struct. Biol.* **135**, 84 (2001).
14. C. Schiene, G. Fischer, *Curr. Opin. Struct. Biol.* **10**, 40 (2000).
15. T. Langer et al., *Nature* **356**, 683 (1992).
16. J. Frydman, E. Nimmesgern, K. Ohtsuka, F. U. Hartl, *Nature* **370**, 111 (1994).
17. K. Siegers et al., *EMBO J.* **18**, 75 (1999).
18. O. Fayet, T. Ziegelhoffer, C. Georgopoulos, *J. Bacteriol.* **171**, 1379 (1989).
19. A. L. Horwich, K. B. Low, W. A. Fenton, I. N. Hirshfield, K. Furtak, *Cell* **74**, 909 (1993).
20. E. Deuerling, A. Schulze-Specking, T. Tomoyasu, A. Mogk, B. Bukau, *Nature* **400**, 693 (1999).
21. S. A. Teter et al., *Cell* **97**, 755 (1999).
22. V. Thulasiraman, C. F. Yang, J. Frydman, *EMBO J.* **18**, 85 (1999).
23. K. L. Ewalt, J. P. Hendrick, W. A. Houry, F. U. Hartl, *Cell* **90**, 491 (1997).
24. W. A. Houry, D. Frishman, C. Eckerskorn, F. Lottspeich, F. U. Hartl, *Nature* **402**, 147 (1999).
25. J. Buchner, *Trends Biochem. Sci.* **24**, 136 (1999).
26. J. C. Young, I. Moarefi, F. U. Hartl, *J. Cell Biol.* **154**, 267 (2001).
27. T. Hestekamp, S. Hauser, H. Lutcke, B. Bukau, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4437 (1996).
28. H. Patzelt et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14244 (2001).
29. C. Schiene-Fischer, J. Habazettl, G. Fischer, personal communication.
30. B. Bukau, E. Deuerling, C. Pfund, E. A. Craig, *Cell* **101**, 119 (2000).
31. B. Wiedmann, H. Sakai, T. A. Davis, M. Wiedmann, *Nature* **370**, 434 (1994).
32. M. R. Leroux, *Adv. Appl. Microbiol.* **50**, 219 (2001).
33. B. Beatrix, H. Sakai, M. Wiedmann, *J. Biol. Chem.* **275**, 37838 (2000).
34. C. Pfund et al., *EMBO J.* **17**, 3981 (1998).
35. M. Gautschi et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3762 (2001).
36. M. Gautschi, A. Mun, S. Ross, S. Rospert, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
37. B. Bukau, A. L. Horwich, *Cell* **92**, 351 (1998).
38. X. T. Zhu et al., *Science* **272**, 1606 (1996).
39. S. Rüdiger, L. Germeroth, J. Schneider-Mergener, B. Bukau, *EMBO J.* **16**, 1501 (1997).
40. C. J. Harrison, M. Hayer-Hartl, M. Di Liberto, F. U. Hartl, J. Kuriyan, *Science* **276**, 431 (1997).
41. M. P. Mayer et al., *Nature Struct. Biol.* **7**, 586 (2000).
42. M. Pellicchia et al., *Nature Struct. Biol.* **7**, 298 (2000).
43. S. Rüdiger, J. Schneider-Mergener, B. Bukau, *EMBO J.* **20**, 1042 (2001).
44. B. D. Sha, S. Lee, D. M. Cyr, *Struct. Fold. Des.* **8**, 799 (2000).
45. B. Misselwitz, O. Staack, T. A. Rapoport, *Mol. Cell* **2**, 593 (1998).
46. J. Höfheld, S. Jentsch, *EMBO J.* **16**, 6209 (1997).
47. H. Sondermann et al., *Science* **291**, 1553 (2001).
48. A. Szabo et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10345 (1994).
49. R. Herbst, K. Gast, R. Seckler, *Biochemistry* **37**, 6586 (1998).
50. S. Kim, B. Schilke, E. A. Craig, A. L. Horwich, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12860 (1998).
51. I. E. Vainberg et al., *Cell* **93**, 863 (1998).
52. S. Geissler, K. Siegers, E. Schiebel, *EMBO J.* **17**, 952 (1998).
53. M. R. Leroux et al., *EMBO J.* **18**, 6730 (1999).
54. R. Siebert, M. R. Leroux, C. Scheufler, F. U. Hartl, I. Moarefi, *Cell* **103**, 621 (2000).
55. W. J. Hansen, N. J. Cowan, W. J. Welch, *J. Cell Biol.* **145**, 265 (1999).
56. K. Siegers, F. U. Hartl, personal communication.
57. P. B. Sigler et al., *Annu. Rev. Biochem.* **67**, 581 (1998).
58. J. Chatellier, A. M. Buckle, A. R. Fersht, *J. Mol. Biol.* **292**, 163 (1999).
59. L. L. Chen, P. B. Sigler, *Cell* **99**, 757 (1999).
60. G. W. Farr et al., *Cell* **100**, 561 (2000).
61. Z. H. Xu, A. L. Horwich, P. B. Sigler, *Nature* **388**, 741 (1997).
62. A. Richardson, S. J. Landry, C. Georgopoulos, *Trends Biochem. Sci.* **23**, 138 (1998).
63. A. M. Roseman, S. X. Chen, H. White, K. Braig, H. R. Saibil, *Cell* **87**, 241 (1996).
64. R. J. Ellis, *Curr. Biol.* **11**, R1038 (2001).
65. M. Mayhew et al., *Nature* **379**, 420 (1996).
66. J. S. Weissman, H. S. Rye, W. A. Fenton, J. M. Beechem, A. L. Horwich, *Cell* **84**, 481 (1996).
67. A. Brinker et al., *Cell* **107**, 223 (2001).
68. J. I. Glass et al., *Nature* **407**, 757 (2000).
69. M. Shtilerman, G. H. Lorimer, S. W. Englander, *Science* **284**, 822 (1999).
70. J. W. Chen, S. Walter, A. L. Horwich, D. L. Smith, *Nature Struct. Biol.* **8**, 721 (2001).
71. T. K. Chaudhuri, G. W. Farr, W. A. Fenton, S. Rospert, A. L. Horwich, *Cell* **107**, 235 (2001).
72. M. Klumpp, W. Baumeister, L. O. Essen, *Cell* **91**, 263 (1997).
73. L. Ditzel et al., *Cell* **93**, 125 (1998).
74. O. Llorca et al., *EMBO J.* **19**, 5971 (2000).
75. O. Llorca et al., *EMBO J.* **20**, 4065 (2001).
76. I. Gutsche, J. Holzinger, N. Rauh, W. Baumeister, R. P. May, *J. Struct. Biol.* **135**, 139 (2001).
77. W. J. Netzer, F. U. Hartl, *Nature* **388**, 343 (1997).
78. A. V. Nicola, W. Chen, A. Helenius, *Nature Cell Biol.* **1**, 341 (1999).
79. C. D. McCallum, H. Do, A. E. Johnson, J. Frydman, *J. Cell Biol.* **149**, 591 (2000).
80. C. Scheufler et al., *Cell* **101**, 199 (2000).
81. J. Demand, S. Alberti, C. Patterson, J. Höfheld, *Curr. Biol.* **11**, 1569 (2001).
82. P. Connell et al., *Nature Cell Biol.* **3**, 93 (2001).
83. J. Luders, J. Demand, J. Höfheld, *J. Biol. Chem.* **275**, 4613 (2000).
84. S. A. Lesley, *Protein Expr. Purif.* **22**, 159 (2001).
85. P. Kazemi-Esfarjani, S. Benzer, *Science* **287**, 1837 (2000).
86. C. J. Cummings et al., *Hum. Mol. Genet.* **10**, 1511 (2001).
87. P. J. Kraluis, *J. Appl. Crystallogr.* **24**, 946 (1991).
88. We thank our colleagues B. Bukau, E. Deuerling, R. J. Ellis, G. Fischer, A. Horwich, S. Rospert, and K. Siegers for providing access to unpublished information, and R. J. Ellis and members of the Hartl group for critically reading the manuscript.

So instant, you don't need water...

NEW! Science Online's Content Alert Service: The only source for instant updates on breaking science news. This free Science Online enhancement e-mails summaries of the latest research articles published weekly in *Science* – **instantly**. To sign up for the Content Alert service, go to Science Online – and save the water for your coffee.

Science
www.sciencemag.org

For more information about Content Alerts go to www.sciencemag.org. Click on Subscription button, then click on Content Alert button.

Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein

F. Ulrich Hartl and Manajit Hayer-Hartl

Science **295** (5561), 1852-1858.
DOI: 10.1126/science.1068408

ARTICLE TOOLS

<http://science.sciencemag.org/content/295/5561/1852>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/307/5706/42.2.full>

REFERENCES

This article cites 60 articles, 19 of which you can access for free
<http://science.sciencemag.org/content/295/5561/1852#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)