Cell, Vol. 101, 119–122, April 14, 2000, Copyright ©2000 by Cell Press

Getting Newly Synthesized Proteins into Shape

Minireview

Bernd Bukau,*‡ Elke Deuerling,* Christine Pfund,† and Elizabeth A. Craig†‡
*Institut für Biochemie und Molekularbiologie
Universität Freiburg
Hermann Herder Str. 7
D-79104 Freiburg
Germany
†Department of Biomolecular Chemistry
University of Wisconsin
1300 University Avenue
Madison, Wisconsin 53706

Cellular synthesis of polypeptides is an amazing, complex, and efficient process. An *E. coli* cell utilizes up to 20,000 ribosomes to produce an estimated total of 30,000 polypeptides per minute. Yet it is only the beginning. Each newly made protein must be folded into its correct tertiary structure. How this is achieved is one of the most basic, and complicated, questions of molecular biology.

When does the process of folding begin in the lifetime of a protein? Crystallographic data of bacterial ribosomes identified a peptide exit tunnel in the large subunit with a length of 100 Å and an average diameter of about 20 Å (Ban et al., 1999). This tunnel is long enough to accommodate extended chains of about 30 residues and perhaps longer peptides in helical conformation, but certainly not peptides with tertiary structure. Once the N-terminal residues are synthesized, the nascent polypeptide emerges into the crowded environment of the cytosol. It is well established that nascent proteins can fold cotranslationally (i.e., while still attached to ribosomes) in cell-free translation systems. Recently, a study using the Semliki Forest virus capsid protein established that cotranslational folding can occur in living bacterial and mammalian cells (Nicola et al., 1999). In both prokaryotic and eukaryotic cells, the N-terminal domain of this protein folds rapidly into its native state during translation, well before termination of synthesis. Such cotranslational stepwise folding may be beneficial to the folding process since it avoids unwanted collisions between domains during the vulnerable early stages of folding (Netzer and Hartl, 1997).

Specialized proteins called molecular chaperones, which bind nonnative states of proteins, are found in all species. Work in past years established that chaperones facilitate the refolding of proteins denatured in the cytosol due to heat or chemical stresses. An understanding of the role of chaperones in assisting folding of newly synthesized cytosolic proteins has been more elusive. This review focuses on the function of chaperones in the folding of polypeptides as they emerge from the ribosome and are released into the cytosol. Coimmunoprecipitation of extracts from pulse-labeled cells has allowed an estimate of the number of proteins interacting with a particular chaperone in vivo. Aggregation of newly synthesized proteins in strains carrying mutations in chaperone genes has been taken as evidence of the in vivo role for a particular chaperone. Together with the analysis of folding of specific endogenous proteins in mutant strains, these studies have led to significant advances in our understanding of protein folding in the complex cellular milieu.

‡To whom correspondence should be addressed (e-mail: ecraig@facstaff.wisc.edu [E. A. C.], bukau@sun2.ruf.uni-freiburg.de [B. B.]).

The Welcoming Committee for Nascent Polypeptide Chains

Nascent chains emerging at the peptide exit tunnel of the ribosome are awaited by a welcoming committee of ribosome-associated chaperones. These factors have been most thoroughly studied in *E. coli*. Trigger factor (TF), which is found in all eubacteria analyzed, is the major protein that cross-links to virtually all nascent chains of secretory and cytosolic proteins tested (Valent et al., 1995; Hesterkamp et al., 1996). It may be positioned close to the exit site, as it cross-links to ribosome-associated chains 57 residues in length. TF, which has a central domain with homology to FK506 binding proteins (FKBPs), displays peptidyl-prolyl-*cis*-trans isomerase and chaperone-like activities in vitro. It is unclear which of these activities is required for its function in vivo.

Recent genetic studies provide clues suggesting a role of TF in protein folding in vivo. Δ*tig* mutants lacking TF have no apparent phenotype and no major defects in folding of newly synthesized proteins (Deuerling et al., 1999; Teter et al., 1999). However, in the absence of the Hsp70 chaperone DnaK, Δ*tig* mutants are inviable. Upon depletion of DnaK in Δ*tig* strains, more than 40 different newly synthesized cytosolic proteins aggregate, indicating a role for TF in folding of cytosolic proteins in cooperation with the DnaK chaperone system (see below).

TF binds in vitro with 1:1 stoichiometry to the large subunits of both translating and nontranslating ribosomes, indicating an intrinsic affinity for the translation apparatus. How the interactions of TF with the ribosome and the nascent chain are coupled with subsequent release of the chain from the ribosome and folding events is unclear. Several pieces of evidence point to a possible mode of action. The in vitro interaction with nontranslating ribosomes is salt sensitive, while the interaction with ribosome–nascent chain complexes is salt resistant. Furthermore, in translation extracts TF fails to cross-link to nascent chains after puromycin-induced release (Hesterkamp et al., 1996). TF may thus first associate with the nascent chain while bound to the ribosome, then migrate transiently with the nascent chain, followed by rapid dissociation and rebinding to ribosomes (Figure 1). Consistent with such cycling, TF is in 2-fold molar excess over ribosomes in the cytosol, with 50% ribosome-bound and 50% free.

The components of the welcoming committee differ among organisms, but similarities in their properties are intriguing. In both yeast and mammalian cells, a complex termed the nascent chain–associated complex (NAC) cross-links to nascent chains as short as 17 residues and protects about 30 C-terminal residues against proteolytic attack (Wang et al., 1995). These findings led to

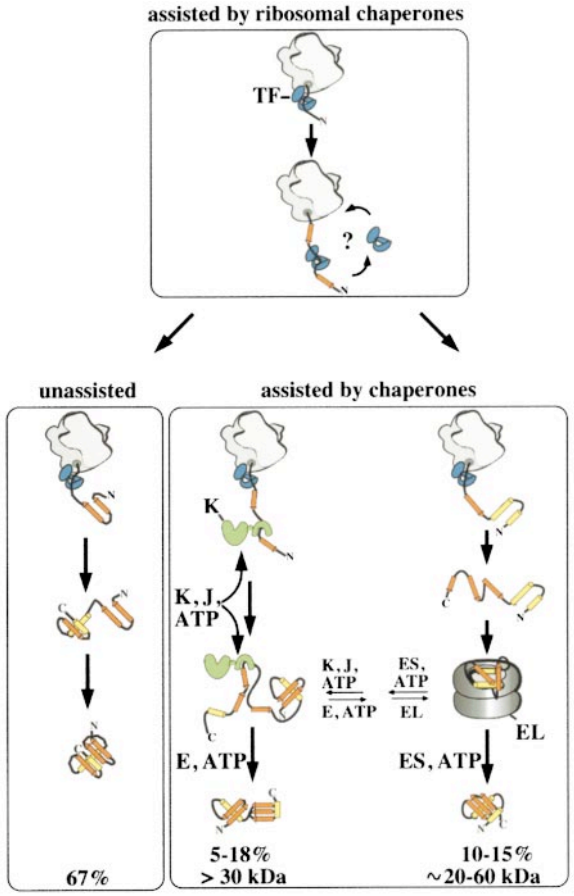

Figure 1. Model for Folding of Newly Synthesized Proteins in the Prokaryotic Cytosol

Ribosome-bound TF associates with emerging polypeptides and may transiently migrate with them. Subsequent folding of newly synthesized proteins occurs through different pathways (estimates of total protein molecules in %). Many proteins may fold without assistance by chaperones or with assistance by unknown factors. Specific subpopulations of substrates fold through assistance by either DnaK (K) with its DnaJ (J) and GrpE (E) cochaperones (co- and posttranslationally) or GroEL (EL) with its GroES (ES) cochaperone (posttranslationally). A small subset of these populations may require transfer between these two chaperone systems through a free folding intermediate to reach the native state.

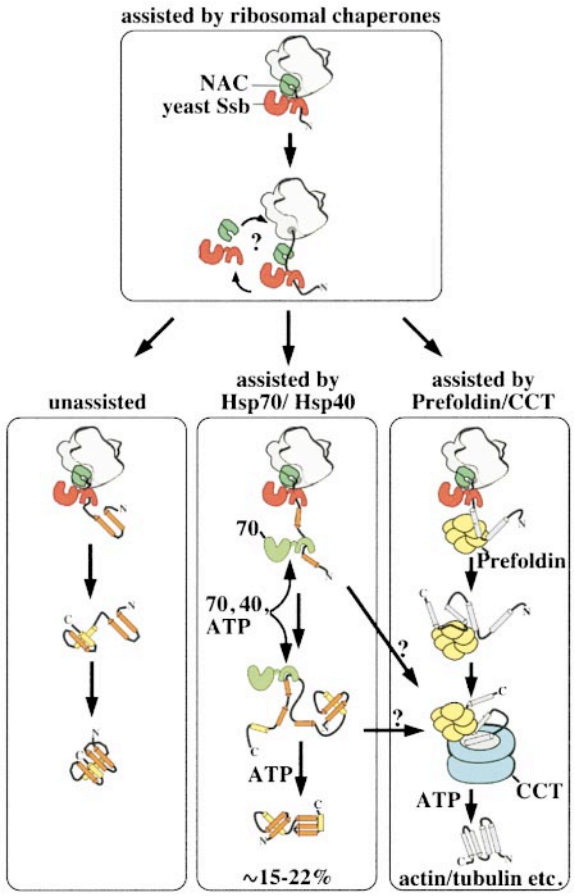

Figure 2. Model for Folding of Newly Synthesized Proteins in the Eukaryotic Cytosol

Ribosome-bound NAC and in yeast the Hsp70 Ssb associate with emerging polypeptides and, similar to TF, may transiently migrate with them. Some proteins may fold without assistance or with assistance by unknown factors. A subpopulation requires co- and posttranslational folding assistance by Hsp70 and Hsp40 (DnaJ) cochaperones. Other proteins, such as actin and tubulin require prefoldin and CCT for folding. Crossover between pathways may occur with Hsp70/Hsp40 passing newly synthesized substrates to CTT (shown) or Hsp90 (not shown).

the suggestion that NAC is deeply embedded in the ribosomal tunnel, constituting a dynamic component of the tunnel that provides a protected environment for nascent peptides. NAC does not form an immobilized roof of the channel. It can interact with regions of nascent chains that are 100 residues away from the peptidyl transferase center, and be released by puromycin together with nascent chains. Intriguingly, NAC is found in stoichiometric excess to ribosomes. Thus, similar to the proposed scenario for TF, NAC may cycle between ribosome-bound and free states (Figure 2).

Two specialized Hsp70 homologs, the 70 kDa Ssb proteins, are found in fungi. Ssbs of *S. cerevisiae* have been shown to associate with nontranslating ribosomes in salt-sensitive manner and with translating ribosomes in stable salt-resistant manner, and can be cross-linked in vitro to a variety of short nascent chains on the ribosome (Pfund et al., 1998). *SSB* mutants are hypersensitive to protein synthesis inhibitors, consistent with a role for Ssb in folding of newly translated proteins or ribosome metabolism. However, direct evidence of a role for Ssb in protein folding has yet to be solidified.

The stable association of Ssb with translating ribosomes is disrupted by release of nascent chains, suggesting that nascent chains are involved in this interaction. This suggestion was supported by demonstration that Ssb can be cross-linked to ribosome-associated nascent chains in vitro, but not to released nascent chains. Ssb may thus also undergo a nascent chain-controlled cycling between ribosome-bound and free states (Figure 2). The abundance of Ssbs, amounting to a 2–4 molar excess over ribosomes, is consistent with this possibility. Intriguingly, as Hsp70s normally function with an Hsp40 (DnaJ) cochaperone, the Hsp40 Zuotin

has been found stoichiometrically associated with ribosomes, and thus may function as a noncycling ribosome-bound partner of Ssb (Yan et al., 1998).

Much remains to be understood about the mechanism of action of ribosome-associated nascent chain binding proteins. In eukaryotes the story is particularly puzzling. Does NAC act as a general chaperone? Does NAC have other roles? Interestingly, roles in the sorting of organellar and membrane proteins have been reported for NAC, as well as TF. Also, there are obvious differences amongst eukaryotes. Fungi have the unusual ribosome-associated Hsp70 chaperone Ssb; mammalian cells apparently do not. Perhaps there are other, yet to be identified, ribosome-associated chaperones in mammalian cells. In addition, at present we can only speculate on the potential roles of the different components of the welcoming committee in protein folding. They may act to physically separate the folding domains from the surface of the ribosome, or postpone folding until a domain that can fold autonomously has entirely emerged from the tunnel, thus facilitating the cotranslational folding of proteins.

Finishing the Folding Job

Beyond their interaction with ribosome-associated factors, newly synthesized chains must continue to fold in the crowded cytosol. To assist folding, a second group of chaperones, acting in solution, begin their work. Nearly all species appear to have Hsp70 (DnaK, Ssa) and Hsp60 (GroEL, CCT) types of chaperones functioning to ensure the proper folding of newly synthesized polypeptides. A clearer picture of protein folding has emerged for *E. coli* than eukaryotes. While a majority of newly synthesized *E. coli* proteins may not require chaperone assistance for reaching native states (Figure 1), genetic and biochemical evidence indicates roles for the DnaK and GroEL systems in folding of many cytosolic proteins. DnaK is found associated with 5%–18% of newly synthesized protein molecules, a fraction of them in statu nascendi, indicating that DnaK interacts both co- and posttranslationally (Deuerling et al., 1999; Teter et al., 1999). The interaction for many proteins is transient ($\leq$2 min), but long-lasting for others, suggesting considerable differences in the number of ATPase-driven cycles of interactions involved in the maturation of different substrates.

Pioneering experiments implicated GroEL as a critical player in in vivo protein folding (Horwich et al., 1993). Shift of a conditional *groEL* mutant to a nonpermissive temperature resulted in the disappearance of a number of newly synthesized cytosolic proteins from the soluble cell fraction, and increased aggregation of at least nine proteins. More recent studies have identified GroEL substrates as pulse-labeled proteins that coimmunoprecipitated with GroEL ex vivo (Ewalt et al., 1997; Houry et al., 1999). 10%–15% of newly synthesized protein molecules, comprising about 300 protein species, interacted posttranslationally with GroEL. The majority of GroEL substrates are between 20 and 60 kDa, with only a minority ($\leq$20%) having higher molecular weights. Many transit GroEL within 10–30 s, indicating that they are associated for only one or a few ATPase-driven cycles of interaction. A significant fraction of the higher molecular weight proteins stay associated with GroEL for minutes. The significance of this observation to in vivo protein folding is unclear, particularly considering the size limitation of the GroEL/GroES folding chamber, which would be expected to only be able to accommodate a 60 kDa protein in its molten globule state having significant tertiary structure. When proteins larger than 60 kDa are associated with GroEL, the GroES cochaperone may be unable to seal this chamber, raising the question whether productive folding of these proteins can occur.

The 52 abundant GroEL substrates identified participate in a variety of basic cellular processes including translation, transcription, and metabolism (Houry et al., 1999). Analysis of structural data for 24 of these substrates revealed a common feature. They contain multiple $\alpha\beta$ domains with $\alpha\beta\alpha$ or $\alpha\beta$ sandwiches more frequently than the average *E. coli* protein. It was proposed that GroEL preferentially interacts with hydrophobic interfaces of the β domains, thereby stabilizing these domains until proper assembly of the sandwiches has occurred. Likely a wider spectrum of structures recruits GroEL for assistance in folding. In any case, the stringency of the GroEL requirement for folding of the identified substrates remains to be determined.

Less is understood about assisted protein folding in the eukaryotic cell. Hsp70 appears to be the major chaperone involved in assisted general folding pathways (Figure 2). The abundant soluble Hsp70 of the mammalian cytosol, which is closely related to the Ssa class of Hsp70s of the yeast cytosol, is the most prominent chaperone found to copurify with translating polysomes prepared from mammalian cells and to associate with stalled ribosome–nascent chains complexes in reticulocyte extracts. In addition, the specific activity of newly translated firefly luciferase is reduced by 60%–70% in reticulocyte lysates upon depletion of Hsp70 or its cochaperone Hsp40 (Frydman et al., 1994), indicating roles for this chaperone system in luciferase folding. However, it should be remembered that these experiments were performed in in vitro translation extracts where proteins fold more slowly, or rely on stalled or truncated nascent chains. Both situations may exacerbate misfolding, thus generating substrates for the repair function of Hsp70.

The most compelling in vivo evidence for an involvement of Hsp70 in the regular folding pathway of a cytosolic protein exists for ornithine transcarbamoylase (OTC). A yeast temperature-sensitive *SSA* mutant has a lowered specific activity of OTC synthesized after temperature upshift, compared to a wild-type strain, suggesting a role of Ssa in the folding of this protein (Kim et al., 1998). However, what fraction of cytosolic proteins in yeast and other eukaryotes recruit nonribosomal Hsp70 homologs as well as other chaperones for folding remains an open question.

Whether a non-Hsp70 type of general chaperone that interacts with a large number of diverse polypeptides exists in the eukaryotic cytosol is a matter of debate. Certainly the class II oligomeric chaperonin CCT (Tcp-1; TRiC; c-cpn), which is distantly related to GroEL, and its cochaperone prefoldin, are essential for folding of actin and tubulin (Figure 2). CCT may also play a more general role, as about 70 different protein species could be coprecipitated with CCT in pulse-labeling experiments (Thulasiraman et al., 1999). In addition, Hsp90 and its cochaperones are involved in the maturation of

at least several proteins involved in signal transduction, such as steroid hormone receptors and protein kinases.

Chaperone Pathways and Networks

How the chaperones described above function together in the cell to facilitate folding, as well as the parameters that determine whether or not a particular protein requires chaperone assistance, is unresolved. Some clues to the relationships among different chaperones have come from genetic analysis of *E. coli*. DnaK is essential for viability at heat shock temperatures, consistent with its essential protein repair function under these conditions. But, at normal growth temperatures Δ*dnaK* mutants grow well, provided suppressor mutations partially restore regulatory defects in heat shock gene expression. Such mutants show little alteration in solubility of cytosolic proteins. However, as mentioned above, DnaK becomes essential in cells lacking TF, suggesting overlapping or cooperative action of both chaperones. Supporting this view, the amount of newly synthesized proteins associated with DnaK doubles in Δ*tig* cells (Deuerling et al., 1999; Teter et al., 1999). The molecular basis for the functional cooperation between TF and DnaK is unclear. DnaK may mechanistically replace TF in its absence, although the lack of ribosome binding and prolyl isomerase activities argue against this possibility. Alternatively, DnaK may act co- and posttranslationally to repair misfolded proteins that may accumulate in Δ*tig* mutants.

The relationship between the DnaK and the GroEL systems also appears complex. These proteins have been proposed to be involved in a sequential folding pathway. This mode of action is supported by data that show similar kinetics for the release of rhodanese from DnaK and its accumulation on GroEL. Additionally, increased expression of GroEL results in both an increased dissociation of rhodanese from DnaK, as well as an increased dissociation from GroEL, suggesting an increase in the flux rate of rhodanese from DnaK to GroEL (Teter et al., 1999). Evidence for a sequential pathway has also been found in mammalian cells. Some nascent polypeptides appear to be "handed over" from Hsp70 to CCT and are not accessible to the bulk cytosol, as they are not available for binding by a GroEL "trap" that can bind, but not release substrate (Thulasiraman et al., 1999). However, such a rigid sequential pathway is not absolutely required, at least in *E. coli*, as Δ*dnaK* cells can achieve normal flux if GroEL levels are increased (Teter et al., 1999).

Furthermore, some data argue for an independent mode of action of DnaK and GroEL. The molecular weights of the proteins with which DnaK and GroEL preferentially interact differ, suggesting that the substrate populations of the two chaperones are, at least in part, distinct. DnaK substrates range from about 30 to >150 kDa, with a notable enrichment of proteins ≥60 kDa relative to the average cytosolic proteins of *E. coli*. This size distribution differs from that of GroEL substrates. Parallel folding pathways, which exhibit a significant degree of specificity may exist, with a subset of substrates recruiting DnaK and others GroEL for folding assistance. Other substrates may partition kinetically as nonnative conformers between DnaK and GroEL with either chaperone being sufficient to facilitate folding. Clearly in some cases two chaperones can substitute for one another, as evidenced by the DnaK and TF studies. A more detailed determination of whether the substrate populations of GroEL, TF, and DnaK overlap or are distinct, and the degree by which these substrates stringently require the corresponding chaperone for folding, will be important in a final understanding of the dynamics and synergism of chaperone action.

Summary

Recent progress made in the analysis of in vivo folding of cytosolic proteins suggests that folding of cytosolic proteins occurs via multiple chaperone-assisted, as well as unassisted pathways. In the case of individual proteins, certain pathways may be highly favored. But, the cellular folding machinery also shows significant redundancy and flexibility, resulting in a variable network of folding pathways having alternative routes and backup systems.

Selected Reading

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B., and Steitz, T.A. (1999). Nature *400*, 841–847.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Nature *400*, 693–696.

Ewalt, K.L., Hendrick, J.P., Houry, W.A., and Hartl, F.U. (1997). Cell *90*, 491–500.

Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F.U. (1994). Nature *370*, 111–117.

Hesterkamp, T., Hauser, S., Lütcke, H., and Bukau, B. (1996). Proc. Natl. Acad. Sci. USA *93*, 4437–4441.

Horwich, A.L., Low, K.B., Fenton, W.A., Hirshfield, I.N., and Furtak, K. (1993). Cell *74*, 909–917.

Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F.U. (1999). Nature *402*, 147–154.

Kim, S., Schilke, B., Craig, E.A., and Horwich, A.L. (1998). Proc. Natl. Acad. Sci. USA *95*, 12860–12865.

Netzer, W.J., and Hartl, F.U. (1997). Nature *388*, 343–349.

Nicola, A.V., Chen, W., and Helenius, A. (1999). Nat. Cell Biol. *1*, 341–345.

Pfund, C., Lopez-Hoyo, N., Ziegelhoffer, T., Schilke, B.A., Lopez-Buesa, P., Walter, W.A., Wiedmann, M., and Craig, E.A. (1998). EMBO J. *17*, 3981–3989.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., and Hartl, F.U. (1999). Cell *97*, 755–765.

Thulasiraman, V., Yang, C.F., and Frydman, J. (1999). EMBO J. *18*, 85–95.

Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). EMBO J. *14*, 5494–5505.

Wang, S., Sakai, H., and Wiedmann, M. (1995). J. Cell Biol. *130*, 519–528.

Yan, W., Schilke, B., Pfund, C., Walter, W., Kim, S., and Craig, E.A. (1998). EMBO J. *17*, 4809–4817.