# Chaperone substrates inside the cell

Biochemistry as a discipline is distinguished by the first act of its practitioners – they convert cells into a soup. Studies of the molecules purified from this soup serve to define the spectrum of possible functions these molecules carry out within the living cell, but such studies rarely duplicate closely the intracellular conditions under which these molecules have evolved to function. For this reason, biochemists have to question continually the relevance of such *in vitro* systems to the living cell. As the late David Green pointed out, a clever engineer can make a vacuum cleaner from the wreck of an automobile, but this does not show that cars contain vacuum cleaners. Only *in vivo* experiments can determine where within the spectrum of possible functions the actual biological functions of a molecule or set of molecules reside. Two recent papers[1,2] address the problem of determining which, among ~4300 polypeptides encoded by the *Escherichia coli* genome, bind to the GroEL and DnaK molecular chaperones in that organism. Why is this an interesting problem?

## Two families of promiscuous chaperones

Some families of molecular chaperones are specific for their substrates. The PapD-like chaperones, for example, provide steric information necessary for the correct folding of the subunits of the adhesive pili of some bacteria[3]. The *papD* gene can be deleted with no effect on the cell other than the loss of the ability to assemble pili.

Other chaperone families prevent identical or similar polypeptide chains from aggregating, the chief hazard faced by some partly folded polypeptides. Aggregation is commonly observed during the *in vitro* refolding of some pure denatured proteins, and is greatly enhanced by the addition of high concentrations of macromolecules that mimic the crowded state of the cytoplasm[4]. Aggregation in the cell results from the transient exposure of hydrophobic regions during polypeptide chain synthesis by polysomes and after the unfolding of mature proteins caused by environmental stresses such as heat shock. This problem is combated partly by the chaperonin and the heat-shock protein 70 families, exemplified in *E. coli* by GroEL and DnaK proteins, respectively.

These chaperones act in a nonsteric fashion, and thus have very little substrate specificity. The oligomeric structure of each GroEL molecule contains a central cage that provides a sequestered environment where a partly folded chain, up to at least 57 kDa in size[5], can bind. Binding of GroES and ATP to GroEL then releases this chain from the walls into the cage lumen, where it continues to fold until the danger of aggregation with similar chains has passed. The more folded chain is then released into the cytoplasm after dissociation of the GroES that caps the cage[6–8]. DnaK, being monomeric, is too small to form a folding cage, but instead binds to short hydrophobic stretches (approximately seven residues) as they appear on nascent polypeptides and on the surface of mature proteins denatured by stress. This binding temporarily shields these exposed stretches, thus reducing their tendency to aggregate[6–8].

Both GroEL and DnaK will bind *in vitro* to a wide variety of partly folded polypeptides as they detect exposed hydrophobic regions rather than specific primary structures. GroEL will even bind to denatured hen egg lysozyme[9], a protein it never normally encounters because lysozyme folds *in vivo* inside the lumen of the endoplasmic reticulum, which contains no chaperonin homologue. Deletion of the GroEL gene from *E. coli* is lethal at all growth temperatures, whereas deletion of the DnaK gene causes loss of the ability of the cell to withstand heat shock, but allows growth at normal temperature. Presumably, among the *in vivo* substrates of these chaperones, some are essential for cell survival under either normal or heat-shock conditions. How can these natural substrates be identified?

An average *E. coli* cell growing logarithmically at 30–37°C on a glucose minimal medium contains ~$2.35 \times 10^6$ polypeptide chains, with an average size of ~35 kDa (317 residues)[10]. The total number of different types of soluble polypeptide chain found in the cytoplasm[11] is ~2600, so the problem addressed by the laboratories of Ulrich Hartl and Bernd Bukau, respectively, was to identify within this large and heterogeneous population those polypeptides that bind either to GroEL immediately after synthesis[1] or to DnaK after heat shock[2]. The long-term aim is to determine which features of these natural substrates are recognized by each chaperone.

## *In vivo* methods

Two methods used to check *in vitro* models against *in vivo* reality employ specific antibodies or mutants. Those proteins that bind inside the cell to the GroEL chaperone were determined by pulse-labelling cells with $^{35}$S-methionine at 30°C, treating cell extracts with an antibody specific for GroEL, and analysing the immunoprecipitated polypeptides on one- and two-dimensional (2D) polyacrylamide gels[1,12]. 2D gels resolve ~1000 polypeptides out of a total of ~4300 encoded in the *E. coli* genome. It is essential in such experiments to determine whether proteins bind to GroEL inside the cell or during the extraction procedures, or both. This was achieved in two ways. First, addition during cell lysis of a mutant form of GroEL that binds but cannot release unfolded polypeptides fails to alter the pattern of polypeptides bound to wild-type GroEL. Second, addition of GroES and ATP to the extract causes bound labelled polypeptides to be released from the walls of GroEL into the cage, and then into the medium; these polypeptides do not rebind to GroEL in the extract.

At the end of a 15-sec pulse with $^{35}$S methionine, 250–300 labelled polypeptides bound to GroEL are resolved on 2D gels. This pulse duration is sufficient to allow the synthesis of chains up to 300 residues in length. Some of these bound polypeptides are released completely from GroEL during a subsequent chase, with time constants between 20 sec and 2 min, but about one-third persist for longer in the bound state. Experiments using much longer pulse and chase times suggest that this latter group of proteins include some mature proteins that continue to cycle on and off GroEL, even after their initial folding is complete. The authors conclude that these are proteins that need continual conformational maintenance because their folded states are only marginally stable; such states readily unfold into conformations recognized by GroEL, especially when the cells are stressed by heat-shock treatment[1].



0968–0004/00/ – See front matter  PII: S0968-0004(00)01576-0

Those proteins that bind to DnaK inside the cell were determined by pulse-labelling as for GroEL, but using mutant cells containing a deletion mutation for DnaK and a heat shock of 42°C for 60 min applied after the pulse[2]. Cell extracts were then centrifuged and the pellet analysed on 2D gels to determine how many proteins have aggregated. Extracts from wild-type cells show no aggregation products attributable to this heat treatment, whereas mutant cells kept at 30°C show slightly increased aggregation of only a few polypeptides. By contrast, in mutant cells exposed to heat shock, 150–200 polypeptides show varying degrees of aggregation[2]. At least 80% of these polypeptides are identical to those that aggregate when cell-free extracts of wild-type cells are exposed to 45°C for 15 min. Such *in vitro* aggregation is prevented by the addition of physiological concentrations of DnaK and ATP to the extract before the heat treatment, but not significantly by addition of other chaperones, including GroEL and GroES. The authors conclude that DnaK is the major chaperone in *E. coli* that prevents the aggregation of a wide range of heat-labile proteins, and that the proteins aggregating in the heat-treated extracts are the natural substrates for DnaK inside the cell[2].

Combination of these observations for the two chaperone systems in *E. coli* confirms earlier suggestions that the primary role of GroEL is to assist the folding of some newly synthesized proteins, whereas that of DnaK is to protect some proteins from aggregating at high temperatures. This conclusion does not exclude the possibility that these chaperones also exhibit some overlap of function; overproduction of GroEL and GroES partly complements the growth defect of DnaK mutant cells at 40°C, but not at 42°C. Another example of functional overlap between chaperones is known for DnaK; this chaperone binds to nascent chains in wild-type cells at normal temperatures, but viability is not affected in DnaK-deletion mutants unless the gene for the trigger-factor protein that binds to nascent chains is also deleted[13,14]. Thus, DnaK does have an essential function in the folding of newly synthesized chains, but only in collaboration with trigger factor. Such redundancy of control systems could be a useful survival strategy for cells, as it is in jumbo jets, for example. In the case of GroEL, the authors point out that, although their studies identify some of the proteins that bind to GroEL *in vivo*, they do not establish to what extent this binding is essential for these proteins to fold correctly inside the cell[1].

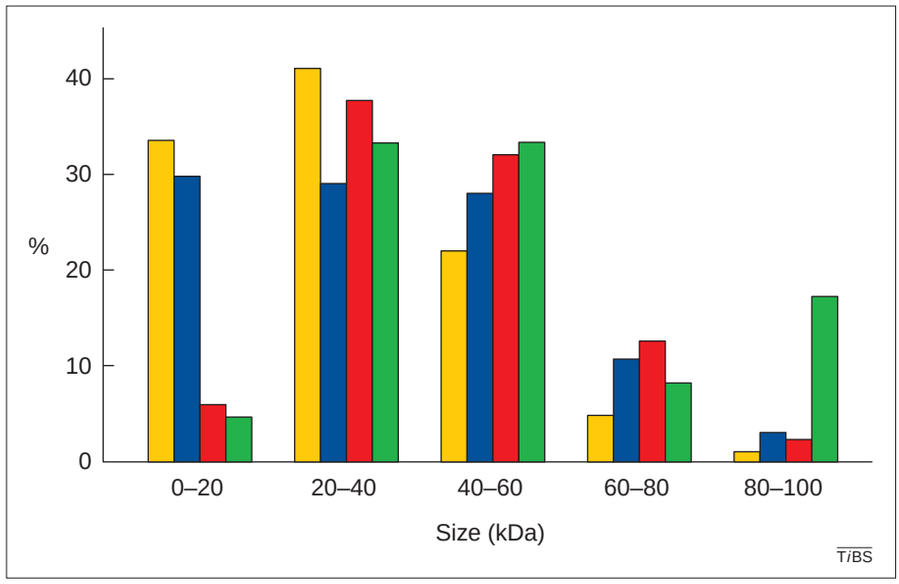


**Figure 1**

Size distribution of polypeptides that bind to either GroEL or DnaK inside *E. coli*. The columns indicate the size distribution of polypeptides that are encoded in the total genome (yellow); are labelled by $^{35}$S-methionine in a 15-sec pulse of cells growing at 30°C (blue); bind to GroEL inside cells growing at 30°C (red); bind to DnaK inside cells after a heat shock of 42°C for 60 min (green). The distributions were calculated from data presented in Refs 1,2 and 11.

## The importance of size

Polypeptides bound to either GroEL or DnaK were analysed by treatment of polypeptides from 2D gels with proteolytic enzymes, determining the exact mass of the peptides produced by mass spectrometry, and comparison with the database derived from the total genome sequence of *E. coli*. These procedures identified 52 of the GroEL-bound proteins and 57 of the proteins that aggregate in DnaK-mutant cells treated at 42°C. Both lists contain proteins essential for transcription, translation and metabolism and, intriguingly, share eleven proteins; these are the products of the *rpoA*, *rpoB*, *metE*, *metF*, *glf*, *gatY*, *minD*, *g3p1*, *cplX*, *thrS* and *nusG* genes. As yet, there has been no identification of those nascent and newly synthesized polypeptides that bind to DnaK at normal growth temperatures; thus, it is possible that the eleven gene products that bind to GroEL after translation also bind to DnaK at the nascent-chain stage.

No correlation is obvious between the chaperone substrates and their pI, cellular function or oligomeric state. However, there is some correlation with the size of the polypeptide chains for both chaperones (see Fig. 1). About one-third of the total proteins encoded in the *E. coli* genome are in the 0–20 kDa range, but only 5–6% of either the GroEL or DnaK substrates fall in this range. Most GroEL and DnaK substrates (72% and 69%, respectively) lie in the 20–60 kDa range, compared with 60% of the total number of polypeptides. In contrast, whereas only 1.5% of all *E. coli* polypeptides are >80 kDa, ~17% of DnaK substrates fall in this range. What can we say about these distributions?

The most striking conclusion from this analysis of DnaK substrates is that the majority of the larger proteins in *E. coli* are thermolabile, even at 42°C, a temperature at which the cells continue to grow. A major function of the DnaK chaperone system is thus to prevent these proteins from aggregating. Larger proteins could be more prone to aggregation than smaller proteins because they fold more slowly, contain larger hydrophobic patches and are composed of more domains than smaller proteins[2]. Surface contacts between domains could expose hydrophobic surfaces more readily than more deeply buried contacts when exposed to high temperatures. Those proteins that aggregate inside the cells despite the best preventative efforts of DnaK can be subsequently solubilized *in vivo* by the combined action of the DnaK system with the ClpB chaperone[2,15]; similar observations have been reported for yeast[16].

Size is important for GroEL substrates in a different sense than for DnaK substrates. The volume of the GroEL folding cage imposes an upper limit of ~57 kDa on the polypeptides it can accommodate[5]. Only ~15% of the observed GroEL-bound polypeptides are larger than this (Fig. 1); these species exhibit very slow release kinetics from GroEL *in vivo*[1]. However, as well as size, there is some evidence for certain preferred structural features in GroEL substrates.

## Structural features of GroEL substrates

No obvious consensus sequences are present in the 52 identified GroEL substrates, but interrogation of the protein structure databases reveals that 24 of them have known tertiary structures or are homologues of such structures[1]. These structures show a preference for two or three αβ domains. It is estimated that *E. coli* contains between 200 and 600 cytoplasmic proteins with multiple αβ domains. Given that less than 10% of the ~300 observed GroEL substrates have known structures, the strength of this preference awaits identification of a larger sample of these substrates. The authors do not rule out that other types of domains will be found in GroEL substrates, and their list includes one all-α protein in the form of ferritin[1]. However, a preference for multidomain proteins can be rationalized in terms of the slow folding rates and propensity to misfolding of such proteins[17]; these proteins might also be especially prone to aggregation as a result of 3D domain swapping, in which one domain binds to its partner domain in another identical molecule instead of in the same molecule[18]. Such a combination of properties suggests that these proteins in particular need the chaperone function of GroEL.

Future work is expected to extend the list of identified substrates for GroEL and DnaK. It will be of special interest to compare the substrates of these two chaperones under the same growth conditions. These reports should encourage more laboratories to try similar *in vivo* methods to help unravel what chaperones actually do inside the cell, rather that what they can be persuaded to do in the test tube.

## Acknowledgements

I thank Bernd Bukau, Chris Dobson, David Jones, Walid Houry, Pierre Goloubinoff, Peter Lund and Ulrich Hartl for helpful discussion, and Howard Dalton for providing encouragement and office facilities.

## References

1 Houry, W.A. *et al.* (1999) Identification of *in vivo* substrates of the chaperonin GroEL. *Nature* 402, 147–154
2 Mogk, A. *et al.* (1999) Identification of thermolabile *E. coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* 18, 6934–6949
3 Sauer, F.G. *et al.* (2000) PapD-like chaperones and pilus biogenesis. *Semin. Cell Develop. Biol.* 11, 27–34
4 van den Berg, B. *et al.* (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* 18, 6927–6933
5 Sakikawa, C. *et al.* (1999) On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J. Biol. Chem.* 274, 21251–21256
6 Ellis, R.J. and Hartl, F.U. (1999) Principles of protein folding in the cellular environment. *Curr. Opin. Struct. Biol.* 9, 102–110
7 Bukau, B. and Horwich, A.L. (1998) The hsp70 and hsp60 chaperone machines. *Cell* 92, 351–366
8 Agashe, V.R. and Hartl, F.U. (2000) Roles of molecular chaperones in cytoplasmic protein folding. *Semin. Cell Develop. Biol.* 11, 15–26
9 Coyle, J.E. *et al.* (1997) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Fold. Design* 2, R93–R104
10 Bremer, H. and Dennis, P.D. (1987) Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology* (Vol. 2) (Neidhardt, F.C., ed.), pp. 1527–1542, American Society for Microbiology, Washington, USA
11 Netzer, W.J. and Hartl, F.U. (1998) Protein folding in the cytosol: chaperonin-dependent and independent mechanisms. *Trends Biochem. Sci.* 23, 68–73
12 Ewalt, K.L. *et al.* (1997) *In vivo* observation of polypeptide flux through the bacterial chaperonin system. *Cell* 90, 491–500
13 Deuerling, E. *et al.* (1999) Trigger factor and DnaK cooperate in the folding of newly synthesized proteins. *Nature* 400, 693–696
14 Teter, S.A. *et al.* (1999) Polypeptide flux through bacterial hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* 97, 755–765
15 Goloubinoff, P. *et al.* (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13732–13737
16 Glover, J.R. and Lindquist, S. (1998) Hsp104, hsp70 and hsp40; a novel chaperone system that rescues previously aggregated proteins. *Cell* 94, 73–82
17 Garel, J-R. (1992) Large multidomain and multisubunit proteins. In *Protein Folding* (Creighton, T.E., ed.), pp. 405–454, W.H. Freeman
18 Schlunegger, M.P. *et al.* (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv. Protein Chem.* 50, 61–122

**R. JOHN ELLIS**

Dept of Biological Sciences, University of Warwick, Coventry, UK CV4 7AL.

# The changing face of mitochondrial research

Biochemical research tends to move forward at an uneven pace that depends on the development of new technologies, on the novel insights of investigators involved, and even on the vagaries of the funding sources available for research projects. Nowhere is this more evident than in the area of mitochondrial research. What started out as a quest to define the principles of biological energy conservation has largely been replaced by attempts to understand the molecular basis of several human diseases. As a result, mitochondria are receiving increased attention from investigators with varied interests and backgrounds. Novel techniques are now being used to study both mitochondrial structure and functioning, and these methods are leading to a redefining of long-held views about the basic biology of the organelle. *TiBS* has covered new developments in mitochondrial research from time to time (see for example Refs 1–3). Over the next few months, the changing face of research on this organelle is to be covered in some depth.

Work in the late 1940s and early 1950s by David Green, Albert Lehninger and others first demonstrated that mitochondria are the centers of energy metabolism within the cell[4]. Since then, isolated mitochondria have been studied extensively to understand the mechanism of energy conservation, especially the process of oxidative phosphorylation. This work has already resulted in several landmarks in biochemistry, beginning with the development of the concept of chemi-osmosis[5], which was rewarded with a Nobel Prize to Peter Mitchell. A more recent *tour de force* has been the structure determination of most of the respiratory chain. Very few membrane proteins have been crystallized, but these include the $bc_1$ complex and cytochrome *c* oxidase[6,7]. Both are large complexes, 200 000 Daltons, and their high-resolution structure determinations have pushed the limits of X-ray crystallographic methods. Additionally, there is now rapid progress in determining a high-resolution structure of the ATP synthase[8], the other



 PII: S0968-0004(00)01584-X