Mechanism of substrate recognition by the chaperonin GroEL

Walid A. Houry

Abstract: The bacterial chaperonin GroEL functions with its cofactor GroES in assisting the folding of a wide range of proteins in an ATP-dependent manner. GroEL–GroES constitute one of the main chaperone systems in the Escherichia coli cytoplasm. The chaperonin facilitates protein folding by enclosing substrate proteins in a cage defined by the GroEL cylinder and the GroES cap where folding can take place in a protected environment. The in vivo role of GroEL has recently been elucidated. GroEL is found to interact with 10–15% of newly synthesized proteins, with a strong preference for proteins in the molecular weight range of 20–60 kDa. A large number of GroEL substrates have been identified and were found to preferentially contain proteins with multiple αβ domains that have α-helices and β-sheets with extensive hydrophobic surfaces. Based on the preferential binding of GroEL to these proteins and structural and biochemical data, a model of substrate recognition by GroEL is proposed. According to this model, binding takes place preferentially between the hydrophobic residues in the apical domains of GroEL and the hydrophobic faces exposed by the β-sheets or α-helices in the αβ domains of protein substrates.

Key words: chaperone, folding, binding, hydrophobic interaction, structure.

Introduction

The successful folding of newly translated proteins is essential to the viability of the cell. Since the environment in the cell is very crowded (Fulton 1982; Goodsell 1991) and viscous (Swaminathan et al. 1997; Partikian et al. 1998; Elowitz et al. 1999), there is a high probability that newly translated proteins will misfold and aggregate. Consequently, the cell employs quality-control mechanisms to ensure that the folding process of any newly synthesized polypeptide chain results in the formation of properly folded protein and that the folded protein is maintained in an active conformation throughout its functional lifetime (Netzer and Hartl 1998; Horwich et al. 1999; Wickner et al. 1999). Molecular chaperones are key components of this quality-control machinery. They are a fundamental group of proteins which have been identified only relatively recently (Ellis and Hemmingsen 1989). They have been shown to play essential roles in cell viability under both normal and stress conditions by assisting in the folding of newly translated proteins and by maintaining the conformational integrity of preexisting proteins (Hartl 1996).

Molecular chaperones are known to be very promiscuous in vitro (Bukau et al. 1996; Coyle et al. 1997). In a test tube, they have the ability to bind to a wide range of proteins in several conformational states. In vivo, however, there is
mounting evidence that chaperones have a high affinity for a defined set of proteins and therefore preferentially assist in the folding and conformational maintenance of only this set of proteins. Key advances have recently been made in characterizing the mechanism of function of the bacterial chaperonin GroEL in vitro and in vivo.

**GroEL function in vitro**

The chaperonin GroEL with its cofactor GroES are the typical representatives of the Hsp 60 and Hsp 10 families of molecular chaperones, respectively (Hsp refers to proteins induced by heat shock). GroEL–GroES is the only essential chaperone system in *Escherichia coli* cytoplasm under all growth conditions (Fayet et al. 1989). The structures of GroEL, GroES, and the GroEL–GroES complex have been extensively investigated and have been visualized at atomic resolution by X-ray crystallography (Braig et al. 1994; Hunt et al. 1996; Xu et al. 1997) (Fig. 1). GroEL is a homooligomer of fourteen 57-kDa subunits. The subunits are arranged into two heptameric rings forming a cylindrical structure with two large cavities. GroES is a heptamer of 10-kDa subunits, and it forms a cap over the cavities in the GroEL cylinder. The GroEL subunit is divided into apical (residues 191–376), intermediate (residues 134–190 and 377–408), and equatorial domains (residues 1–133 and 409–547). Protein binding takes place at the outermost apical domains. Residues in the apical domain which are important for substrate binding are predominantly hydrophobic (Fenton et al. 1994). These residues form a hydrophobic patch at the rim of each of the GroEL rings. The intermediate domain forms a short flexible linker between the apical and equatorial domains, and the equatorial domains form the contact region between the two GroEL rings. ATP binding and hydrolysis take place at the equatorial domains. The apical and intermediate domains of GroEL undergo large conformational changes as a result of nucleotide and GroES binding (Fig. 1).

The mechanism of GroEL function is well established in vitro (Fenton and Horwich 1997; Martin and Hartl 1997; Sigler et al. 1998; Rye et al. 1999) (Fig. 2). The substrate–acceptor state of GroEL is thought to be an ADP bullet in which GroES and seven ADPs are bound to the same ring of GroEL. This ring might also contain a previously encapsulated polypeptide. Another substrate protein then binds to the apical domains of the opposite free ring of GroEL in a non-native compact conformation mainly through hydrophobic interactions to hydrophobic residues in helices H (residues 234–243) and I (residues 256–268) and a loop region (residues 196–211) (Fenton et al. 1994). ATP and then GroES subsequently bind to the newly occupied ring of GroEL, resulting in displacement of the protein into an enclosed cavity defined by the GroEL cylinder and the GroES cap. This also results in the release, from the opposite ring, of ADP and GroES and any previously encapsulated polypeptide. The folding process of the new substrate then takes place in the enclosed cage in which aggregation and other possible unproductive interactions are prevented. After about 10–20 s of GroES binding, ATP on the polypeptide-containing ring is hydrolyzed to ADP. This results in the formation of a new ADP bullet which can act as the acceptor state for another substrate molecule. ATP and GroES then bind to the opposite ring, resulting in the release of ADP, GroES, and substrate protein from the distal ring. The cycle then repeats as described above. Some proteins require multiple rounds of binding and release to reach their native state.

There is some controversy regarding the role that GroEL plays in the actual folding process of substrate proteins. GroEL is usually observed to retard the folding of proteins which can fold rapidly in vitro (Itzhaki et al. 1995; Katsumata et al. 1996; Tsurupa et al. 1998). It has also been observed to accelerate the refolding rate of those proteins which fold slowly and are prone to aggregation (Martin et al. 1991; Ben-Zvi et al. 1998; Beissinger et al. 1999). Furthermore, it has been proposed that GroEL can actively unfold substrate proteins during its functional cycle (Walter et al. 1996; Shiltzerman et al. 1999). In any case, there seems to be a general consensus that, even though GroEL might affect the folding rate of proteins, it does not change the folding mechanism of these proteins (Coyle et al. 1999). It only provides a folding cage without contributing any steric information to the folding process itself.

**GroEL function in vivo**

Although the in vitro mechanism of GroEL function is well studied, the role of GroEL in folding newly translated proteins or in the maintenance of preexisting proteins in vivo is poorly understood, until recently. My colleagues and I wanted to quantitatively analyze the contribution of GroEL to protein folding in the *E. coli* cytoplasm under normal and heat-shock conditions and identify those *E. coli* proteins which utilize GroEL for folding and (or) maintenance. The strategy employed was to visualize those proteins which directly interacted with GroEL in vivo. To this end, affinity-purified anti-GroEL antibodies were used to immunoprecipitate GroEL-substrate complexes from *E. coli* cells in mid-log phase (Ewalt et al. 1997; Houry et al. 1999). Cells were converted to spheroplasts by treatment with lysozyme and then lysed gently in a hypertonic buffer in the presence of EDTA, which stopped the ATP-dependent polypeptide release from GroEL. Several control experiments were carried out to ensure that immunoprecipitations were specific. Antibodies used were shown to be specific in recognizing only the GroEL band by treating the cell lysate with 1% SDS, diluting to 0.1% SDS, and then immunoprecipitating using anti-GroEL antibodies. Only GroEL was immunoprecipitated after this procedure and no other protein. Moreover, protein bands which communoprecipitated with GroEL were shown to be true substrates of GroEL, since upon addition of GroES and ATP regenerating system to the cell lysate, the majority of these proteins were released from GroEL. Also, the addition of noncycling GroEL mutant, which can bind substrates but does not release them (Weissman et al. 1994), during cell lysis did not change the amount or pattern of GroEL-bound polypeptides, indicating that the binding to GroEL occurred in the intact cell and not during or after lysis (Ewalt et al. 1997). It should be noted that no GroES was detected in our anti-GroEL immunoprecipitations in the presence of EDTA, in agreement with in vitro observa-
tions made using purified proteins which demonstrated that the binding of GroES to GroEL requires the presence of nucleotides (Martin et al. 1993; Todd et al. 1994).

To investigate how newly translated proteins utilize GroEL for folding, the kinetics of passage of these proteins through GroEL (flux kinetics) was followed by carrying out pulse-chase experiments on *E. coli* cells in mid-log phase under normal and heat-shock conditions. Cells were pulsed with radiolabelled methionine for a short time of about 15 s and then chased with nonradiolabelled methionine. At different time points after the chase, cells were lysed rapidly and GroEL–substrate complexes were immunoprecipitated as indicated previously. Complexes were then separated on SDS-PAGE (one-dimensional) or on two-dimensional gels and visualized by phosphorimaging analysis. One-dimensional gels enabled us to follow the flux kinetics of the total set of proteins, and two-dimensional gels enabled us to follow the flux kinetics of individual proteins. Similar experiments were carried out on steady-state labelled cells to determine the role of GroEL in the conformational maintenance of pre-existing proteins. Also, model proteins were overexpressed in vivo or in *E. coli* translation extracts to verify observations made with bulk proteins. The techniques employed proved to be very powerful in elucidating GroEL function in vivo. The following major conclusions were reached:

1. Three classes of proteins are present in the cell cytoplasm: proteins with a chaperonin-independent folding pathway, proteins with an intermediate chaperonin dependence for which normally only a small fraction transits GroEL, and highly chaperonin dependent proteins which require sequestration of aggregation-sensitive intermediates within the GroEL cavity for successful folding.

2. About 300 proteins, representing 10–15% of total cytoplasmic *E. coli* proteins, utilize GroEL for de novo folding under normal growth conditions and about twice as much under heat-stress conditions.

3. GroEL preferentially interacts with proteins of molecular weight between 20 and 60 kDa, as compared to cytoplasmic *E. coli* proteins. About two-thirds of these proteins flux through GroEL with time constants between 20 s and 2 min, reflecting the requirement of one to several rounds of GroEL binding and release to reach their native state. For the other substrates, a fraction of the population of a particular protein remains associated with GroEL after the chase. This later set of proteins represents preexisting proteins that require repeated cycles of GroEL binding and release for conformational maintenance during their lifetime in the cell. This set of proteins was found to be relatively unstable in vivo, since, under heat shock, the same set interacted more extensively with the chaperonin. Hence, GroEL is important for folding newly translated proteins and for conformational maintenance of preexisting proteins.

**Substrate recognition by GroEL**

Although purified GroEL is able to bind in vitro to ~50% of soluble *E. coli* proteins in their unfolded or partially folded states (Viitanen et al. 1992), the finding that only ~300 proteins
require the chaperonin for folding in vivo indicates that GroEL, in the cell, preferentially recognizes features present mainly in this small fraction of proteins. It was necessary to identify these proteins to determine the features that distinguish GroEL substrates from the rest of *E. coli* proteins. To this end, large-scale immunoprecipitation of GroEL–substrate complexes from *E. coli* cells at mid-log phase was carried out. These complexes were separated on two-dimensional gels and protein spots were identified by peptide-mass fingerprint analysis using MALDI-TOF (Houry et al. 1999). Fifty-two proteins were identified unequivocally using this procedure. Substrates included essential components of the transcription translation machinery and enzymes involved in several metabolic pathways.

Sequence analysis failed to reveal statistically significant consensus sequences among the 52 identified substrates, in agreement with in vitro studies which also did not find a consensus sequence among model peptides that bind to GroEL (Coyle et al. 1997). However, 24 out of the 52 GroEL substrates were amenable to structural classification using domain classification databases such as SCOP (Hubbard et al. 1999) or CATH (Orengo et al. 1999). This classification demonstrated that GroEL substrates preferentially contained two or more αβ domains compared to soluble *E. coli* proteins (Fig. 3). The preference was statistically significant at the 95% confidence limit. The most common domain architectures in GroEL substrates were those of the three-layer αββ and the two-layer αβ sandwiches. These domains typically have α-helices and β-sheets with extensive hydrophobic surfaces. These surfaces are ideally suited for recognition by the hydrophobic residues in the apical binding domains of GroEL.

For αβ proteins, the formation of the β-sheet is expected to be a difficult step in the folding process since, unlike the formation of an α-helix, the assembly of the sheet requires the formation of proper long-range contacts in the proper orientation. This is further complicated by the fact that the hydrophobic surface of the assembled sheet has to be buried away from the surrounding solvent and hence has to be stabilized by the packing of helices, or some other secondary structure elements, against the sheet surface through hydrophobic interactions. Misfolding can occur during the folding process either through the improper packing of helices against the sheet within one domain or the packing of helices of one domain against a sheet in another domain. Alternatively, helices or sheets with exposed hydrophobic surfaces in one protein molecule might pack against a sheet in another protein molecule, leading to aggregation.

Since the identified GroEL substrates have preferentially multiple αβ domains compared to *E. coli* proteins, they are expected to fold slowly (Plaxco et al. 1998) and to be structurally labile. Mogk et al. (1999) recently identified 57 thermolabile *E. coli* proteins. Significantly, 11 of these proteins are also GroEL substrates: CLPX_ECOLI, G3P1_ECOLI, GATY_ECOLI, GLF_ECOLI, METE_ECOLI, METF_ECOLI, MIND_ECOLI, NUSG_ECOLI, RPOA_ECOLI, RPOB_ECOLI, and SYT_ECOLI. Furthermore, 24 of the thermolabile proteins were amenable to structural classification using the SCOP database (Hubbard et al. 1999). It was striking to find that their set of thermolabile proteins was structurally very similar to the set of 24 GroEL substrates (Fig. 3). Both sets were enriched with proteins containing multiple αβ domains compared to *E. coli* proteins. It should be noted, however, that the average protein size of the 24 thermolabile *E. coli* proteins and the 24 GroEL substrates is 425 and 328 amino acids, respectively. This might indicate that GroEL’s preference for binding substrates with multiple αβ domains is even stronger than that dictated by random selection of thermolabile proteins. The significance of this finding will be verified once the structures of more *E. coli* proteins are known.

The apical domains of GroEL can assist in the folding of such multidomain αβ proteins in several ways. Residues implicated in polypeptide binding in the apical domains of GroEL could provide a scaffold on which the hydrophobic β-sheet(s) or amphipathic helices of the substrate protein can assemble; subsequent folding of the protein could then proceed in the enclosed GroEL–GroES cavity. Alternatively, if a wrong or improper packing occurred between helices and β-sheet(s) in a protein, the apical domains of GroEL could interfere by displacing the helices and β-sheet(s) and then binding to the exposed hydrophobic surfaces. Upon GroES binding to GroEL, the apical domains would then be released from these hydrophobic faces, and the protein would attempt to refold to achieve the right packing interactions.

In this regard, it has been demonstrated biochemically that the binding of unfolded rhodanese, malate dehydrogenase, and rubisco to the chaperonin requires recognition by multiple GroEL apical domains (Farr et al. 2000). All three
proteins are stringent substrates of GroEL, requiring GroES and ATP for efficient folding, and all three consist of two \(\alpha\beta\) domains. In support of the biochemical data, the structural data of Wang et al. (2000) indicate that the large size of the \(\alpha\beta\) domains in the identified substrates would indeed require more than three GroEL apical domains for binding.

The scheme shown in Fig. 4 provides our working model for the binding between the apical domains of GroEL and the hydrophobic face of the sheet(s) or helices in \(\alpha\beta\) protein domains. At this level, the model does not distinguish between the binding of GroEL to the hydrophobic surface as a whole or to individual strands and helices which make up that surface.

**Evidence for the GroEL binding model**

Recent structural data have demonstrated that only those peptides with a high tendency to present a hydrophobic surface exhibit high affinity for GroEL binding, irrespective of whether the peptide adopts a helical or a strand conformation (Wang et al. 1999; Chatellier et al. 1999). In the X-ray structure of the apical domain of GroEL (residues 191–376), Buckle et al. (1997) observed that seven residues in the N-terminal tag of one molecule bind through hydrophobic interactions to helices H and I of a second molecule in a mainly extended conformation. Chen and Sigler (1999) solved the crystal structure of the apical domain of GroEL in complex with a model peptide. They found that the peptide bound to helices H and I in a \(\beta\) hairpin conformation. Although no helical substrates have yet been seen bound to GroEL at atomic resolution, transferred nuclear Overhauser enhancements experiments have demonstrated that some peptides adopt a helical conformation when bound to GroEL (Landry et al. 1992; Kobayashi et al. 1999). These different structural studies reveal the ability of the apical domain of GroEL of promiscuous binding to a wide range of nonnative substrates provided a hydrophobic surface is available for binding.

GroEL has been observed to be able to bind proteins in their native conformation. Upon incubation of native pre-\(\beta\)-lactamase (Lamint et al. 1990) or native human dihydrofolate reductase (DHFR) (Viitanen et al. 1991) with GroEL, both proteins become inactivated and are bound by GroEL. Both enzymes are \(\alpha\beta\) proteins which contain \(\alpha\beta\) (pre-\(\beta\)-lactamase) or \(\alpha\beta\alpha\) (DHFR) sandwiches. The observed binding of the native proteins to the chaperonin could be the result of the binding of helices H and I in the apical domain of GroEL to the buried sheet in these proteins after displacement of the endogenous helices. This would lead to the destabilization and inactivation of the proteins. Interestingly, Frieden and coworkers (Clark et al. 1996; Clark and Frieden 1997) have observed that *E. coli* DHFR interacts only transiently with GroEL upon refolding, whereas murine DHFR exhibits a strong interaction with the chaperonin. The two proteins are very homologous and are expected to have similar structures consisting of \(\alpha\beta\alpha\) sandwiches. The major difference between the two structures is the presence of long loop regions in the murine DHFR which are absent in the *E. coli* protein. Upon insertion of these loops into *E. coli* DHFR, Clark et al. (1996) observed strong binding to GroEL. Significantly, one of the inserted loops replaces a tight turn region in *E. coli* DHFR which links the buried sheet to the opposite helix. Hence, based on our model, the longer loop would allow for easier access of the apical
domain of GroEL to the buried sheet and would explain the increased binding of the mutated *E. coli* DHFR.

In addition, there have been many investigations where model proteins bound to GroEL were subjected to limited proteolysis to determine what protein fragments were bound by GroEL. In one such study, Hlodan et al. (1995) identified two fragments of rhodanese which remain bound to GroEL after limited proteolysis of the rhodanese–GroEL complex. In the native structure, these fragments contain hydrophobic and amphiphilic α-helices and extensive β-sheet regions. In another study, Gervasoni et al. (1998) proposed a model for the β-lactamase–GroEL complex in which a carboxy-terminal helix and other secondary structure elements in β-lactamase are melted out upon binding to GroEL, causing a partial uncovering of the large buried β-sheet. Their result is consistent with our proposed model for the binding between GroEL and its substrates.

Hence, most of the observations in the literature support the model shown in Fig. 4. It should be emphasized, however, that the model represents only what is proposed to be the predominant mode of binding between GroEL and its substrates, especially that most GroEL substrates are expected to have αβ domains. Other possible modes of binding by GroEL are not excluded because the chaperonin can also recognize and help in the folding of all alpha (e.g., citrate synthase; Mendoza and Campo 1996) and all beta proteins (e.g., Fab fragment of a monoclonal antibody; Schmidt and Buchner 1992) in vitro and has also been shown to interact with an all alpha protein in vivo (ferritin; Houry et al. 1999).

**GroEL versus DnaK**

The other major chaperone system in the *E. coli* cytoplasm, in addition to the Hsp 60 system, is the Hsp 70 system. This system consists of DnaK (Hsp 70), its cofactor DnaJ (Hsp 40), and a nucleotide exchange factor GrpE. DnaK has been studied extensively in vitro and in vivo. Like GroEL, DnaK has an ATPase activity and undergoes ATP-dependent conformational changes during its functional cycle, which is regulated by DnaJ and GrpE (Szabo et al. 1994; Pierpaoli et al. 1997; Laufen et al. 1999; Russell et al. 1999). The binding of DnaK to target polypeptides occurs
through a channel defined mainly by loop regions in the C-terminal domain of DnaK (Zhu et al. 1996). The preferred peptide-binding motif of DnaK is proposed to consist of a linear stretch of five hydrophobic residues that are flanked on both sides by regions enriched in basic residues (Rudiger et al. 1997).

The role of DnaK in vivo has been the subject of some controversy. Under normal growth conditions, deletion of the dnaKJ gene in a wild-type strain results in lethality (Teter et al. 1999), whereas in other strain backgrounds or under certain conditions the deletion results only in the rapid accumulation of suppressor mutations (Bukau and Walker 1990; Kang and Craig 1990; Teter et al. 1999). Under stress conditions, however, DnaK seems to be essential for viability under all genetic backgrounds. Recently, DnaK has been shown to assist in vivo in the de novo folding of 5–10% of total soluble E. coli proteins at 30°C, including ribosome-bound nascent chains (Deuerling et al. 1999; Teter et al. 1999). These proteins are preferentially in the size range of 30–75 kDa. However, the identities of these substrates remain largely unknown. Only four proteins are currently known to directly interact with DnaK under normal growth conditions at 30°C (Mogk et al. 1999).

Since DnaK recognizes polypeptide segments in extended conformation, whereas GroEL recognizes compact intermediates, and since DnaK has the ability to bind to ribosome-associated nascent chains, whereas no such binding has been observed for GroEL, it is reasonable to assume that DnaK acts in the cell at an earlier stage in the folding of target proteins than does GroEL. The possibility that the set of DnaK substrates might overlap with that of GroEL substrates comes from the observation that both DnaK and GroEL preferentially bind to proteins in a similar molecular weight range: 30–75 kDa for DnaK and 20–60 kDa for GroEL. Whether there is a direct transfer of substrates from DnaK to GroEL requires further investigation. Such a direct transfer has been demonstrated in vitro (Langer et al. 1992) and has been suggested to occur in vivo (Gaitanaris et al. 1994; Teter et al. 1999). It is most likely that a large fraction of newly synthesized polypeptides would require both the Hsp 70 and Hsp 60 systems for proper folding and, hence, the two chaperone systems have to cooperate in folding these proteins.

Future directions

Although much is now known about GroEL structure and function, the identification of the natural substrates of the chaperonin raises several major issues which remain to be resolved. The extent of interaction of the identified proteins with the chaperonin has to be determined, both in vitro and in vivo. The mode of binding of these substrates to GroEL also has to be further investigated by using a battery of biophysical and imaging techniques. Lastly, the role that GroEL plays as part of a multichaperone network has to be investigated by the identification and characterization of substrates of other chaperone systems. Work is currently in progress on all of these tracks. The goal, ultimately, is to be able to elucidate at the molecular level the function of different chaperone systems in the in vivo folding process of E. coli proteins.

Acknowledgements

I would like to thank F.U. Hartl, D. Frishman, C. Eckerskorn, F. Lottspeich, and K. Anderson for their support during the course of this work.

References


