Chapter 7
Evolution of Bacterial Chaperonin 60 Paralogues and Moonlighting Activity

Shekhar C. Mande, C.M. Santosh Kumar, and Aditi Sharma

Abstract Around two thirds of genome sequenced bacteria encode one chaperonin 60 protein with the other third encoding between two and eight chaperonin 60 paralogues. A surprising finding is that these bacterial proteins have a wide, and growing, range of additional functions both within the bacterium, but principally when the Cpn60 protein exits the cell and exists on the bacterial cell wall or in the bacterium's external milieu. These findings have occurred at the same time that it has been realised that bacterial Cpn60 proteins can assume lower oligomeric forms than that of the prototypic tetradecameric *E. coli* GroEL. It is possible that lower oligomeric forms of Cpn60 may more readily be secreted and interact with biopolymers in a distinct manner to that of the tetradecameric homologues and paralogues. How the Cpn60 moonlighting functions evolved is a key question to be addressed. To address this question we postulate that the chaperonin genes have been subject to different selective constraints over evolutionary time. Gene duplication, followed by sequence divergence, resulted in the evolution of paralogous Cpn60 proteins that have distinct moonlighting activities. Moreover, these functional variations might be acquired by incorporating chemically dissimilar substitutions at functionally important residues.

7.1   Introduction

The central dogma of molecular biology outlines three steps in the transfer of information present in the genetic material to the synthesis a polypeptide chain. These polypeptides perform a predestined function. However, for the proteins to be functional, they need to be folded in an appropriate conformation. Cells have

S.C. Mande (✉) • C.M.S. Kumar • A. Sharma
National Centre for Cell Science (NCCS), NCCS Complex,
University of Pune Campus, Ganeshkhind, Pune 411007, India
e-mail: shekhar@nccs.res.in

evolved a family of proteins, aptly known as molecular chaperones, to assist the folding of other proteins. The chaperone proteins are ubiquitous, highly conserved and have been demonstrated to bind naïve and unfolded substrate proteins via exposed hydrophobic patches, which otherwise are buried in the native folded conformation (Hartl 1996; Hartl and Hayer-Hartl 2003). Molecular chaperones were initially identified as a group of proteins that are abundantly expressed during heat stress. Therefore, these proteins are also known as Heat Shock Proteins (HSPs) and consequently classified according to their molecular masses such as Hsp100, Hsp90, Hsp70, Hsp60, small Hsp families. See Chaps. 1 and 2 for more detailed discussions of the molecular chaperones of prokaryotes and eukaryotes.

One of the best characterized families of HSPs is the Hsp60 family, which is commonly referred to as the chaperonin (GroEL or Cpn60) family. Chaperonins form a characteristic cylindrical assembly whereby they assist misfolded or unfolded substrate proteins to reach their native state upon encapsulation in the cylindrical cavity in ATP dependent cycles of binding and release (Hartl and Martin 1995). Chaperonins are classified into two classes based on their systematic occurrence, cellular localization and co-chaperonin dependence. The group I class includes members from the cytosol of prokaryotes and endo-symbiotically related membrane bound eukaryotic organelles such as mitochondria and chloroplasts. These chaperones require assistance of the co-chaperonin, Hsp10 as a lid for encapsulating the substrates. On the other hand, group II chaperonins include members from the eukaryotic cytosol and from archaea. These chaperonins possess a protrusion in their apical domain, which acts like a built-in lid for substrate encapsulation and therefore are independent of the co-chaperonin (Fig. 7.1). Examples of group I chaperonins include GroEL from *Escherichia coli* and several eubacteria (Horwich et al. 2007), while CCT chaperonins and the well-studied thermosome from the archaeal branch of life are members of group II chaperonins. Whereas the group I chaperonins form homo-tetradecameric complexes, some members of the group II family form hetero oligomers (Fig. 7.1a). Moreover, the two groups differ in their subunit movements; while the subunit movement in group I chaperonins is sequential, for group II it is concerted. This review focuses on the structure, function and evolutionary aspects of the group I chaperonins.

7.2   Structure and Function of the Chaperonins

Understanding of the biology of chaperonin function is dominated by studies on *E. coli* GroEL and its co-chaperonin GroES. Since the discovery of chaperonin function, genetic, biochemical and structural studies on *E. coli* GroEL-GroES have led to systematic knowledge on various aspects of its function (Hartl 1996; Hartl and Hayer-Hartl 2003; Hartl and Martin 1995; Horwich et al. 2007). From these studies GroEL-GroES have emerged as the panacea in preventing all the undesirable consequences of intracellular protein misfolding.
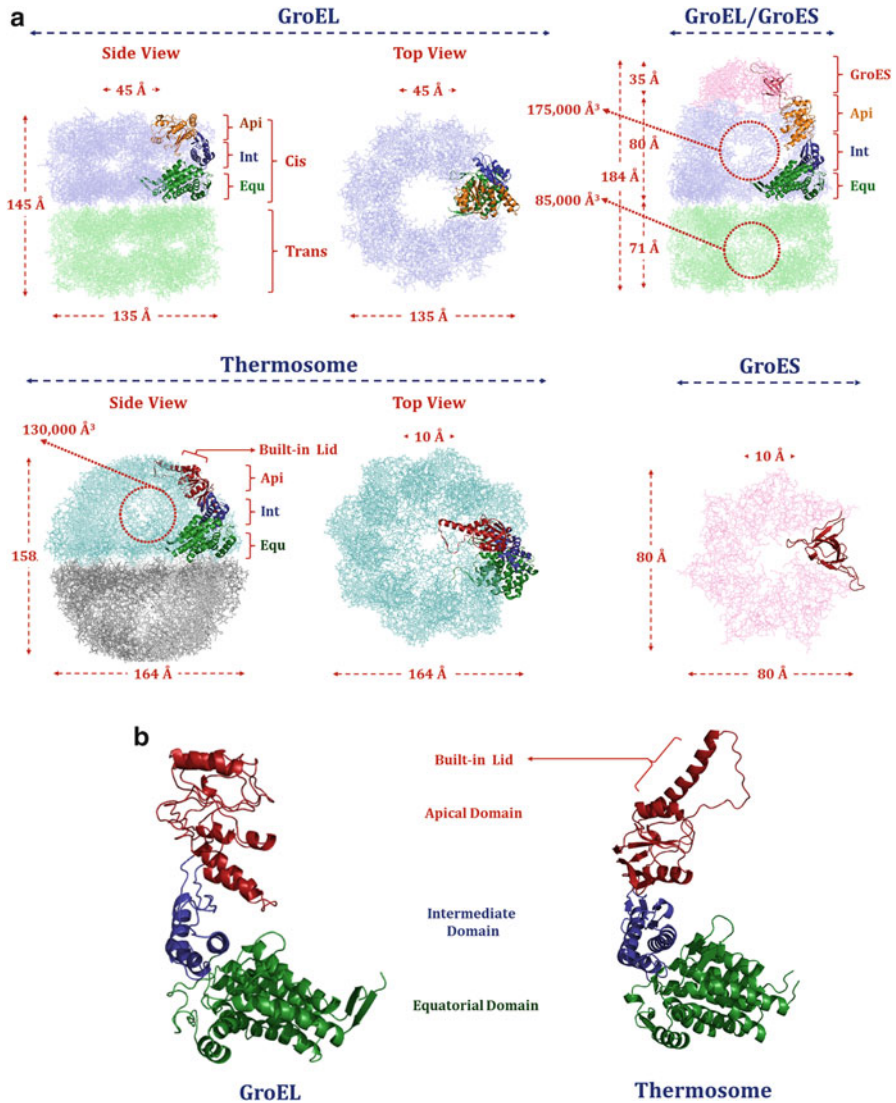
Fig. 7.1 Architecture of group I and II chaperonins. (**a**) Crystallographic models of *E. coli* GroEL, GroES and GroEL-GroES representing the Group I chaperonins and the thermosome from *T. acidophilum*, representing the Group II chaperonins, are presented. Individual domains in one subunit of GroEL and thermosome are indicated. *Api* apical domain, *Int* intermediate domain, *Equ* equatorial domain. GroES acting as a lid binds to GroEL asymmetrically, at the cis GroEL ring, wherein the general substrate polypeptides are encapsulated. Other open ring is termed the trans ring. Thermosome forms a symmetric complex, showing a "closed" cavity. Illustrations are generated using Pymol 1.3 (DeLano Scientific LLC, USA). Co-ordinates for the molecules GroEL (1OEL), GroEL-GroES and GroES (1AON) and thermosome (1A6E) were obtained from PDB. (**b**) Crystallographic models of GroEL and thermosome monomers as indicated depicting colour coded individual domains. The built-in lid in thermosome is shown that replaces GroES as cap

Structurally, GroEL possesses a three-domain architecture. The central region of the polypeptide, spanning amino acid residues 191–376, constitutes the apical domain. This domain is rich in hydrophobic residues and thus binds substrates and GroES. The equatorial domain spans two extremities of the GroEL polypeptide, that is, residues 1–133 and 409–523. This domain is responsible for the ATPase activity and bulk of the inter-subunit and inter-ring interactions. The hinge forming intermediate domain spans two regions on the polypeptide namely, residues 134–190 and 377–408, and transmits signals between the equatorial domain to the apical domain owing to nucleotide and substrate binding, respectively (Hayer-Hartl et al. 1995; Kumar et al. 2009; Mayhew et al. 1996; Xu et al. 1997) (Fig. 7.1b).

GroEL's chaperoning ability is dictated by the formation of a tetradecamer constituting two isologous heptameric rings, each enclosing a cavity for substrate proteins to bind (Fig. 7.1 see also Chap. 2). The substrate-bound *cis* cavity is expanded upon capping by GroES and thus forms a sequestered environment of ~175,000 Å^3 to enable the substrate polypeptides to fold. The *trans* cavity, on the other hand, remains constrained with a capacity of ~85,000 Å^3 (Fig. 7.1). The volume of the cavity in group II chaperonins is ~130,000 Å^3, which is significantly smaller than the *cis* cavity of GroEL but larger than the *trans* cavity (Lars et al. 1998). GroEL's wide substrate repertoire constitutes about 10–15 % of the cellular proteins in *E. coli*. Thus, the major intracellular function of GroEL is understood to be as an essential folding machine (Horwich et al. 2007).

7.3 Multiple Copies of GroELs/Chaperonin 60 in Bacteria

The ubiquitous occurrence of GroEL, or more correctly chaperonin (Cpn)60 (as GroEL is the Cpn60 protein of *E. coli*), across species might be explained by its essential role in the protein folding process. High architectural conservation among Cpn60s from different species indicates that the mechanism of GroEL is universally conserved. Moreover, several Cpn60 homologues from other bacteria have been shown to function in *E. coli*, suggesting that the substrate spectrum of the respective Cpn60s must overlap considerably (Goyal et al. 2006; Lund 2009). Since Cpn60 needs to interact with a wide range of substrate proteins, sequence analysis of the substrate-binding apical domain has revealed significant plasticity in its sequence and structure (Dekker et al. 2000; Goyal et al. 2006) (Fig. 7.3). On the other hand, the equatorial domain required for oligomerization through inter-subunit interactions exhibits higher conservation.

The availability of complete genome sequences of various bacteria has revealed the presence of multiple copies of *groEL* genes; with one (or more) of the multiple genes in operonic arrangement with the cognate co-chaperonin, *groES*. Examples of bacteria hosting multiple *groEL* genes include actinobacteria, α-proteobacteria, cyanobacteria and chlamydiae (Lund 2009). The discovery of multiple copies of Cpn60s poses interesting questions about the function of these extra copies. First,

whether all the copies of Cpn60 retain their function as protein chaperones? Second, do they distribute the substrate pool either temporally or based on factors such as function and/or composition of the substrates? Such a distribution has been demonstrated in the case of the rhizobial Cpn60s, wherein different Cpn60 copies play different roles in nitrogen fixation, probably by encountering different substrates (see Lund 2009). Third, do the extra copies of chaperonins acquire additional functions, owing to their substrate promiscuity?

Multiple chaperonin genes in different bacteria and subcellular organelles might have arisen through either gene duplication in different lineages or by horizontal gene transfer (Goyal et al. 2006; Lund 2009). Invariably such evolutionary events lead to functional divergence, and it is interesting to address if paralogous Cpn60s too have acquired new functions. It is well known that apart from the role as a protein-folding chaperone, several Cpn60 proteins exhibit non-chaperonin activities, such as binding to other biopolymers, eliciting immune responses and functioning as insecticides as explained in the following sections (Joshi et al. 2008) and in other chapters of this volume. The presence of multiple copies of chaperonins in different bacterial lineages are therefore believed to have introduced new functional roles for the different copies of the chaperonins (Piatigorsky and Vistow 1989). Exerting multiple roles by a single protein is known as moonlighting and has been widely observed in Nature and thus understanding the molecular basis of moonlighting is important to appreciate its implications for the function of Cpn60 proteins.

7.4   Moonlighting in Proteins

Moonlighting, initially known as gene-sharing (Piatigorsky and Vistow 1989), is defined as the ability of a single polypeptide to perform multiple unrelated functions (Jeffery 2004a). The definition, however, excludes the polypeptides that have resulted from gene-fusion, alternative splicing or inherently promiscuous enzymes. One of the earliest moonlighting functions was discovered by identification of carbohydrate metabolic functions in certain vertebrate eye lens proteins (Piatigorsky and Vistow 1989). Since then, it is estimated that several hundred moonlighting proteins have been identified with around two dozen of these proteins performing multiple unrelated biological functions. Accumulation of the number of moonlighting proteins identified has added another dimension to the complexity of cellular networks. Although, moonlighting in eukaryotic proteins is well documented, evidences for the prokaryotic proteins have only been described recently (Muro-Pastor et al. 1997; Ostrovsky de Spicer and Maloy 1993; Walden et al. 2006). Different mechanisms have been proposed for moonlighting, including secretion into extracellular space, interactions with nucleic acids, changes in physico-chemical parameters such as temperature or redox condition of the cell, changes in oligomeric status or changes in the cellular concentration of ligands, substrates, co-factors or products (Jeffery 2009; Kumar and Mande 2011).

The evolution of moonlighting might be a result of two fundamental cellular phenomena. One, moonlighting is typically exhibited by proteins which occur ubiquitously (Jeffery 1999). For example, many of the enzymes involved in carbohydrate metabolism are ubiquitous and hence during evolution an extra function might have been incorporated into these proteins. Examples include several glycolytic enzymes (Fig. 7.2a). Two, moonlighting has been proposed for the proteins that are constitutively overproduced in the cell, such as crystallins (Piatigorsky 1998; Jeffery 2004b) (Fig. 7.2b). Chaperonin 60 turns out to be a good candidate for exhibiting moonlighting activity since it fits both of these features well: high sequence conservation and elevated cellular expression levels.

In addition, two models of the evolution of moonlighting activity have been proposed which are based on the fundamental structural aspects of the proteins: (i) the **allostery model** and (ii) the **adaptability model**. The allostery model arises out of the common observation in proteins, namely the larger-than-required size of many of enzymes. Since the majority of proteins exhibit a larger structure than is necessary for performing their specific function, the apparent unused large surface areas exposed on proteins might have evolved new pockets and active sites for performing novel function (Jeffery 1999). Examples for the allosteric model include the glycolytic enzymes (Huberts and van der Klei 2010; Jeffery 2004b). On the other hand, the adaptability model is based on the location of a protein in different cell types. Proteins expressed in different cell types might exhibit different activities owing to the local necessity and the binding partners encountered Piatigorsky 1998). The examples for this model include the classical lens proteins and several molecular chaperones such as Hsp60, Hsp70 etc. Therefore, one protein performing several functions would be advantageous for the cell in terms of energy conservation (Jeffery 2004b). We discuss here two examples of moonlighting proteins: crystallins and phosphoglucoisomerase (PGI). Complete discussion on the biology of moonlighting with more examples is provided in Chap. 3, written by Connie Jeffery.

7.4.1   Crystallins

Crystallins were the first proteins to be discovered to show moonlighting function (Piatigorsky and Vistow 1989). Crystallins are the principal structural proteins in the lens of the vertebrate eye, constituting about 90 % of the lens protein content. Crystallins are principally represented by α, β and γ variants, while other variants assist assembly of the principal crystallins in certain vertebrates. However, certain crystallins also have been demonstrated to display additional functions in other locations of the body. These examples include the δ-crystallin from ducks, which is the enzyme, argininosuccinate lyase, while the ε and η-crystallins have been shown to exhibit lactate dehydrogenase activity (Bateman et al. 2003; Hendriks et al. 1988) (Fig. 7.2b). Moreover, turtle τ-crystallin shows α-enolase activity (Wistow et al. 1998). Due to their high concentrations in the eye lens, however, such
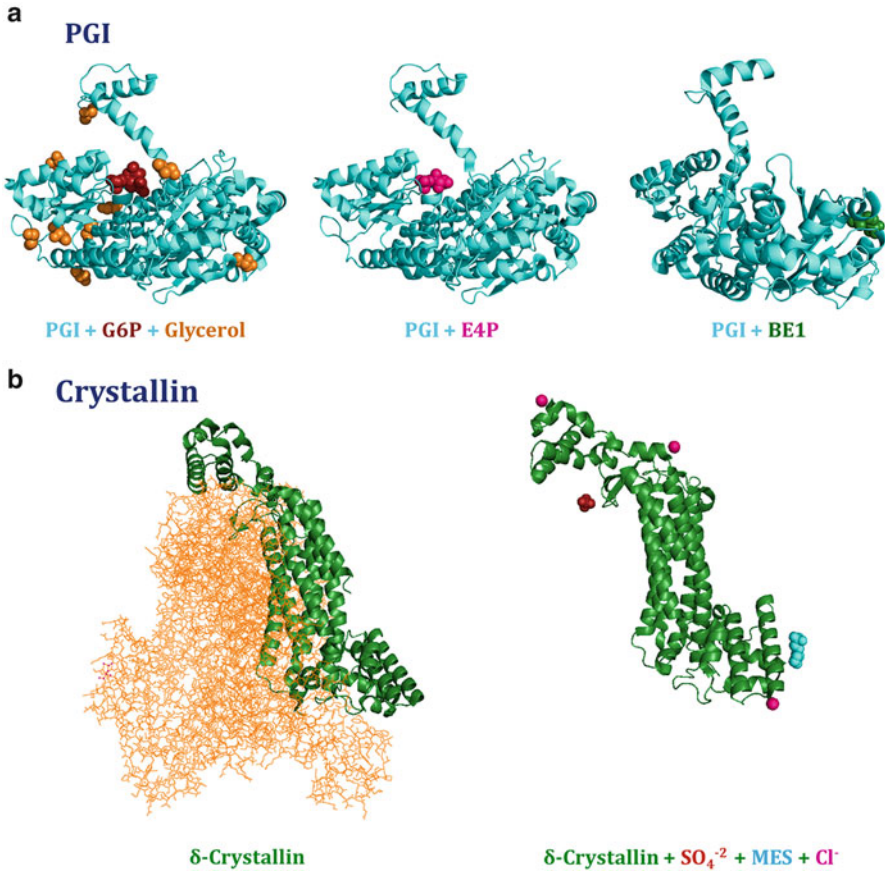
Fig. 7.2 Examples of moonlighting. (**a**) Phosphoglucoisomerase (*PGI*) is glycolytic enzyme involved in the reversible isomerization of glucose-6-phosphate to fructose-6-phosphate. PGI assumes several extracellular cytokine functions. Presented is PGI in complex with its natural substrate, glucose-6-phosphate (1U0F), the cytokine activity inhibitor, erythrose-4-phosphate iso-citrate (1IRI) and n-bromoacetyl-aminoethyl phosphate (1C7Q). The ligands glucose-6-phosphate (G6P), glycerol, inhibitors Erythrose-4-phosphate isocitrate (E4P) and n-bromoacetyl-aminoethyl phosphate (BE1) are presented as space filled in red/blue, gold, pink and green, respectively. (**b**) Moonlighting activity of crystallins. Cartoon representations of duck δ-crystalline monomers are presented in the apo (1U15) and holo (1U16) conformations. Apo form exists as a tetramer and is involved in the lens function, while the holo form is involved in the argino succinate lyase activity. Ligands 2-(N-morpholino)-ethanesulfonic acid (MES), chloride ions and sulphate ion are presented as space filled in cyan, pink and maroon, respectively

metabolic role is unlikely and thus they might have only the structural function in the lens (Wistow and Piatigorsky 1988). Therefore, the crystallins expressed in different cell types might have evolved additional functions in accordance with the cell type requirement.

7.4.2   *Phosphoglucoisomerase*

Phosphoglucoisomerase (PGI) is a glycolytic enzyme which catalyses the reversible isomerization of glucose-6-phosphate to fructose-6-phosphate (Read et al. 2001). Moonlighting activity in PGI has been widely documented (Cao et al. 2000; Chaput et al. 1988; Gurney et al. 1986; Hansen et al. 2005; Schulz and Bahr 2003; Tanaka et al. 2002; Watanabe et al. 1996; Xu et al. 1996). Apart from its original glycolytic function in the cytosol, PGI has been demonstrated to perform several unrelated cytokine-like extracellular functions: (a) When secreted from T-cells, PGI promotes the survival of nerve cells, and thus is known as a neuroleukin (Chaput et al. 1988; Gurney et al. 1986); (b) as the well-known autocrine motility factor (AMF), which promotes cell migrations, and is believed to be involved in cancer metastasis (Watanabe et al. 1996; Tanaka et al. 2002) (Fig. 7.2a); (c) as a maturation factor (MF), promotes myeloid cell differentiation and may play some role in leukaemia (Xu et al. 1996); (d) as a serine protease inhibitor when bound to myofibrils (Cao et al. 2000) and; (e) as an implantation factor activity in the ferret (Schulz and Bahr 2003, 2004). PGI, therefore, appears to have acquired several diverse biological functions depending on its location in different cell types.

With the examples such as those described above, proteins are believed to evolve moonlighting functions either due to the development of additional binding sites, other than their original active site (allostery model), or alteration of functions based on the requirements of the expressing cell type (adaptability model).

7.5   Moonlighting in Chaperonin 60 Proteins

Chaperonin 60, as described above, is a 900 kD cylindrical assembly of 14 monomers in two rings, which encloses cavities for the non-native protein substrates to bind (Fig. 7.1). The non-specific binding is a consequence of recognition of substrates by the hydrophobic surfaces of Cpn60 presented by the apical domains. Therefore Cpn60 has been attributed the protein folding function of the bacterial cytosol and intracellular organelles. However, recent discoveries on the Cpn60 molecules from other bacteria and eukaryotic organelle have been expanding the substrate repertoire and imparting new functions to Cpn60 (Basu et al. 2009; Kumar et al. 2009; Kumar and Mande 2011). Yeast mitochondrial Hsp60 has been shown to act as a protein chaperone *in vitro* and *in vivo*. However, this same chaperone has been demonstrated to bind single stranded DNA *in vitro* and play a role in stability and transmission of nucleoids *in vivo* (Kaufman et al. 2003). Moreover, Cpn60 homologues from several insect symbionts such as *Enterobacter aerogenes* (Yoshida et al. 2001) and *Xenorhabdus nematophila* (Joshi et al. 2008) have been shown to exhibit insect toxicity. The toxicity of Cpn60 from *X. nematophila* has been demonstrated to be alleviated upon its interaction with α-chitin. Mutational analysis, followed by biochemical characterizations of these Cpn60s showed that the amino-acid residues

critical for toxicity are distinct from those essential for chaperone activity, suggesting that the two functions are independently operated. Furthermore, several pathogen-borne Cpn60s, such as the surface-associated Cpn60 from the proteobacterial pathogen, *Legionella pneumophila* (Garduño et al. 1998 – see Chap. 9) and several mollicute pathogens (Clark and Tillier 2010), have been implicated in host cell invasion, owing to their association with the cell surface. A more detailed analysis of the Cpn60 of *L. pneumophila* is given in Chap. 9. Moreover, Cpn60 from pathogenic *E. coli* (Reddi et al. 1998) and *from Aggregatibacter actinomycetemcomitans* (Kirby et al. 1995) have been implicated in bone resorption. Taken together, these observations suggest that Cpn60s from different organisms have evolved moonlighting functions, probably to support the organism either at different stages of growth or for the purposes of virulence.

The substrate repertoire of Cpn60 has been found to be expanded from polypeptides to various biopolymers. Since binding to different biopolymers is independent of each other, and that Cpn60 might encounter different macromolecules in the cell, this protein might have evolved the capability to distinguish between the protein and non-protein substrates. The ability to differentiate between polypeptides and other biopolymers might be conferred upon Cpn60 by its specific conformational features such as the hydrophobic surfaces of the apical and equatorial domains. Moreover, employing mass spectrometry and NMR coupled with hydrogen-exchange techniques, it has been shown that *E. coli* GroEL is inefficient in binding to extended polypeptides but is able to bind the collapsed, molten globule-like folding intermediates of the substrates effectively (Gervasoni et al. 1996; Goldberg et al. 1997; Robinson et al. 1995; Zahn et al. 1994). Furthermore, Cpn60's preference to interact with α/β proteins without any sequence similarity suggests that the discrimination among the substrates by this protein might be by the formation of central cavities enclosed in the heptameric rings and thus a lower oligomeric form might not differentiate different biopolymers (Kumar and Mande 2011). The moonlighting Cpn60s from different prokaryotic species might indeed exhibit different oligomeric species and thereby be involved in distinct biochemical functions, as observed in *Helicobacter pylori*, *Mycobacterium tuberculosis* and several intracellular pathogenic bacteria (Basu et al. 2009; Kumar et al. 2009; Lin et al. 2009). The moonlighting functions of Cpn60 homoguess might arise due to subtle alterations in substrate specificity as a consequence of differences in the oligomeric states of these proteins. Such an ability to differentiate between polypeptides and non-protein substrates by modulating their oligomeric properties appears promising to us from our recent studies on the paralogous Cpn60s of *M. tuberculosis* (Basu et al. 2009; Kumar et al. 2009; Qamra and Mande 2004; Qamra et al. 2004).

7.5.1 *Moonlighting by* M. tuberculosis *Cpn60 Proteins*

Mycobacterium tuberculosis, the causative pathogen of tuberculosis, encodes two chaperonin homologs; GroEL1 (Cpn60.1 or Hsp60) and GroEL2 (Cpn60.2 or

Hsp65). Sequence identity between the two GroELs is about 60 % (Kong et al. 1993). GroEL2, the first to be discovered (Henderson and Martin 2011), is essential (Stewart et al. 2002) and has been implicated in inducing host cytokine responses (Friedland et al. 1993; Lewthwaite et al. 2001). On the other hand, Cpn60.1 is dispensable (Stewart et al. 2002). Deletion of the *cpn60.1* gene in mycobacteria shows similar growth patterns as the wild type, *in vitro* and *in vivo* (Hu et al. 2008; Ojha et al. 2005). A Cpn60.1 deletion mutant in *Mycobacterium smegmatis* is deficient in formation of biofilms (Ojha et al. 2005). In contrast, deletion of the same gene in *M. tuberculosis* fails to have much effect on biofilm formation but the mutant fails to be able to induce granuloma formation in mice or guinea pigs (Hu et al. 2008). In *M. smegmatis*, biofilm formation is implicated for its chaperone function and thus requires the interaction of Cpn60.1 with KasA, an enzyme involved in membrane lipid metabolism (Ojha et al. 2005). Moreover, both myco-bacterial Cpn60 proteins are secreted and consequently are involved in eliciting several host cell immune responses via macrophage and monocyte stimulation (Friedland et al. 1993; Khan et al. 2008; Lewthwaite et al. 2001; Riffo-Vasquez et al. 2012). For more information on the extracellular roles of *M. tuberculosis* Cpn60.2 the reader should consult Chap. 8 by Richard Stokes.

Biochemical and biophysical studies showed that the recombinant *M. tuberculosis* Cpn60s exist as dimers unlike *E. coli* GroEL and consequently, are ineffective as chaperones (Kumar et al. 2009; Qamra and Mande 2004; Qamra et al. 2004). Moreover, recombinant Cpn60.1, the dimeric form, was demonstrated to co-localize with the nucleoids isolated from *M. tuberculosis* extracts (Basu et al. 2009). In addition, structural studies on the apical domains of *E. coli* GroEL (Xu et al. 1997), *M. tuberculosis* GroEL1 (Sielaff et al. 2011) and GroEL2 (Qamra and Mande 2004) showed identical substrate binding cleft (Fig. 7.3). Furthermore, using peptide arrays, GroEL1 has been shown to bind polypeptides derived from GroEL1's native substrate, KasA, indicating that GroEL1 is a *bona fide* chaperonin (Ojha et al. 2005).

Chaperonin 60.1 has been demonstrated to exist in different oligomeric forms, as a dimer, heptamer and tetradecamer and the conversion between the heptamer and the tetradecamer is mediated by a phosphorylation switch (Kumar et al. 2009). Considering that the tetradecameric form of Cpn60.1 might be an active chapero-nin, it has been proposed that the phosphorylation event might act as an energy (ATP pool) conservation mechanism in slow growing *M. tuberculosis* (Kumar et al. 2009; Kumar and Mande 2011). Furthermore, such multiple oligomeric forms of chaperonins were observed in the chloroplast (Dickson et al. 2000) and mito-chondrial chaperonins (Levy-Rimler et al. 2001), wherein they existed in the monomeric, single ring heptameric and double ring tetradecameric forms and the conversion from single ring to the double ring form is concentration and GroES dependent. Moreover, as described earlier, yeast mitochondrial Cpn60 proteins were demonstrated to associate with the stability and transmission of the nucleoid DNA (Kaufman et al. 2003). Taken together, these observations tend to suggest that the Cpn60 protein might switch between different functional forms by modu-lation of its oligomeric status.
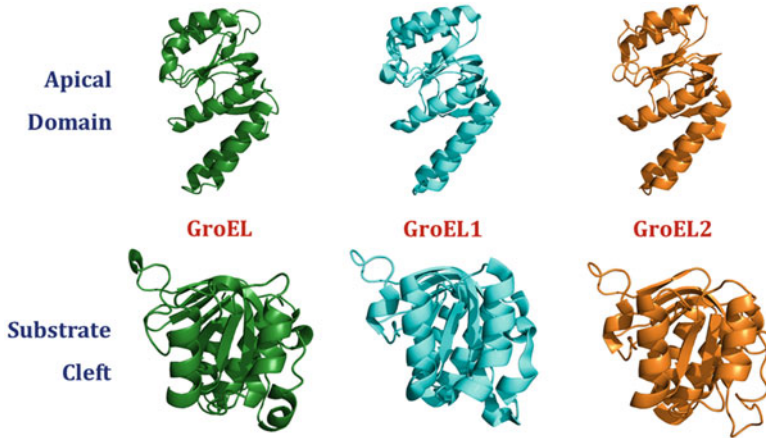
Fig. 7.3 **Conserved substrate binding clefts in GroELs**. Apical domains and the substrate binding clefts of the indicated GroEL homologues are presented. Co-ordinates for *E. coli* GroEL (1AON), *M. tuberculosis* GroEL1 (3M6C) and GroEL2 (1SJP) were obtained from PDB

7.5.2   *Functional Dichotomy in* M. tuberculosis *Cpn60 Proteins*

Structural features which confer substrate dichotomy on Cpn60 proteins would be interesting to study. The equatorial domain in GroEL is responsible for its oligomerization (Xu et al. 1997). While the equatorial domain is principally buried in the tetradecameric form, it gets increasingly exposed in the lower oligomeric forms. However, the substrate interacting face of the apical domain remains exposed independent of the oligomeric status. Therefore, it might be possible that different functions of *M. tuberculosis* Cpn60.1 arise out of different oligomeric forms; dimer interacting with DNA via its nucleotide binding equatorial domain and the tetradecamer being involved in the chaperone function. Chaperonin 60.1 therefore presents a unique way of moonlighting – distributing its distinct functions into different oligomeric forms. The evolutionary significance of the functional divergence is therefore essential to understand the basis such a behaviour.

7.6   The Evolution of the Chaperonin 60 Protein

As stated before, the chaperonins are distributed into two different groups based on their phylogenetic distribution. While the group I belongs to prokaryotes and eukaryotic organelles, the group II belongs to archaea and eukaryotes. Although the evolution of group II chaperonins from eukaryotes (Archibald et al. 2000, 2001) and archaea (Archibald et al. 1999; Archibald and Roger 2002) is well documented, that

of group I chaperonins is still in its infancy (Dekker et al. 2011; Goyal et al. 2006; Hughes 1993; Levy-Rimler et al. 2001; Techtmann and Robb 2010). We have attempted to address evolutionary aspects of GroELs in our laboratory based on sequence analysis, functional studies and biochemical experiments (Goyal et al. 2006; Kumar et al. 2009; Kumar and Mande 2011).

Two experimental studies on the directed evolution of GroEL have delineated the functional significance of individual domains of GroEL. In one study, employing random mutagenesis to derive GFP-folding GroEL variants resulted in mutants with diminished ability to recognize its natural substrates and, consequently deficient in functioning as a general chaperone (Wang et al. 2002). In another study based on random mutagenesis of GroELs, apical domain has been shown to be capable of absorbing large insertions or deletions, unlike the highly conserved equatorial domain (Kumar et al. 2009). Therefore, GroEL, even as it encounters a wide range of substrates by maintaining plasticity in the apical domain, it requires conserved architecture anchored by the equatorial domain to be functional as a chaperone.

Conserved residues among group I chaperonins map onto ATP binding pseudo-Walker motif and the substrate binding sites on *E. coli* GroEL structure (Brocchieri and Karlin 2000). Highly conserved charge clusters line the central cavity and the intra subunit interfaces, and thus are presumed to play a role in interacting with the substrate and constituting the quaternary structure. Interestingly, the less conserved segments were mapped outside wall of the cylinder (Brocchieri and Karlin 2000). Furthermore, the residues mapped to the functionally important regions on GroEL have been predicted to be selected by positive selection (Fares et al. 2002a, 2005).

Phylogenetic studies on the two groups of chaperonins predicted the divergence of the two groups at a Last Universal Common Ancestor (LUCA), which further diverged in terms of type and number of subunits in a ring and requirement of the co-chaperonin (Dekker et al. 2011). Co-occurrence of paralogous chlorophyll and nucleomorph chaperonins in the same eukaryotic cells has been implicated in the divergent evolution of the organellar chaperonins (Wast et al. 1999). Similarly, several mycobacterial species host multiple paralogues of GroEL (Kong et al. 1993; Qamra and Mande 2004). Phylogenetic studies on chaperonins predicted single gene duplication event in the common ancestor of *M. tuberculosis*, *M. leprae*, and *Streptomyces albus* that might have duplicated the chaperonin genes. In addition to evolving novel functions, Cpn60 has also been implicated to play role in evolution of several proteins by buffering mutations and elevated temperatures (Fares et al. 2002b; Rudolph et al. 2010).

To understand the basis for the divergence of multiple copies of chaperonins in bacteria we have performed divergence analysis on GroEL homologues from completely sequenced bacterial genomes (Table 7.1). Homologues of *M. tuberculosis* Cpn60.1 were identified using BLAST against 1,129 bacterial genomes (Cummings et al. 2002), aligned using ClustalW and an unrooted phylogenetic tree was generated from the 1859 Cpn60 homologues using MEGA 5.0 (Tamura et al. 2011). The tree depicts Cpn60 proteins from bacteria belonging to eight phyla, wherein multiple copies of Cpn60 proteins are identified for actinobacteria, cyanobacteria and chlamydia.

Table 7.1 Number and distribution of *M. tuberculosis* GroEL1 homologues in Bacteria

Phylum	Number of completely sequenced genomes	Number of sequences with expect value less than e^{-10}
Actinobacteria	138	326
Bacteriodetes/Chlorobi	81	98
Chlamydiae	12	83
Cyanobacteria	28	62
Firmicutes	212	415
Mollicutes	32	10
Proteobacteria	500	1,011
Spirochaetales	35	36
Total number of sequences with Expect Value less than e^{-10}		2,041

***Phylum Proteobacteria*:** The phylum proteobacteria is divided into five classes: α-, β-, γ-, δ- and ε- proteobacteria. Three of the five classes of proteobacteria; γ-, δ- and ε- proteobacteria show the presence of a single copy of the *groEL* gene, while α- and β-proteobacteria possess two or more copies of this gene. Interestingly, the conserved C terminal $(GGM)_4M$ repeat sequence has been observed in all the Cpn60 proteins in this phylum (Farr et al. 2007; Suzuki et al. 2008; Tang et al. 2006). Examples for this phylum include rhizobiaceae family, members of which have GroEL copies ranging from one to seven (Lund 2009). Interestingly, *Bradyrhizobium japonicum* hosts five *groESL* operons. Among these, *groESL2*, *groESL4* are constitutive and the other stress induced (Fischer et al. 1993). Moreover, *Buchnera aphidicola*, a member of γ-proteobacteria, hosts a single copy of *groEL* in operonic arrangement with *groES*, typically overexpressed at elevated temperatures (Baumann et al. 1996) (85).

***Phylum Actinobacteria*:** Actinobacteria clade represents another Gram positive bacterial phylum hosting multiple copies of *groEL* gene (Goyal et al. 2006). The comparison of the protein sequences across different lineages showed that the duplicated copies of chaperonin 60s, Cpn60.1 and Cpn60.2, are distributed to different phylogenetic branches (Fig. 7.4), suggesting that the duplication event might have occurred in the common ancestor of actinobacteria. Interestingly, unlike the rhizobiaceae family, only one of the actinobacterial *groEL* genes is in operonic arrangement with *groES* (Goyal et al. 2006). The absence of a *cpn10* copy with the other *cpn60*(s) could be either due to a duplication (or multiplication) of *cpn60* alone or loss of a *cpn10* copy after duplication of the operon. For example, *M. smegmatis* hosts three copies of *cpn60*, namely *cpn60.1*, *cpn60*.2 and *cpn60.3*, but only a single copy of *cpn10* that is associated with *cpn60.1* (Ojha et al. 2005; Rao and Lund 2010). *M. smegmatis* Cpn60.2 exhibits greater sequence identity (93 %) with *M. tuberculosis* Cpn60.2 than with other copies. Some of the actinobacteria host single *cpn60* genes, such as in *Bifidobacterium longum* and *Tropheryma whipplei*, which is homologus to the actinobacterial Cpn60.2 and the gene is located away from *cpn10* on the chromosome. Moreover, *Gordonia bronchialis DSM 43247* hosts three Cpn60 copies, which are distributed to three different clades of phylogenetic tree. *G. bronchialis* Cpn60.1
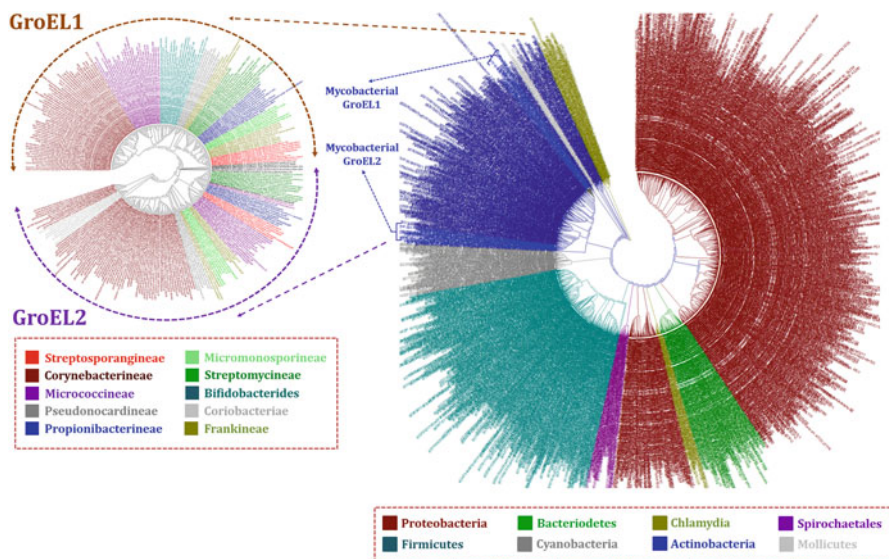
Fig. 7.4 Unrooted phylogenetic tree of bacterial chaperonins. A phylogenetic tree was generated using the 1,129 complete bacterial genomes. Protein sequences that are homologus to *M. tuberculosis* GroEL1 were aligned using the ClustalW program and an unrooted tree was constructed using MEGA 5.0. The bacterial phyla are colour coded as indicated. Regions corresponding to mycobacterial GroEL1 and GroEL2 are indicated in blue. Inset shows the actinobacterial branch expanded with individual bacterial families color-coded. Branches of the tree corresponding to GroEL1 and GroEL2 sequences are indicated in orange and purple, respectively

and Cpn60.2 are homologous to the mycobacterial Cpn60.1 and Cpn60.2, respectively, while Cpn60.3 is phylogenetically distant from these two copies.

***Phylum Cyanobacteria*:** Members of cyanobacteria phylum are characterized by multiple copies of *cpn60*s and a single copy of *cpn10* (Huq et al. 2010). *Nostoc punctiforme* PCC 73102 has three copies of *cpn60* with only one *cpn10* gene (Ran et al. 2007). Analogous to actinobacterial Cpn60s, these copies are similar in sequence to the corresponding homologues across the species, but are distant from the other copies within the species. Likewise, bacteria belonging to synechocystis species host two copies of *cpn60*, encoding Cpn60.1 and Cpn60.2, and one copy of *cpn10* (Lehel et al. 1993). Cpn60.1, but not Cpn60.2, has been demonstrated to complement the loss of GroEL in an *E. coli* mutant (Tanaka et al. 1997), while Cpn60.2 is essential under stress conditions (Tanaka et al. 1997).

7.7 Inference on Cpn60 Evolution

The size of the bacterial genomes ranges between 1.5 and 13 Mb (Cummings et al. 2002; McCutcheon et al. 2009; Schneiker et al. 2007). Owing to the absence of introns, the size of a bacterial genome can be logically correlated with the

number of cistrons. Since the number of protein families is limited, it is reasonable to imagine that the genome size variation probably correlates with the number of paralogous genes. While several bacteria with average genome size host one *cpn60* gene, bacteria with large genome sizes have been shown to encode multiple copies of the *cpn60* gene; up to five copies in *B. japonicum* harbouring a 9.2 Mb genome (Fischer et al. 1993).

The process of evolution of multiple genes might be due to either horizontal gene transfer (xenologues) or gene duplication (paralogues). To understand the basis of the evolution of Cpn60 paralogues we need to understand the process of gene duplication. Two models have been proposed on the consequences of gene duplication: (i) neofunctionalization model and (ii) subfunctionalization model. The neofunctionalization model, as the name suggests, assumes that the duplicated gene acquires a new function upon acquiring adaptive mutations. The model is supposed to be free of the selection pressure since the ancestral gene continues to function normally (Ohno 1970). Contrariwise, the subfunctionalization model assumes that the duplicated gene acquires neutral mutations and consequently retains one of the ancestral functions. Therefore, each of the ancestral functions (sub-function) is acquired by the duplicated genes (Force et al. 1999; He and Zhang 2005; Lynch and Force 2000). Although, examples for both the models have been identified, the subfunctionalization model has been widely observed (He and Zhang 2005). The phylogenetic distribution of the Cpn60 sequences has been observed in agreement with the 16S rRNA tree. The Cpn60 sequences were clustered according to the host bacteria in the phylogenetic tree, suggesting duplication and rapid evolution of the genes (Fig. 7.4). Thus, the duplication event of *cpn60* genes seem to have occurred in ancestors of certain clades, rather than being horizontally transferred across different species.

7.8 Domain Conservation in Cpn60

In the duplicated Cpn60 proteins of clades such as mycobacteriacea, the two paralogous classes of the Cpn60s: Cpn60.1 and Cpn60.2, are distinguished by the characteristic C-terminal sequence. While the proteobacterial and cyanobacterial Cpn60s display a hydrophobic GGM tripeptide repeat motif, the actinobacterial Cpn60.1 and Cpn60.2 display histidine rich and GGM repeat motifs, respectively.

The apical domain, which spans the central part of the Cpn60 primary structure, is responsible for binding a wide range of substrate molecules (Kumar and Mande 2011). Owing to its wide range of substrate interactions, this domain appears less conserved. Only six residues in proteobacterial Cpn60s and 12 residues in actinobacterial Cpn60s are conserved (Table 7.2). Interestingly, the apical domains of cyanobacterial Cpn60s displayed 40 conserved residues. Moreover, the equatorial domain, which spans two extremes of the Cpn60 polypeptide, and is responsible for the inter-subunit interactions, the formation of essential Anfinsen cage and ATPase activity, exhibits higher conservation. Five point residues and three peptide stretches in proteobacterial Cpn60s and 15 residues and one peptide stretch in actinobacterial

Table 7.2 Conserved residues among GroEL1 homologues

Phylum	GroEL domains	Conserved residues
Proteobacteria	Apical domain	Ser200, Pro276, Gly279, Asp288, Gly315, Arg365
	Equatorial domain	Gly52, Gly436, Asn454, Gly489, Val496, Ser506, Gly85-Asp-Gly-Thr-Thr-Thr90, Gly411-Gly-Gly413 Asp492-Pro493
	Intermediate domain	Gly172, Gly191, Gly377, Gly407, Ala402, Ala403
Actinobacteria	Apical domain	Gly197, Gly281, Gly317; Leu220, Leu258; Pro278; Arg284, Arg344, Arg367 and Asp290
	Equatorial domain	Gly69, Ala77, Gly85-Asp-Gly-Thr-Thr-Thr90, Gly102, Gly109, Gly118, Gly409, Gly413, Gly414, Pro448, Asn455, Gly457, Gly475, Asp493, Val497, Ser507
	Intermediate domain	Gly158, Gly381, Thr384, Asp397
Cyanobacteria	Apical domain	Gly196, Ser199-Pro-Tyr201, Arg208, Leu219, Lys224, Leu231, Leu235-Glu236, Ala256-Leu257, Thr259-Leu-Val-261, Asn263, Gly267, Ala276-Pro277, Phe279-Gly280, Arg283, Asp289, Leu293-Thr294, Glu301, Gly316, Lys325, Arg343, Ser356, Lys362, Arg366, Ala368, Leu370, Gly372-Gly373, Ala375
	Equatorial domain	Arg12, Leu15-Glu-Gly17, Thr29, Gly31-Pro32, Ala56, Asn67, Gly69, Leu72, Thr80, Gly85-Asp-Thr-Thr-Thr-Ala91, Gly102, Asn105, Gly109, Gly118, Gly412-Gly-Gly414, Pro450, Asn457, Gly459, Gly462, Gly477, Gly492, Asp495, Val499, Ser509, Thr516-Thr-Glu518
	Intermediate domain	Asp153, Gly157, Ala163, Val167, Gly171-Val-Ile173, Glu176-Glu177, Ser180, Thr183, Glu189-Gly-Met191, Lys378-Val-Gly-Ala381, Thr383-Glu-Thr-Glu386, Lys388, Lys391-Leu392, Glu395-Asp396, Asn399, Thr401, Ala403-Ala404, Glu407-Gly408

Cpn60s are conserved. Moreover, 48 residues are conserved in equatorial domains of cyanobacterial Cpn60s. The intermediate domain, which connects the apical and equatorial domains in the primary and tertiary structure, shows moderate conservation. Since the domain is responsible for the *en bloc* movement of the substrate recognition domain, a few conserved residues were identified in Cpn60s from three phyla. Eight residues in the proteobacterial Cpn60s and four residues in actinobacterial Cpn60s were conserved. As expected, cyanobacterial Cpn60s displayed 33 conserved residues in intermediate domains (Table 7.2).

In a nut shell, these observations imply that the apical domain, owing to its promiscuity in substrate interactions, is less conserved. Moreover, displaying structural similarity, this domain has been attributed the promiscuous peroxiredoxin origin (Dekker et al. 2011). The equatorial domain, on the other hand, is greatly conserved owing to the inter-subunit interactions, thereby, the formation of the essential Anfinsen cage

and the ATP binding for its activity. The intermediate domain is fairly conserved, since it needs to regulate the *en bloc* movement of the apical domain in response to the presence of nucleotide in the equatorial domain. Therefore the conservation profile for the Cpn60 monomer is the highest for equatorial domain, followed by the intermediate domain, while the apical domain shows the least conservation.

7.9 Conclusions

Chaperonin 60 functions as the constitutional protein chaperone in several bacteria. Recent studies have discovered additional novel functions for paralogous Cpn60 proteins in certain pathogenic bacteria. Moreover, these studies have expanded the substrate catalogue for Cpn60s from the polypeptides to other biopolymers and therefore, the resulting functional divergence has been attributed to different oligomeric status and cellular localization. Lower oligomeric forms are implicated in binding and transport of extended polymers such as DNA, while the higher oligomeric forms might be involved in the protein folding function. Considering the dimensions, it is reasonable to assume that the lower oligomeric form might be secretory and involved in eliciting the immunological responses and transport of biopolymers, while the higher oligomeric forms might be confined to the cytoplasm and protein folding actions.

In addition, we propose that chaperonin genes have been subjected to different selective constraints during evolution. Gene duplication followed by sequence divergence resulted in paralogous Cpn60s that can perform different functions. Moreover, these functional variations might be acquired by incorporating chemically dissimilar substitutions at functionally important residue positions.

References

Archibald JM, Roger AJ (2002) Gene duplication and gene conversion shape the evolution of archaeal chaperonins. J Mol Biol 316:1041–1050

Archibald JM, Logsdon JM, Doolittle WF (1999) Recurrent paralogy in the evolution of archaeal chaperonins. Curr Biol 9:1053–1056

Archibald JM, Logsdon JM, Doolittle WF (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. Mol Biol Evol 17:1456–1466

Archibald JM, Blouin C, Doolittle WF (2001) Gene duplication and the evolution of group II chaperonins: implications for structure and function. J Struct Biol 135:157–169

Basu D, Khare G, Singh S, Tyagi A, Khosla S, Mande SC (2009) A novel nucleoid-associated protein of *Mycobacterium tuberculosis* is a sequence homolog of GroEL. Nucleic Acids Res 37:4944–4954

Bateman OA, Purkiss AG, van Montfort R, Slingsby C, Graham C, Wistow G (2003) Crystal structure of η-crystallin: adaptation of a class 1 aldehyde dehydrogenase for a new role in the eye lens. Biochemistry 42:4349–4356

Baumann P, Baumann L, Clark MA (1996) Levels of *Buchnera aphidicola* Chaperonin GroEL during growth of the aphid *Schizaphis graminum*. Curr Microbiol 32:279–285

Brocchieri L, Karlin S (2000) Conservation among HSP60 sequences in relation to structure, function, and evolution. Prot Sci 9:476–486

Cao MJ, Osatomi K, Matsuda R, Ohkubo M, Hara K, Ishihara T (2000) Purification of a novel serine proteinase inhibitor from the skeletal muscle of white croaker (*Argyrosomus argentatus*). Biochem Biophys Res Commun 272:485–489

Chaput M, Claes V, Portetelle D, Cludts I, Cravador A, Burny A, Gras H, Tartar A (1988) The neurotrophic factor neuroleukin is 90 % homologous with phosphohexose isomerase. Nature 332:454–455

Clark GW, Tillier ER (2010) Loss and gain of GroEL in the mollicutes. Biochem Cell Biol 88:185–194

Cummings L, Riley L, Black L, Souvorov A, Resenchuk S, Dondoshansky I, Tatusova T (2002) Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. FEMS Microbiol Lett 216:133–138

Dekker C, Willison KR, Taylor WR (2011) On the evolutionary origin of the chaperonins. Proteins 79:1172–1192

Dickson R, Weiss C, Howard RJ, Alldrich SP, Ellis RJ, Lorimer GH, Azem A, Viitanen PV (2000) Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. J Biol Chem 275:11829–11835

Fares MA, Ruiz-Gonza´ lez MX, Moya A, Elena SF, Barrio E (2002a) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. Nature 417:398

Fares MA, Barrio E, Sabater-Mun˜oz B, Moya A (2002b) The evolution of the heat-shock protein GroEL from Buchnera, the primary endosymbiont of aphids, is governed by positive selection. Mol Biol Evol 19:1162–1170

Fares MA, Moya A, Barrio E (2005) Adaptive evolution in GroEL from distantly related endosymbiotic bacteria of insects. J Evol Biol 18:651–660

Farr GW, Fenton WA, Horwich AL (2007) Perturbed ATPase activity and not "close confinement" of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. Proc Natl Acad Sci U S A 104:5342–5347

Fischer HM, Babst M, Kaspar T, Acuña G, Arigoni F, Hennecke H (1993) One member of a *groESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. EMBO J 12:2901–2912

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerate mutations. Genetics 151:1531–1545

Friedland JS, Shattock R, Remick DG, Griffin GE (1993) Mycobacterial 65-kD heat shock protein induces release of proinflammatory cytokines from human monocytic cells. Clin Exp Immunol 91:58–62

Garduño RA, Garduño E, Hoffman PS (1998) Surface-associated Hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. Infect Immun 66:4602–4610

Gervasoni P, Staudenmann W, James P, Gehrig P, Plückthun A (1996) β-Lactamase binds to GroEL in a conformation highly protected against hydrogen/deuterium exchange. Proc Natl Acad Sci U S A 93:12189–12194

Goldberg MA, Zhang J, Sondek S, Matthews CR, Fox RO, Horwich AL (1997) Native-like structure of a protein-folding intermediate bound to the chaperonin GroEL. Proc Natl Acad Sci U S A 94:1080–1085

Goyal K, Qamra R, Mande SC (2006) Multiple gene duplication and rapid evolution in the *groEL* gene: functional implications. J Mol Evol 63:781–787

Gurney ME, Heinrich SP, Lee MR, Yin HS (1986) Molecular cloning and expression of neuroleukin, a neurotrophic factor for spinal and sensory neurons. Science 234:566–574

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571–579

Hartl FU, Hayer-Hartl M (2003) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295:1852–1858

Hartl F, Martin J (1995) Molecular chaperones in cellular protein folding. Curr Opin Struct Biol 5:92–102

Hayer-Hartl MK, Martin J, Hartl FU (1995) Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. Science 269:836–841

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neo-functionalization in duplicate gene evolution. Genetics 169:1157–1164

Henderson B, Martin A (2011) Bacterial virulence in the moonlight: multitasking bacterial moon-lighting proteins are virulence determinants in infectious disease. Infect Immun 79:3476–3491

Hendriks W, Mulders JW, Bibby MA, Slingsby C, Bloemendal H, de Jong WW (1988) Duck lens ε-crystallin and lactate dehydrogenase B_4 are identical: a single-copy gene product with two distinct functions. Proc Natl Acad Sci U S A 85:7114–7118

Horwich AL, Fenton WA, Chapman E, Farr GW (2007) Two families of chaperonin: physiology and mechanism. Annu Rev Cell Dev Biol 23:115–145

Huberts DH, van der Klei IJ (2010) Moonlighting proteins: an intriguing mode of multitasking. Biochim Biophys Acta 1803:520–525

Hughes L (1993) Contrasting evolutionary rates in the duplicate chaperonin genes of *Mycobacterium tuberculosis* and *M. leprae*. Mol Biol 10:1343–1359

Huq S, Sueoka K, Narumi S, Arisaka F, Nakamoto H (2010) Comparative biochemical character-ization of two GroEL homologs from the *Cyanobacterium Synechococcuselongatus* PCC 7942. Biosci Biotechnol Biochem 74:2273–2280

Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci 24:8–11

Jeffery CJ (2004a) Molecular mechanisms for multitasking: recent crystal structures of moonlight-ing proteins. Curr Opin Struct Biol 14:663–668

Jeffery CJ (2004b) Moonlighting proteins: complications and implications for proteomics research. DDT: Targets 3:71–78

Jeffery CJ (2009) Moonlighting proteins – an update. Mol Biosyst 5:345–350

Joshi MC, Sharma A, Kant S, Birah A, Gupta GP, Khan SR, Bhatnagar R, Banerjee N (2008) An insecticidal GroEL protein with chitin binding activity from *Xenorhabdus nematophila*. J Biol Chem 283:28287–28296

Kaufman BA, Kolesar JE, Perlman PS, Butow RA (2003) A function for the mitochondrial chap-eronin Hsp60 in the structure and transmission of mitochondrial DNA nucleoids in *Saccharomyces cerevisiae*. J Cell Biol 163:457–461

Khan N, Alam K, Mande SC, Valluri VL, Hasnain SE, Mukhopadhyay S (2008) *Mycobacterium tuberculosis* heat shock protein 60 modulates immune response to PPD by manipulating the surface expression of TLR2 on macrophages. Cell Microbiol 10:1711–1722

Kirby AC, Meghji S, Nair SP, White P, Reddi K, Nishihara T, Nakashima K, Willis AC, Sim R, Wilson M, Henderson B (1995) The potent bone resorbing mediator of *Actinobacillus actino-mycetemcomitans* is homologous to the molecular chaperone GroEL. J Clin Invest 96:1185–1194

Kong TH, Coates AR, Butcher PD, Hickman CJ, Shinnick TM (1993) *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. Proc Natl Acad Sci U S A 90:2608–2612

Kumar CMS, Mande SC (2011) Protein chaperones and non-protein substrates: on substrate pro-miscuity of GroEL. Curr Sci 100:1646–1653

Kumar CMS, Khare G, Srikanth CV, Tyagi AK, Sardesai AA, Mande SC (2009) Facilitated oligo-merization of mycobacterial GroEL: evidence for phosphorylation-mediated oligomerization. J Bacteriol 191:6525–6538

Lars Ditzel L, Lo we J, Stock D, Stetter K, Huber H, Huber R, Steinbacher S (1998) Crystal struc-ture of the thermosome, the archaeal chaperonin and homolog of CCT. Cell 93:125–138

Lehel C, Los D, Wada H, Györgyei J, Horváth I, Kovács E, Murata N, Vigh L (1993) A second *groEL*-like gene, organized in a *groESL* operon is present in the genome of *Synechocystis* sp. PCC 6803. J Biol Chem 268:1799–1804

Levy-Rimler G, Viitanen P, Weiss C, Sharkia R, Greenberg A, Niv A, Lustig A, Delarea Y, Azem A (2001) Type I chaperonins: not all are created equal. Eur J Biochem 268:3465–3472

Lewthwaite JC, Coates AR, Tormay P, Singh M, Mascagni P, Poole S, Roberts M, Sharp L, Henderson B (2001) *Mycobacterium tuberculosis* chaperonin 60.1 is a more potent cytokine stimulator than chaperonin 60.2 (Hsp 65) and contains a CD14-binding domain. Infect Immun 69:7349–7355

Lin CY, Huang YS, Li CH, Hsieh YT, Tsai NM, He PJ, Hsu WT, Yeh YC, Chiang FH, Wu MS, Chang CC, Liao KW (2009) Characterizing the polymeric status of *Helicobacter pylori* heat shock protein 60. Biochem Biophys Res Commun 388:283–289

Lund PA (2009) Multiple chaperonins in bacteria – why so many? FEMS Microbiol Rev 33:785–800

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473

Mayhew M, da Silva AC, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU (1996) Protein folding in the central cavity of the GroEL-GroES chaperonin complex. Nature 379:420–426

McCutcheon JP, McDonald BR, Moran NA (2009) Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. PLoS Genet 5:e1000565

Muro-Pastor AM, Ostrovsky P, Maloy S (1997) Regulation of gene expression by repressor localization: biochemical evidence that membrane and DNA binding by the PutA protein are mutually exclusive. J Bacteriol 179:2788–2791

Ohno S (1970) Evolution by gene duplication. Springer, Berlin/Heidelberg/New York, pp 59–87

Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR Jr, Hatfull GF (2005) GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell 123:861–873

Ostrovsky de Spicer P, Maloy S (1993) PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. Proc Natl Acad Sci U S A 90:4295–4298

Piatigorsky J (1998) Multifunctional lens crystallins and corneal enzymes. More than meets the eye. Ann N Y Acad Sci 842:7–15

Piatigorsky J, Vistow GJ (1989) Enzyme/Crystallins: gene sharing as evolutionary strategy. Cell 57:197–199

Qamra R, Mande SC (2004) Crystal structure of the 65-kDa heat shock protein, chaperonin 60.2 of *Mycobacterium tuberculosis*. J Bacteriol 186:8105–8113

Qamra R, Srinivas V, Mande SC (2004) *Mycobacterium tuberculosis* GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. J Mol Biol 342:605–617

Ran L, Huang F, Ekman M, Klint J, Bergman B (2007) Proteomic analyses of the photo auto- and diazotrophically grown cyanobacterium Nostoc sp. PCC 73102. Microbiology 153:608–618

Rao T, Lund PA (2010) Differential expression of the multiple chaperonins of *Mycobacterium smegmatis*. FEMS Microbiol Lett 310:24–31

Read J, Pearce J, Li X, Muirhead H, Chirgwin J, Davies C (2001) The crystal structure of human phosphoglucose isomerase at 1.6 Å resolution: implications for catalytic mechanism, cytokine activity and haemolytic anaemia. J Mol Biol 309:447–463

Reddi K, Meghji S, Nair SP, Arnett TR, Miller AD, Preuss M, Wilson M, Henderson B, Hill P (1998) The *Escherichia coli* chaperonin 60 (*groEL*) is a potent stimulator of osteoclast formation. J Bone Miner Res 13:1260–1266

Riffo-Vasquez Y, Coates AR, Page CP, Spina D (2012) *Mycobacterium tuberculosis* chaperonin 60.1 inhibits leukocyte diapedesis in a murine model of allergic lung inflammation. Am J Respir Cell Mol Biol 47:245–252

Robinson CV, Groß M, Eyles SJ, Ewbank JJ, Mayhew M, Hartl F-U, Dobson CM, Radford SE (1995) Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry. Nature 372:646–651

Rudolph B, Gebendorfer KM, Buchner J, Winter J (2010) Evolution of *Escherichia coli* for growth at high temperatures. J Biol Chem 285:19029–19034

Schneiker S et al (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat Biotechnol 25:1281–1289

Schulz LC, Bahr JM (2003) Glucose-6-phosphate isomerase is necessary for embryo implantation in the domestic ferret. Proc Natl Acad Sci U S A 100:8561–8566

Schulz LC, Bahr JM (2004) Potential endocrine function of the glycolytic enzyme glucose-6-phosphate isomerase during implantation. Gen Comp Endocrinol 13:283–287

Sielaff B, Lee KS, Tsai FT (2011) Structural and functional conservation of *Mycobacterium tuberculosis* GroEL paralogs suggests that GroEL1 is a chaperonin. J Mol Biol 405:831–839

Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. Microbiology 148:3129–3138

Suzuki M, Ueno T, Iizuka R, Miura T, Zako T, Akahori R, Miyake T, Shimamoto N, Aoki M, Tanii T, Ohdomari I, Funatsu T (2008) Effect of the C-terminal truncation on the functional cycle of chaperonin GroEL: implication that the C-terminal region facilitates the transition from the folding-arrested to the folding-competent state. J Biol Chem 283:23931–23939

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tanaka N, Hiyama T, Nakamoto H (1997) Cloning, characterization and functional analysis of *groESL* operon from thermophilic cyanobacterium *Synechococcus vulcanus*. Biochem Biophys Acta 1343:335–348

Tanaka N, Haga A, Uemura H, Akiyama H, Funasaka T, Nagase H, Raz A, Nakamura KT (2002) Inhibition mechanism of cytokine activity of human autocrine motility factor examined by crystal structure analyses and site-directed mutagenesis studies. J Mol Biol 318:985–997

Tang YC, Chang HC, Roeben A, Wischnewski D, Wischnewski N, Kerner MJ, Hartl FU, Hayer-Hartl M (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. Cell 125:903–914

Techtmann SM, Robb FT (2010) Archaeal-like chaperonins in bacteria. Proc Natl Acad Sci U S A 107:20269–20274

Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla-Camps JC, Theil EC, Volz K (2006) Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. Science 314:1903–1908

Wang J, Herman C, Tipton K, Gross C, Weissman J (2002) Directed evolution of substrate-optimized GroEL/S chaperonins. Cell 111:1027–1039

Wast J, Fraunholz M, Zauner S, Douglas S, Maier UG (1999) Ancient gene duplication and differential gene flow in plastid lineages: the GroEL/Cpn60 example. J Mol Evol 48:112–117

Watanabe H, Takehana K, Date M, Shinozaki T, Raz A (1996) Tumor cell autocrine motility factor is the neuroleukin/phosphohexose isomerase polypeptide. Cancer Res 56:2960–2963

Wistow GJ, Piatigorsky J (1988) Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Annu Rev Biochem 57:479–504

Wistow GJ, Lietman T, Williams LA, Stapel SO, de Jong WW, Horwitz J, Piatigorsky J (1998) τ-crystallin/α-enolase: one gene encodes both an enzyme and a lens structural protein. J Cell Biol 107:2729–2736

Xu W, Seiter K, Feldman E, Ahmed T, Chiao JW (1996) The differentiation and maturation mediator for human myeloid leukemia cells shares homology with neuroleukin or phosphoglucose isomerase. Blood 87:4502–4506

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)$_7$ chaperonin complex. Nature 388:741–750

Yoshida N, Oeda K, Watanabe E, Mikami T, Fukita Y, Nishimura K, Komai K, Matsuda K (2001) Protein function. Chaperonin turned insect toxin. Nature 411:44

Zahn R, Spitzfaden C, Ottiger M, Wüthrich K, Plückthun A (1994) Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. Nature 368:261–265