Heat Shock Proteins 11
Series Editors: Alexzander A. A. Asea · Stuart K. Calderwood

C. M. Santosh Kumar
Shekhar C. Mande *Editors*

Prokaryotic Chaperonins

Multiple Copies and Multitude Functions

Springer

Heat Shock Proteins

Volume 11

Series editors

Alexzander A. A. Asea, Ph.D.
Professor and VD for Research Innovations
Deanship for Scientific Research, University of Dammam
Dammam, Saudi Arabia

Stuart K. Calderwood, Ph.D.
Professor and Director
Division of Molecular and Cellular Radiation Oncology
Department of Radiation Oncology, Harvard Medical School of Beth Israel
Deaconess Medical Center
Boston, MA, USA

Heat Shock Proteins: key mediators of Health and Disease. Heat shock proteins (HSP) are essential molecules conserved through cellular evolution required for cells to survive the stresses encountered in the environment and in the tissues of the developing and aging organism. These proteins play the essential roles in stress of preventing the initiation of programmed cell death and repairing damage to the proteome permitting resumption of normal metabolism. Loss of the HSP is lethal either in the short-term in cases of acute stress or in the long-term when exposure to stress is chronic. Cells appear to walk a fine line in terms of HSP expression. If expression falls below a certain level, cells become sensitive to oxidative damage that influences aging and protein aggregation disease. If HSP levels rise above the normal range, inflammatory and oncogenic changes occur. It is becoming clear that HSP are emerging as remarkably versatile mediators of health and disease. The aim of this series of volumes is to examine how HSP regulation and expression become altered in pathological states and how this may be remedied by pharmacological and other interventions.

More information about this series at http://www.springer.com/series/7515

C. M. Santosh Kumar • Shekhar C. Mande
Editors

Prokaryotic Chaperonins

Multiple Copies and Multitude Functions

Springer

Editors
C. M. Santosh Kumar
Institute of Microbiology and
Infection – School of Biosciences
University of Birmingham
Birmingham
West Midlands
UK

Shekhar C. Mande
Laboratory of Structural Biology
National Centre for Cell Science
Pune
Maharashtra
India

ISSN 1877-1246 ISSN 1877-1254 (electronic)
Heat Shock Proteins
ISBN 978-981-10-4650-6 ISBN 978-981-10-4651-3 (eBook)
DOI 10.1007/978-981-10-4651-3

Library of Congress Control Number: 2017945831

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Chaperonins are a fascinating class of molecular chaperones that form a double toroidal architecture, consisting of two isologous rings of 7–9 subunits, each with a large central cavity for binding and encapsulation of naïve or unfolded substrate proteins. Chaperonin-bound substrates are encapsulated with the help of the co-chaperonin that acts as a lid for the central cavity. The bound substrates are allowed to fold upon encapsulation in an ATP-dependent manner. This enables sequestering the substrates from adverse folding environment and consequently facilitates their folding. Physiological function and mechanism of action of the chaperonins have been well-studied using *Escherichia coli* chaperonin, GroEL, as the model. The biological significance of chaperonins stems from the fact that they assist folding of about 10–15% of cellular proteins, including many essential proteins. A plethora of information through structural and functional studies on GroEL has enabled the proposition of a generalized mechanism of action and regulation of prokaryotic chaperonins.

However, the discovery of multiple chaperonins with a multitude of functions, in about 30% of the newly sequenced bacteria, has dramatically shifted the paradigm of chaperonin function. The presence of multiple chaperonins introduced new conundrums on whether they enhance general chaperoning ability in the cell or have deviated to undertake any specific novel cellular roles. Although the latter view is widely supported, evidence for the former is beginning to appear. While some of these multiple-copy chaperonins can functionally replace *E. coli* GroEL and thus are essential, the others are ineffective and likewise are non-essential. Surprisingly, several of these non-essential paralogues have been demonstrated to have acquired novel moonlighting functions, including antigenic and pathogenic functions. Notably, the GroEL1 in mycobacteria has been implicated in the formation of granulomas and disease progression, while GroEL2 acts as a general chaperone. Furthermore, in different classes of bacteria such as myxobacteria, cyanobacteria and rhizobia, the chaperonins have exhibited life-phase specific functional regulation. These observations directly imply functional variation amongst these chaperonin paralogues.

Therefore, the *extra* copies of chaperonins in several bacteria are believed to assist the organism during different phases of its life cycle. Furthermore, studies on the phylogenetic distribution of multiple GroELs revealed a specific pattern of distribution, rather than a random distribution, suggesting a strong biological correlation for the presence of multiple genes. Moreover, evolutionary analysis suggested that acquisition of multiple chaperonins followed case-specific evolutionary paths. For example, while multiple copies of the chaperonins resulted from a gene-duplication event in mycobacteria, in methanogens they resulted from horizontal acquisition. Therefore, in a nutshell, with multitude functions and diverse biological roles, these multiple chaperonins are changing the outlook of chaperonin biology. These studies, therefore, have suggested larger functional roles for chaperonins and consequently necessitated a comprehensive understanding of the structural, biochemical, functional and phylogenetic attributes of this class of molecular chaperones. Gaining evidences of the diverse roles of chaperonins would enable translating the biological significance of the multiple copies towards human welfare. Therefore, in this book, we present the current perception on the multiple chaperonins and their physiological and functional specificities.

Since the book deals with chaperonins, proteins that assist folding of other proteins in the cell, we have begun the book with an introductory note on the current advances in the understanding of structure-function relations and mechanism of action of chaperonins in Chap. 1. In addition, a brief sketch on the classification of the chaperonins into Group I, Group II and newly identified Group III, with an emphasis on their physiological features, has also been discussed in this chapter. Since different chaperonin genes are regulated differently, Chap. 2 has been devoted to review classical and novel modes of regulation of heat shock response in different bacteria. Moving on to the multiple chaperonins, Chap. 3 presents an overview on the functional diversity of multiple chaperonins in prokaryotes and will introduce subsequent chapters, 4 through 9, each of which comprehensively reviews different fascinating cases of multiple chaperonins. To understand how these multiple chaperonin genes have emerged, evolution and phylogenetic distribution of the multiple chaperonins are presented in Chap. 10. This chapter, with interesting activities for the readers, discusses possible modes of evolution and pathways of distribution of multiple chaperonins.

Therefore, we are convinced that this book, by bringing together leading experts in the field of chaperone biology, presents enthusiastic readers with a comprehensive review on the current advances in the understanding of the functional diversity of chaperonins, particularly multiple chaperonins. This is followed by an exciting and novel discussion on the possible modes of evolution and distribution of these multiple chaperonins. Therefore, we believe this book will serve as a reference for life science researchers, particularly those in the field of protein folding and molecular chaperones. Santosh is Newton International Fellow at the University of Birmingham, UK, sponsored by the The Royal Society, The British Academy and the Academy of Medical Sciences, UK. Further, we wish to acknowledge the support of Department of Biotechnology, India.

Birmingham, UK — C. M. Santosh Kumar
Pune, India — Shekhar C. Mande

Contents

About the Editors

C. M. Santosh Kumar started his research career as a molecular geneticist, understanding the functions of molecular chaperones in health and disease. His research aimed at understanding the functions of the bacterial chaperonins, especially that of the multiple chaperonins. He has demonstrated that the activity of a mycobacterial chaperonin is regulated by a phosphorylation switch that facilitates oligomerization. His current interests include unravelling the role of mycobacterial chaperonins in the establishment and progression of tubercular diseases.

Shekhar C. Mande is a structural biologist interested in understanding the molecular attributes of mycobacterial stress proteins. He started his career as a structural biologist in understanding the structural features of peanut lectin. Further, he became interested in the structure-function relations of mycobacterial stress proteins, such as the redox proteins and heat shock proteins. He led the way towards the understanding of structural features of the mycobacterial chaperonin proteins. His work demonstrated that mycobacterial chaperonins exhibit noncanonical attributes that are evolved to assist the pathogen in its disease establishment and progression. Concurrently, he began to explore the system-wide functional interactions amongst the mycobacterial proteins. These investigations have led to the identification of several novel interactions that are currently being examined.

Part I
Structure–Function of Chaperonins

Chapter 1
Structure, Function and Evolution of the Hsp60 Chaperonins

Sara E. Rowland and Frank T. Robb

Abstract In 1973, Christian Anfinsen and coworkers noted that accelerated protein folding in intact cells and cell extracts suggested that a "disulfide interchange enzyme" might be present in vivo. This concept of catalyzed folding foreshadowed the discovery of ubiquitous protein chaperones. The chaperonin GroEL/GroES was identified serendipitously when *GroE* mutants of *E. coli* failed to grow bacteriophage λ and were also temperature sensitive. The GroEL/GroES proved to be a ubiquitous chaperone and heat shock protein in bacteria and eukaryotic organelles, with two back-to-back rings of seven subunits each, forming a cavity that enclosed nonnative proteins, capped by the separate GroES lid complex. Group II chaperonins were subsequently discovered in all of the Archaea and in the Eukaryote cytoplasm with a similar cage-like shape, only with a "built-in" lid instead of the GroES module of Group I chaperonins. These chaperones have been intensely studied for three decades and have provided deep insights into protein-folding mechanisms. Despite this, some aspects of chaperonin-induced protein folding remain controversial.

The shared architecture and sequence similarity of two classes of chaperonins implies that they share a common ancestor. A recently identified, deeply branching clade of archaeal-like chaperonins encoded in bacteria may shed light on the early history of chaperonins. This clade shares many molecular properties with Group II chaperones; however, their phylogeny suggests that they arose early in prokaryotic evolution and may represent a vestige of the common ancestor of Group I and Group II chaperonins.

S.E. Rowland • F.T. Robb (✉)
Department of Microbiology and Immunology, School of Medicine, University of Maryland, Baltimore, and Institute of Marine and Environmental Technology (IMET), Columbus Center, Baltimore, MD, USA
e-mail: frobb@som.umaryland.edu

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_1

1.1 Protein Folding: Early Days

Christian Anfinsen discovered that a unique polypeptide sequence does not give rise to a functional protein until a specific folding pattern defines the native structure (Anfinsen 1973). The Anfinsen group established this with crucial ribonuclease renaturation experiments (Sela et al. 1957). Apart from rare intrinsically unstructured proteins (Wright and Dyson 1999), the majority of proteins must adopt a specific conformation to function appropriately in their biological roles. These studies showed that the final fold was imprinted in the unique amino acid sequence of each protein but also showed that spontaneous, unassisted folding was extremely slow. Levinthal concluded that ordered folding by a random walk through all possible conformations of a polypeptide chain would take far too long to be biologically relevant, but instead the native fold is likely to be achieved through a sequence of folding intermediates comprising a pathway to the production of mature, active protein (Levinthal 1968; Kim and Baldwin 1982). The first clues to catalyzed protein folding came when Anfinsen observed that the rates of refolding specifically due to the reoxidation of RNase in vitro were much slower compared with the much faster rates of active RNase production predicted in vivo. Adjusting variables including temperature, redox, and pH of the reaction affected the efficiency of spontaneous refolding; however tweaking these constraints alone failed to match in vivo rates of active RNase production. Anfinsen surmised that an additional factor must be present in vivo to facilitate these accelerated folding rates. Confirming this, reactivation of RNase was accelerated fivefold by adding rat liver homogenate to the refolding reaction (Goldberger et al. 1963; Venetianer and Straub 1963). Anfinsen inferred that a "disulfide interchange enzyme" present in vivo was acting to assist the RNase in achieving proper, native folds (Anfinsen 1973). This pioneering work founded the contemporary understanding that folding catalysis in trans is critical for maintaining the proteome, leading to the discovery of protein chaperones acting in all cells to assist their client proteins to rapidly mature into their native folds.

Pivotal experiments with *Escherichia coli* bacteriophage-resistant mutants revealed that the *groE* gene encoded a product that was essential for bacteriophage λ head morphogenesis at elevated temperatures (Sternberg 1973). The *groE*-negative *E. coli* strains were temperature sensitive and could not produce viable λ heads at 41 °C nor could they form colonies at 43 °C (Georgopoulos and Hohn 1978). Bacteriophages that were genetically modified to contain a functional *groE* bacterial gene and then transduced into *groE*$^-$ *E. coli*, however, were able to form colonies at 43 °C and recovered the ability to produce functional bacteriophage λ heads (Georgopoulos and Hohn 1978). The *groE* locus was soon established as a dicistronic operon (Tilly et al. 1981), and the *groEL* and *groES* genes were demonstrated to act in concert during phage λ head formation and growth of *E. coli* at elevated temperatures. These results implicated helper proteins in the proper maturation of polypeptide chains as predicted by Anfinsen but went further in suggesting that GroEL and GroES proteins were crucial for extending growth of *E. coli* beyond its optimal growth temperature. This insight

pointed to the existence of a global bacterial response to heat stress. Serendipitous studies on phage λ thus identified the chaperonin, a molecular chaperone subsequently identified across all domains of life.

The term "molecular chaperone" was first coined to describe a nucleosome assembly protein that facilitated the correct interaction of histones with DNA in *Xenopus laevis* eggs (Laskey et al. 1978). Chaperones are an eclectic group of proteins primarily responsible for maintaining protein homeostasis, or proteostasis within cells by preventing aggregation of unfolded proteins and promoting proper protein folding and assembly. Chaperonins are central players in the proteostasis network (Fayet et al. 1989). They are ubiquitous 1 MDa ATP-dependent complexes that interact with denatured proteins and cyclically encapsulate these polypeptides in a chamber to promote refolding and prevent folding intermediates from straying off the proper fold pathway. The discovery of bacterial *groEL* and *groES* in the 1970s was followed by the characterization of a homologous complex in higher plant chloroplasts that interacted directly with the newly synthesized large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase, or RuBisCo (Zweig and Cummings 1973; Sternberg 1973; Georgopoulos and Hohn 1978; Barraclough and Ellis 1980; Hemmingsen et al. 1988). Hemmingsen et al. first recognized the ubiquity of this protein complex, identified its broad role as a molecular chaperone acting on a diverse set of oligomeric protein complexes, and named it as the chaperonin (Hemmingsen et al. 1988).

1.2 Structure and Function of GroEL and GroES

The *E. coli* groEL chaperonin complex was identified as having sevenfold cylindrical symmetry with two stacked heptameric rings, 125 Å in diameter and 100 Å in height, formed with 14 identical ~65 kDa subunits (Georgopoulos and Hohn 1978; Hendrix 1978, 1979; Hohn et al. 1979). Later studies demonstrated the direct interaction of groEL with groES, a heptameric ring-like complex with a molecular weight of 80,000 (Chandrasekhar et al. 1986; Fig. 1.2 (top)). Observations including the co-purification of GroEL with several proteins including RNA polymerase and RuBisCo (Ishihama et al. 1976; Barraclough and Ellis 1980), and the requirement of a functional GroEL for *E. coli* growth at elevated temperature (Sternberg 1973; Georgopoulos and Hohn 1978), further clarified the biological role of these chaperonins. GroEL abundance in *E. coli* increased from 1.6% during growth at 37 °C to 15% of total cell protein mass when the culture was shifted to 46 °C (Herendeen et al. 1979). GroEL homologues in other systems including human cell lines and protozoa also exhibited similar heat shock-inducible responses, prompting the classification of this chaperonin as a major component of the heat shock response (Waldinger et al. 1989; McMullin and Hallberg 1987; Neidhardt et al. 1984). GroEL was found to have high and specific affinity for unfolded or misfolded protein forms (Bochkareva et al. 1988), which supported previous experiments that the presence of intracellular unfolded proteins directly stimulated the induction of heat shock

proteins (Parsell and Sauer 1989). Integrating previously collected data, Bochkareva et al. proposed that, during both normal growth conditions and stress conditions, the chaperonin's role in vivo is to assist nonnative proteins including newly synthesized and translocated polypeptides as well as stress-denatured proteins in reaching their native states by transient ATP-driven interactions (Bochkareva et al. 1988).

1.3 Chaperonin Cycling

While this prediction of the general mechanism for chaperonin action proved correct, and the chaperonin has been studied intensely for 30 years, the mechanism of chaperonin cycling is still controversial. Several competing models have been proposed regarding the productive interaction of the substrate protein and the chaperonin complex and the enclosure of substrate molecules in the aptly named Anfinsen cage (Gupta et al. 2014; Ellis 2013). The passive caging model proposes that the chaperonin complex merely provides an environment in which the unfolded substrate is sequestered from the crowded cytosol and given an opportunity to refold spontaneously without any work being done on the substrate by the complex. The inner lining of the cavity transitions form a hydrophobic surface to a hydrophilic one, promoting the exposed hydrophobic regions of the substrate to fold (Fenton and Horwich 2003; Apteri and Horwich 2008; Horwich et al. 2009). In the active caging model, it is believed that the chaperonin complex is capable of destabilizing internalized folding intermediates and can thus accelerate substrate folding (Sparrer et al. 1997; Gupta et al. 2014). Recent work supports yet a third model called the iterative annealing mechanism, whereby the chaperonin complex interacts with the substrate protein for multiple rounds of encapsulation and helps to forcibly unfold it to initiate refolding cycles. The proponents of the last model established that the polypeptide has a half time of 1 s inside the cavity and can be partially protruding from the chaperonin and may achieve native structure either inside or outside the complex (Shtilerman et al. 1999; Yang et al. 2013; Motojima and Yoshida 2010).

In addition to the various models of chaperonin-mediated substrate protein refolding, conflicting reports describe different modes of inter-ring interaction, the timing of their nucleotide binding and hydrolysis, and substrate binding and encapsulation. Currently, chaperonins are grouped by their structure and phylogeny (Kim et al. 1994; Woese et al. 1990). Group I chaperonins encompass those present in bacteria and organelles of eukaryotes and are composed of two homomeric seven-membered rings that interact with the co-chaperone known either as GroES or Hsp10 adapted for substrate displacement insertion, acting as a lid during closure of the cavities. Group II chaperonins are present in archaea and the cytosol of eukaryotes, form homo- or heteromeric eight- or nine-membered rings, and function independent of a co-chaperone, by closing off the central cavities with use of a helical protrusion in the apical domains of subunits known as a built-in lid (Fig. 1.1).

Both groups of chaperonins form double stacked ring complexes, with each ring forming a cavity for substrate protein encapsulation and refolding. The actual mechanism of protein folding within the cage remains controversial. With Group I

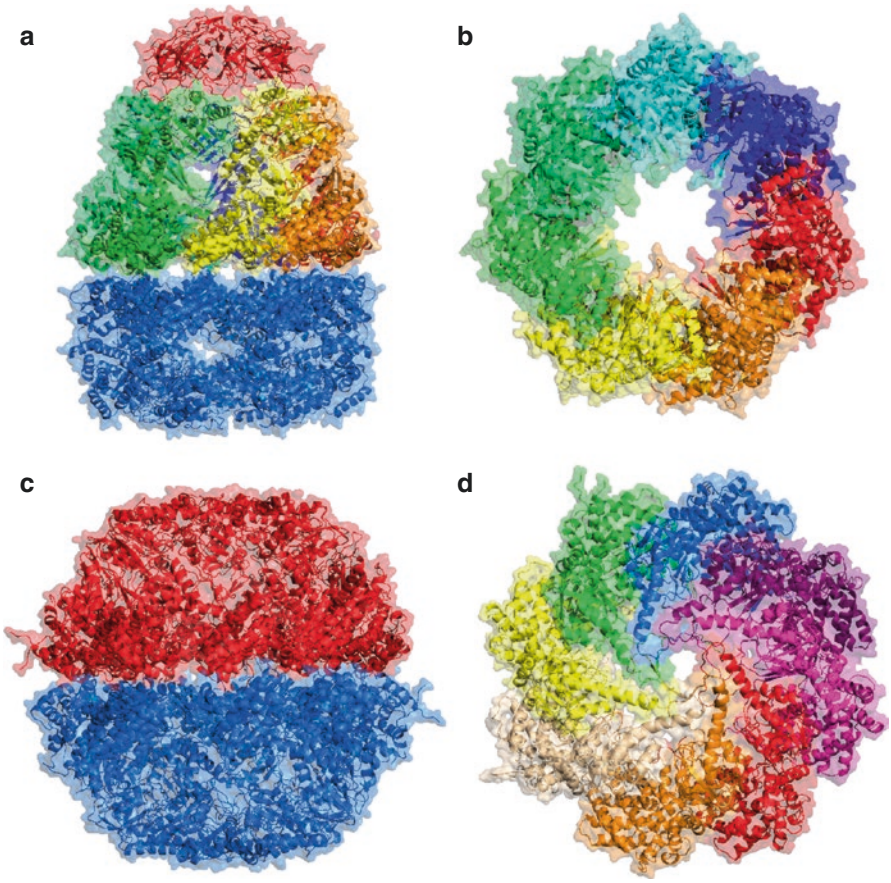


Fig. 1.1 Comparison of Group I and Group II chaperonin architecture. *Top*: Crystal structure of *Thermus thermophilus* Group I chaperonin complexed with GroES and seven ADP molecules. (**a**) GroEL as seen from the side with one ring in *blue*, and the other ring with each individual subunit shown in a different color interacting with a *red* GroES heptamer. (**b**) A single monomeric GroEL ring as seen from the apical domain down showing the central cavity of the CPN when complexed with GroES (hidden), (PDB 4V4O; Shimamura et al. 2004). *Bottom*: Crystal structure of *Saccharomyces cerevisiae* Group II chaperonin (CCT). (**c**) Stacked rings shown in *red* and *blue*. (**d**) Single ring shown with apical domain on top and with each subunit displayed in a unique color. This view (PDB 4 V81; Dekker et al. 2011a, b) clearly shows the apical protrusion or built-in lid. Figures generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC.)

chaperonins, the *cis* binding model states that the substrate protein binds at the apical domain of one ring, ATP binds at the same ring, then a conformational change causes the substrate to be released into the chamber upon simultaneous binding of GroES co-chaperone cap. Binding of ATP to the *trans* ring causes displacement of the GroES from the *cis* ring and the cycle repeats (Xu et al. 1997; Horwich et al. 2009). An alternative model proposes that ATP and GroES can bind both rings simultaneously, forming a symmetric complex that is favored in the presence of

substrate and allows for many more iterations of work on a substrate than estimated with the *cis* model (Sparrer et al. 1997; Yang et al. 2013).

For Group II chaperonins present in archaea and in the cytosol of eukaryotes, ATP hydrolysis is required for complete closure of the built-in lid and introduction of substrate into the central cavity (Douglas et al. 2011). Depending on the size and structure of the substrate, the chaperonin may be able to encapsulate the entire protein, or in the case of larger, multidomain substrates, it may partially enclose a segment of the protein leaving the remainder exposed to the cytosol. This explains the ability of Group II chaperonins to interact with clients that are considerably larger than the predicted cavity space (Ruessmann et al. 2012, Kurouski et al. 2012, 2013). Chaperonin subunits undergo profound conformational changes during the opening and closing cycle (Pereira et al. 2010). As with Group I chaperonins, the complete cycle is still unclear. It is suggested that ATP concentrations alter the allosteric function of the rings where low to intermediate concentrations of ATP require ADP dissociation in the trans ring for ATP hydrolysis to occur in the *cis* ring, while high concentrations of ATP permit simultaneous ring closure of the whole complex (Pereira et al. 2010). The built-in lid also appears to play a critical allosteric role in coordinating synchronized movement of intra-ring subunits as well as in inter-ring communication and is essential for substrate refolding but not ATP hydrolysis (Reissmann et al. 2007). Recent studies with yeast CCT with ATP hydrolysis synchronized by stopped-flow methods showed that the initial encounter of CCT with ATP results in a burst of hydrolysis, followed by a second burst suggesting that a conformational change is required before a second round of ATP binding can be initiated. This is evidence for inter-ring coordination of allosteric choreography in response to the initial binding and hydrolysis of ATP (Korobko et al. 2016).

1.4 The Evolutionary History of Chaperonins

Chaperonins are thought to have evolved initially from a rearrangement of a thioredoxin superfamily protein, specifically a member with a peroxiredoxin fold (Dekker et al. 2011a, b). Peroxiredoxins can form ring-like decamers and are reported to switch from peroxide detoxification to holdase molecular chaperoning activity via phosphorylation during stress events such as low pH and increase in temperature (Jang et al. 2006; Saccoccia et al. 2012; Angelucci et al. 2013; Teixeira et al. 2015). This is an example of protein moonlighting, in which a single polypeptide can adopt a new function involving a different set of interactions than the original role required. It is possible that modern chaperonins were not even present in the last universal common ancestor and were instead present as a bifunctional progenitor such as the peroxiredoxins.

Due to the pervasiveness of chaperonins across domains, larger evolutionary theories can be supported based on related chaperonin groups in much the same way that ribosomal gene sequences have been used to restructure the tree of life (Woese and Fox 1977; Woese et al. 1990). Since the early 1990s, chaperonins have been categorized into two distinct groups based on their structure and phylogeny (Kim

et al. 1994). Like the ribosomal RNA sequences, which are sequence-related but are separated into 16S and 18S classes, the dichotomy of the chaperonin classes begs the question of what the common ancestor of the chaperonins looked like. This ubiquitous protein family can inform on domain evolution (Woese et al. 1990). For instance, early observations that mitochondrial and chloroplast chaperonin sequences were most similar to those in purple bacteria and cyanobacteria, respectively, further supported the endosymbiont origin theory of these organelles in eukaryotic organisms (Gupta et al. 1989).

As more Cpn homologs are continually emerging from the deluge of new genomic and metagenomic sequence data, it has become evident that the chaperonin groups by and large adhere to the three domains of life proposed by Carl Woese and colleagues (Woese et al. 1990). Previously characterized polypeptides were found to be distantly related to the bacterial-type or groEL chaperonins including the cytosolic eukaryotic TCP-1 (tailless complex polypeptide 1) or CCT (chaperonin containing TCP-1), but these sequences are distantly related to GroEL compared to the Group I chaperonins of eukaryotic organelles. The cytosolic chaperonins form heterooligomeric complexes, mostly with eight subunits per ring, and are only about 20% identical and 40–60% similar to the chaperonins of bacteria and eukaryotic organelles (Gupta 1990; Rommelaere et al. 1993). In archaea, heat shock inducible homo- and heteromeric eight- to nine-membered rings that form double stacks with ATPase activity and ~40% identity to TCP-1 were discovered (Trent et al. 1991; Phipps et al. 1991, 1993; Marco et al. 1994; Guagliardi et al. 1994). The functional similarities of eukaryotic and archaeal Group II chaperonins are striking, despite the differences in the lifestyles of these organisms. It has been possible to model a human heritable pathogenic mutation in the human cctγ gene encoding the CCT5 subunit of the octameric human complex in the chaperonin from a hyperthermophilic archaeon, *Pyrococcus furiosus* (Min et al. 2014). The archaeal chaperonin is built from a single subunit and thus the subtle biochemical defects in the human CCT complex with one defective subunit are amplified eightfold in the archaeal model.

In the eukaryotic cytosol GroES or hsp10, homologs are notably absent (Gupta 1995) although bacterial-type GroES/GroEL copies occur in some archaea, most likely representing lateral gene transfers from bacteria. The *groEL* and *groES* genes are typically linked in dicistronic operons; their protein products form the Group I chaperonins, also known as Hsp60 or Cpn60 and Hsp10 or Cpn10, respectively, and are present in all bacteria and eukaryotic organelles except for a few Mycoplasma species (Kim et al. 1994; Lund 2009). Characterized Group I chaperonins generally form homomeric tetradecamers like the *E. coli* GroEL with the exception of higher plant chloroplasts which form heteromeric complexes of alpha and beta subunits or beta alone in the presence of MgATP (Martel et al. 1990; Dickson et al. 2000).

In contrast, the Group II chaperonins are primarily present in archaea and in the cytosol of eukaryotes. These chaperonins can form homo- or heteromeric eight- to nine-membered double ring structures with domain arrangements similar to the Group I chaperonins, although the apical domain does not rely on interaction with a GroES homolog for complete cycling and instead has a helix-turn-helix protrusion with a large hydrophobic surface referred to as a built-in lid responsible for binding substrate and facilitating folding (Fig. 1.1; Klumpp et al. 1997; Yoshida et al. 2002;

Reissmann et al. 2007). Eukaryotic CCT rings consists of eight distinct subunits that complex in a specific pattern, while archaeal thermosomes consist of eight- to nine-membered rings ranging from a single to as many as five different subunits (Dekker et al. 2011a, b; Leitner et al. 2015; Bigotti and Clarke 2008). The heterooligomeric complexes are thought to have arisen via gene duplications. In the archaea, paralogs within species are typically more closely related to each other than with other archaeal chaperonin genes, suggesting intraspecies gene duplication. Duplications are thought to have occurred independently many times, with some lineages being lost in archaeal genomes (Archibald et al. 1999). Repeated gene conversions are thought to have slowed differentiation in the substrate-binding domain of these paralogous archaeal genes (Archibald and Roger 2002). In contrast, the CCT subunits in the eukarya are thought to have evolved from a very early multiple duplication event in a proto-eukaryote leading to eight distinct subunits, with subunit-specific variation concentrated in the apical domain sequences (Archibald et al. 2001). Specific substrates have been identified for eukaryotic chaperonins including the cytoskeletal components, tubulin, and actin (Gao et al. 1992; Sternlicht et al. 1993). Mutagenesis and crosslinking-mass spectrometry studies identified the residues in each unique subunit that makes direct contact with these physiologically relevant substrates. Results suggest that several low-affinity interactions across subunits are required to discern folded versus nonfolded substrate (Joachimiak et al. 2014) (Figs. 1.2 and 1.3).

A third group of chaperonins was recently discovered in the genomes of a select set of Firmicutes (Techtmann and Robb 2010; Williams et al. 2010). These chapero-

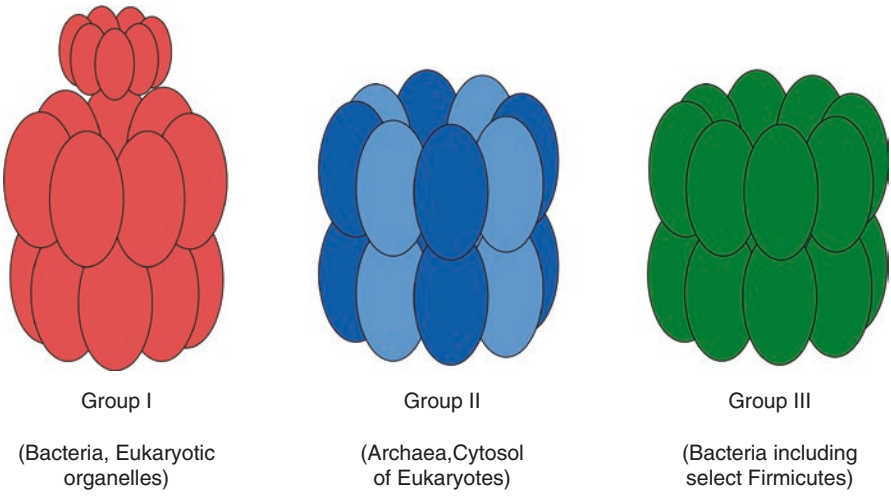


Fig. 1.2 Cartoon depiction of the structural and phylogenetic variation of known chaperonins. Group I (*red*) cpns are found in bacteria and eukaryotic organelles, interact with a co-chaperone lid (Hsp10/GroES), and are monomeric 14-mer complexes. Group II (*blue*) cpns are found in archaea and the cytosol of eukaryotes, have a built-in lid, and can form homo- or heteromeric complexes of eight- or nine-membered rings. Group III (*green*) cpns are found mostly in Firmicutes, form homomeric complexes of eight-mer rings, and contain a built-in lid

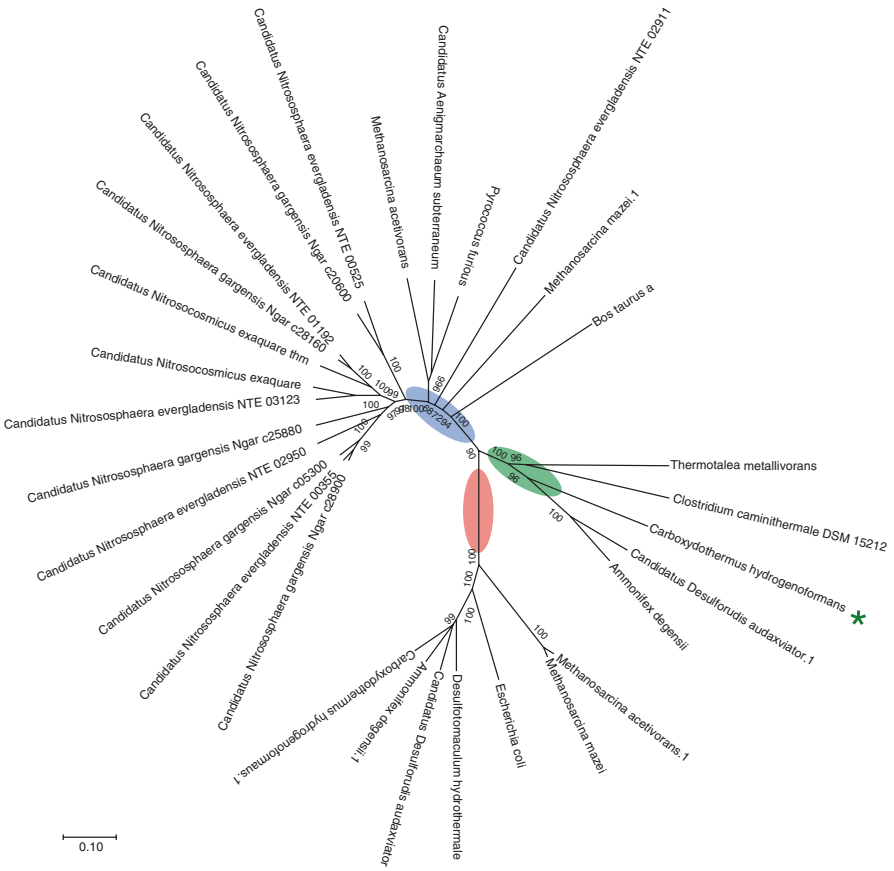


Fig. 1.3 Molecular phylogenetic analysis of chaperonin groups by maximum likelihood. *Red*: Group I sequences. *Blue*: Group II sequences. *Yellow*: Non-Group II clustering archaeal sequences. *Green*: Bacterial archaeal-like sequences. The evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model (Jones 1992). The tree with the highest log likelihood (−58873.9569) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 190 full length chaperonin amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 370 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar 2015). The initial alignment was constructed using ClustalW default settings

nins are structurally similar to Group II chaperonins but phylogenetically (Techtman and Robb 2010) and structurally distinct from both groups (An et al. 2016). All bacterial genomes with these archaeal-like chaperonins also contain at least one copy of a Group I chaperonin and its co-chaperone. While several lone and divergent archaeal-like chaperonin genes are present in other bacterial species including the cyanobacteria *Gloeobacter violaceus* PCC 7421 and *Gloeobacter kilaueensis* JS1,

and the proteobacterium *Oligotropha carboxidovorans* DSM 1227, the majority of Group III chaperonin genes appear to be under *hrcA* regulation and are situated downstream of *grpE-dnaK-dnaJ* chaperones (Williams et al. 2010; Techtmann 2009; Techtmann and Robb 2010). These novel chaperone operons are conserved across many Firmicutes, and the synteny of the chromosome neighborhood of the DnaK/DnaJ operon is conserved in most Firmicutes, including those lacking the archaeal-like chaperonin gene. CIRCE (controlling inverted repeat of chaperone expression) elements have been found upstream of both the *groESL* and *cpn/hsp70* operons in *C. hydrogenoformans*. This conserved sequence element binds the negative repressor HrcA (heat regulation at CIRCE) during stress-free conditions, inhibiting transcription of downstream chaperones (Schultz and Schumann 1996; Narberhaus 1999; Techtmann and Robb 2010). This suggests a coordinated co-regulation and an overlap of function of both the Group I and Group III chaperonins in vivo.

The continued exponential accumulation of sequences in the global microbiome will in time provide more evidence regarding the origin of this protein complex. The group is either the result of an interdomain gene transfer from archaea to bacteria or else they represent the relics of an ancient chaperonin that existed in the last universal common ancestor prior to the splitting of archaeal and bacterial lineages. The horizontal acquisition of Group I chaperonins by methanogens and haloarchaea are clearly discernible considering the tight grouping of their GroEL/GroES with bacterial Group I cpn sequences (Deppenmeier et al. 2002; Klunker et al. 2003). The green and yellow branches of Fig. 1.4 represent sequences from many organisms

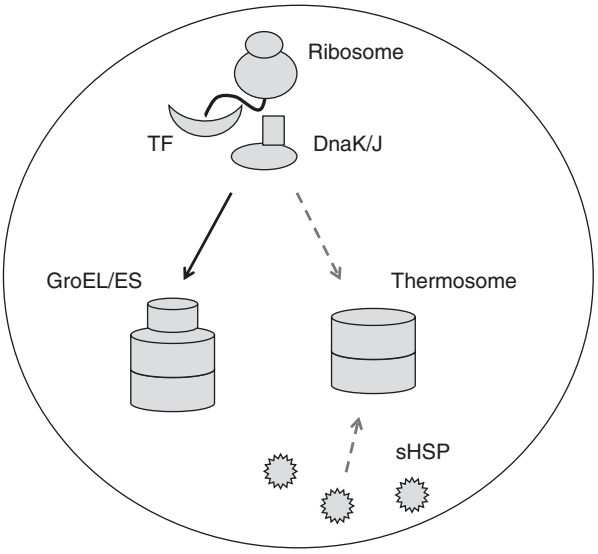


Fig. 1.4 The chaperones of *C. hydrogenoformans* and their potential substrate folding workflow. *Dashed arrows* represent proposed chaperone-mediated substrate delivery pathways to the archaeal-like chaperonin in the absence of prefoldin; *solid arrow* represents a widely accepted substrate flow between DnaK/J and GroEL/ES. *TF* Trigger Factor, *sHSP* small heat shock proteins

that are difficult to sample and/or culture in the laboratory. Techtmann argued that the absence of the conserved clustering of the Group III chaperonin within the DnaK/DnaJ operon in a combination that is not represented in any archaeal genomes to date suggests that this sequence was not obtained from an archaeal genome (Techtmann 2009; Techtmann and Robb 2010). Thus the hypothesis that this class of chaperones represents a relic of the ancestral class of chaperonins is supported.

1.5 Chaperonin Interaction with Co-Chaperones and Chaperone Networks

In addition to GroEL's association with the co-chaperone GroES, chaperonins have been reported to interact directly with other molecular chaperones within the cell. The Group II-containing systems interact with a hexameric holdase chaperone known as prefoldin (PFD or GimC), which off-loads nonnative substrates such as eukaryotic actin and tubulin into chaperonin chambers (Vainberg et al. 1998). Substrate interaction can occur as the polypeptide chain is being synthesized, allowing for efficient, protected delivery to chaperonin cavities for proper folding while minimizing the incidence of aggregation in the cytosol (Hansen et al. 1999). Prefoldin is a heteromeric complex composed of a double beta barrel body with six coiled-coil protrusions each containing hydrophobic patches at the ends for unfolded substrate recognition and interaction (Siegart et al. 2000). The termini of the beta subunit interact directly with chaperonins during substrate delivery, at least in vitro (Okochi et al. 2004). Other holdases such as small heat shock proteins are thought to function in a similar substrate delivery method as demonstrated with in vitro protein folding and salvage assays with chaperonins, although direct contact between the two chaperones has not yet been validated (Laksanalamai et al. 2008).

DnaK or Hsp70, a dimeric ATP-dependent chaperone that interacts with nascent polypeptide chains using its C-terminal substrate-binding domain in coordination with co-chaperones DnaK and GrpE, is a crucial component of the cellular proteostasis network. In proteomic studies using *E. coli*, DnaK demonstrated nonspecific interaction with approximately 700 cytosolic proteins (Calloni et al. 2012). DnaK is described as forming productive folding pathways with other chaperones including chaperonins (Beckmann et al. 1990). Cytosolic DnaK has been shown to form a stable complex with CCT subunits in eukaryotic systems, although this direct interaction is not demonstrable using the bacterial DnaK homologs (Cuellar et al. 2008). DnaK is absent in several species of Aquificales and all of the Archaea apart from instances of horizontal gene transfer with bacteria (Warnecke 2012; Macario et al. 1991).

Thus, chaperonins are thought to participate in a network with other chaperones in the cell to maintain proteostasis (Fig. 1.4). These network members vary across domains and include proteins that assist in polypeptide transport, aggregate circumvention, folding, and degradation.

1.6 Chaperonin Substrates

The foundational understanding of Group I chaperonin substrate recognition and binding is based on the *E. coli* GroEL system. GroEL structures reveal hydrophobic residues in alpha-helices of the apical domain that are crucial for establishing interactions with substrates and the co-chaperone GroES, which upon binding in the presence of MgATP displaces substrate into the central cavity (Fenton et al. 1994; Fenton and Horwich 2003). Structural studies of bound substrates report findings ranging from loosely organized proteins with variable amounts of secondary structure to random coils devoid of any tertiary structure to disordered polypeptides that associate weakly at a single hydrophobic region but likely become more disordered upon binding to GroEL, promoting exposure of more hydrophobic segments of the unfolded substrate, and ultimately permitting a stable substrate interaction across multiple GroEL subunits (Fenton and Horwich 2003; Horst et al. 2005; Libich et al. 2013). While most substrates are fully encapsulated by GroEL/ES complex during the refolding cycle, there are reports documenting the productive folding of substrates exceeding the size capacity of the internal GroEL cavity (Chaudhuri et al. 2001; Kurouski et al. 2012, 2013).

Proteome-wide obligate GroEL substrates in *E. coli* have been identified by molecular association studies. The obligate client protein set is responsible for a diverse range of cellular functions, and attempts to generally characterize the requirements for obligate chaperonin interaction include a polypeptide size of 20–60 kDa, and the presence of maintained, exposed hydrophobic regions mediated either by a slow folding rate, an affinity for a co-chaperone to deliver substrate to the chaperonin, and/or presence of beta-sheets normally buried in the native structure (Ewalt et al. 1997; Houry et al. 1999; Kerner et al. 2005). Substrates with native folding motifs including TIM barrel domains and knotted proteins are also thought to require interaction with the chaperonin system for proper, efficient folding, possibly due to a propensity for proteins with these similar tertiary structures to get stuck more frequently in similar off-pathway intermediates (Kerner et al. 2005; Fujiwara et al. 2010; Lim and Jackson 2015). A similar in vivo GroEL substrate study was completed in the Gram-positive bacterium *Bacillus subtilis*, and several substrates absent in *E. coli* including sporulation proteins were identified, indicating that chaperonin interactions may be promiscuous but are still variable across species depending on the host proteome (Endo and Kurusu 2007). Protein folding should be considered in the crowded in vivo cellular environment as the polypeptide chain is being synthesized and simultaneously exiting the ribosome. Larger continuous proteins may have domains of variable stabilities, proteins may form different secondary structural elements before others, N and C-termini cannot interact until translation is complete, and ribosomal pausing resulting in rare codon insertion into beta-sheet sequences are all factors that can influence protein-folding intermediate formation and potentially their interaction with chaperonins (Tsytlonok and Itzhaki 2013).

The enigma of chaperonin discrimination of substrates has been the topic of several unsuccessful attempts to identify any sequence similarity in *E. coli* chaperonin substrates apart from nonspecific, hydrophobic interactions. The GroES mobile loop sequence that binds to the H and I helices of the GroEL apical domain consists of a hydrophobic region, GGIVLTGAA. Because GroES and unfolded substrate bind the same region of GroEL, researchers looked for a similar sequence pattern in *E. coli* proteins. Sixty percent of *E. coli* proteins contain one hydrophobic patch, and approximately 30% contain multiple hydrophobic patches (Chaudhuri and Gupta 2005; Stan et al. 2005). Using this sequence-based approach in combination with previous experimentally derived in vivo substrates, it was estimated that four to five contact points should be established by the denatured substrate with GroEL across two to four subunits (Horovitz 1998). The substrate-binding motifs of identified substrate proteins with available structures are buried in their native state, thereby establishing a modus operandi for GroEL to recognize denatured versus folded substrates (Stan et al. 2005, 2006; Houry et al. 1999; Kerner et al. 2005). Additional studies are required to make more informed predictions regarding GroEL interactors, especially across species (Azia et al. 2012). While final folds of client proteins may have little influence on selection as chaperonin substrate, it is still an interesting trend that the more diverse proteomes, in terms of multiple fold families, typically contain the most copies of chaperonin genes and subunits including proteobacteria and eukaryotes (Stan et al. 2006; Lund 2009; Kim and Caetano-Anollés 2012).

References

An YJ, Rowland SE, Robb FT, Cha SS (2016) Purification, crystallization and preliminary X-ray crystallographic analysis of group III chaperonin from *Carboxydothermus hydrogenoformans*. J Microbiol 54(6):440–444., Accepted for publication

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

Angelucci F, Saccoccia F, Ardini M, Boumis G, Brunori M, Di Leandro L, Ippoliti R, Miele AE, Natoli G, Scotti S, Bellelli A (2013) Switching between the alternative structures and functions of a 2-Cys peroxiredoxin, by site-directed mutagenesis. J Mol Biol 425:4556–4568

Apetri AC, Horwich AL (2008) Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proc Natl Acad Sci 105:17351–17355

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

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

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

Azia A, Unger R, Horovitz A (2012) What distinguishes GroEL substrates from other *Escherichia coli* proteins? FEBS J 279:543–550

Barraclough R, Ellis RJ (1980) Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. Biochim Biophys Acta 608:19–31

Beckmann RP, Mizzen LE, Welch WJ (1990) Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. Science 18:850–854

Bigotti MG, Clarke AR (2008) Chaperonins: the hunt for the group II mechanism. Arch Biochem Biophys 474:331–339

Bochkareva ES, Lissin NM, Girshovich AS (1988) Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature 336:254–257

Calloni G, Chen T, Schermann SM, Chang HC, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl M, Hartl FU (2012) DnaK functions as a central hub in the *E. coli* chaperone network. Cell Rep 1:251–264

Chandrasekhar GN, Tilly K, Woolford C, Hendrix R, Georgopoulos C (1986) Purification and properties of the groES Morphogenetic protein of *Escherichia coli*. J Biol Chem 261(26):12414–12419

Chaudhuri TK, Gupta P (2005) Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. Cell Stress Chaperones 10:24–36

Chaudhuri TK, Farr GW, Fenton WA, Rospert S, Horwich AL (2001) GroEL/ES-mediated folding of a protein too large to be encapsulated. Cell 107:235–246

Cuellar J, Martin-Benito J, Scheres SHW, Sousa R, Moro F, Lopez-Vinas E, Gomez-Puertas P, Muga A, Carrascosa JL, Valpuesta JM (2008) The structure of CCT-Hsc70 NBD suggests a mechanisms for Hsp70 delivery of substrates to the chaperonin. Nat Struct Mol Biol 15:858–864

Dekker C, Roe SM, McCormack EA, Beuron F, Willison KR (2011a) The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. EMBO J 30:3078–3090

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

Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baeumer S, Jacobi C, Brueggemann H, Lienard T, Christmann A, Boemeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol 4:453–461

Dickson R, Weiss C, Howard RJ, Alldrick SP, Ellis RJ, Lorimer G, Azem A, Vitanen PV (2000) Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. J Biol Chem 275(16):11829–11835

Douglas NR, Reissmann S, Zhang J, Chen B, Jakana J, Kumar R, Chiu W, Frydman J (2011) Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. Cell 144(2):240–252

Ellis RJ (2013) Assembly chaperones: a perspective. Philos Trans R Soc B: Biol Sci 368(1617): 20110398

Endo A, Kurusu Y (2007) Identification of *in vivo* substrates of the chaperonin GroEL from *Bacillus subtilis*. Biosci Biotechnol Biochem 71:1073–1077

Ewalt KL, Hendrick JP, Houry WA, Hartl FU (1997) *In vivo* observation of polypeptide flux through the bacterial chaperonin system. Cell 90:491–500

Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J Bacteriol 171:1379–1385

Fenton WA, Horwich AL (2003a) Chaperonin-mediated protein folding: fate of substrate polypeptide. Q Rev Biophys 36:229–256

Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371:614–619

Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of *in vivo* obligate chaperonin-dependent substrates. EMBO J 29:1552–1564

Gao Y, Thomas JO, Chow RL, Lee GH, Cowan NJ (1992) A cytoplasmic chaperonin that catalyzes B-actin folding. Cell 69:1043–1050

Georgopoulos CP, Hohn B (1978) Identification of a host protein necessary for bacteriophage morphogenesis (the *groE* gene product). Proc Natl Acad Sci 75:131–135

Goldberger RF, Epstein CJ, Anfinsen CB (1963) Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. J Biol Chem 238:628

Guagliardi A, Cerchia L, Bartolucci S, Rossi M (1994) The chaperonin from the archaeon Sulfolobus solfataricus promotes correct refolding and prevents thermal denaturation in vitro. Protein Sci 3:1436–1443

Gupta RS (1995) Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. Mol Microbiol 15(1):1–11

Gupta RS (1990) Sequence and structural homology between a mouse T-complex protein TCP-1 and the 'chaperonin' family of bacterial (GroEL, 60-65kDa heat shock antigen) and eukaryotic proteins. Biochem Int 20:833–841

Gupta RS, Picketts DJ, Ahmad S (1989) A novel ubiquitous protein 'chaperonin' supports the endosymbiotic origin of mitochondrion and plant chloroplast. Biochem Biophys Res Commun 163:780–787

Gupta AJ, Halder S, Milicic G, Hartl FU, Hayer-Hartl M (2014) Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. J Mol Biol 426:2739–2754

Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333:330–334

Hendrix RW (1979) Purification and properties of groE, a host protein involved in bacteriophage assembly. J Mol Biol 129:375–392

Hendrix RW, Tsui L (1978) Role of the host in virus assembly: cloning of the *Escherichia coli* groE gene and identification of its protein product. Proc Natl Acad Sci U S A 75(1):136–139

Hansen WJ, Cowan NJ, Welch WJ (1999) Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. J Cell Biol 145:265–277

Herendeen SL, VanBogelen RA, Neidhardt FC (1979) Levels of major proteins of *Escherichia coli* during growth at different temperatures. J Bacteriol 139:185–194

Hohn T, Hohn B, Engel A, Wurtz M (1979) Isolation and characterization of the host protein groE involved in bacteriophage Lamba assembly. J Mol Biol 129:359–373

Horovitz A (1998) Structural aspects of GroEL function. Curr Opin Struct Biol 8(1):93–100

Horst R, Bertelsen EB, Fiaux J, Wider G, Horwich AL, Wuethrich K (2005) Direct NMR observation of a substrate protein bound to the chaperonin GroEL. Proc Natl Acad Sci 102:12748–12753

Horwich AL, Apetri AC, Fenton WA (2009) The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS Lett 583:2654–2662

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of *in vivo* substrates of the chaperonin GroEL. Nature 401:147–154

Ishihama AT, Ikeuchi T, Yura T (1976) A novel adenosine triphosphatase isolated from RNA polymerase preparation of *Escherichia coli* I. Copurification and separation. J Biochem 79(5):917–925

Jang HH, Kim SY, Park SK, Jeon HS, Lee YM, Jung JH, Lee SY, Chae HB, Jung YJ, Lee KO, Lim CO, Chung WS, Bahk JD, Yun DJ, Cho MJ, Lee SY (2006) Phosphorylation and concomitant structural changes in human 2-Cys peroxiredoxin isotype I differentially regulate its peroxidase and molecular chaperone functions. FEBS Lett 580:351–355

Joachimiak LA, Walzthoeni T, Liu CW, Aebersold R, Frydman J (2014) The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. Cell 159:1042–1055

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122:209–220

Kim PS, Baldwin RL (1982a) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51:459–489

Kim KM, Caetano-Anollés G (2012) The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. BMC Evol Biol 12:13

Kim S, Willison KR, Horwich AL (1994) Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. Trends Biochem Sci 19:543–548

Klunker D, Haas B, Hirtreiter A, Figueiredo L, Naylor DJ, Pfeifer G, Mueller V, Deppenmeier U, Gottschalk G, Hartle FU, Hayer-Hartl M (2003) Coexistence of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. J Biol Chem 278:33256–33267

Klumpp M, Baumeister W, Essen LO (1997) Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. Cell 91:263–270

Korobko I, Nadler-Holly M, Horovitz A (2016) Transient kinetic analysis of ATP hydrolysis by the CCT/TRiC chaperonin. J Mol Biol 428(22):4520–4527. Sep 26. Epub.

Kumar CMS, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria- novel functions and non-canonical behaviors. Cell Stress Chaperones 20:555–574

Kurouski D, Luo H, Sereda V, Robb FT, Lednev IK (2012) Rapid degradation kinetics of amyloid fibrils under mild conditions by an archaeal chaperonin. Biochem Biophys Res Commun 422:97–102

Kurouski D, Luo H, Sereda V, Robb FT, Lednev IK (2013) Deconstruction of stable cross-Beta fibrillar structures into toxic and nontoxic products using a mutated archaeal chaperonin. ACS Chem Biol 8(9):2095–2101

Laksanalamai P, Narayan S, Luo H, Robb FT (2008) Chaperone action of a versatile small heat shock protein from *Methanococcoides burtonii*, a cold adapted archaeon. Proteins Struct Funct Bioinf 75:275–282

Laskey RA, Honda BM, Mills AD, Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature 275:416–420

Leitner A, Joachimiak LA, Bracher A, Moenkemeyer L, Walzthoeni T, Chen B, Pechmann S, Holmes S, Cong Y, Ma B, Lim NCH, Jackson SE (2015) Mechanistic insights into the folding of knotted proteins *in vitro* and i*n vivo*. J Mol Biol 427:248–258

Levinthal C (1968) Are there pathways for protein folding? J Chim Phys 65:44–45

Libich DS, Fawzi NL, Ying J, Clore GM (2013) Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. Proc Natl Acad Sci 110:11361–11366

Lim NCH, Jackson SE (2015) Mechanistic insights into the folding of knotted proteins in vitro and in vivo. J Mol Biol 427:248–258

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

Macario AJL, Dugan CB, Macario EC (1991) A *dnaK* homolog in the archaeabacterium *Methanosarcina mazei* S6. Gene 108:133–137

Marco S, Urena D, Carrascosa JL, Waldmann T, Peters J, Hegerl R, Pfeifer G, Stack-Kongehl H, Baumeister W (1994) The molecular chaperone TF55: assessment of symmetry. FEBS Lett 341:152–155

Martel R, Cloney LP, Pelcher LE, Hemmingsen SM (1990) Unique composition of plastid chaperonin-60: alpha and beta polypeptide-encoding genes are highly divergent. Gene 94:181–187

McMullin TW, Hallberg RL (1987) A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophila*. Mol Cell Biol 7:4414–4423

Min W, Angileri F, Luo H, Lauria A, Shanmugasundaram M, Almerico AM, Cappello F, de Macario EC, Lednev IK, Macario AJ, Robb FT (2014) A human CCT5 gene mutation causing distal neuropathy impairs hexadecamer assembly in an archaeal model. Sci Rep 4:6688–6695

Motojima F, Yoshida M (2010) Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. EMBO J 29:4008–4019

Narberhaus F (1999) Negative regulation of bacterial heat shock genes. Mol Microbiol 31:1–8

Neidhardt FC, VanBogelen RA, Vaughn V (1984) The genetics and regulation of heat-shock proteins. Annu Rev Genet 18:295–329

Okochi M, Nomura T, Zako T, Arakawa T, Iizuka R, Ueda H, Funatsu T, Leroux M, Yohda M (2004) Kinetics and binding sites for interaction of the prefoldin with a group II chaperonin. J Biol Chem 279:31788–31795

Parsell DA, Sauer RT (1989) Induction of a heat shock-like response by unfolded protein in *Escherichia coli*: dependence on protein level not protein degradation. Genes Dev 3:1226–1232

Pereira JH, Ralston CY, Douglas NR, Meyer D, Knee KM, Goulet DR, King JA, Frydman J, Adams PD (2010) Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. J Biol Chem 285:27958–27966

Phipps BM, Hoffmann A, Stetter KO, Baumeister W (1991) A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. EMBO J 10:1711–1722

Phipps BM, Typke D, Hegerl R, Volker S, Hoffmann A, Stetter KO, Baumeister W (1993) Structure of a molecular chaperone from a thermophilic archaebacterium. Nature 361:475–477

Reissmann S, Parnot C, Booth CR, Chiu W, Frydman J (2007a) Essential fu nction of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. Nat Struct Mol Biol 14:432–440

Rommelaere H, Van Troys M, Gao Y, Melki R, Cowan NJ, Vandekerckhove J, Ampe C (1993) Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven related subunits. Proc Natl Acad Sci 90:11975–11979

Ruessmann F, Stemp MJ, Moenkemeyer L, Etchells SA, Bracher A, Hartl FU (2012) Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. Proc Natl Acad Sci 109:21208–21215

Saccoccia F, Di Micco P, Boumis G, Brunori M, Koutris I, Miele AE, Morea V, Sriratana P, Williams DL, Bellelli A, Angelucci F (2012) Moonlighting by different stressor: crystal structure of the chaperone species of a 2-Cys peroxiredoxin. Structure 20:429–439

Schulz A, Schumann W (1996) *HrcA*, the first gene of the Bacillus subtilis *dnaK* operon encodes a negative regulator of class I heat shock genes. J Bacteriol 178:1088–1093

Sela M, White FH, Anfinsen CB (1957) Reductive cleavage of disulfide bridges in ribonuclease. Science 125:691

Shimamura T, Koike-Takeshita A, Yokoyama K, Masui R, Murai N, Yoshida M, Taguchi H, Iwata S (2004) Crystal structure of the native chaperonin complex from Thermus thermophilus revealed unexpected asymmetry at the cis-cavity. Structure 12:1471–1480

Shtilerman M, Lorimer GH, Englander SW (1999) Chaperonin function: folding by forced unfolding. Science 284:822–825

Siegart R, Leroux MR, Scheufler C, Hartl FU, Moarefi I (2000) Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell 103:621–632

Sparrer H, Rutkat K, Buchner J (1997) Catalysis of protein folding by symmetric chaperone complexes. Proc Natl Acad Sci 94:1096–1100

Stan G, Brooks BR, Lorimer GH, Thirumalai D (2005) Identifying natural substrates for chaperonins using a sequence-based approach. Protein Sci 14:193–201

Stan G, Brooks BR, Lorimer GH, Thirumalai D (2006) Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. Proc Natl Acad Sci 103:4433–4438

Sternberg N (1973) Properties of a mutant of *Escherichia coli* defective in bacteriophage λ head formation (*groE*). J Mol Biol 76:1–23

Sternlicht H, Farr GW, Sternlicht ML, Driscoll JK, Willison K, Yaffe MB (1993) The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin *in vivo*. Proc Natl Acad Sci 90:9422–9426

Techtmann SM (2009) A genomic and biochemical charactertization of carbon monoxide utilizing thermophilic bacteria. Dissertation, University of Maryland, Baltimore

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

Teixeira F, Castro H, Cruz T, Tse E, Koldewey P, Southworth DR, Tomas AM, Jakob U (2015) Mitochondrial peroxiredoxin functions as a crucial chaperone reservoir in *Leishmani infantum*. Proc Natl Acad Sci 112(7):E616–E624

Tilly K, Murialdo H, Georgopoulos C (1981) Identification of a second *Escherichia coli* groE gene whose product is necessary for bacteriophage morphogenesis. Proc Natl Acad Sci 78:1629–1633

Trent JD, Nimmesgern E, Wall JS, Ulrich Hartl F, Horwich AL (1991) A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. Nature 354:490–493

Tsytlonok M, Itzhaki LS (2013) The how's and why's of protein folding intermediates. Arch Biochem Biophys 531:14–23

Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 93:863–873

Venetianer P, Straub FB (1963) The enzymatic reactivation of reduced ribonuclease. Biochim Biophys Acta 67:166

Waldinger D, Subramanian AR, Cleve H (1989) The polymorphic human chaperonin protein HuCha60 is a mitochondrial protein sensitive to heat shock and cell transformation. Eur J Cell Biol 50:435–441

Warnecke T (2012) Loss of the DnaK-DnaJ-GrpE chaperone system among the Aquificales. Mol Biol Evol 29:3485–3495

Williams WT, Codoner FM, Toft C, Fares MA (2010) Two chaperonin systems in bacterial genomes with distinct ecological roles. Trends Genet 26:47–51

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci 74:5088–5090

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc Natl Acad Sci 87:4576–4579

Wright PE, Dyson JH (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-ES-(ADP)7 chaperonin complex. Nature 388:741–750

Yang D, Ye X, Lorimer GH (2013) Symmetric GroEL: GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. Proc Natl Acad Sci 110(46):E4298–E4305

Yoshida T, Kawaguchi R, Taguchi H, Yoshida M, Yasunaga T, Wakabayshi T, Yohda M, Maruyama T (2002) Archaeal group II chaperonin mediates protein folding in the cis-cavity without a detachable GroES-like co-chaperonin. J Mol Biol 315:73–85

Zweig M, Cummings DJ (1973) Cleavage of head and tail protein during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. J Mol Biol 80:505–518

Chapter 2
Regulation of the Heat Shock Response in Bacteria

Wolfgang Schumann

Abstract Bacteria sense temperature changes in many ways and have developed different strategies to respond to these changes. A sudden increase in temperature results in protein unfolding, and the level of unfolded proteins seems to be the primary signal that triggers the heat shock response. Four different systems have been described so far involved in temperature sensing: alternative sigma factors, transcriptional repressors, and RNA and DNA thermosensors. Furthermore, titration of molecular chaperones serves as mediators in some cases. All four mechanisms will be described in detail and illustrated by prominent examples.

2.1 Introduction

In their natural environments, bacteria are constantly exposed to changing conditions including oxidative stress induced by reactive oxygen species and sudden increases and decreases in temperature called heat and cold shock, changes in their external pH (alkaline and acid shock), and changes in the concentration of nutrients and toxins to mention the most important ones (Storz and Hengge-Aronis 2000). Bacteria use different strategies to adapt to varying environmental conditions. Stress factors typically induce a stress response resulting in a characteristic change in the pattern of gene expression, where some genes are induced and others repressed. The stress response protects cells from severe damage and restores cellular homeostasis. Research carried out over the about last 50 years revealed that bacteria code for genetic programs allowing them to cope with stressful situations, and they consist of three major steps: (1) the stress factor is registered either directly or indirectly by a sensor; (2) the sensor leads to the induction and repression of the appropriate set of genes; and (3) in many cases, expression of the stress genes is turned off after adaptation through feedback inhibition.

W. Schumann
Institute of Genetics, University of Bayreuth, 95440 Bayreuth, Germany
e-mail: wschumann@uni-bayreuth.de

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_2

One of the best-studied examples is the heat shock response (HSR) in *Escherichia coli*, where more than 100 genes are under heat shock control (Guisbert et al. 2008). A heat shock causes protein unfolding and misfolding followed by formation of protein aggregates which can lead to cell death. To counteract these reactions, cells developed the HSR, which maintains protein homeostasis by ensuring appropriate expression levels of heat shock proteins (HSPs). The HSR results in the transient increased expression of genes called heat shock genes (HSGs) coding HSPs. Important HSPs are molecular chaperones and proteases involved in the folding of unfolded proteins and degradation of terminally misfolded proteins, respectively. As mentioned above, the expression of some genes is repressed after a heat shock in each organism, but the mechanisms responsible for repression have been rarely studied and will not be analyzed in this review article.

Many studies showed that roughly 50–200 genes are significantly induced after a heat shock in different model organisms, from archaea to human cell lines. Based on their functions, HSPs can be grouped into several classes. The two most prominent classes are the molecular chaperones and the proteases, both of which are needed to clear misfolded and irreversibly aggregated proteins from cells (Ellis et al. 1989; Meyer and Baker 2011). How do bacteria sense a heat shock? Multiple ways have been described, and very often, they sense the consequences of a temperature shift rather than the temperature increase itself.

This review article will concentrate on the question how bacteria regulate their HSGs after a sudden heat shock. So far, four different heat shock sensors have been described: (1) alternative sigma factors, (2) transcriptional repressors, (3) RNA thermosensors, and (4) DNA thermosensors. In addition, molecular chaperones play an important role as outlined below. So far, a multitude of review articles covering this field in bacteria have been published over the last 30 years, and I will cite only the most recent ones (Ho and Ellermeier 2012; Kortmann and Narberhaus 2012; Rajaram et al. 2014; Schumann 2012).

2.2 Regulation of the Heat Shock Response by Alternative Sigma Factors

All bacterial species code for only one multiunit RNA polymerase responsible for the transcription of genes coding either for proteins or for RNA. All genes are preceded by normally one and sometimes by two or even more promoters recognized by a sigma factor. Two different types of sigma factors have been described, housekeeping and alternative sigma factors. While all bacterial species code for one housekeeping sigma factor only, they code for a various number of alternative sigma factors depending on the species. So far, three different alternative sigma factors have been described as regulators of HSGs, namely, Sig32, SigE, and SigH. Sig32 and SigE have been discovered in *E. coli* first and later shown to be present in several bacterial species and SigH in *Mycobacterium tuberculosis* and other *Mycobacterium* species (Figs. 2.1, 2.2, 2.3, and 2.4).

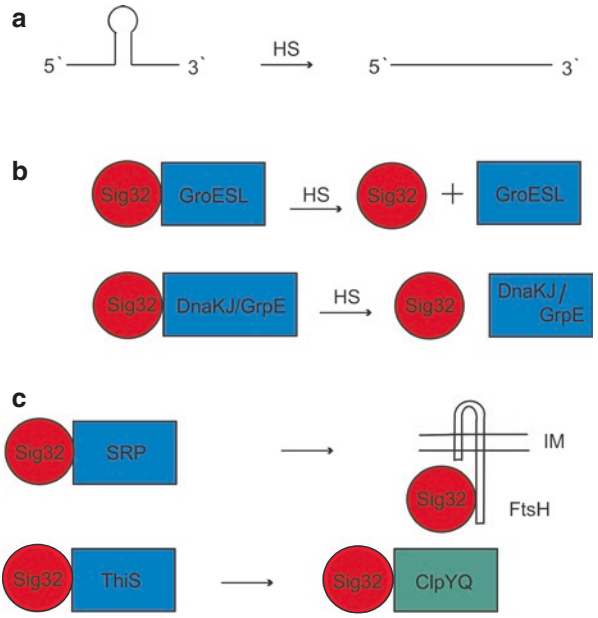


Fig. 2.1 Regulation of the E. coli sigma factor Sig32 encoded by the rpoH gene. (**a**) At low temperature, most rpoH mRNA molecules adopt a secondary structure preventing ribosome binding. After a heat shock, the secondary structure will open allowing binding of the ribosomes and translation. (**b**) The activity of Sig32 is regulated by the major chaperones GroESL and DnaKJ/GrpE. They bind to Sig32 and thereby block interaction with the core RNA polymerase. (**c**) The stability of Sig32 is influenced by the two proteases FtsH and ClpXP. Interaction with the inner membrane protease FtsH is mediated by the signal recognition particle (SRP), which binds to Sig32 and guides it to the FtsH protein. Degradation by the ClpXP protease needs first covalent bindingof the ubiquitin-like protein ThiS to Sig32 followed by interaction with ClpXP

2.2.1 The Alternative Sigma Factor 32

The Sig32 factor is encoded by the *rpoH* gene and the first regulator of the HSR to be described in *E. coli* (for a review, see Yura and Nakahigashi 1999). At 30 °C, Sig32 is extremely unstable and is normally present at very low levels (10–30 molecules per cell). After 5 min of temperature upshift, the amount of Sig32 first increases about 15-fold and thereafter gradually declines to a new steady-state level (Lesley et al. 2003). The amount of active Sig32 is regulated by three different mechanisms at the level of translation, activity, and stability. Regulation at the level of translation is exerted by two distinct regions termed A and B present within the coding region of *rpoH* (Nagai et al. 1991). Secondary structure formation between these two regions largely represses translation at low temperatures, whereas a temperature upshift leads to disruption of this structure. This has been proven experimentally by deletion analysis, point mutagenesis, and structural probing of the *rpoH* transcript (Morita et al. 1999a, b). A close inspection of *rpoH* sequences from γ-proteobacteria strongly suggests that translational control of *rpoH* is conserved among these

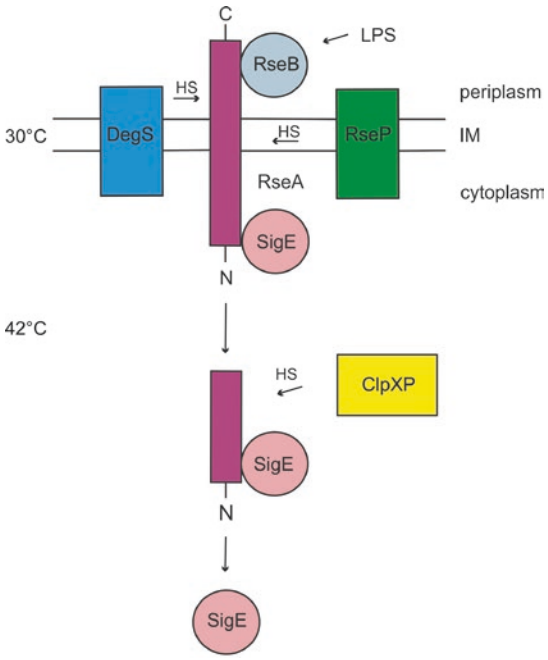


Fig. 2.2 Regulation of the E. coli sigma factor SigE. In the absence of heat stress, SigE is bound to the antisigma factor RseA, an inner membrane protein. Two types of stresses will lead to the degradation of RseA. After a heat shock, denatured outer membrane proteins activate the DegS protease which will cleave within the periplasmic domain of RseA followed by a second cleavage reaction within the periplasmic domain by the RseP protease. This second cleavage reaction results in the release of the shortened RseA into the cytoplasm, where it will be completely degraded by the ClpXP protease finally resulting in the release of SigE into the cytoplasm. The second type of stress is exerted by lipopolysaccharides (LPS) accumulating in the periplasm. This LPS will be recognized by the sensor protein RseB resulting in its dissociation from RseA followed by the stepwise cleavage of RseA as already described

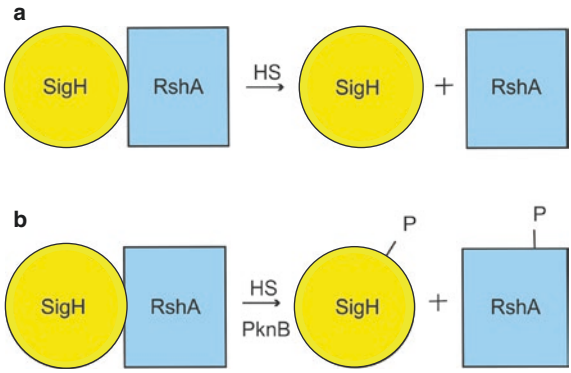


Fig. 2.3 The HrcA repressor protein of B. subtilis. The groESL and the dnaK operons are controlled at the transcriptional level by the HrcA repressor. (**a**) After synthesis, HrcA interacts with the GroEL chaperone assisting correct folding thereby allowing binding to its operators present in front of both operons. (**b**) After a heat shock, the GroEL chaperone is titrated by the denatured proteins having a higher affinity than HrcA. This will lead to the accumulation of inactive HrcA unable to bind to its operators

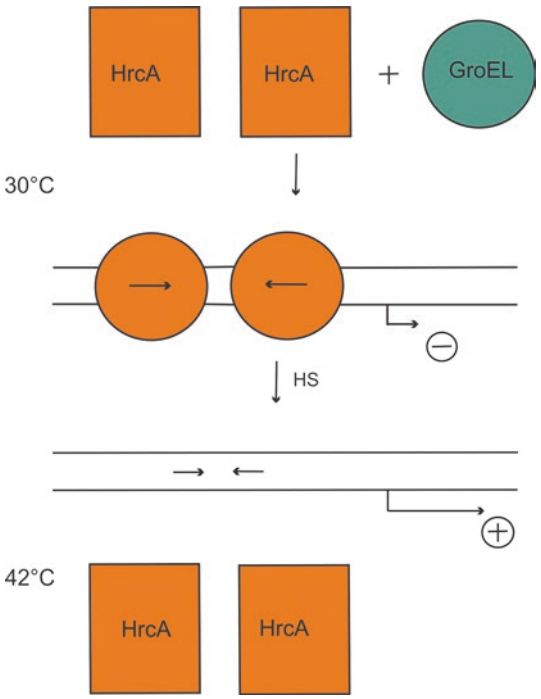


Fig. 2.4 Regulation of Mycobacterium tuberculosis sigma factor SigH. At low temperature, SigH is sequestered by the antisigma factor RshA. After a heat shock, SigH dissociates from RshA. Furthermore, both proteins can be phosphorylated by the PknB kinase resulting in decreased interaction of both proteins

bacteria including *Citrobacter freundii*, *Enterobacter cloacae*, *Serratia marcescens*, *Proteus mirabilis*, and *Pseudomonas aeruginosa* (Nakahigashi et al. 1995).

The activity of Sig32 is regulated by the two major chaperone systems DnaK/J/GrpE and GroEL/S. These chaperones bind to free Sig32 at low temperatures and are titrated by unfolded proteins accumulating after a heat shock. While depletion of either chaperone system or overexpression of chaperone substrates leads to an increase in Sig32 activity, overexpression of either chaperone system on the other hand leads to a decrease in Sig32 activity (Guisbert et al. 2004; Tomoyasu et al. 1998). Since the two chaperone systems bind Sig32 in vitro and inhibit its activity in a purified in vitro transcription system, inhibition is most likely also direct in vivo (Gamer et al. 1992, 1996; Guisbert et al. 2004). Recently, an in vivo physical association between GroEL and Sig32 in *E. coli* at physiological temperatures has been described (Patra et al. 2015). Since neither DnaK nor GroEL singly can modulate Sig32 stability in vivo, there is an ordered network between both chaperone systems, where GroEL acts upstream of DnaK. The stability of Sig32 is controlled by the inner membrane protease FtsH (Herman et al. 1995; Kanemori et al. 1997; Tomoyasu et al. 1995). Recently, it could be shown that Sig32 interacts with the signal recognition particle (SRP), which consists of the Ffh protein and the 4.5S RNA. Using in vivo cross-linking approaches, it could be demonstrated that the region 2.1 directly interacts with the SP-binding site in the M domain of Ffh (Miyazaki et al. 2016). Sig32 is transported to the cytoplasmic membrane and then delivered to the FtsH protease (Lim et al. 2013). In another study, it could be shown that the ubiquitin-like protein ThiS is covalently attached to Sig32 and subsequently targeted to degradation by the ATP-dependent protease ClpYQ (Xu et al. 2015).

2.2.2 The Alternative Sigma Factor E

In *E. coli*, SigE controls transcription of the regulon required for keeping proteins correctly folded present in the periplasm and in the outer membrane. In the absence of heat stress (and other conditions which lead to denaturation of these proteins), SigE is kept inactive by tight interaction with the C-terminal domain of the anti-sigma factor RseA (*r*egulation of *s*igma *E*). The RseA protein is inserted into the inner membrane by one transmembrane domain, where the C-terminal domain is exposed to the cytoplasm and the N-terminal domain to the periplasm. The co-antisigma factor RseB is bound to its N-terminal domain. Deletion of the *rseA* gene leads to full induction of the SigE regulon at physiological temperatures (Campbell et al. 2003; De Las Peñas et al. 1997; Missiakas et al. 1997).

What happens after a severe heat shock? The RseA antisigma factor is degraded sequentially by three different proteases, two of them (DegS and RseP) are inner membrane proteins, while the third one is present in the cytoplasm. Two different mechanisms have been described resulting in the degradation of RseA and thereby release of SigE. One involves proteins present in the periplasm and the outer membrane, while the second one the accumulation of LPS in the periplasm. While DegS acts as a sensor of unfolded and unassembled proteins in the periplasm, RseB recognizes LPS accumulating the periplasm.

In the absence of protein stress, the DegS protease is kept inactive. It consists of an N-terminal transmembrane domain, a central protease domain, and a C-terminal PDZ domain (Alba et al. 2001). PDZ domains are present in a large variety of proteins and are known to recognize specific C-terminal amino acid sequences (Doyle et al. 1996). The PDZ domain of DegS recognizes the motif YxF present at the C-terminal end of many outer membrane proteins (Mecsas et al. 1993; Walsh et al. 2003). In the absence of protein stress, the PDZ domain of DegS keeps its proteolytic activity inactive most probably by direct interaction. Upon exposure to the YxF motif present at the C-terminal end of partially unfolded proteins, the PDZ domain interacts with this motif and thereby liberates the proteolytic activity. This in turn will cleave within the C-terminal domain of RseA resulting in a conformational change within the transmembrane domain (Mecsas et al. 1993; Walsh et al. 2003). This will be recognized by the second protease, RseP, anchored in the cytoplasmic membrane with four transmembrane domains, and RseA cleaves within or near the transmembrane domain. This will result in the release of the remaining part of RseA with SigE still bound into the cytoplasm (Kanehara et al. 2002, 2003). Next, the remaining part of RseA will be completely degraded by one of different cytosolic proteases including ClpAP, ClpXP, HslUV, or Lon (Chaba et al. 2007). The last proteolytic step results in the release of SigE into the cytoplasm where it will bind to the RNA polymerase core enzyme leading to the transcriptional activation of 89 operons.

As already mentioned, the SigE regulon can also be activated by LPS accumulating in the periplasm. Here, RseB acts as a sensor protein of LPS signals. RseB binds to the RseA antisigma factor with a 1:1 stoichiometry (Cezairliyan and Sauer 2007; Kim et al. 2007, 2010). LPS or LPS fragments containing lipid-A can dissociate RseB from RseA, thereby inducing cleavage of RseA by the DegS protease (Lima et al. 2013).

2.2.3 The Alternative Sigma Factor H

Mycobacterium tuberculosis is a slow-growing facultative intracellular parasite and the causative agent of tuberculosis. Genome sequencing of *M. tuberculosis* strain H37Rv has identified 13 putative sigma factors (Gomez et al. 1997). Using a real-time relative RT-PCR assay resulted in the identification of the three heat shock-responsive sigma factors SigB, SigE, and SigH (Manganelli et al. 1999). The alternative sigma factor SigH is a central regulator of the response to heat and oxidative stresses (Fernandes et al. 1999; Raman et al. 2001; Song et al. 2003). SigH activity is regulated at the transcriptional level via autoregulation of the *sigH* promoter and, in addition, posttranslationally via interaction with the antisigma factor RshA (Song et al. 2003). The RshA-SigH complex is disrupted in in vitro experiments by elevated temperatures and under oxidizing conditions (Singh et al. 2014). In addition, both SigH and RshA are phosphorylated by PknB, a seine/threonine protein kinase. Phosphorylation of RshA, but not of SigH, affects the SigH/RshA interaction, resulting in decreased binding of SigH by RshA adding a further level of regulation (Park et al. 2008).

It could be shown that a *sigH* mutant is more susceptible to heat and oxidative stress (Raman et al. 2001). Expression from its promoter was strongly induced by a heat shock at 50 °C (Fernandes et al. 1999). The gene encoding SigH is induced by heat shock and oxidative stress (Fernandes et al. 1999; Manganelli et al. 1999). While the *sigH* gene is subjected to autoregulation, the SigH factor is posttranslationally regulated by the cytoplasmic antisigma factor RshA.

2.3 Regulation of the Heat Shock Response by Transcriptional Repressors

Several transcriptional repressor proteins have been described being involved in the regulation of HSGs. After a sudden heat shock, the repressor proteins have to dissociate from their operators to allow the RNA polymerase to bind to promoters and start with transcription of the HSGs. Different strategies have been developed to dissociate the repressor from its operator immediately after a heat shock which will be discussed.

2.3.1 The HrcA Repressor

The very first repressor to be described was the HrcA repressor of *Bacillus subtilis* which controls expression of the heptacistronic *dnaK* and the bicistronic *groESL* operon (Homuth et al. 1997; Schmidt et al. 1992) by binding to an operator we called CIRCE (for *c*ontrolling *i*nverted *r*epeat of *c*haperone *e*xpression; in the Greek mythology, CIRCE is the daughter of the god of the sun Helios) (Zuber and Schumann 1994). We assumed that the HrcA repressor is present in two different

conformations, an active and an inactive one. The equilibrium between these two conformations is modulated by GroEL. Our model is based on the following experimental data: (1) An increase in the amount of GroEL reduced the basal level of the nine proteins being part of the HrcA regulon, and a decrease resulted in their increase; (2) purified HrcA retarded more DNA in the presence of GroEL; and (3) GroEL specifically binds to immobilized HrcA (Mogk et al. 1997; Reischl et al. 2002). Based on these three observations, we suggested the following model: Free HrcA is present in its inactive conformation, and its interaction with GroEL converts it into its active conformation able to bind to the CIRCE element. After a heat shock, GroEL is titrated by the non-native proteins leaving HrcA inactive. The more non-native proteins have been either refolded or degraded, the more GroEL molecules will become available to convert inactive HrcA into its active conformation. Meanwhile, the *hrcA* gene has been discovered in more than 130 bacterial species making it the most widespread system of heat shock regulation.

2.3.2 The CtsR Repressor

The CtsR (*c*lass *t*hree *s*tress *r*epressor) regulon of *B. subtilis* consists of three operons: the tetracistronic *clpC* (*ctsR-mcsA-mcsB-mcsC*) and the two monocistronic *clpP* and *clpE* operons (Krüger et al. 1996, 1997). CtsR possesses a DNA-binding helix-turn-helix motif and binds as a dimeric protein to its operator called CtsR box (Derré et al. 1999). The CtsR repressor is inactivated after a sudden temperature upshift by a three-step process. First, it can sense temperature changes directly via a glycine-rich loop (RGGGGY) present near the DNA-binding domain (Derré et al. 2000). Second, the McsB (*m*odulator of *C*ts*R* activity) adaptor protein is released from ClpC and phosphorylates CtsR at several conserved arginine residues, thereby preventing rebinding to its operators (Elsholz et al. 2011; Fuhrmann et al. 2009). Third, autophosphorylated McsB delivers CtsR-P to the ClpCP protease for degradation (Kirstein et al. 2007).

2.3.3 The RheA Repressor

The RheA repressor of *Streptomyces albus* is involved in the regulation of two monocistronic operons only, one coding for the small 18 kDa HSP18 protein and the second for the repressor protein (Guglielmi et al. 1991). Both genes are adjacent but transcribed in opposite directions. In the absence of RheA, cells produce a large amount of *hsp18* mRNA at physiological temperatures, but only small amounts of HSP18 protein. Only after a heat shock, large amounts of HSP18 are present in the cells. These data suggest two independent mechanisms involved in the regulation of the *hsp18* gene.

The first mechanism is based on the temperature-sensing RheA autorepressor. At low temperatures, it is present in its active conformation negatively regulating transcription of its own gene and of the *hsp18* gene. After a heat shock, RheA unfolds into

an inactive conformation. Whether additional proteins are involved in changing and/or keeping RheA in its inactive conformation is unknown. The second mechanism involves the *hsp18* mRNA. As already mentioned, it cannot be translated at low temperature. At least three possibilities can be envisaged preventing binding of ribosomes and thereby translation: The mRNA adopts a three-dimensional configuration, or a regulatory RNA or a protein binds in the region of the ribosome-binding site.

2.3.4 The HspR Repressor

The *hspR* repressor gene has been discovered in *Streptomyces coelicolor* (Bucca et al. 1993). It is part of the tetracistronic *dnaK* operon and binds to an operator designated as HAIR (for *H*spR-*a*ssociated *i*nverted *r*epeat) which is present in front of the *dnaK* operon and some other heat shock genes. The activity of the HspR repressor is modulated by the DnaK chaperone (Bucca et al. 2000) based on four observations: (1) In a band-shift assay, HspR is able to bind to the HAIR element only in the presence of DnaK; (2) addition of anti-DnaK antibodies to the HspR-DnaK-DNA complex resulted in a supershift, thereby proving that DnaK really interacts with HspR; (3) induction of the DnaK operon in the presence of overproduced HspR repressor is partially decreased; and (4) HspR copurified with DnaK during column chromatography. In summary, the DnaK chaperone acts as a corepressor by assisting HspR in binding to the HAIR operator.

2.4 Regulation of the Heat Shock Response by RNA Thermosensors

Bacterial RNA thermosensors are temperature-sensing RNA sequences present in the 5′-UTR or the intercistronic regions of some mRNA molecules (Kortmann and Narberhaus 2012). They can form secondary structures that partially or entirely include the 5′ untranslated region of the mRNA and thereby the Shine-Dalgarno (SD) sequence as part of the ribosome-binding site (RBS) (Kortmann and Narberhaus 2012). Therefore, typical RNA thermosensors control initiation of translation by forming a secondary structure trapping the RBS. A heat shock destabilizes this structure, liberates the RBS, and permits formation of the translation initiation complex.

RNA thermosensors are zipper-like structures present in the 5′-UTR of some mRNA molecules masking the RBS. Increases in temperature result in melting of the secondary structure, thereby exposing the RBS to the ribosomes followed by translation of the mRNA molecules. These RNA structures are sensitive to temperature changes in a way to detect variations on the scale of 1 °C (Rinnenthal et al. 2010).

RNA-mediated regulation is independent of protein factors and less cost-intensive for the cell. It is faster than protein-based regulation because of the mRNA which directly controls gene expression without the need of translation of

a regulatory factor. With increasing temperatures, the RNA thermosensor structure unfolds in a zipper-like manner and translation can be initiated.

The first RNA thermosensor to be described regulates synthesis of the *E. coli* Sig32 (Morita et al. 1999b). In the absence of heat shock, the *rpoH* mRNA is folded into a secondary structure that occludes the SD sequence and the downstream initiation codon. Two segments called A and B form an extensive secondary structure blocking binding of the ribosomes to the SD sequence (Morita et al. 1999b). A sudden temperature increase disrupts the secondary structure and allows binding of the ribosomes followed by synthesis of Sig32.

Another RNA thermosensor designated as ROSE (for *r*epression *o*f heat *s*hock gene *e*xpression) was discovered first in *Bradyrhizobium japonicum* (Narberhaus et al. 1998) and was later found in different *Rhizobium* species and in *Agrobacterium tumefaciens* (Balsiger et al. 2004; Nocker et al. 2001). These ROSE elements are located in the 5′-UTR transcripts coding for small heat shock genes. They are 70–120 nucleotides long and acquire a complex structure including 2–4 stem-loops, where the 3′-proximal hairpin contains the SD sequence and sometimes in addition the start codon.

Another type of RNA thermosensor has been called the fourU element. Members of this element are characterized by the presence of a stretch of four uridines, which pair with AGGA in the SD sequence. This structure is stable at low temperatures and prevents binding of the ribosomes to the SD sequence. At high temperatures, the stem-loop structure unfolds; ribosomes bind and start translation.

The best-studied fourU elements play an important role in the thermoregulation of translation of small HSGs. The first and so far best characterized fourU element was described located upstream of the small HSG *agsA* in *Salmonella* (Waldminghaus et al. 2007). The predicted structure of this element consists of two hairpins, where the second hairpin contains the four uridine residues able to base-pair with the SD sequence. This hairpin opens in a temperature-controlled way and allows binding of the ribosomes to the SD sequence only after a heat shock.

Additional fourU thermosensors control translation of the virulence factor LcrF in *Yersinia* ss. (Bohme et al. 2012), iron-acquisition genes in *Shigella dysenteriae* (Kouse et al. 2013), and the transcriptional activator protein ToxT of *Vibrio cholerae* (Weber et al. 2014), to mention a few more examples.

2.5 Regulation of the Heat Shock Response by DNA Thermosensors

Temperature sensing through DNA involves local changes in DNA topology followed by transcriptional and translational events (Falconi et al. 1998; Prosseda et al. 2004). Three different principles have been described involving DNA molecules as thermosensors: DNA supercoiling, promoter curvature, and DNA-associated proteins.

2.5.1 DNA Supercoiling

It has been shown that plasmids from mesophilic and hyperthermophilic bacteria can undergo changes in their supercoiling level depending on the temperature (Lopez-Garcia and Forterre 2000). A heat shock leads to a transient increase in positive supercoils. Recovery to the normal supercoiling level occurs within 10 min after the heat shock and is catalyzed by DNA gyrase, the nucleoid-binding protein HU, and the DnaK chaperone (Lopez-Garcia and Forterre 2000). The level of DNA supercoiling can influence the transcription efficiency and therefore acts as an important parameter in temperature-dependent gene regulation (Pruss and Drlica 1989).

2.5.2 Promoter Curvature

Another important DNA thermosensor is intrinsic bends. These curved DNA regions are characterized by AT-tracts (Mizuno 1987) located upstream of a promoter and influencing binding of the RNA polymerase (Nickerson and Achberger 1995). Temperature-induced changes affect the topology of the curved DNA and thereby transcription.

The best-studied organism is *Shigella flexneri*, a facultative intracellular pathogen where most of the virulence genes are located on the 230 kb plasmid pINV (Maurelli et al. 1984; Sasakawa et al. 1988). These cells are able to penetrate into and replicate within human colonic epithelial cells. Expression of the invasive phenotype is regulated by the growth temperature (Maurelli et al. 1984). It has been shown that bacteria growing at 37 °C express the virulent genes and are invasive, whereas the same cells are noninvasive when grown at 30 °C. Using transposon mutagenesis, a gene has been identified responsible for the growth-dependent phenotype. This gene codes for a protein that has been designated as H-NS (for *h*eat-stable *n*ucleoid-*s*tructuring). The H-NS protein silences expression of the *virF* gene coding for a transcriptional activator of several virulence genes. The H-NS protein binds at two sites separated by a region of DNA curvature. Binding of H-NS to these regions occurs cooperatively at temperatures below 32 °C but not at 37 °C. It follows that bent DNA acts as a sensor of temperature (Falconi et al. 1998).

2.5.3 Nucleoid-Associated Proteins

Nucleoid-associated proteins not only influence its conformation but also DNA replication, recombination, and transcription (Dame 2005; Dorman 2004). H-NS is the best characterized nucleoid-associated protein present in different enteric bacteria and prefers AT-rich DNA sequences and is itself subject to temperature control. While the DNA-binding capacity is reduced at 37 °C, the H-NS to DNA

ratio increases several fold during growth at low temperatures (Atlung and Ingmer 1997; Ono et al. 2005). The temperature-dependent accessibility of promoters kept inactive by binding of H-NS plays a key role in the regulation of virulence genes in many human pathogens. One example is the Pap pili which are encoded by the *papBA* (for *p*yelonephritis-*a*ssociated *p*ili) operon and expressed by uropathogenic *E. coli* cells. They allow these cells to attach to uroepithelial cells and subsequent colonization of the host upper urinary tract. Optimal expression of the *papBA* operon occurs at 37 °C, and a 52-fold reduction in *papBA* transcription was measured at 32 °C (Blyn et al. 1989; Bolotin et al. 2001). Regulation of transcription of the *papBA* operon occurs by two different proteins, H-NS and RimJ. While H-NS binds to the *pap* regulatory region at 23 °C but not at 37 °C, RimJ is an N-terminal acetyltransferase of the ribosomal protein S5 (Cumberlidge and Isono 1979). Deletion of the *rimJ* gene leads to the loss of thermoregulation resulting in expression of the *papBA* operon at 32 and 37 °C (White-Ziegler et al. 1990). The mechanism by which RimJ represses *papBA* transcription is still unknown.

Another example where H-NS silences expression of virulence factors at low temperature is a type III secretion system (T3SS) present in *Salmonella enterica* and encoded by the pathogenicity island2 (SPI-2). This genomic island codes for genes involved in activating and assembling the T3SS. These gene products activate and assemble the T3SS required during intracellular infection and injection of effector proteins into host cells required for intracellular survival (Ochman et al. 1996; Shea et al. 1996). When cells are grown at 30 °C, they are unable to express the T3SS. It has been shown that expression of the virulence genes is controlled by Hha and H-NS, and both proteins are needed to silence the virulence genes at temperatures below 30 °C (Duong et al. 2007). While H-NS silences expression of the response regulator SsrR, which activates several genes responsible for host infection, Hha silences the SPI-2 gene transcription.

The Hha protein has been identified in *E. coli* as a modulator of the expression of the toxin α-hemolysin. It could be shown that synthesis of this toxin is repressed both under conditions of high osmolarity and at low temperatures (Mourino et al. 1994). Hemolysin expression is derepressed in *hha* mutants during growth at low temperatures or in the presence of high osmolarity (Daguer et al. 2005).

References

Alba BM, Zhong HJ, Pelayo JC, Gross CA (2001) *degS* (*hhoB*) is an essential *Escherichia coli* gene whose indispensable function is to provide σ^E activity. Mol Microbiol 40:1323–1333

Atlung T, Ingmer H (1997) H-NS: a modulator of environmentally regulated gene expression. Mol Microbiol 24:7–17

Balsiger S, Ragaz C, Baron C, Narberhaus F (2004) Replicon-specific regulation of small heat shock genes in *Agrobacterium tumefaciens*. J Bacteriol 186:6824–6829

Blyn LB, Braaten BA, White-Ziegler CA, Rolfson DH, Low DA (1989) Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. EMBO J 8:613–620

Bohme K, Steinmann R, Kortmann J, Seekircher S, Heroven AK, Berger E, Pisano F, Thiermann T, Wolf-Watz H, Narberhaus F, Dersch P (2012) Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. PLoS Pathog 8:e1002518

Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A (2001) The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res 11:731–753

Bucca G, Smith CP, Alberti M, Seidita G, Passantino R, Puglia AM (1993) Cloning and sequencing of the *dnaK* region of *Streptomyces coelicolor* A3(2). Gene 130:141–144

Bucca G, Brassington AME, Schönfeld H-J, Smith CP (2000) The HspR regulon of *Streptomyces coelicolor*: a role for the DnaK chaperone as a transcriptional co-repressor. Mol Microbiol 38:1093–1103*Streptomyces coelicolor*

Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, Darst SA (2003) Crystal structure of *Escherichia coli* sigma (E) with the cytoplasmic domain of its anti-sigma factor. Mol Cell 11:1067–1078

Cezairliyan BO, Sauer RT (2007) Inhibition of regulated proteolysis by RseB. Proc Natl Acad Sci U S A 104:3771–3776

Chaba R, Grigorova IL, Flynn JM, Baker TA, Gross CA (2007) Design principles of the proteolytic cascade governing the σ^E-mediated envelope stress response in *Escherichia coli*: keys to graded, buffered, and rapid signal transduction. Genes Dev 21:124–136

Cumberlidge AG, Isono K (1979) Ribosomal protein modification in *Escherichia coli*. I. A mutant lacking the N-terminal acetylation of protein S5 exhibits thermosensitivity. J Mol Biol 131:169–189

Daguer JP, Chambert R, Petit-Glatron MF (2005) Increasing the stability of *sacB* transcript improves levansucrase production in *Bacillus subtilis*. Lett Appl Microbiol 41:221–226

Dame RT (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol Microbiol 56:858–870

De Las PA, Connolly L, Gross CA (1997) The σ^E-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of σ^E. Mol Microbiol 24:373–385

Derré I, Rapoport G, Msadek T (1999) CtsR, a novel regulator of stress and heat shock response, controls *clp* and molecular chaperone gene expression in gram-positive bacteria. Mol Microbiol 31:117–131

Derré I, Rapoport G, Msadek T (2000) The CtsR regulator of stress response is active as a dimer and specifically degraded *in vivo* at 37°C. Mol Microbiol 38:335–347

Dorman CJ (2004) H-NS: a universal regulator for a dynamic genome. Nat Rev Microbiol 2:391–400

Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85:1067–1076

Duong N, Osborne S, Bustamante VH, Tomljenovic AM, Puente JL, Coombes BK (2007) Thermosensing coordinates a cis-regulatory module for transcriptional activation of the intracellular virulence system in *Salmonella enterica* serovar Typhimurium. J Biol Chem 282:34077–34084

Ellis RJ, Van der Vies SM, Hemmingsen SM (1989) The molecular chaperone concept. Biochem Soc Symp 55:145–153

Elsholz AK, Hempel K, Pother DC, Becher D, Hecker M, Gerth U (2011) CtsR inactivation during thiol-specific stress in low GC, Gram+ bacteria. Mol Microbiol 79:772–785

Falconi M, Colonna B, Prosseda G, Micheli G, Gualerzi CO (1998) Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. EMBO J 17:7033–7043

Fernandes ND, Wu QL, Kong D, Puyang X, Garg S, Husson RN (1999) A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. J Bacteriol 181:4266–4274

Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T (2009) McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. Science 324:1323–1327

Gamer J, Bujard H, Bukau B (1992) Physical interactions between heat shock proteins DnaK, DnaJ, and GrpE and the bacterial heat shock transcription factor σ^{32}. Cell 69:833–842

Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rüdiger S, Schönfeld HJ, Schirra C, Bujard H, Bukau B (1996) A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor σ^{32}. EMBO J 15:607–617

Gomez JE, Chen JM, Bishai WR (1997) Sigma factors of *Mycobacterium tuberculosis*. Tuber Lung Dis 78:175–183

Guglielmi G, Mazodier P, Thompson CJ, Davies J (1991) A survey of the heat shock response in four *Streptomyces* species reveals two *groEL*-like genes and three GroEL-like proteins in *Streptomyces albus*. J Bacteriol 173:7374–7381

Guisbert E, Herman C, Lu CZ, Gross CA (2004) A chaperone network controls the heat shock response in *E-coli*. Genes Dev 18:2812–2821

Guisbert E, Yura T, Rhodius VA, Gross CA (2008) Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. Microbiol Mol Biol Rev 72:545–554

Herman C, Thévenet D, D'Ari R, Bouloc P (1995) Degradation of σ^{32}, the heat shock regulator in *Escherichia coli*, is governed by HflB. Proc Natl Acad Sci U S A 92:3516–3520

Ho TD, Ellermeier CD (2012) Extra cytoplasmic function σ factor activation. Curr Opin Microbiol 15:182–188

Homuth G, Masuda S, Mogk A, Kobayashi Y, Schumann W (1997) The *dnaK* operon of *Bacillus subtilis* is heptacistronic. J Bacteriol 179:1153–1164

Kanehara K, Ito K, Akiyama Y (2002) YaeL (EcfE) activates the σ^{E} pathway of stress response through a site-2 cleavage of anti-σ^{E}, RseA. Genes Dev 16:2147–2155

Kanehara K, Ito K, Akiyama K (2003) YaeL proteoylsis of RseA is controlled by the PDZ domain of YaeL and a Gln-rich region of RseA. EMBO J 22:6389–6398

Kanemori M, Nishihara K, Yanagi H, Yura T (1997) Synergistic roles of Hs1VU and other ATP-dependent proteases in controlling in vivo turnover of σ^{32} and abnormal proteins in *Escherichia coli*. J Bacteriol 179:7219–7225

Kim DY, Jin KS, Kwon E, Ree M, Kim KK (2007) Crystal structure of RseB and a model of its binding mode to RseA. Proc Natl Acad Sci U S A 104:8779–8784

Kim DY, Kwon E, Choi J, Hwang HY, Kim KK (2010) Structural basis for the negative regulation of bacterial stress response by RseB. Protein Sci 19:1258–1263

Kirstein J, Dougan DA, Gerth U, Hecker M, Turgay K (2007) The tyrosine kinase McsB is a regulated adaptor protein for ClpCP. EMBO J 26:2061–2070

Kortmann J, Narberhaus F (2012) Bacterial RNA thermometers: molecular zippers and switches. Nat Rev Microbiol 10:255–265

Kouse AB, Righetti F, Kortmann J, Narberhaus F, Murphy ER (2013) RNA-mediated thermoregulation of iron-acquisition genes in Shigella dysenteriae and pathogenic *Escherichia coli*. PLoS One 8:e63781

Krüger E, Msadek T, Hecker M (1996) Alternate promoters direct stress-induced transcription of the *Bacillus subtilis clpC* operon. Mol Microbiol 20:713–724

Krüger E, Msadek T, Ohlmeier S, Hecker M (1997) The *Bacillus subtilis clpC* operon encodes DNA repair and competence proteins. Microbiology 143:1309–1316

Lesley SA, Thompson NE, Burgess RR (2003) Studies of the role of the *Escherichia coli* heat shock regulatory protein sigma factor 32 by the use of monoclonal antibodies. J Biol Chem 262:5404–5407

Lim B, Miyazaki R, Neher S, Siegele DA, Ito K, Walter P, Akiyama Y, Yura T, Gross CA (2013) Heat shock transcription factor sigma(32) co-opts the signal recognition particle to regulate protein homeostasis in *E. coli*. PLoS Biol 11:e1001735

Lima S, Guo MS, Chaba R, Gross CA, Sauer RT (2013) Dual molecular signals mediate the bacterial response to outer-membrane stress. Science 340:837–841

Lopez-Garcia P, Forterre P (2000) DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. BioEssays 22:738–746

Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. Mol Microbiol 31:715–724

Maurelli AT, Blackmon B, Curtiss R III (1984) Temperature-dependent expression of virulence genes in *Shigella* species. Infect Immun 43:195–201

Mecsas J, Rouviere PE, Erickson JW, Donohue TJ, Gross CA (1993) The activity of σ^E, an *Escherichia coli* heat-inducible σ-factor, is modulated by expression of outer membrane proteins. Genes Dev 7:2618–2628

Meyer AS, Baker TA (2011) Proteolysis in the *Escherichia coli* heat shock response: a player at many levels. Curr Opin Microbiol 14:194–199

Missiakas D, Mayer MP, Lemaire M, Georgopoulos C, Raina S (1997) Modulation of the *Escherichia coli* σ^E (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. Mol Microbiol 24:355–371

Miyazaki R, Yura T, Suzuki T, Dohmae N, Mori H, Akiyama Y (2016) A novel SRP recognition sequence in the homeostatic control region of heat shock transcription factor σ^{32}. Sci Rep 6:24147

Mizuno T (1987) Random cloning of bent DNA segments from *Escherichia coli* chromosome and primary characterization of their structures. Nucleic Acids Res 15:6827–6841

Mogk A, Homuth G, Scholz C, Kim L, Schmid FX, Schumann W (1997) The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. EMBO J 16:4579–4590

Morita M, Kanemori M, Yanagi H, Yura T (1999a) Heat-induced synthesis of σ^{32} in *Escherichia coli*: structural and functional dissection of *rpoH* mRNA secondary structure. J Bacteriol 181:401–410

Morita MT, Tanaka Y, Kodama TS, Kyogoku Y, Yanagi H, Yura T (1999b) Translational induction of heat shock transcription factor σ^{32}: evidence for a built-in RNA thermosensor. Genes Dev 13:655–665

Mourino M, Munoa F, Balsalobre C, Diaz P, Madrid C, Juarez A (1994) Environmental regulation of alpha-haemolysin expression in *Escherichia coli*. Microb Pathog 16:249–259

Nagai H, Yuzawa H, Yura T (1991) Interplay of two *cis*-acting mRNA regions in translational control of σ^{32} synthesis during the heat shock response of *Escherichia coli*. Proc Natl Acad Sci U S A 88:10515–10519

Nakahigashi K, Yanagi H, Yura T (1995) Isolation and sequence analysis of *rpoH* genes encoding σ^{32} homologs from gram negative bacteria: conserved mRNA and protein segments for heat shock regulation. Nucleic Acids Res 23:4383–4390

Narberhaus F, Käser R, Nocker A, Hennecke H (1998) A novel DNA element that controls bacterial heat shock gene expression. Mol Microbiol 28:315–323

Nickerson CA, Achberger EC (1995) Role of curved DNA in binding of *Escherichia coli* RNA polymerase to promoters. J Bacteriol 177:5756–5761

Nocker A, Hausherr T, Balsiger S, Krstulovic NP, Hennecke H, Narberhaus F (2001) A mRNA-based thermosensor controls expression of rhizobial heat shock genes. Nucleic Acids Res 29:4800–4807

Ochman H, Soncini FC, Solomon F, Groisman EA (1996) Identification of a pathogenicity island required for *Salmonella* survival in host cells. Proc Natl Acad Sci U S A 93:7800–7804

Ono S, Goldberg MD, Olsson T, Esposito D, Hinton JC, Ladbury JE (2005) H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. Biochem J 391:203–213

Park ST, Kang CM, Husson RN (2008) Regulation of the SigH stress response regulon by an essential protein kinase in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 105:13105–13110

Patra M, Roy SS, Dasgupta R, Basu T (2015) GroEL to DnaK chaperone network behind the stability modulation of sigma at physiological temperature in *Escherichia coli*. FEBS Lett 589(24 Pt B):4047–4052

Prosseda G, Falconi M, Giangrossi M, Gualerzi CO, Micheli G, Colonna B (2004) The *virF* promoter in *Shigella*: more than just a curved DNA stretch. Mol Microbiol 51:523–537

Pruss GJ, Drlica K (1989) DNA supercoiling and prokaryotic transcription. Cell 56:521–523
Rajaram H, Chaurasia AK, Apte SK (2014) Cyanobacterial heat-shock response: role and regulation of molecular chaperones. Microbiology 160:647–658
Raman S, Song T, Puyang X, Bardarov S, Jacobs WR Jr, Husson RN (2001) The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. J Bacteriol 183:6119–6125
Reischl S, Wiegert T, Schumann W (2002) Isolation and analysis of mutant alleles of the *Bacillus subtilis* HrcA repressor with reduced dependency on GroE function. J Biol Chem 277:32659–32667
Rinnenthal J, Klinkert B, Narberhaus F, Schwalbe H (2010) Direct observation of the temperature-induced melting process of the *Salmonella* fourU RNA thermometer at base-pair resolution. Nucleic Acids Res 38:3834–3847
Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M (1988) Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. J Bacteriol 170:2480–2484
Schmidt A, Schiesswohl M, Völker U, Hecker M, Schumann W (1992) Cloning, sequencing, mapping, and transcriptional analysis of the *groESL* operon from *Bacillus subtilis*. J Bacteriol 174:3993–3999
Schumann W (2012) Thermosensor systems in eubacteria. Adv Exp Med Biol 739:1–16
Shea JE, Hensel M, Gleeson C, Holden DW (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. Proc Natl Acad Sci U S A 93:2593–2597
Singh R, Anil KV, Das AK, Bansal R, Sarkar D (2014) A transcriptional co-repressor regulatory circuit controlling the heat-shock response of *Mycobacterium tuberculosis*. Mol Microbiol 94:450–465
Song T, Dove SL, Lee KH, Husson RN (2003) RshA, an anti-sigma factor that regulates the activity of the mycobacterial stress response sigma factor SigH. Mol Microbiol 50:949–959
Storz G, Hengge-Aronis R (2000) Bacterial stress responses. American Society for Microbiology, Washington, DC
Tomoyasu T, Gamer J, Bukau B, Kanemori M, Mori H, Rutman AJ, Oppenheim AB, Yura T, Yamanaka K, Niki H, Hiraga S, Ogura T (1995) *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor σ^{32}. EMBO J 14:2551–2560
Tomoyasu T, Ogura T, Tatsuta T, Bukau B (1998) Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. Mol Microbiol 30:567–582
Waldminghaus T, Heidrich N, Brantl S, Narberhaus F (2007) FourU: a novel type of RNA thermometer in *Salmonella*. Mol Microbiol 65:413–424
Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT (2003) OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. Cell 113:61–71
Weber GG, Kortmann J, Narberhaus F, Klose KE (2014) RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae*. Proc Natl Acad Sci U S A 111:14241–14246
White-Ziegler CA, Blyn LB, Braaten BA, Low DA (1990) Identification of an *Escherichia coli* genetic locus involved in thermoregulation of the *pap* operon. J Bacteriol 172:1775–1782
Xu X, Niu Y, Liang K, Wang J, Li X, Yang Y (2015) Heat shock transcription factor σ^{32} is targeted for degradation via an ubiquitin-like protein ThiS in *Escherichia coli*. Biochem Biophys Res Commun 459:240–245
Yura T, Nakahigashi K (1999) Regulation of the heat-shock response. Curr Opin Microbiol 2:153–158
Zuber U, Schumann W (1994) CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *Bacillus subtilis*. J Bacteriol 176:1359–1363

Part II
Multiple Chaperonins of Bacterial System

Chapter 3
Prokaryotic Multiple Chaperonins: The Mediators of Functional and Evolutionary Diversity

C. M. Santosh Kumar

Abstract Chaperonins are a class of molecular chaperones that form large multimeric assemblies for encapsulation of substrate proteins. Surprisingly, 30% of newly sequenced bacterial genomes encode multiple copies of the chaperonins. The distribution of these multiple copies appears to follow a phylum-specific pattern. Functional and structural studies on several of these chaperonins have delineated how these *extra* chaperonins evolved functional diversity and contributed towards the biological adaptation of the hosting organisms. Since several of these bacteria are either pathogenic or economically important, and the chaperonins regulate the pathogenic processes in these organisms, it is important to understand their biology. This chapter is aimed to act as a primer for the subsequent chapters that describe different examples of multiple chaperonins and the plethora of their functional diversity.

3.1 Introduction

Advancements in genomic technologies have yielded wealth of information from completely sequenced genomes. The startling revelation of the presence of several eukaryotic-like features in bacteria, such as the protein kinases (Kumar et al. 2009; Perez et al. 2008), different classes of intronic regions (Ferat and Michel 1993; Hausner et al. 2014; Martinez-Abarca and Toro 2000) and protein-protein interaction mediating ankyrins (Price et al. 2010), has provided interesting insights into understanding the biology of these organisms. Likewise, the presence of multiple copies of genes encoding chaperonins in 30% of the bacterial genomes (Barreiro et al. 2005; Fischer et al. 1993; Karunakaran et al. 2003; Kong et al. 1993), another well-known eukaryotic feature, encoding 2–3 copies of chaperonin genes (Nishio

C. M. Santosh Kumar
School of Biosciences and Institute of Microbiology and Infection,
University of Birmingham, Birmingham B15 2TT, UK
e-mail: S.Kumar.7@bham.ac.uk

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_3

et al. 1999; Vitlin Gruber et al. 2013), has gained a lot of interest in recent times. Interestingly, many of the bacteria that possess multiple copies of chaperonin genes are either pathogenic to human, livestock and crops or economically important. In addition, these excess chaperonin copies have been demonstrated to be involved in the pathogenic or economically important biological functions in those bacteria. These observations, therefore, have propelled intense investigations to unravel the functional diversity of these chaperonins, thereby aiming to provide tools for either curbing the pathogens or tuning beneficial bacteria towards human well-being.

3.2 Distribution of Multiple Chaperonins

Comprehensive phylogenetic analyses on the multiple chaperonins have revealed that their distribution follows a phylum-specific pattern (Kumar et al. 2015; Lund 2009). While many bacterial phyla possess a single copy of the chaperonin gene, the presence of multiple copies of chaperonin genes predominates in five phyla: (a) phylum *Actinobacteria* that constitutes high-G + C Gram-positive species, (b) phylum *Firmicute*s that constitutes low-G + C Gram-positive species, (c) phylum *Cyanobacteria* that constitutes photosynthetic bacteria, (d) phylum *Chlamydia* that constitutes obligate intracellular pathogens and (e) alpha subdivision of phylum *Proteobacteria* that constitutes root-nodulating symbionts (Table 3.1). I will briefly review below the current understanding of the salient features of the multiple chaperonins, such as gene organisation, regulation, essentiality, sequence and functional diversity and the possible modes of evolution in the following sections. For detailed description, the readers are advised to read a comprehensive review by Peter Lund (Lund 2009).

3.2.1 Functional Diversity Among the Chaperonins of Actinobacteria

Actinobacteria constitutes a phylum of Gram-positive bacteria that are characterised by high-G + C content genomes, such as *Mycobacterium tuberculosis*, *M. leprae*, *Streptomyces albus* and *Bifidobacterium longum*. The presence of multiple chaperonins was first reported in *Actinobacteria*, in the genome of *M. tuberculosis* (Kong et al. 1993). About 70% of the sequenced actinobacterial genomes possess two copies of GroEL genes, with instances of three or four copies occurring at a lower frequency (Table 3.1). While the first copy is in operonic arrangement with the co-chaperonin gene, the second and subsequent copies exist singly (Kong et al. 1993; Rinke de Wit et al. 1992). Interestingly, the major difference between these copies lies at their carboxy-terminal segments (CTS). While the chaperonin encoded by the copy in operonic arrangement bears a non-canonical histidine-rich carboxy terminus (Ferreyra et al. 1993; Kumar and Mande 2011; Mande et al. 2013), the other copy bears characteristic glycine-methionine-rich carboxy terminus, probably

Table 3.1 Phylum-wide distribution of multiple chaperonin genes among completely sequenced bacteria

	Number of chaperonin homologues						
Phyla	1	2	3	4	5	6	7
Actinobacteria	34	119	10	6	0	0	0
Aquificae	11	0	0	0	0	0	0
Bacteroidetes/Chlorobi group	92	2	1	0	0	0	0
Caldiserica	1	0	0	0	0	0	0
Chlamydiae/Verrucomicrobia group	3	3	10	0	0	0	0
Chloroflexi	5	7	0	0	0	0	0
Chrysiogenetes	1	0	0	0	0	0	0
Cyanobacteria	1	49	3	0	0	0	0
Deferribacteres	4	0	0	0	0	0	0
Deinococcus-Thermus	16	0	0	0	0	0	0
Dictyoglomi	2	0	0	0	0	0	0
Elusimicrobia	2	0	0	0	0	0	0
Fibrobacteres/Acidobacteria group	8	0	0	0	0	0	0
Firmicutes	243	7	2	0	1	0	0
Fusobacteria	5	0	0	0	0	0	0
Gemmatimonadetes	1	1	0	0	0	0	0
Nitrospirae	4	0	0	0	0	0	0
Planctomycetes	0	0	6	0	0	0	0
Proteobacteria	470	79	25	8	6	2	1
Spirochaetes	35	0	0	0	0	0	0
Synergistetes	4	0	0	0	0	0	0
Tenericutes	9	0	0	0	0	0	0
Thermodesulfobacteria	2	0	0	0	0	0	0
Thermotogae	14	1	0	0	0	0	0

The table is adopted from Kumar et al. (2015), with permission

conferring essentiality to this copy (Mazodier et al. 1991; Rinke de Wit et al. 1992). Interestingly, in the organisms with more than two chaperonin genes, the third and subsequent copies possess a pattern-free CTS (Fig. 3.1). Since all these bacteria are fast-growing, these chaperonin copies are implicated in enhancing the growth rate of those organisms (Kumar et al. 2015). Surprisingly, in the organisms where only one copy of chaperonin is present, such as *B. breve*, *B. longum* and *B. animalis lactis*, the chaperonin and co-chaperonin genes exist separately on genome (Maiwald et al. 2003; Ventura et al. 2004). Notably, such a situation is observed in 20–22% of the *Actinobacteria*, and interestingly in these organisms, in addition to the loss of operonic arrangement, the expression of chaperonin and co-chaperonin is differentially regulated. Generally, the expression of actinobacterial chaperonin genes is regulated via repression by HrcA (de Leon et al. 1997; Duchene et al. 1994; Grandvalet et al. 1998) or, in rare cases, by HspR (Barreiro et al. 2005), which bind the upstream inverted repeat elements, CIRCE and HAIR, respectively. Functionally, while the essential chaperonin copy has been proposed to act as the generalist chaperonin, the dispensable copy has been demonstrated to have diverged to attain

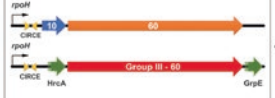
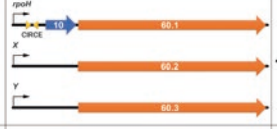
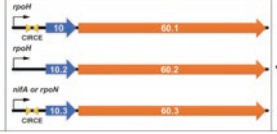
Phylum	Gene Organization	Essentiality	Moonlighting Functions	CTS	Modes of Evolution
Actinobacteria		• Cpn60.2 is essential • Cpn60.1 in Mtb - required for Granuloma Formation	• Mtb Cpns – Immunogenic	• Operonic Cpn60 – His-rich • Essential Cpn60 – GGM-like • Third and fourth copies – Pattern-less tail	• Gene Duplication at the LUCA
Firmicutes		• Not determined	• Group III Cpn60 - Folds proteins involved in anaerobic carbon monoxide oxidation	• Pattern-less tail	• Horizontal Transfer
Cyanobacteria		• Cpn60.1 is essential	• Cpn60.2 – Provides thermo-protection for photosynthesis related proteins	• Cpn60.1 – Smaller GGM-like tail • Cpn60.2 – Longer GGM-like tail	• Gene Duplication
Chlamydia		• Cpn60.1 is essential	• Cpn60.2 and Cpn60.3 – regulate Host's molecular events during infection.	• Cpn60.1 – smaller GGM-like tail • Other copies – pattern-less tail	• Successive Gene Duplication Events
α-Proteobacteria (Rhizobia)		• Cpn60.1 is essential	• NifA regulated Cpn – Dedicated for folding nitrogen fixation related proteins	• Generally GGM-like tail • Some have pattern-less tail	• Gene Duplication at the LUCA • NifA regulated Cpn – Horizontal Transfer

Fig. 3.1 Salient features of the multiple chaperonins in different phyla. Numbers 10 and 60 represent Cpn10 and Cpn60 homologues. *CTS* stands for carboxy-terminal segments and *LUCA* stands for last universal common ancestor. *CIRCE* stands for controlling inverted repeat of chaperone expression. In the cartoon depicting cyanobacterial chaperonins, H and K represent the upstream enhancing elements; H-box and K-box are induced by heat and light, respectively. In the cartoon depicting *Chlamydia* chaperonins, X and Y represent the two uncharacterised promoter elements

atypical functions to assist the organism during specific life stages, principally, the pathogenic stages (Basu et al. 2009; Ojha et al. 2005). This argument is supported by the evolutionary studies where a faster rate of evolution was observed for the dispensable copy (Goyal et al. 2006; Hughes 1993; Kumar et al. 2015). In addition, phylogenetic studies have shown that the modes of origin of multiple chaperonins in actinobacterial species have resulted due to a gene duplication event at the last common ancestor of *Actinobacteria* (Goyal et al. 2006; Hughes 1993; Kumar et al. 2015; Mande et al. 2013). A detailed description on the current advances in mycobacterial chaperonins is given in Chap. 5. Surprisingly, in *S. lividans* the second chaperonin copy can function independent of a co-chaperonin (de Leon et al. 1997). This observation provided a probable explanation for non-operonic location and independent regulation of the second chaperonin gene and suggested that this copy might play a different cellular role. Taken together, *Actinomycetes* provide a fascinating picture of genetic and functional diversity among the multiple chaperonins.

3.2.2 *Unique Chaperonins in* Firmicutes

Firmicutes constitute several Gram-positive bacteria, such as *Carboxydothermus hydrogenoformans*, *Staphylococcus aureus* and *Desulfitobacterium dehalogenans*, which are characterised by a low-G + C content genome. Surprisingly, in addition to the classical group I chaperonin genes, unlike their high-G + C phylogenetic neighbours, some of the *Firmicutes* encode archaeal-like chaperonins that are classified as group III chaperonins owing to their primary, tertiary and quaternary structural features, peculiar genomic location alongside the *dnaK* operon and unique mode of regulation (Techtmann and Robb 2010). Majority of the *Firmicutes* encode multiple copies of chaperonins (Table 3.1). Apparently, the group I and group III chaperonin genes are regulated by HrcA-mediated heat shock response. Surprisingly, all the chaperonin copies possess pattern-free CTS (Fig. 3.1). Since the Gly-Met-rich tail is supposed to determine the substrate pool, this observation suggests that the substrate pool of *Firmicutes* chaperonins is different from the other bacteria. Moreover, since these bacteria dwell in carbon monoxide-rich environments and thus rely on anaerobic oxidation of CO, the extra chaperonin copy is believed to fold the proteins involved in this pathway. In addition, in several *Firmicutes*, the location of chaperonin genes is peculiarly in operonic arrangements with either the Hsp70 system or with the gene encoding trigger factor (Smidt et al. 2000), suggesting a unified and temporal mode of regulation for the genes encoding different chaperone systems. Owing to such genomic organisation, phylogenetic analysis, therefore, proposed that the group III chaperonins might have been acquired horizontally from ancient archaea. Since the two phylogenetically diverse chaperonins coexist and share substrate pools, *Firmicutes*, therefore, present a unique coitus among chaperonin groups. There is therefore a need for comprehensive structural and functional studies to delineate their functional and phylogenetic diversity.

3.2.3 Functional Distribution Among the Chlamydial Chaperonins

Chlamydiae phylum constitutes several obligate intracellular pathogens such as *Chlamydia trachomatis*, *C. psittaci* and *C. pneumoniae* that are characterised by complex developmental cycles through different host cell types. *Chlamydiae* portray an extremely complex and unique scenario of chaperonins (Table 3.1). While majority of *Actinobacteria* possess 2 copies of chaperonin genes, majority of chlamydial species (10 out of 16 completely sequenced species) possess 3 chaperonin genes (Kumar et al. 2015; McNally and Fares 2007). However, similar to *Actinobacteria*, only one of the chaperonin genes is in operonic arrangement with the co-chaperonin gene (Fig. 3.1). This chaperonin bears the characteristic Gly-Met-rich CTS, which is essential and thus believed to function as the generalist chaperonin (Fig. 3.1). The other two chaperonin copies deviate from characteristic features, such as unusual ATP-binding site and lack of Gly-Met-rich CTS, and thus are believed to have diverged to acquire different non-canonical functions. Such a notion is further supported by the complex lifestyle-specific expression patterns of these chaperonins. Intriguingly, the expression of only the first copy is heat shock regulated and is thus repressed by the HrcA-CIRCE system (Karunakaran et al. 2003). However, the second copy is induced when the bacterium is in pathogenic mode, either inside a monocyte for a persistent infection or in synovial macrophages during reactive arthritis (Kol et al. 1999). On the other hand, the expression of the third copy is induced when the bacterium is in Hep-2 cells (Gerard et al. 2004). These observations suggest a life-cycle-specific expression patterns for these chaperonins. Additionally, low sequence identity among these chaperonins and the observation that the second and third copy deviate further in sequence from the first copy (Karunakaran et al. 2003) suggested the possibility of two independent gene duplication events during the evolution of chlamydial chaperonins (McNally and Fares 2007). Taken together, the chlamydial chaperonins present a complex interplay with sequence divergence, differential expression patterns and genome locations that have aided these chaperonin copies to perform specific functions during different life stages of chlamydia.

3.2.4 Rhizobial Chaperonins: The Aristocrats of Chaperonin Biology

Alphaproteobacteria constitute several legume symbionts that engage in nitrogen fixation in root nodules. This class of bacteria, called the rhizobia, harbours the highest number of copies for chaperonins, with the *Bradyrhizobium japonicum* hosting seven genes. Rhizobia, therefore, present a perfect division of labour among the chaperonins (Fischer et al. 1993). In the most well-characterised example, *Rhizobium leguminosarum*, the bacteria harbour three copies of chaperonin genes with all of

them forming separate operons along with the respective co-chaperonin genes (George et al. 2004; Gould et al. 2007). Interestingly, one among the three operons exhibits unique features; it is located in a genomic island that hosts nitrogen fixation genes, unlike the regular heat shock, it is regulated by NiF that regulates expression of nitrogen fixation gene, and as a chaperone it assists the folding and assembly of several Nod proteins (Ogawa and Long 1995). These observations added credence to the notion that one copy of chaperonin in rhizobia is dedicated to fold the proteins involved in nitrogen fixation (Kumar et al. 2015). Among the other two operon copies, one of them is essential, regulated by HrcA and thus is believed to act as a generalist chaperonin (Gould et al. 2007). Although considerable literature on the second copy is not available, this copy is demonstrated to act as a chaperone in folding several model substrates albeit possessing a pattern-free CTS. A detailed description on the rhizobial chaperonins is given in Chap. 6.

3.2.5 *Multiple Chaperonins in* Cyanobacteria*: One Copy is Green!*

Cyanobacteria phylum largely constitutes photosynthetic bacteria such as *Synechococcus platensis*, *Synechocystis* sp., *Anabaena variabilis* and *Prochlorococcus marinus*. About 90% of the currently available cyanobacterial genomes encode two chaperonin genes (Table 3.1), with one copy in operonic arrangement with the co-chaperonin and the other located separately (Kumar et al. 2015). Although this situation appears similar to that of *Actinobacteria*, the difference shows up in the species with three chaperonin genes (Fig. 3.1), where two of the chaperonins are in operonic arrangement with their co-chaperonin genes, while one is independent (Lund 2009). In contrast to the *Actinobacteria*, in *Cyanobacteria* the chaperonin(s) in operonic arrangement is (are) essential, while the individual one is dispensable (Sato et al. 2008). Interestingly, both chaperonins bear a Gly-Met-rich CTS, although CTS of the independent and dispensable chaperonin is very long (Lund 2009). Interestingly, since *Cyanobacteria* is photosynthetic, the extra copy is believed to offer thermo-tolerance to the photosynthetic system during heat shock. This notion is strongly supported by the way the chaperonin genes are regulated. Although both copies are regulated positively by RpoH and negatively by HrcA, the expression of the operon is rapidly induced upon heat shock due to the presence of the upstream enhancer elements known as the H, K and N boxes, while the expression of the second gene is induced gradually (Kojima and Nakamoto 2007; Rajaram and Apte 2010). In addition, the observation that even upon heat shock the second gene remains repressed during several photosynthesis-diminishing circumstances, such as when the bacteria are cultured in dark, when the photosystem's electron transfer is obstructed or when intracellular nitrate levels are increased (Kojima and Nakamoto 2007; Rajaram and Apte 2010), suggested that this chaperonin might have a direct connection with photosynthesis, probably by providing thermo-protection to the proteins involved in the light reaction.

Notably, similar dual copies of chaperonins are observed in chloroplasts of higher organisms, such as plants, suggesting ancient connections between the chaperonins and the evolution of photosynthesis (Nishio et al. 1999). Moreover, phylogenetic studies observed that the extra copies might have emerged by a single gene duplication event at the LUCA of cyanobacteria (Goyal et al. 2006). Moreover, functional studies on these chaperonins lead to interesting insights on the role of CTS in chaperonin function. While the copy in operonic arrangement that has optimal CTS could complement readily, the second gene albeit with a longer Gly-Met-rich CTS failed to complement *E. coli* GroEL (Furuki et al. 1996; Kovacs et al. 1992; Tanaka et al. 1997). Since a longer CTS has been shown to fill the chaperonin cavity, limit encapsulation to only smaller proteins and consequently decrease the client repertoire (Tang et al. 2006), the inability of the second chaperonin to complement *E. coli* GroEL could be due to its longer CTS and consequent smaller cavity. However, this limitation might have been evolutionarily driven to sequester only the photosynthesis-related proteins that are populated by smaller-sized proteins (Nakamura et al. 1998). A comprehensive chaperonin-client interaction studies are therefore required to comprehend the functional diversity in these chaperonins. Taken together, although the current understanding indicates that the cyanobacterial chaperonins have diverse functions and that the second chaperonin is linked to the photosynthesis, the precise characterisation of these chaperonins is required to delineate their functional diversity. Comprehensive description of cyanobacterial chaperonin system is presented in Chap. 7.

3.3 Why Multiple Chaperonins: Specific Examples

The existence of multiple genes for chaperonin has led to several hypotheses:

(a) Functional diversity: if all the copies work as intracellular chaperonins or have diverged to perform different functions.
(b) Evolutionary lineage: if these copies have resulted by horizontal acquisition from niche neighbours or due to gene duplication within the organism and do these multiple copies have any phylogenetic signature.
(c) Substrate spectrum: do the multiple chaperonins share the substrates or they have distinct substrate pools?
Primarily it was proposed that the organisms with multiple chaperonins might benefit either from the dosage effect (Kondrashov and Kondrashov 2006) or from the functional divergence of different chaperonins (Goyal et al. 2006). The former seems unlikely as the intracellular levels of chaperonins are always high. Moreover, as elaborated in the following chapters, multiple GroELs have been characteristic of organisms with complex lifestyle, suggesting the plausibility of the latter scenario. The following chapters will, therefore, review the current advances in understanding on the functional dictum of multiple chaperonins by presenting fascinating examples of bacteria and archaea with multiple

chaperonin genes. Chapter 4 will review the functional redundancy observed in chaperonins of myxobacteria and how the two dispensable chaperonins distribute their substrates and functions in life-stage-specific fashion (Chap. 4). Chapter 5 presents the current understanding in the functional diversity of mycobacterial chaperonin paralogues, where only one copy is essential and thus might function as the generalist chaperonin (Chap. 5). The other copies, on the other hand, have diverged in sequence and have been demonstrated to play important roles in the establishment and progression of the pathogenesis. Chapter 6 reviews the fascinating division of labour among the rhizobial chaperonins, where one set of chaperonins functions exclusively to fold the proteins involved in nitrogen fixation (Chap. 6). Likewise, Chap. 7 illustrates how one copy of chaperonin is dedicated to photosynthesis (Chap. 7). Notably, rhizobia are the bacteria which harbour the highest number of chaperonins. Chapter 8 reviews the situation of multiple chaperonins in thermoresistant archaea (Chap. 8) and reviews how the coexistence of evolutionarily diverse group I and group II chaperonins shaped the proteomes of the mesophilic methanogens (Chap. 8) and how this understanding can be translated to therapeutic approaches. The final chapter will review probable means of evolution of the multiple chaperonins (Chap. 9). These chapters are scientifically scintillating and reveal how the multiple chaperonin copies have been tuned according to the species-dependent requirement.

3.4 A Note on Chaperonin Nomenclature

Apart from the functional diversity that multiple chaperonins display, diversity prevails even in their nomenclature, leading to a conundrum. The purpose of this note is to explain the basis of the conundrum and try to unify different ways the chaperonins are referred to. Molecular chaperones are classified according to their molecular masses as Hsp100, Hsp90, Hsp70, Hsp60 and small Hsps (Kumar et al. 2015). Thus, the 60 kD chaperones are named as Hsp60 chaperones. Further, since they form rings, they were called chaperonins and thus were abbreviated as Cpn60 (Hemmingsen et al. 1988). Incidentally, since the chaperonin homologue of *E. coli* was identified as a gene required for the growth of bacteriophage lambda (Georgopoulos et al. 1973), it was named as GroEL (or GroL). Therefore, the same protein has been given in different names by different researchers as Hsp60, Cpn60 and GroEL. Likewise, the 10 kD co-chaperonins are called Hsp10, Cpn60 and GroES, respectively. The situation with multiple chaperonins is even more complicated. The copies of the chaperonins are named either as GroEL1, GroEL2 and so on or as Cpn60.1, cpn60.2 and so on. The Hsp60-type nomenclature, Hsp60_1 and Hsp60_2, is less common in multiple chaperonins. Peculiarly, some researchers prefer to name the chaperonin copy that forms an operon with its co-chaperonin as GroEL while the independent copy as Cpn60, as seen with a few cyanobacteria

(Lehel et al. 1993). Such a diversity in the nomenclature, obviously, leads to confusion to the readers, and a unified code for naming chaperonins, especially in the case of multiple chaperonins, has been proposed (Coates et al. 1993). According to this proposal, the GroEL name should be limited to the *E. coli* GroEL since this implicates a function in bacteriophage maturation, and since the chaperonins in other bacteria have not been demonstrated a bacteriophage maturation role, they should be termed as Cpn60 (Coates et al. 1993; Lund 2009). Hsp60 type of naming, however, is generally used for the mitochondrial chaperonins. The diversity still remains, since the researchers tend to continue to follow the names they are comfortable with. Therefore, while editing this book, we have acknowledged the nomenclature styles that the respective authors are comfortable with. Therefore, the purpose of this note is to make the readers familiar with the variety in chaperonin nomenclature that can be encountered in the subsequent chapters and thus have a lucid reading.

3.5 Conclusions

Multiple chaperonins are becoming common in prokaryotes that go through either several growth stages or hosts during their life cycle. In several organisms, these chaperonins have been demonstrated to assist either a particular life phase or a process (Fig. 3.1). Examples for the former appear in the chlamydial chaperonins, where the different chaperonins conquest as the bacterium passes through different host cells. Examples for the latter, however, appear in the rhizobia, mycobacteria and cyanobacteria where one of the copies of chaperonins is dedicated to assist the nitrogen fixation, pathogenesis and photosynthesis, respectively (Fig. 3.1). Taken together, such observations suggest a strong correlation to the biological significance for the existence of these multiple chaperonin copies and therefore compel a need for comprehensive investigations to unravel the biology of these fascinating molecules.

Acknowledgments Santosh is Newton International Fellow at the University of Birmingham, UK, sponsored by The Royal Society, The British Academy and the Academy of Medical Sciences, UK. Further, we wish to acknowledge the support of Department of Biotechnology, India.

References

Barreiro C, Gonzalez-Lavado E, Brand S, Tauch A, Martin JF (2005) Heat shock proteome analysis of wild-type Corynebacterium glutamicum ATCC 13032 and a spontaneous mutant lacking GroEL1, a dispensable chaperone. J Bacteriol 187(3):884–889

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(15):4944–4954

Coates AR, Shinnick TM, Ellis RJ (1993) Chaperonin nomenclature. Mol Microbiol 8(4):787

de Leon P, Marco S, Isiegas C, Marina A, Carrascosa JL, Mellado RP (1997) Streptomyces lividans groES, groEL1 and groEL2 genes. Microbiology 143(Pt 11):3563–3571

Duchene AM, Thompson CJ, Mazodier P (1994) Transcriptional analysis of groEL genes in Streptomyces coelicolor A3(2). Mol Gen Genet 245(1):61–68

Ferat JL, Michel F (1993) Group II self-splicing introns in bacteria. Nature 364(6435):358–361

Ferreyra RG, Soncini FC, Viale AM (1993) Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (groESL) genes from the phototrophic sulfur bacterium Chromatium vinosum. J Bacteriol 175(5):1514–1523

Fischer HM, Babst M, Kaspar T, Acuna 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(7):2901–2912

Furuki M, Tanaka N, Hiyama T, Nakamoto H (1996) Cloning, characterization and functional analysis of groEL-like gene from thermophilic cyanobacterium Synechococcus vulcanus, which does not form an operon with groES. Biochim Biophys Acta 1294(2):106–110

George R, Kelly SM, Price NC, Erbse A, Fisher M, Lund PA (2004) Three GroEL homologues from Rhizobium leguminosarum have distinct in vitro properties. Biochem Biophys Res Commun 324(2):822–828

Georgopoulos CP, Hendrix RW, Casjens SR, Kaiser AD (1973) Host participation in bacteriophage lambda head assembly. J Mol Biol 76(1):45–60

Gerard HC, Whittum-Hudson JA, Schumacher HR, Hudson AP (2004) Differential expression of three chlamydia trachomatis hsp60-encoding genes in active vs. persistent infections. Microb Pathog 36(1):35–39

Gould P, Maguire M, Lund PA (2007) Distinct mechanisms regulate expression of the two major groEL homologues in Rhizobium leguminosarum. Arch Microbiol 187(1):1–14

Goyal K, Qamra R, Mande SC (2006) Multiple gene duplication and rapid evolution in the groEL gene: functional implications. J Mol Evol 63(6):781–787

Grandvalet C, Rapoport G, Mazodier P (1998) hrcA, encoding the repressor of the groEL genes in Streptomyces albus G, is associated with a second dnaJ gene. J Bacteriol 180(19):5129–5134

Hausner G, Hafez M, Edgell DR (2014) Bacterial group I introns: mobile RNA catalysts. Mob DNA 5(1):8

Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333(6171):330–334

Hughes AL (1993) Contrasting evolutionary rates in the duplicate chaperonin genes of Mycobacterium tuberculosis and M. leprae. Mol Biol Evol 10(6):1343–1359

Karunakaran KP, Noguchi Y, Read TD, Cherkasov A, Kwee J, Shen C, Nelson CC, Brunham RC (2003) Molecular analysis of the multiple GroEL proteins of Chlamydiae. J Bacteriol 185(6):1958–1966

Kojima K, Nakamoto H (2007) A novel light- and heat-responsive regulation of the groE transcription in the absence of HrcA or CIRCE in cyanobacteria. FEBS Lett 581(9):1871–1880

Kol A, Bourcier T, Lichtman AH, Libby P (1999) Chlamydial and human heat shock protein 60s activate human vascular endothelium, smooth muscle cells, and macrophages. J Clin Invest 103(4):571–577

Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. J Theor Biol 239(2):141–151

Kong TH, Coates AR, Butcher PD, Hickman CJ, Shinnick TM (1993) Mycobacterium tuberculosis expresses two chaperonin-60 homologs. Proc Natl Acad Sci USA 90(7):2608–2612

Kovacs E, Lehel C, Mustardy L, Gombos Z, Torok Z, Horvath I, Vigh L (1992) Heat-stress induces association of the GroEL-analog chaperonin with thylakoid membranes in cyanobacterium, Synechocystis PCC-6803. Kluwer Academic, Dordrecht, p 214

Kumar CM, Mande SC (2011) Protein chaperones and non-protein substrates: on substrate promiscuity of GroEL. Curr Sci 100(11):1646–1653

Kumar P, Kumar D, Parikh A, Rananaware D, Gupta M, Singh Y, Nandicoori VK (2009) The Mycobacterium tuberculosis protein kinase K modulates activation of transcription from the promoter of mycobacterial monooxygenase operon through phosphorylation of the transcriptional regulator VirS. J Biol Chem 284(17):11090–11099

Kumar CM, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. Cell Stress Chaperones 20(4):555–574

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(3):1799–1804

Lund PA (2009) Multiple chaperonins in bacteria—why so many? FEMS Microbiol Rev 33(4):785–800

Maiwald M, Lepp PW, Relman DA (2003) Analysis of conserved non-rRNA genes of Tropheryma whipplei. Syst Appl Microbiol 26(1):3–12

Mande SC, Kumar CM, Sharma A (2013) Evolution of bacterial chaperonin 60 paralogues and moonlighting activity. In: Henderson B (ed) Moonlighting cell stress proteins in microbial infections. Springer, Netherlands, pp 101–121

Martinez-Abarca F, Toro N (2000) Group II introns in the bacterial world. Mol Microbiol 38(5):917–926

Mazodier P, Guglielmi G, Davies J, Thompson CJ (1991) Characterization of the groEL-like genes in Streptomyces albus. J Bacteriol 173(22):7382–7386

McNally D, Fares MA (2007) In silico identification of functional divergence between the multiple groEL gene paralogs in Chlamydiae. BMC Evol Biol 7:81

Nakamura Y, Kaneko T, Hirosawa M, Miyajima N, Tabata S (1998) CyanoBase, a www database containing the complete nucleotide sequence of the genome of Synechocystis sp. strain PCC6803. Nucleic Acids Res 26(1):63–67

Nishio K, Hirohashi T, Nakai M (1999) Chloroplast chaperonins: evidence for heterogeneous assembly of alpha and beta Cpn60 polypeptides into a chaperonin oligomer. Biochem Biophys Res Commun 266(2):584–587

Ogawa J, Long SR (1995) The Rhizobium meliloti groELc locus is required for regulation of early nod genes by the transcription activator NodD. Genes Dev 9(6):714–729

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(5):861–873

Perez J, Castaneda-Garcia A, Jenke-Kodama H, Muller R, Munoz-Dorado J (2008) Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. Proc Natl Acad Sci U S A 105(41):15950–15955

Price CT, Al-Khodor S, Al-Quadan T, Abu Kwaik Y (2010) Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of Legionella pneumophila within macrophages and amoebae. Infect Immun 78(5):2079–2088

Rajaram H, Apte SK (2010) Differential regulation of groESL operon expression in response to heat and light in Anabaena. Arch Microbiol 192(9):729–738

Rinke de Wit TF, Bekelie S, Osland A, Miko TL, Hermans PW, van Soolingen D, Drijfhout JW, Schoningh R, Janson AA, Thole JE (1992) Mycobacteria contain two groEL genes: the second Mycobacterium leprae groEL gene is arranged in an operon with groES. Mol Microbiol 6(14):1995–2007

Sato S, Ikeuchi M, Nakamoto H (2008) Expression and function of a groEL paralog in the thermophilic cyanobacterium Thermosynechococcus elongatus under heat and cold stress. FEBS Lett 582(23–24):3389–3395

Smidt H, van Leest M, van der Oost J, de Vos WM (2000) Transcriptional regulation of the cpr gene cluster in ortho-chlorophenol-respiring Desulfitobacterium dehalogenans. J Bacteriol 182(20):5683–5691

Tanaka N, Hiyama T, Nakamoto H (1997) Cloning, characterization and functional analysis of groESL operon from thermophilic cyanobacterium Synechococcus vulcanus. Biochim Biophys Acta 1343(2):335–348

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(5):903–914

Techtmann SM, Robb FT (2010) Archaeal-like chaperonins in bacteria. Proc Natl Acad Sci U S A 107(47):20269–20274

Ventura M, Canchaya C, Zink R, Fitzgerald GF, van Sinderen D (2004) Characterization of the groEL and groES loci in bifidobacterium breve UCC 2003: genetic, transcriptional, and phylogenetic analyses. Appl Environ Microbiol 70(10):6197–6209

Vitlin Gruber A, Nisemblat S, Azem A, Weiss C (2013) The complexity of chloroplast chaperonins. Trends Plant Sci 18(12):688–694

Chapter 4
Dynamic Interplay of the Myxobacterial Chaperonins

Li Zhuo, Yan Wang, Zheng Zhang, and Yue-zhong Li

Abstract Most of the sequenced myxobacterial genomes possess duplicated *groEL* genes: one is in a complete bicistronic *groESL* operon, while the other(s) are in a *groESL* operon or stand alone. The two *groEL* genes are subneofunctionalized in *Myxococcus xanthus* DK1622, the model strain of myxobacteria. Although alternatively essential for cell survival, *groEL1* is required for the developmental process, while *groEL2* participates in the predation process and the biosynthesis of secondary metabolite myxovirescin. The divergent functions of GroEL1 and GroEL2 are majorly resulted from the differences of the apical and C-terminal equatorial regions of the two paralogous chaperonins. The stand-alone *groEL2* gene still requires *groES* for functions, and the expression levels of *groELs* and *groES* genes could be synergic and self-regulated. There is a complicated dynamic interplay between duplicated GroEL proteins and the single cofactor GroES in *M. xanthus*.

4.1 Introduction

The Gram-negative myxobacteria are phylogenetically located in the delta-division of the *Proteobacteria* (Reichenbach 2004; Shimkets et al. 2006). Myxobacteria are unique among the *Proteobacteria* by their complex multicellular social behaviors. The myxobacterial cells are able to move in swarms on solid surfaces, feed on macromolecules or other microbial cells cooperatively, and, when foods are exhausted, develop multicellular fruiting body structures, inside which embodied adversity-resistant myxospores (Dworkin and Kaiser

L. Zhuo • Z. Zhang • Y.-z. Li (✉)
State Key Laboratory of Microbial Biotechnology, School of Life Science, Shandong University, Jinan 250100, China
e-mail: lilab@sdu.edu.cn

Y. Wang
College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_4

1993; Shimkets 1990). Many proteins participate in the social behavior processes, and chaperonins are required for the involving protein refolding. There are several reports related to the functions of chaperonins in the sociality of *Myxococcus*. For example, SglK, a homolog of DnaK, was shown to be essential for social motility and multicellular development in *M. xanthus* DK1622 (Weimer et al. 1998; Yang et al. 1998); and the duplicated *groEL* genes play divergent functions in the predation and development processes (Li et al. 2010), as well as in secondary metabolisms (Wang et al. 2014).

Chaperonins are essential cellular components that are responsible for protein refolding, assembly, transportation, and degradation (Lund 2001; Ranson et al. 1998) and are important in many cellular physiological processes (Houry et al. 1999). Chaperonins are also a major group of heat-shock proteins that are overexpressed at high temperatures to help proper refolding of lots of proteins denatured by heat shock, thus playing fundamental roles for the survival at nonpermissive temperatures (Kerner et al. 2005; Fayet et al. 1989; VanBogelen et al. 1987). GroEL is a type I chaperonin. The GroEL chaperonin proteins have been confirmed to participate in various physiological processes in *Escherichia coli* (Kerner et al. 2005). The GroEL proteins are characterized by forming 14-mer homopolymers, arranged as two back-to-back stacked rings, and each of the rings comprises seven GroEL subunits (Horwich et al. 2006). Classically, GroEL strictly depends on the dome-shaped heptamer of GroES proteins to complete the folding task in an ATP-dependent manner (Fenton et al. 1994). In the presence of ATP, the co-chaperonin GroES heptamer binds to the GroEL homopolymer, forming a large central cavity that encapsulates the substrate proteins and enables correct refolding through multiple cycles of binding and release (Liu et al. 2009; Saibil and Ranson 2002; Weissman et al. 1995).

Because of its important roles in many cellular processes, the *groEL* gene is ubiquitously distributed in bacteria, normally arranged with a *groES* gene to form a bicistronic *groESL* operon. While most bacteria possess single *groEL* genes, some others possess two or more highly conserved copies of the *groEL* gene (Craig et al. 1993; Jiang et al. 2008; Karlin and Brocchieri 2000; Gould et al. 2007). Genome sequencing showed that approximately 30% of bacterial genomes have evolved duplicated *groEL* genes (Lund 2001; Kumar et al. 2015). The duplicated *groEL* genes play divergent roles in many different cellular processes in different bacterial species (George et al. 2004; Ojha et al. 2005; Bittner et al. 2007; Li et al. 2010). *Myxococcus xanthus* DK1622 is a model strain of myxobacteria. The strain has a large genome of 9.14 Mbp (Goldman et al. 2006), possessing two *groEL* genes—one in a *groESL* cluster while the other with no neighboring *groES*. In this chapter, we described the compositions of *groEL* genes in sequenced myxobacterial genomes. We compared the divergent characteristics of the duplicated *groEL* genes in *Myxococcus xanthus* DK1622. The molecular evolution of *groEL1* and *groEL2* for functional divergence was analyzed. We assayed the requirement of *groES* by the stand-alone *groEL* gene and expressions of the duplicated *groEL* and single *groES* genes. The present results indicate that the duplicated GroEL proteins and the single cofactor GroES proteins have complicated dynamic interplays in *M. xanthus*.

4.2 Composition of the *groEL* and *groES* Genes in Myxobacteria

Myxobacteria have the largest genomes among different bacteria. Myxobacterial genomes normally possess many duplicated genes, including the chaperone genes (Goldman et al. 2006; Han et al. 2013). Figure 4.1 is a phylogenomic tree of 24 sequenced myxobacterial genomes. Bioinformatics analysis indicated that, of the 24 sequenced myxobacterial genomes, 19 have two *groEL* genes, 2 have three copies of *groEL*, and 3 have single *groEL* genes (Fig. 4.2a). There is at least a complete bicistronic *groESL* operon in a myxobacterial genome. The second *groEL* gene is in a *groESL* operon or stands alone. It is noted that the loss of neighboring *groES* is rather a characteristic for the duplicated *groEL* genes in the *Cystobacterineae* suborder, whereas the four sequenced genomes of *Sorangium* and *Chondromyces* in the *Sorangineae* suborder of myxobacteria have an additional complete *groESL* operon. Similar to that of the *Sorangineae* genomes, *Haliangium ochraceum* DSM14365, a marine halophilic strain that presently belongs to the *Nannocystineae* suborder, also has two complete *groESL* operons. Although the occurrences of *groES* and *groEL* genes are normally coincident within high taxonomic units, they may be varied at

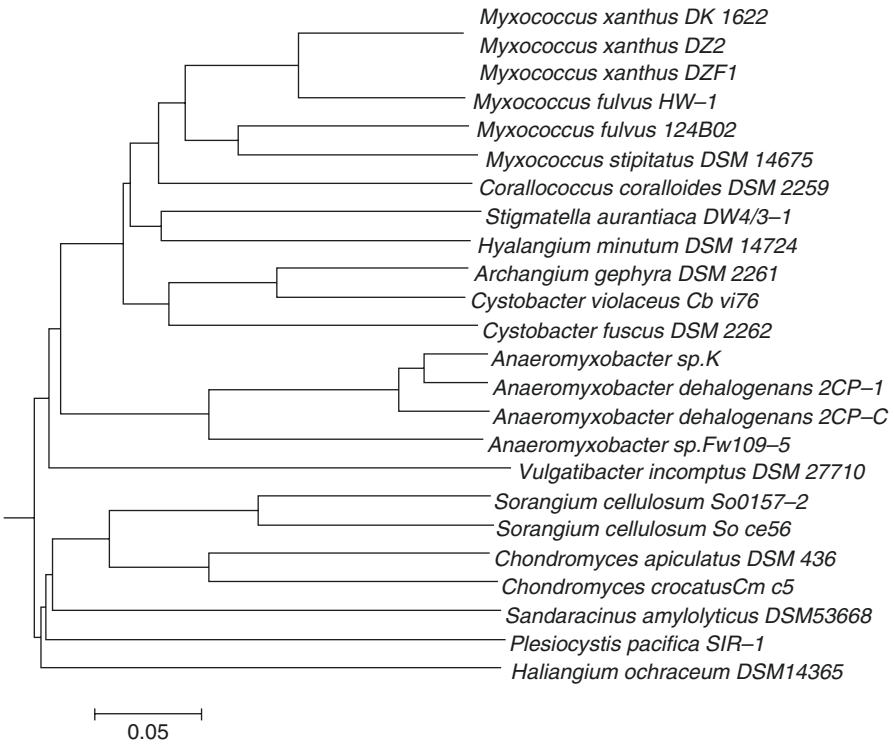


Fig. 4.1 A non-rooted phylogenomic tree based on the complete genome sequences of 24 sequenced myxobacteria

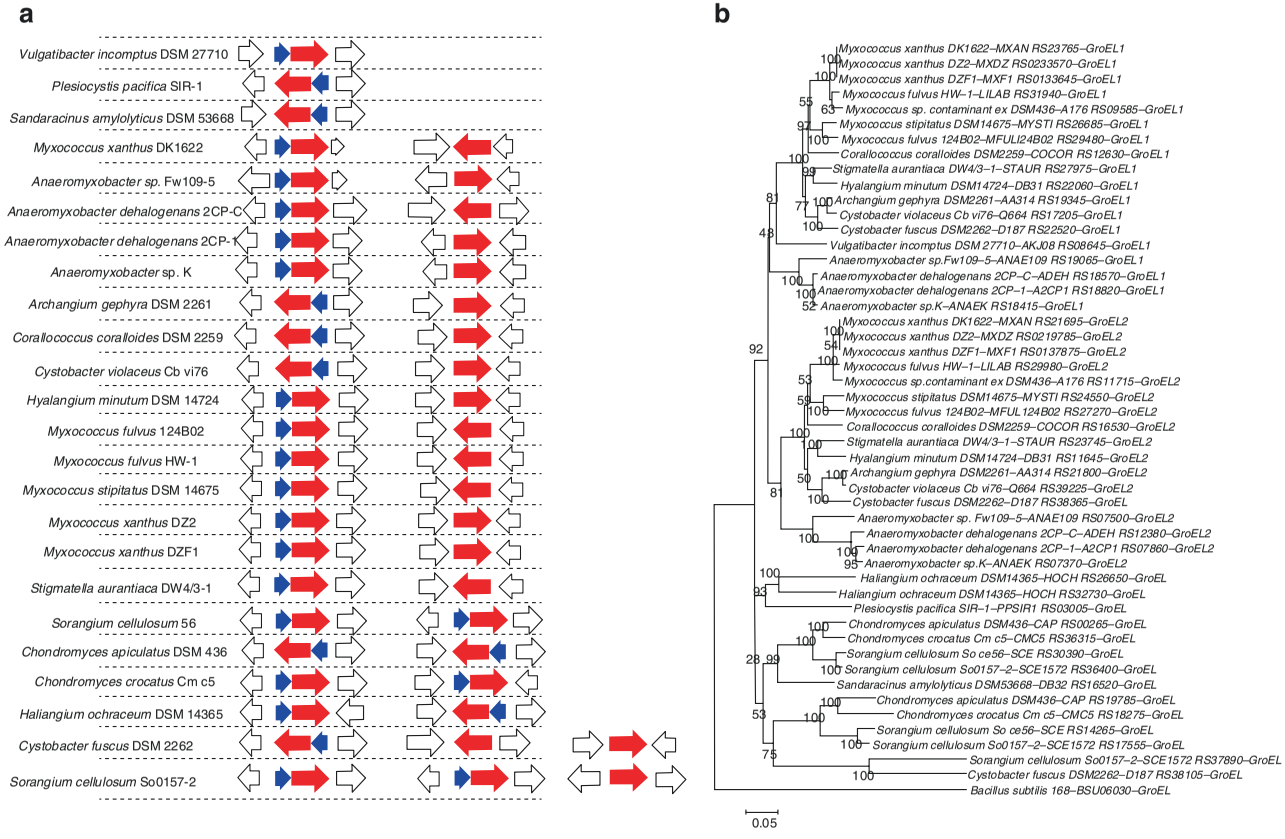


Fig. 4.2 The composition and organization (**a**) of the *groEL* and *groES* genes and a phylogenetic analysis of the GroEL protein sequences (**b**) in the 24 sequenced myxobacterial genomes. The *red* and *blue arrows* are the *groEL* and *groES* genes, respectively, showing the transcription directions of the genes

genus or even species levels. For example, *Sorangium cellulosum* So ce56 has two copies of the *groEL* gene; the So0157-2 strain has three copies. In the two genomes containing three *groEL* copies, the *S. cellulosum* strain So0157-2 has two complete *groESL* operons and a stand-alone *groEL*, while the *Cystobacter fuscus* strain DSM 2262 has a single complete *groESL* operon and two stand-alone *groEL* genes. The three myxobacteria with single *groEL* genes are some newly identified myxobacterial strains, i.e., *Vulgatibacter incomptus* DSM 27710, *Sandaracinus amylolyticus* DSM 53668, and *Plesiocystis pacifica* SIR-1, which are all in far phylogenetic distances from those "classical myxobacteria" (refer to Fig. 4.1).

Phylogenetic analysis indicated that the duplication events of the *groEL* genes happened in the early days after the separation of myxobacterial suborders, and the *groEL* genes thus form three major phylogenetic groups according to the suborder taxonomy (Fig. 4.2b). In the *Cystobacterineae* suborder, the two *groEL* genes were clearly separated into the *groEL1* and *groEL2* groups. Similarly, duplicated *groEL* genes were also separated in *Sorangineae* suborder. Interestingly, the phylogenetic analysis indicated that the third *groEL* genes in *S. cellulosum* So0157-2 (SCE1572_RS37890) and *C. fuscus* DSM 2262 (D187_RS38105) were significantly distinct from the other two sympatric *groEL* genes in the two strains, even different from the other myxobacterial ones, which indicated that they were clearly derived from horizontal gene transfers. Thus, in each of the myxobacterial genomes, there was at best a single duplication event of the *groEL* gene. The three genomes containing single *groEL* genes are near the roots of the suborders. It is unclear whether the duplicated *groEL* genes were lost during the evolution or the *groEL* genes were not duplicated in these myxobacterial genomes.

4.3 Divergent Functions of the Two *groEL* Genes in *M. xanthus* DK1622

Of the duplicated *groEL* genes, one is normally essential, while the others are detectable in many bacterial species (George et al. 2004; Ojha et al. 2005; Bittner et al. 2007). However, the two *groEL* paralogous genes in *M. xanthus* DK1622 are alternatively detectable (Li et al. 2010). Single deletion of *groEL1* (the mutant was named YL0301) or *groEL2* (YL0302) did not affect the growth abilities of the mutants, and the mutants showed almost no difference of their growth abilities, comparing with the wild-type strain DK1622. However, the two *groEL* genes could not be deleted at the same time, which indicated that the two *groEL* genes were alternatively essential for the cell survivals. The deletion mutants of the *groEL1* and *groEL2* genes exhibited different survival abilities after a heat shock at 42 °C, which were determined to be due to the differences in the total expression levels of the *groEL1* and/or *groEL2* genes (Wang et al. 2013). The results indicated that the functions of *groEL1* and *groEL2* genes for the survival in response to heat shock were also substituted with each other.

Different from their replaceable functions in growth and heat-shock response, functions of the *groEL1* and *groEL2* genes are divergent in the predation and development processes in *M. xanthus* DK1622 (Li et al. 2010). The deletion of the *groEL2* gene made

the mutant (YL0302) to be deficient in the predation process. However, when placed on the TPM development medium, the mutant had similar abilities in the development of fruiting body structures and sporulation as the wild-type strain DK1622. Comparatively, the deletion of the *groEL1* gene made the mutant (YL0301) to be deficient in the development of fruiting body structures and the sporulation, but has a normal predation behavior on *E. coli* cell mat. When the *groEL1* deletion mutant was complemented with the *groEL1* gene (the YL0901 mutant), the formation ability of fruiting body structures was mostly reversed in the complementary mutant, and the sporulation ability reached 70–80% of that of the wild-type strain DK1622 (Wang et al. 2013). In contrast, when the *groEL1* deletion mutant was complemented with *groEL2*, forming a mutant containing two copies of *groEL2* but no *groEL1*, the strain (YL0902) still displayed developmental defects similar to that of YL0301, and the sporulation ability of YL0902 was approximately 20% of that of the wild-type strain DK1622.

Similarly, the introduction of *groEL1* into YL0302 does not improve the predation-feeding ability of *Myxococcus* cells on *E. coli* cells, but the complementation of the *groEL2* gene into YL0302 was able to recover the predation abilities of the mutant significantly. However, the total expression levels of *groEL1* and *groEL2* are similar in the YL0906 and YL0907 mutants, derived from the *groEL2* deletion mutant complemented with the *groEL1* and *groEL2* genes, respectively. Extracts of wild-type DK1622 and the deletion mutants of *groEL1* and *groEL2* genes were analyzed by using high-pressure liquid chromatography. One major peak present in DK1622 or YL0301, but not in YL0302, was identified as myxovirescin, a major secondary compound produced in *M. xanthus* DK1622. The myxovirescin compound was determined to be the key factor for the predation defection of YL0302 on *E. coli* cells (Wang et al. 2014). This result suggested that the *groEL2* gene was required for the correct folding of some key enzymes in the biosynthesis of myxovirescin, which is consistent with the results of the immunoprecipitation assays with the *groEL*1 and *groEL2* deletion mutants (Wang et al. 2013).

It is noted that the alternative deletions of the duplicated *groEL* genes do not affect the social motility (S-motility). Instead, the chaperone HSP70 proteins (DnaK and its homolog) are involved in the control of S-motility (Yang et al. 1998). Thus, the chaperonins probably play separate but synergistic functions in the social behaviors in *M. xanthus* DK1622 cells.

4.4 Molecular Evolution of *groEL1* and *groEL2* for Functional Divergence

The divergent functions of *groEL1* and *groEL2* in the predation and development processes were determined due to the sequence differences, rather than their expression levels (Wang et al. 2013, 2014). There are five domains in a GroEL protein sequence, i.e., an N-terminal equatorial region, an N-intermediate region, an apical region, a C-intermediate region, and a C-terminal equatorial region (Brocchieri and Karlin 2000) (demonstrated in Fig. 4.3 using the *M. xanthus* GroELs as an example). The two intermediate regions are the highly conserved between the two

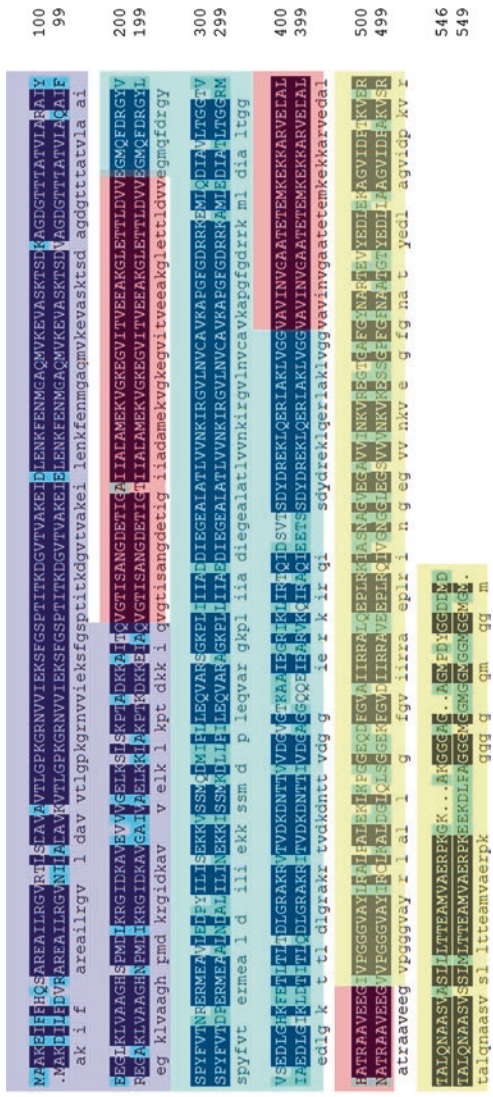


Fig. 4.3 A diagrammatic sketch of the duplicated GroEL protein sequences, showing the domain compositions. *Violet-filled square* N-terminal equatorial region, *orange-filled square* intermediate regions, *blue-filled square* apical region, and *yellow-filled square* C-terminal equatorial region

M. xanthus DK1622 GroELs (>97% identities), but the identities of the other three regions are ranged from 62.6% for the C-terminal equatorial region to 81% for the N-terminal equatorial region. The duplicated GroEL sequences in myxobacteria have similar characteristics of the five domains.

A series of single region-swapping experiments were performed with the N-terminal equatorial, the apical, and the C-terminal equatorial regions between the *groEL*1 and *groEL*2 genes of *M. xanthus* DK1622. The swapping revealed that the apical domains of GroEL1 and GroEL2 are majorly responsible for their functional divergence in the development and predation processes of *M. xanthus* DK1622, respectively (Wang et al. 2013). In addition, the C-terminal equatorial regions also involve in the functional divergence, i.e., the GroEL1 C-terminal region is apt for the development process, while the GroEL2 C-terminal region involves in the predation process. Comparatively, the N-terminal equatorial region has almost no effect for the functional divergence of the two GroEL proteins in predation and development processes. The substrate specificities of the GroEL1 and GroEL2 were exhibited by the pulldown and mass spectrometry results (Wang et al. 2013), which determine the functional divergence of the duplicate *groEL*s in the predation and development processes.

4.5 Both GroELs Require GroES for Their Functions

In the genome of *M. xanthus* DK1622, the *groEL1* gene is neighbored with the single *groES* gene, while the *groEL2* gene stands alone. We determined that the single *groES* gene was undeletable (Zhuo et al. 2017). The two GroELs and the single GroES proteins were isolated and purified from their hetero-expressions in *E. coli* cells. The GroES was designed containing a His-tag at the C-terminus, which was used to bind the protein to a nickel column. When the GroEL proteins with no His-tag were flowed through a GroES-binding column, either GroEL1 or GroEL2 was able to be retained in the column. As the concentration of the column-bound GroES increased, the column recoveries of either GroEL1 or GroEL2 proteins were also increased. Malate dehydrogenase (MDH) is a protein that requires a GroEL chaperonin for refolding after denaturalization (Martin et al. 1991). In the presence of GroEL1 or GroEL2 without GroES, denatured MDH showed slight renaturalization *in vitro*. However, when GroES was added, approximately 25% of MDH were renatured by either GroEL1 or GroEL2 of *M. xanthus* DK1622. GroEL is an ATPase, and the binding of GroES can reduce the ATPase activity of GroEL proteins (Viitanen et al. 1991; Martin et al. 1991). Similarly, the assays of the ATPase activities of the two GroELs, with or without the presence of GroES proteins, also determined that the GroES proteins were able to bind to either GroEL1 or GroEL2 *in vitro* (Zhuo et al. 2017). It is noted that GroEL1 had a higher *in vitro* ATPase activity than GroEL2 in either the absence or presence of GroES, probably suggesting their discrepant refolding activities *in vivo*.

The *in vivo* polymerization characteristics of the DK1622 GroEL1 and GroEL2 proteins were estimated in *E. coli* cells with the presence or absence of the DK1622 GroES. As the native PAGE shown, the extracts of *E. coli* cells containing DK1622 *groEL1* or *groEL2* genes but not DK1622 *groES* gene had similar molecular weights (Mw) of the GroEL complexes. The molecular weights (Mw) of the GroEL complexes are both approximately 800 kDa, 14 times of the molecular weights of the GroEL proteins (Mw of GroEL1 and GroEL2 are 57.9 kDa and 58.1 kDa, and their 14-mer polymers are 810.6 kDa and 813.4 kDa, respectively). Co-expression of DK1622 *groEL1*/*groEL2* and *groES* genes similarly produced the copolymers of GroEL and GroES, forming lagged bands with similar molecular sizes. Thus, the two types of GroEL proteins also have the same binding manner with the GroES proteins *in vivo*. This conclusion was also confirmed by refolding activities of the DK1622 HrcA proteins in *E. coli* cells containing the *groEL1* or *groEL2* gene, with co-expression of the *groES* gene or not. HrcA is a negative regulatory protein, which is able to bind to the CIRCE (Controlling Inverted Repeat of Chaperone Expression) regions in front of the *groEL* operons and is thus able to reduce the expression of *groEL*s or *groESL*s (Wilson et al. 2005). The HrcA protein relies on GroEL for refolding. The DK1622 HrcA protein is a potential substrate of the GroEL proteins from DK1622 in *E. coli* cells. As expected, in the absence of the DK1622 *groES* gene, the HrcA proteins existed mostly in inclusion bodies, which became resolvable when the DK1622 *groES* gene was also co-expressed together in *E. coli* cells (Zhuo et al. 2017).

4.6 Synergic Expressions of the Single *groES* and the Double *groEL*s

Assayed by using the quantitative PCR technique, the expression levels of the two *groEL* genes were significantly different in *M. xanthus* DK1622, but in a similar tendency (Fig. 4.4 shows the curves in the CTT growth medium). Similarly, the single *groES* gene had a similar curve as that of the *groEL* genes. These three genes increased their expressions markedly in the early stage of the incubation at 30 °C and reached the summits at approximately 24 h of incubation (Fig. 4.4a). After 24 h of incubation, the existence of their transcriptional products decreased and reached the lowest levels after approximately 42–48 h of incubation. Heat-shock treatments at 42 °C for 1 h at different incubation times made the expression levels of the three genes to increase approximately ten times. The three genes also had a similar tendency of the curves after the heat shock (Fig. 4.4b). Thus, the expression curve of *groES* was almost always highly consistent with that of the *groEL* genes at different times under different conditions. The expression amounts of the *groES* gene were almost equivalent to the total expressions of *groEL1* and *groEL2* in the wild-type strain.

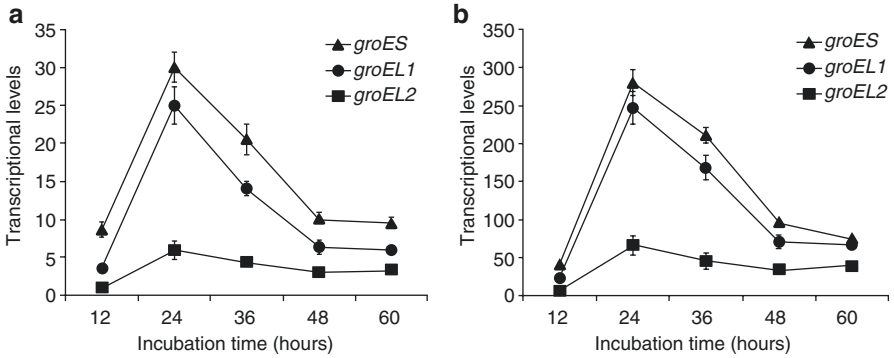


Fig. 4.4 The expression levels of the *groEL1*, *groEL2*, and *groES* genes in *M. xanthus* DK1622 cells cultivated in the CTT growth medium at 30 °C for different incubation times (**a**) or after a heat shock at 42 °C for 1 h at each incubation time (**b**). The transcriptional level of the *groEL2* gene at the 12 h of incubation was set as one unit

When *groEL1* or *groEL2* gene was deleted, expressions of the two *groEL* genes changed complicatedly (Wang et al. 2013; Zhuo et al. 2017; Li et al. 2010). Interestingly, the deletion of the *groEL2* gene, which has no neighboring *groES* gene, caused expressional decreases of not only the *groEL* gene but also the *groES* gene at each testing time points (Zhuo et al. 2017). This result is similar to that caused by the deletion of the *groEL1* gene. When *groEL1* or *groEL2* was complemented allopatrically at the *attB* site in the genome of YL0301 (YL0901) or YL0302 (YL0902), the mutants recovered the expressions of either *groEL* or *groES* to the levels in DK1622. An excess *groES* gene was introduced into DK1622 using the shuttle plasmid pZJY41 under the promoter of either *groEL1* or *groEL2*, forming the mutants of YL1103 and YL1102, respectively. Compared with that in the wild-type strain DK1622, the expressions of the *groES* gene were increased markedly in either the YL1102 or YL1103 mutants. The expressions of the *groEL* genes (*groEL1* and *groEL2*) were also increased in the YL1102 or YL1103 mutants correspondingly. The expression amounts of the *groES* gene were also almost equivalent to the total expressions of *groEL1* and *groEL2* in the mutants. When a second copy of the *groES* gene was inserted in front of *groEL2* in DK1622 (mutant YL1101), forming an artificial complete bicistronic *groESL2* operon, the total expression levels of *groES* in this mutant were still similar to that of DK1622. The results demonstrate that the single *groES* gene is molecularly equivalent for the expression of double *groEL*s in *M. xanthus*. Consistently, the survival rates of the YL1101 mutant had almost no differences from that of the wild-type strain DK1622. We concluded that the expression levels of *groELs* and *groES* genes could be self-regulated in *M. xanthus*, even if they were under different promoters or were distributed in different places in genome. The above results also support that the stand-alone *groEL* gene functions in a *groES*-dependent pattern in *M. xanthus* DK1622.

4.7 Conclusion

Type I chaperonin GroEL commonly exists in bacterial species. GroES, the cofactor of GroEL, commensurably binds with GroEL to carry out the refolding process of protein substrates. In the bacteria containing single *groEL* and *groES* genes, the two synergic genes are neighbors to form a bicistronic *groESL* operon. In bacteria containing duplicated *groEL* genes, stand-alone *groELs* exist broadly. In presently sequenced myxobacterial genomes, the second *groEL* gene stands alone in the *Cystobacterineae* suborder, but normally forms a complete *groESL* operon in the *Sorangineae* suborder. The two *groEL* genes are subneofunctionalized for their cellular functions, showing complex dynamic interplays in *M. xanthus* DK1622. The genetic and biochemical analyses determined that the single *groES* gene is indispensable and the stand-alone *groEL* gene still requires *groES* for its function. The duplicated *groEL* and single *groES* genes expressed and functioned interdependently in a quantum synergy, also showing dynamic interplays between the GroEL and GroES proteins in *M. xanthus* DK1622. The *M. xanthus* cells are able to self-regulate the expressions of *groEL* and *groES* genes to meet the commensurable requirement of these proteins.

The *groES*-dependent stand-alone *groEL* exists in a strange but clever way in the evolutionary process. Dependence of stand-alone *groEL* genes on the allopatrically occurred *groES* is easily understandable for the functioning mechanism of GroEL and its cofactor GroES. We infer that missing the *groES* from the duplicated ancient *groESL2* operon makes the genome to be more succinct, which is probably an important evolutionary mode. However, it is hard to understand how the single *groES* gene expresses to meet the requirement of two allopatric *groEL* genes simultaneously, which requires further investigations.

References

Bittner AN, Foltz A, Oke V (2007) Only one of five groEL genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. J Bacteriol 189:1884–1889

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

Craig EA, Gambill BD, Nelson RJ (1993) Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol Rev 57:402–414

Dworkin M, Kaiser D (1993) Myxobacteria II. American Society for Microbiology, Washington

Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J Bacteriol 171:1379–1385

Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371:614–619

George R, Kelly SM, Price NC, Erbse A, Fisher M et al (2004) Three GroEL homologues from *Rhizobium leguminosarum* have distinct in vitro properties. Biochem Biophys Res Commun 324:822–828

Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS et al (2006) Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci U S A 103:15200–15205

Gould PS, Burgar HR, Lund PA (2007) Homologous cpn60 genes in *Rhizobium leguminosarum* are not functionally equivalent. Cell Stress Chaperones 12:123–131

Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG et al (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. Sci Rep 3:2101

Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. Chem Rev 106:1917–1930

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of *in vivo* substrates of the chaperonin GroEL. Nature 402:147–154

Jiang DM, Zhao L, Zhang CY, Li J, Xia ZJ, Wang J, Wu ZH, Li YZ (2008) Taxonomic analysis of *Sorangium* strains based on HSP60 and 16S rRNA gene sequences and morphology. Int J Syst Evol Microbiol 58:2654–2659

Karlin S, Brocchieri L (2000) Heat shock protein 60 sequence comparisons: duplications, lateral transfer, and mitochondrial evolution. Proc Natl Acad Sci U S A. 97:11348–11353

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC et al (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122:209–220

Kumar CMS, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. Cell Stress Chaperones 20:555–574

Li J, Wang Y, Zhang CY, Zhang WY, Jiang DM et al (2010) *Myxococcus xanthus* viability depends on groEL supplied by either of two genes, but the paralogs have different functions during heat shock, predation, and development. J Bacteriol 192:1875–1881

Liu H, Kovacs E, Lund PA (2009) Characterisation of mutations in GroES that allow GroEL to function as a single ring. FEBS Lett 583:2365–2371

Lund PA (2001) Microbial molecular chaperones. Adv Microb Physiol 44:93–140

Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU (1991) Chaperonin-mediated protein folding at the surface of *groEL* through a 'molten globule'-like intermediate. Nature 352:36–42

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

Ranson NA, White HE, Saibil HR (1998) Chaperonins. Biochem J 333(Pt 2):233–242

Reichenbach H (2004) The myxococcales. In: Garrity GM (ed) Bergey's manual of systematic bacteriology. Springer-Verlag, New York, pp 1059–1143

Saibil HR, Ranson NA (2002) The chaperonin folding machine. Trends Biochem Sci 27:627–632

Shimkets LJ (1990) Social and developmental biology of the myxobacteria. Microbiol Rev 54:473–501

Shimkets L, Dworkin M, Reichenbach H (2006) The myxobacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds) The prokaryotes, vol 7, 3rd edn. Springer, Heidelberg, pp 31–115

VanBogelen RA, Acton MA, Neidhardt FC (1987) Induction of the heat shock regulon does not produce thermotolerance in *Escherichia coli*. Genes Dev 1:525–531

Viitanen PV, Donaldson GK, Lorimer GH, Lubben TH, Gatenby AA (1991) Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. Biochemistry 30:9716–9723

Wang Y, Zhang WY, Zhang Z, Li J, Li ZF, Tan ZG, Zhang TT, Wu ZH, Liu H, Li YZ (2013) Mechanisms involved in the functional divergence of duplicated GroEL chaperonins in *Myxococcus xanthus* DK1622. PLoS Genet 9:e1003306

Wang Y, Li X, Zhang W, Zhou X, Li YZ (2014) The groEL2 gene, but not groEL1, is required for biosynthesis of the secondary metabolite myxovirescin in *Myxococcus xanthus* DK1622. Microbiology 160:488–495

Zhuo L, Wang Y, Zhang Z, Li J, Zhang XH, Li YZ (2017) *Myxococcus xanthus* DK1622 coordinates expressions of the duplicate groEL and single groES genes for synergistic functions of GroELs and GroES. Front Microbiol 8:733

Weimer RM, Creighton C, Stassinopoulos A, Youderian P, Hartzell PL (1998) A chaperone in the HSP70 family controls production of extracellular fibrils in *Myxococcus xanthus*. J Bacteriol 180:5357–5368

Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, Saibil HR, Fenton WA, Horwich AL (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. Cell 83:577–587

Wilson AC, Wu CC, Yates JR 3rd, Tan M (2005) Chlamydial GroEL autoregulates its own expression through direct interactions with the HrcA repressor protein. J Bacteriol 187:7535–7542

Yang Z, Geng Y, Shi W (1998) A DnaK homolog in *Myxococcus xanthus* is involved in social motility and fruiting body formation. J Bacteriol 180:218–224

Chapter 5
Functional Diversity in Mycobacterial Chaperonins: The Generalists and the Specialists

Shekhar C. Mande and C. M. Santosh Kumar

Abstract Chaperonins are a class of molecular chaperones that form cylindrical assemblies and for sequestering the non-native protein and thereby assisting their folding. Genomic annotation studies have identified multiple copies of chaperonin genes in about 30% of the bacteria. *Mycobacterium tuberculosis* was the first organism where two copies were observed, and similarly other mycobacteria bear two and rarely three copies. Owing to the pathogenic lifestyle, the chaperonins of mycobacteria have been demonstrated to be secretory and possess antigenic properties. Furthermore, biochemical and structural studies have demonstrated that these chaperonins are unusual. One of the chaperonins that exists in operonic arrangement with the co-chaperonin gene has been shown to be required in several pathogenic stages of the bacteria. The other copy that exists independently is essential and thus might be functioning as a general chaperone. Several groups have worked to unravel the functional diversity of these mysterious molecules employing structural, immunochemical, cell-biological, computational and genetic tools. We review the current understanding on the mycobacterial chaperonins and the new paradigms that have arisen.

5.1 Introduction

The genus *Mycobacterium* belongs to the phylum *Actinobacteria* that is characterized by high-G+C content bacteria and includes several pathogenic species such as *Mycobacterium tuberculosis*, *M. leprae*, *M. bovis*, *M. avium*, *M. marinum*, etc., that cause tubercular diseases in several vertebrate hosts. *Mycobacterium*, the genus where multiple copies of chaperonins were first observed, in the genome of *M.*

S.C. Mande (✉)
National Centre for Cell Science, NCCS Complex, Ganeshkhind, Pune 411 007, India
e-mail: shekhar@nccs.res.in

C. M. Santosh Kumar
School of Biosciences & Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_5

tuberculosis (Kong et al. 1993), have two (rarely three) copies of the chaperonin genes (Kumar et al. 2015). Although several mycobacteria have multiple chaperonin genes, most work has been done on the *M. tuberculosis* chaperonins, principally because they are antigenic (Henderson et al. 2013; Lewthwaite et al. 2001) and that they were discovered first. Therefore, we limit our discussion in this chapter to *M. tuberculosis* chaperonins GroEL1 and GroEL2 and compare their properties with *Escherichia. coli* chaperonin GroEL.

Chaperonins are a class of molecular chaperones that typically form large double-toroid assemblies, encapsulate the substrate proteins and thereby favour their folding (Hayer-Hartl et al. 1995; Horwich et al. 2006; Weber et al. 1998). Insights into the biochemical and physiological functions of chaperonins have been obtained principally from studies carried out on *E. coli* chaperonin, GroEL (Sigler et al. 1998; Hartl 1996). *E. coli* genome possesses one copy each of *groEL* and *groES* that form the *groE* operon and encode the chaperonin GroEL and co-chaperonin GroES, respectively. These two genes are essential for *E. coli* under all conditions known (Fayet et al. 1989). Although the current understanding on the mechanism of action of chaperonins emerged from the studies on GroEL, the presence of multiple copies of the genes encoding chaperonins in about 30% of the newly sequenced bacterial genomes (Kumar et al. 2015) has perplexed the understanding of chaperonin function. Since GroEL does not exhibit any specificity to amino acid sequences in their substrates, the occurrence of multiple copies has been perplexing. It has been hypothesized that the multiple copies either might be required for distinct specificities to certain class of substrates or might be differentially regulated (Kumar et al. 2015; Lund 2009; Kumar 2017). Another interesting hypothesis has been proposed that there might be functional divergence in the duplicated copies of the GroELs (Kumar and Mande 2011). Thus, the occurrence of multiple copies of chaperonins in different prokaryotes and addressing their functions has been an interesting aspect of chaperonin research.

5.2 Diversity in Mycobacterial Chaperonins: Sequence Features

Among the multiple copies of mycobacteria, one of the copies is found in the canonical operonic arrangement with the *groES* gene, while the second and the third copies exist separately. Therefore, as per the convention, the chaperonin gene in operonic arrangement is termed *cpn60.1* or *groEL1*, and the copies that exist separately are termed *cpn60.2* and *cpn60.3* or *groEL2* and *groEL3*, respectively. Interestingly, GroEL2, the copy that exists separately, is essential for survival, while GroEL1 is dispensable (Stewart et al. 2002).

Regulation of expression of the chaperonin genes has been investigated with intriguing results. The stress-inducible promoter regulating expression of *groES/L1* operon has been demonstrated to be stronger than that of *groEL2* (Aravindhan et al. 2009). In addition to the regular positive regulation by SigH and negative regulation by CIRCE-binding HrcA, several factors regulate the expression of these genes. Although all the

chaperonin genes are induced maximally by heat shock, they are induced during several other stresses, such as hyperosmolarity, starvation and oxidative stress (Hu et al. 2009; Rao and Lund 2010). Expression of *groEL2* is repressed by WhiB, a nitric oxide-responsive transcription factor, which occludes *groEL2* promoter from the activating CRP-family transcription factor, Cmr (Stapleton et al. 2012). Surprisingly, the two genes, *groES* and *groEL1* in the bicistronic *groES/EL1* operon, were shown to be regulated independently (Kong et al. 1993). Interestingly, although all the chaperonins are known as antigens, none of their genes were induced when *M. tuberculosis* was grown inside the phagosomes (Schnappinger et al. 2003). The situation in *M. leprae*, the mycobacteria that lacks heat-shock response, is complicated since the organism is depleted of a functional SigE but has a functional HrcA that is predicted to regulate the chaperonin genes (Williams et al. 2007), indicating that the heat-shock response in this organism requires comprehensive investigation using genetic and biochemical tools. These studies, therefore, indicate that different mechanisms of regulation, in addition to the heat shock, operate in inducing the expression of chaperonin genes in mycobacteria, suggesting that the roles played by the two chaperonins might be more than the typical heat-shock proteins that are required either temporally in the life of the pathogen.

Evolutionary studies towards the understanding of the rate of divergence and evolvability have observed that mycobacterial chaperonin paralogues have resulted due to a gene duplication event (Mande et al. 2013) at the last common ancestor of the actinobacterial phylum (Kumar et al. 2015) and followed different rates of evolution thereafter (Goyal et al. 2006; Hughes 1993), probably due to differential interactions with the human proteins (Lund 2009). Interestingly, a third chaperonin gene has been observed in certain mycobacteria with a faster doubling time (García-Agudo and García-Martos 2011), such as *Mycobacterium chubuence*, *M. smegmatis* and *M. goodie* (Fig. 1), suggesting that the third chaperonin may be functioning towards enhancing the growth rate. Therefore, the chaperonin paralogues in mycobacteria indicated a functional diversity and have opened up the doors for understanding their functions.

An interesting aspect of the paralogous copies of chaperonins is related to their carboxy-terminal sequences (CTS). Chaperonin sequences are characterized by Gly-Met-rich CTS. Although the function of the CTS is not completely defined, since the N- and C-terminal tails of GroEL protrude into the cavity of the tetradecamer, and that they have been demonstrated to interact with the substrates in cryo-EM studies (Chen et al. 2013; Weaver and Rye 2014), the CTS have been proposed to play a significant role in substrate folding, either by defining the size of the substrates that can be encapsulated (Tang et al. 2006) or by perturbing the ATPase activity and thereby providing the substrates with a longer sequestration time (Farr et al. 2007). Since CTS has not been resolved in any of the GroEL crystal structures, and owing to its Gly-rich nature, it is assumed to be highly disordered in the central cavity of GroEL. Interestingly, the diversity in the mycobacterial chaperonins is principally defined by their CTS (Fig. 5.1). While GroEL2 homologues possess the canonical Gly-Met-rich CTS, GroEL1 homologues surprisingly have a His-rich CTS (Kong et al. 1993). Notably, GroEL3 homologues, on the other hand, possess a patternless CTS. Since the CTS is believed to be important in chaperonin function, their diversity is indicative of functional diversity in mycobacterial chaperonins. Moreover, since the CTS is demonstrated to determine the substrate pool (Tang et al. 2006), it is presumed that the three

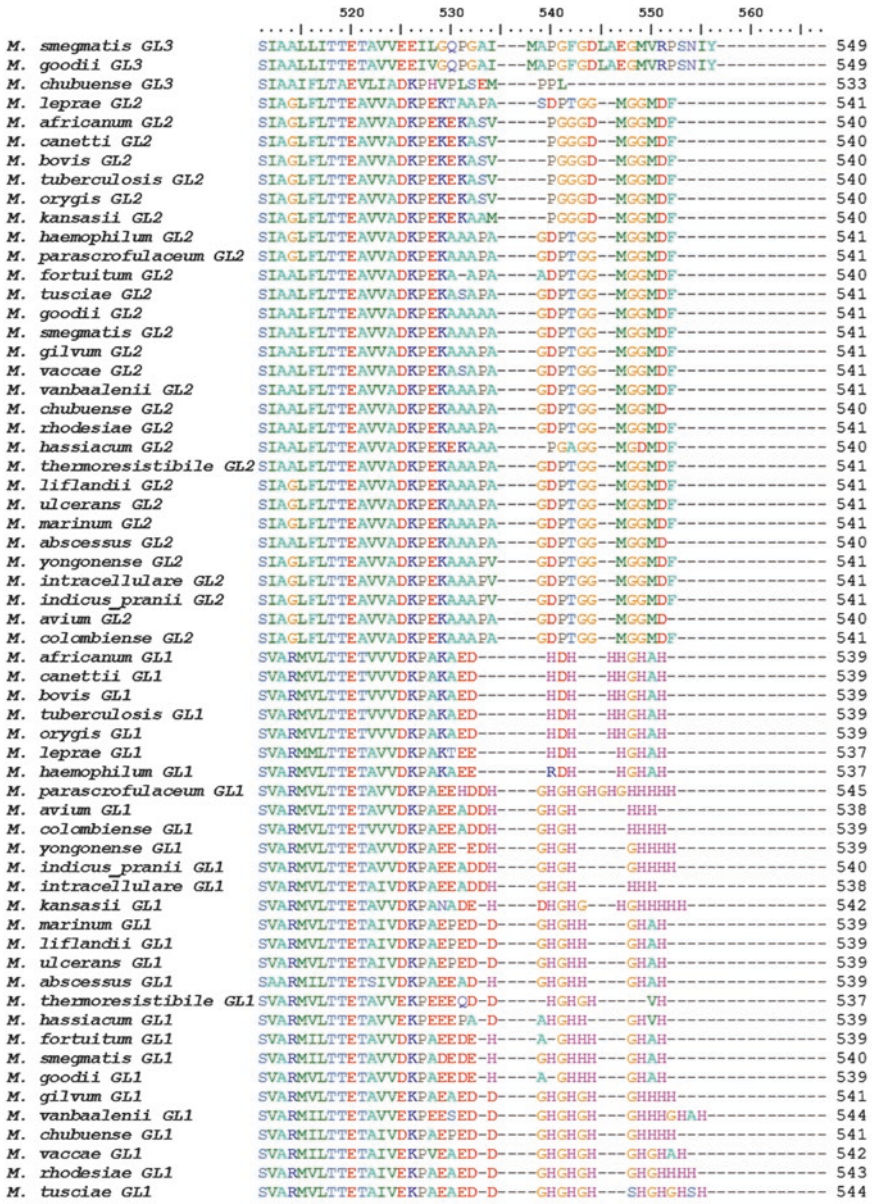


Fig. 5.1 Mycobacterial chaperonins principally differ at their carboxy termini. Sequence alignment showing the carboxy-terminal regions of the indicated mycobacterial GroEL paralogues with the *E. coli* GroEL. Polypeptide sequences of the GroEL paralogues were retrieved from www.uniprot.org, aligned using MUSCLE, and the alignment was formatted using BioEdit

different CTS in mycobacterial chaperonin paralogues might define the respective chaperonin pools. For example, *M. smegmatis groEL1* knockout strain when supplemented with a GroEL1 variant that is devoid of its His-rich C-terminal sequence has been shown to be defective in biofilm formation (Ojha et al. 2005) lending credence to the hypotheses on the connection between chaperonin function and their CTS. Thus, the variety in CTS of mycobacterial GroELs appears to influence their function.

5.3 Structural Investigations on Mycobacterial Chaperonins

One of the intriguing features of mycobacterial chaperonins has been their quaternary structure. Structural and functional studies on GroEL demonstrated that it requires to exist as a tetradecamer to perform the chaperonin function (Braig et al. 1995; Xu et al. 1997). GroEL forms two large cavities for the encapsulation of the substrates, and one of the cavities is capped by the heptameric GroES. Surprisingly, recombinant versions of *M. tuberculosis* GroEL1 and GroEL2, when purified from *E. coli*, existed as lower oligomers (Fig. 5.2), and their assembling into tetradecamers required phosphorylation for GroEL1 (Kumar et al. 2009) and abundance for GroEL2 (Fan et al. 2012). These observations are suggestive of regulated oligomeric assembly of *M. tuberculosis* chaperonins.

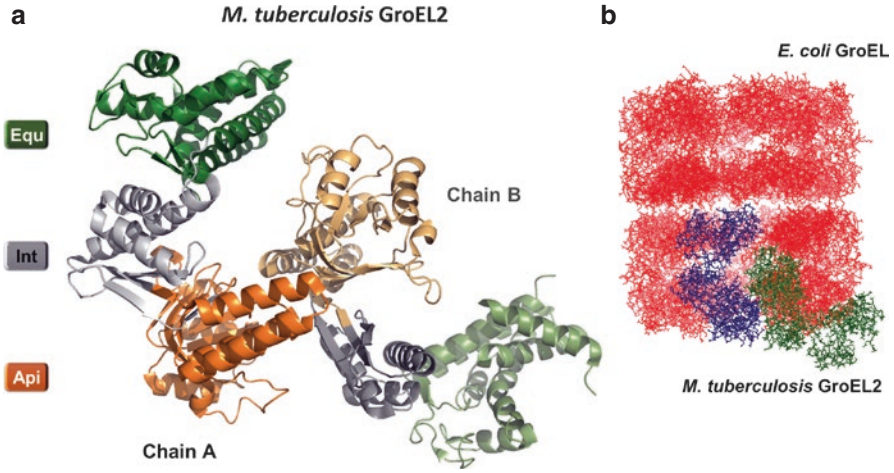


Fig. 5.2 Unusual oligomeric nature of mycobacterial chaperonins. (**a**) Cartoon representation showing the dimeric form of *M. tuberculosis* GroEL2. Interestingly, inter-subunit interactions are mediated via the apical domains, unlike by the equatorial domains in *E. coli* GroEL. Api, Int and Equ denote the apical, intermediate and equatorial domains. (**b**) Structural alignment of the tetradecameric *E. coli* GroEL (shown in *red*) with the dimeric *M. tuberculosis* GroEL2 (two subunits shown in *green* and *blue*). Differences in the inter-subunit interactions, between the two GroEL homologues, are apparent. Coordinates for the dimeric GroEL2 and tetradecameric GroEL structures were sourced from the PDB IDs, 1RTK and 1OEL, respectively

5.3.1 *Structural Studies on* M. tuberculosis *GroEL1*

Structural studies on the GroEL1 homologues did not meet success except for the apical domain (Sielaff et al. 2010). One of the reasons why GroEL1 could not crystallize despite several crystallization trials can be attributed to its conformationally heterogeneous lower oligomeric forms. Therefore, understanding the structural features of the individual domains has been attempted with a range of success. Except the apical domain that runs through a continuous polypeptide, neither the equatorial nor the intermediate domain, each of which are made of two discontinuous segments of the polypeptide, could be crystallized. The crystal structure of the substrate-binding apical domain of GroEL1 shows a high similarity with that of *E. coli* GroEL and *M. tuberculosis* GroEL2 (Fig. 5.3) indicating that GroEL1 might follow the

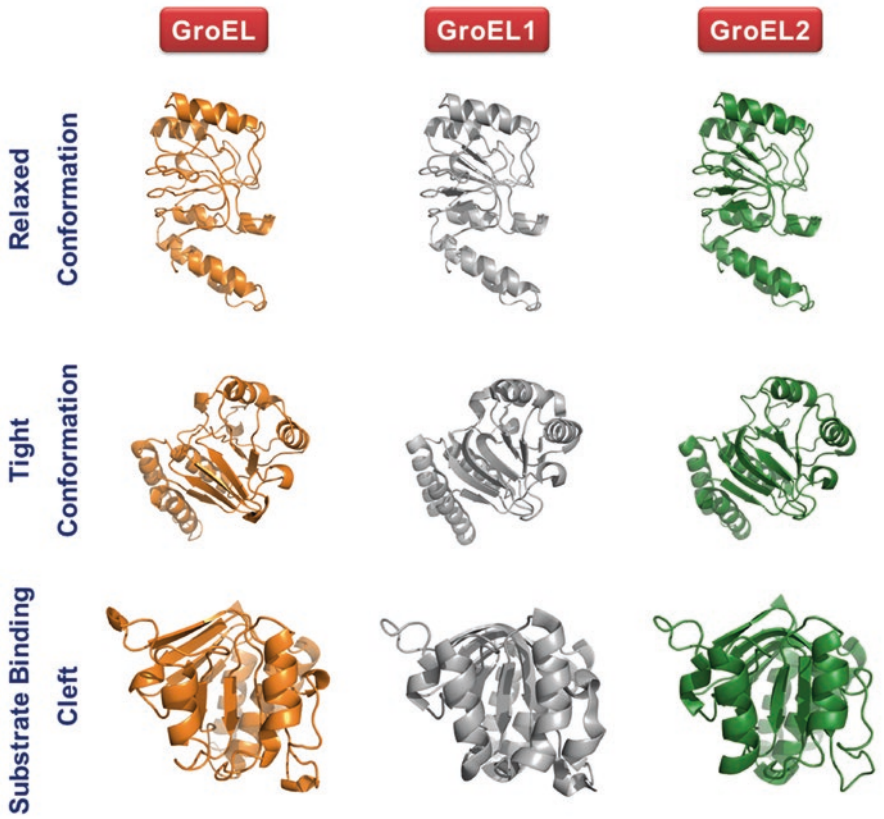


Fig. 5.3 Substrate-binding clefts in mycobacterial chaperonins are conserved. Structural comparison of the apical domains and the substrate-binding clefts of the GroEL homologues. Coordinates for *E. coli* GroEL (1AON), *M. tuberculosis* GroEL1 (3M6C) and GroEL2 (1SJP) were obtained from PDB. Coordinates for the relaxed states of *M. tuberculosis* GroEL1 and GroEL2 were modelled using Modeller 9.13

conserved mode of substrate interaction. Drawing parallels with the mini-chaperone concept of GroEL, owing to the structural similarities and the ability to interact with peptides from a model substrate KasA, GroEL1 has been proposed to function as a molecular chaperone (Sielaff et al. 2011). This, in connection with the demonstration that GroEL1 exists as a tetradecamer in *M. tuberculosis* lysates, indicates that GroEL1 would function as a chaperonin.

5.3.2 Structural Studies on M. tuberculosis *GroEL2*

Unlike GroEL1 whose structural information is limited, GroEL2 of *M. tuberculosis* has been crystallized successfully by different groups, and the structural information revealed several interesting features. The first structure determined at 3.2 Å resolution (Qamra and Mande 2004) and the second structure determined at 2.8 Å resolution independently by another group (Shahar et al. 2011) presented a dimeric molecule with N-terminal truncations. Although the tertiary structure of GroEL2 is similar to that of *E. coli* GroEL (Xu et al. 1997), the quaternary structure is quite different. Principally, the mode of dimerization of GroEL2 appears to be different than that of *E. coli* GroEL (Fig. 5.2). Surprisingly, in the two GroEL2 structures, the apical domain appears to be involved in the dimerization unlike in *E. coli* GroEL where the equatorial domain plays a dominant role in mediating the inter-subunit contacts (Fig. 5.2). Furthermore, the two subunits in GroEL2 are in the tight (*T*) conformation, suggesting that this oligomer might not function as a chaperonin. Since both structures of GroEL2 depict the *T* state, homology models were designed to represent relaxed state, and the conformational transitions between the two states were examined using normal mode analyses (Chilukoti et al. 2016). Normal mode projections of GroEL2 models revealed that inter-domain hinges are very critical for mediating the transition and chaperonin action (Chilukoti et al. 2016). Taken together, structural studies revealed that although a canonical tetradecameric assembly is not observed, with the similarity in the characteristic tertiary structure, conformational transitions and the mode of substrate interaction, mycobacterial GroELs might function as chaperonins in vivo.

5.4 Mycobacterial Chaperonins Are Functionally Diverse

Functional investigations on mycobacterial chaperonins from several groups have highlighted the functional diversity of these chaperonins with several moonlighting functions and unusual biochemical and structural features that could contribute the diverse roles that these proteins appear to play. In the following sections, we review the current understanding on the functional diversity of mycobacterial chaperonins.

5.4.1 Mycobacterial Chaperonins Function as Antigens

Since the mycobacterial chaperonins were identified as prominent proteins of culture filtrates, initial investigations towards understanding their biological function focused on identifying their immunological roles (Henderson et al. 2013; Lamb et al. 1989). These studies consequently demonstrated that the chaperonins GroEL1 (Lewthwaite et al. 2001) and GroEL2 (Hickey et al. 2010) and the co-chaperonin GroES (Sonnenberg and Belisle 1997) are secreted and immunogenic, with GroEL1 being the most immunopotent (Lewthwaite et al. 2001). The polypeptide segment that is responsible for this moonlighting function has been identified in the equatorial domain of GroEL1 (Hu et al. 2013). Moreover, in a quest to identify the predominant intra-macrophage *M. tuberculosis* proteins, both GroEL1 (Chande et al. 2015) and GroEL2 (Kruh-Garcia et al. 2014) were found to be enriched. Interestingly, enrichment of GroEL1 from the virulent *M. tuberculosis* strain persisted for longer periods (Chande et al. 2015) and thus led to the proposal that the GroEL1, as a chaperone that suppressed protein aggregation (Qamra et al. 2004), might suppress ER stress response in macrophages and consequently prevent the activation of apoptosis. Interestingly, mycobacterial chaperonins have been isolated from the nucleus of infected human cells, where the chaperonins accumulate in time-dependent manner (Agrawal et al. 2016), suggesting moonlighting functions for these chaperonins and pressing the need for further investigations to unravel their multitude roles.

5.4.2 GroEL1 Works as a Specialized Chaperonin for Folding Pathogenic Proteins

The biochemical and structural studies were initiated around the same time. Biochemical studies on the recombinant mycobacterial GroELs revealed significant deviations in several characteristic chaperonin features. The first study that GroELs of *M. tuberculosis* exist in lower oligomeric forms was reported in 2004 (Qamra et al. 2004), although the co-chaperonin GroES has been demonstrated to readily assemble into a heptamer (Taneja and Mande 2001). Since any deviation from the tetradecameric structure would seriously hamper chaperonin function, this intriguing observation triggered a series of investigations towards understanding the biochemical features of these chaperonins. Interestingly, both chaperonins in their lower oligomeric form were shown to suppress aggregation of substrate proteins, but lacked the required ATPase activity and consequently failed to refold model substrates (Qamra et al. 2004). Although both chaperonins exhibited similar biochemical properties, the differences were observed in their oligomerization propensities. While GroEL1 required a posttranslational modification to attain the tetradecameric status (Kumar et al. 2009), GroEL2 required either enhanced expression or chemical collocation (Fan et al. 2012). Although these studies employed purified proteins to convincingly demonstrate that GroEL1 fails to act as chaperone,

further genetic studies demonstrated that GroEL1 can regain chaperonin activity upon facilitated oligomerization (Henderson et al. 2013; Kumar and Mande 2011; Kumar et al. 2009). Directed evolution studies on *M. tuberculosis* GroEL1 employing gene shuffling and domain swapping followed by the functional studies of the resulted GroEL1 variants in different *E. coli groEL* mutant strains established that impaired oligomerization is the basis for the weakened chaperonin activity of GroEL1. Furthermore, employing immunochemical and mass spectrometric techniques on fractionated *M. tuberculosis* cellular extracts, tetradecameric form of GroEL of *M. tuberculosis* has been demonstrated to be phosphorylated at a serine residue, whereas the lower oligomeric forms were not (Kumar et al. 2009). This observation has led to the proposal that *M. tuberculosis* GroEL1 follows regulated oligomerization mediated by seryl-phosphorylation to regain activity (Fig. 5.4) and thereby conserve ATP pools in nutrient-deprived and slow-growing mycobacteria by limiting GroEL1 activity (Kumar and Mande 2011). The model proposed based on these observations was that different oligomeric forms exist in the cells—the tetradecamers function as chaperonins, while the lower oligomers might be used as chaperonin stocks that assemble into tetradecamer upon phosphorylation as per the cellular requirement (Fig. 5.4).

The necessity of an extra copy of GroEL and the presence of noncanonical CTS remained mysterious till the discovery that mycobacterial GroEL1 homologues contribute to the formation of biofilms (Ojha et al. 2005) and granulomas (Hu et al. 2008), probably by assisting the folding of proteins involved in membrane biogenesis—such as KasA—or pathogenesis (Sielaff et al. 2011). Interestingly, chaperonins in pathogenic bacteria such as *Bacillus anthracis* (Arora et al. 2017) and *Leptospira interrogans* (Vinod Kumar et al. 2016) have been implicated in the formation of biofilms. These observations, therefore,

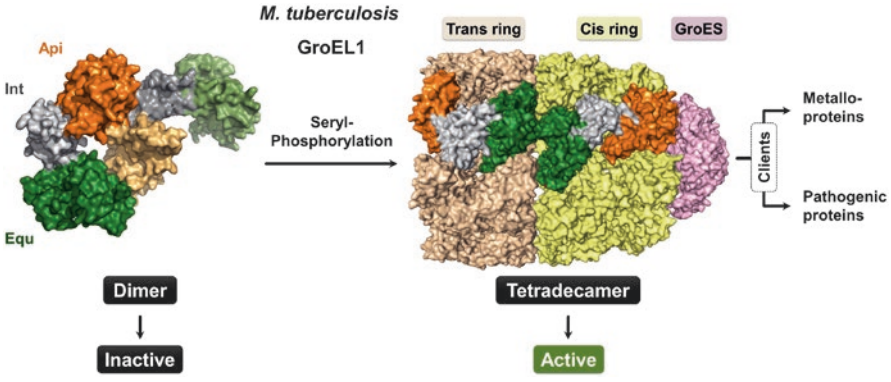


Fig. 5.4 Model for the functional adaptation in *M. tuberculosis* GroEL1. Phosphorylation on serine residues enables *M. tuberculosis* GroEL1 to assume functional double-ringed (*cis* and *trans* rings) tetradecameric form from an inactive dimeric form. Active tetradecamer, owing to its peculiar histidine-rich carboxy terminus, assists the folding of special classes of client proteins, such as the metalloproteins and pathogenicity-related proteins. Api, Int and Equ represent the apical, intermediate and equatorial domains. GroES acts as a cap for the encapsulated cavity

reinforced GroEL1's functional divergence via the involvement in disease establishment or progression and suggested that the two chaperonins of mycobacteria encounter different substrate pools. Moreover, unusual nature of GroEL1 led to the investigations towards identifying other moonlighting functions (Mande et al. 2013) such as the DNA-binding ability of the recombinant dimeric form of GroEL1 (Basu et al. 2009). The observation that *groEL1* knockout of *M. tuberculosis* exhibited growth defects under low aeration condition (Sharma et al. 2016) led to the proposal that GroEL1, owing to His-rich CTS, might be involved in oxygen sensing via assisting the folding of certain metalloproteins (Fig. 5.4). These observations, along with the demonstration that GroEL1 is required for the formation of biofilms and granulomas, faster rate of evolution, higher immunogenicity and the presence of mysterious His-rich CTS, tempt us to speculate that GroEL1 might have evolved to function as a specialized chaperonin to assist the folding of certain classes of proteins, such as the pathogenicity related or metalloproteins (Kumar et al. 2015).

5.4.3 GroEL2 Functions as a Generalist Chaperonin

While the above studies functionally characterized GroEL1 as an ineffective chaperonin, GroEL2 has been shown to functionally replace *E. coli* GroEL in vivo and in vitro upon sufficient accumulation (Chilukoti et al. 2016; Fan et al. 2012; Hu et al. 2008). Since GroEL2 is active as a chaperone under certain conditions (Chilukoti et al. 2016; Fan et al. 2012), domain swapping experiments on GroEL2 were performed to understand the significance of chaperonin hinges, which connect different domains, in the conformational transitions. Functional assays on the resulted variants using genetic and structural tools revealed that hinge 2 that connects the intermediate domain to the apical domain and thereby mediates the large en bloc movements of the apical domain is highly conserved and thus resists any change, while hinge 1 that connects the intermediate domain to the equatorial domain and thus is required to mediate smaller movements can accommodate variations (Fig. 5.5). This study, therefore, established that in addition to the formation of tetradecameric assembly, proper inter-domain communication and precise conformational transitions are essential for chaperonin activity (Chilukoti et al. 2016). Taking together, these observations suggested that the determinants of oligomerization of GroEL1 and GroEL2 are distinct from each other and from *E. coli* GroEL, suggesting that each chaperonin might get into the active form according to the cellular need. Furthermore, in a quest to map the molecular mechanism of secretion of GroEL2, it was demonstrated that GroEL2 localizes at the cell wall (Hickey et al. 2010) in the form of multimers of about 245 kDa in size, which upon cleavage by Hip1 serine protease separate into the secretion-competent monomers. This Hip1-mediated cleavage of GroEL2 has been demonstrated to pacify macrophage response (Naffin-Olivos et al. 2014) and thereby aid *M. tuberculosis* in evading human immunological responses. Interestingly, the site of cleavage in GroEL2 by Hip1 is between

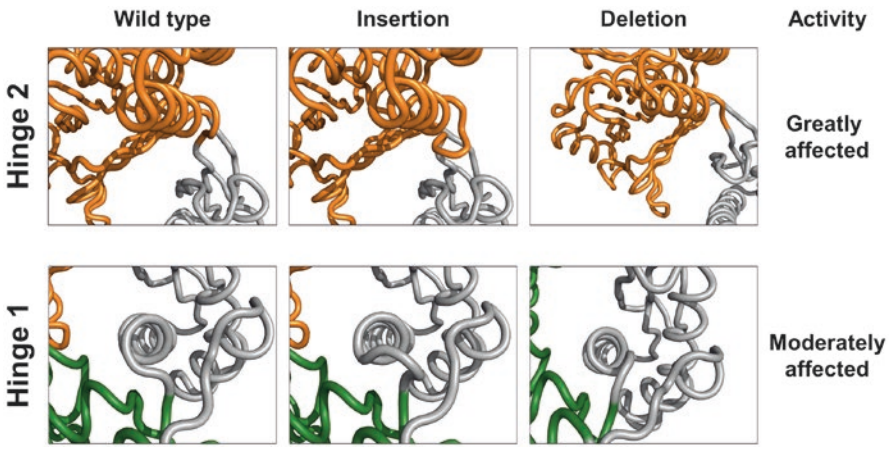


Fig. 5.5 Effects of hinge variations on chaperonin activity. Cartoons representing the indels at the hinges in GroEL2 homology models. The effects of the hinge indels are compared to those of the wild type as represented. *Black arrows* point to the predicted structural variations at the hinges resulting from the indels. Variations at hinge 2 have been shown to greatly affect the activity. Reproduced from Chilukoti et al. (2016), with permission

the 12th and 13th residue and thus is different from those observed in the dimeric crystal structures (Qamra and Mande 2004; Shahar et al. 2011) indicating that the crystal structures do not represent the active form and that GroEL2 might have a different active oligomeric form in mycobacterial milieu. These observations along with the demonstration that GroEL2 is essential for survival (Stewart et al. 2002) and the presence of canonical CTS (Fig. 5.1) have strongly supported the notion that GroEL2 could function as a generalist chaperonin (Kumar et al. 2015).

5.5 Conclusions

The mycobacterial chaperonins have served as models for exploring the functional diversity in multiple chaperonins. One of the two copies, the GroEL1, evolved faster to function as a specialized chaperonin during pathogenic phases of the bacterium, while the other copy, GroEL2, remained as a generalist chaperonin and thus is essential for survival. Unusually, these chaperonins are secreted into the culture and into the organelles of host cells where they modulate the cellular environment to favour the pathogen. Structural and biochemical studies have aided in understanding the biology of these chaperonins. Notably, the discovery that *M. tuberculosis groEL1* knockout fails to form granulomas has opened up several avenues for active research to unravel the mechanism by which GroEL1 influences the granuloma formation and use GroEL1 as a drug target. Interestingly, it remains to be investigated if GroEL1's chaperone function or antigenic function is responsible for the granuloma formation. Taken together, mycobacterial GroELs present a fantastic picture

of functional diversity among chaperonins. However, a comprehensive investigation employing structural, biochemical and cell-biological tools is required to understand their moonlighting functions and role in the disease progression and cure.

Acknowledgements Santosh is Newton International Fellow at the University of Birmingham, UK, sponsored by The Royal Society, The British Academy and the Academy of Medical Sciences, UK. Further, we wish to acknowledge the support of Department of Biotechnology, India.

References

Agrawal AK, Ranjan R, Chandra S, Rout TK, Misra A, Reddy TJ (2016) Some proteins of M. tuberculosis that localise to the nucleus of THP-1-derived macrophages. Tuberculosis (Edinb) 101:75–78

Aravindhan V, Christy AJ, Roy S, Ajitkumar P, Narayanan PR, Narayanan S (2009) Mycobacterium tuberculosis groE promoter controls the expression of the bicistronic groESL1 operon and shows differential regulation under stress conditions. FEMS Microbiol Lett 292(1):42–49

Arora G, Sajid A, Virmani R, Singhal A, Kumar CM, Dhasmana N, Khanna T, Maji A, Misra R, Molle V, Becher D, Gerth U, Mande SC, Singh Y (2017) Ser/Thr protein kinase PrkC-mediated regulation of GroEL is critical for biofilm formation in Bacillus anthracis. NPJ Biofilms Microbiomes 7:1–12. doi:10.1038/s41522-017-0015-4

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(15):4944–4954

Braig K, Adams PD, Brunger AT (1995) Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. Nat Struct Biol 2(12):1083–1094

Chande AG, Siddiqui Z, Midha MK, Sirohi V, Ravichandran S, Rao KV (2015) Selective enrichment of mycobacterial proteins from infected host macrophages. Sci Rep 5:13430

Chen DH, Madan D, Weaver J, Lin Z, Schroder GF, Chiu W, Rye HS (2013) Visualizing GroEL/ES in the act of encapsulating a folding protein. Cell 153(6):1354–1365

Chilukoti N, Kumar CM, Mande SC (2016) GroEL2 of Mycobacterium tuberculosis reveals the importance of structural pliability in chaperonin function. J Bacteriol 198(3):486–497

Fan M, Rao T, Zacco E, Ahmed MT, Shukla A, Ojha A, Freeke J, Robinson CV, Benesch JL, Lund PA (2012) The unusual mycobacterial chaperonins: evidence for in vivo oligomerization and specialization of function. Mol Microbiol 85(5):934–944

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(13):5342–5347

Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J Bacteriol 171(3):1379–1385

García-Agudo L, García-Martos P (2011) Clinical significance and antimicrobial susceptibility of rapidly growing mycobacteria. In: Memdez-Vilas A (ed) Science against microbial pathogens: communicating current research and technological advances. Formatex Research Centre, Badajoz, pp 363–377

Goyal K, Qamra R, Mande SC (2006) Multiple gene duplication and rapid evolution in the groEL gene: functional implications. J Mol Evol 63(6):781–787

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571–579. doi:10.1038/381571a0

Hayer-Hartl MK, Martin J, Hartl FU (1995) Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. Science 269(5225):836–841

Henderson B, Fares MA, Lund PA (2013) Chaperonin 60: a paradoxical, evolutionarily conserved protein family with multiple moonlighting functions. Biol Rev Camb Philos Soc 88(4): 955–987

Hickey TB, Ziltener HJ, Speert DP, Stokes RW (2010) Mycobacterium tuberculosis employs Cpn60.2 as an adhesin that binds CD43 on the macrophage surface. Cell Microbiol 12(11): 1634–1647

Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. Chem Rev 106(5):1917–1930

Hu Y, Coates AR, Liu A, Lund PA, Henderson B (2013) Identification of the monocyte activating motif in Mycobacterium tuberculosis chaperonin 60.1. Tuberculosis (Edinb) 93(4): 442–447

Hu Y, Henderson B, Lund PA, Tormay P, Ahmed MT, Gurcha SS, Besra GS, Coates AR (2008) A mycobacterium tuberculosis mutant lacking the GroEL homologue cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. Infect Immun 76(4): 1535–1546. doi:10.1128/IAI.01078-07

Hughes AL (1993) Contrasting evolutionary rates in the duplicate chaperonin genes of Mycobacterium tuberculosis and M. leprae. Mol Biol Evol 10(6):1343–1359

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(7):2608–2612

Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, Davis JL, Dobos KM (2014) Detection of Mycobacterium tuberculosis peptides in the exosomes of patients with active and latent M. tuberculosis infection using MRM-MS. PLoS One 9(7):e103811

Kumar CM (2017) Prokaryotic multiple chaperonins: the mediators of functional and evolutionary diversity. Chapter 5. In: Santosh Kumar CM, Mande SC (eds) Prokaryotic chaperonins. Springer, Heidelberg

Kumar C, Mande S (2011) Protein chaperones and non-protein substrates: on substrate promiscuity of GroEL. Curr Sci 100:1646–1653

Kumar CM, Khare G, Srikanth CV, Tyagi AK, Sardesai AA, Mande SC (2009) Facilitated oligomerization of mycobacterial GroEL: evidence for phosphorylation-mediated oligomerization. J Bacteriol 191(21):6525–6538

Kumar CM, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria – novel functions and non-canonical behaviors. Cell Stress Chaperones 20(4):555–574

Lamb JR, Bal V, Rothbard JB, Mehlert A, Mendez-Samperio P, Young DB (1989) The mycobacterial GroEL stress protein: a common target of T-cell recognition in infection and autoimmunity. J Autoimmun 2(Suppl 1):93–100

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(12):7349–7355

Lund PA (2009) Multiple chaperonins in bacteria – why so many? FEMS Microbiol Rev 33(4):785–800

Mande S, Kumar C, Sharma A (2013) Evolution of bacterial Chaperonin 60 paralogues and moonlighting activity. In: Henderson B (ed) Moonlighting cell stress proteins in microbial infections. Springer Netherlands, Netherlands, pp 101–121

Naffin-Olivos JL, Georgieva M, Goldfarb N, Madan-Lala R, Dong L, Bizzell E, Valinetz E, Brandt GS, Yu S, Shabashvili DE, Ringe D, Dunn BM, Petsko GA, Rengarajan J (2014) Mycobacterium tuberculosis Hip1 modulates macrophage responses through proteolysis of GroEL2. PLoS Pathog 10(5):e1004132

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(5):861–873

Qamra R, Mande SC (2004) Crystal structure of the 65-kilodalton heat shock protein, chaperonin 60.2, of Mycobacterium tuberculosis. J Bacteriol 186(23):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(2):605–617

Rao T, Lund PA (2010) Differential expression of the multiple chaperonins of mycobacterium smegmatis. FEMS Microbiol Lett 310:24–31. doi:10.1111/j.1574-6968.2010.02039.x

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J Exp Med 198(5):693–704

Shahar A, Melamed-Frank M, Kashi Y, Shimon L, Adir N (2011) The dimeric structure of the Cpn60.2 chaperonin of Mycobacterium tuberculosis at 2.8 Å reveals possible modes of function. J Mol Biol 412(2):192–203

Sharma A, Rustad T, Mahajan G, Kumar A, Rao KV, Banerjee S, Sherman DR, Mande SC (2016) Towards understanding the biological function of the unusual chaperonin Cpn60.1 (GroEL1) of Mycobacterium tuberculosis. Tuberculosis (Edinb) 97:137–146

Sielaff B, Lee KS, Tsai FT (2010) Crystallization and preliminary X-ray crystallographic analysis of a GroEL1 fragment from Mycobacterium tuberculosis H37Rv. Acta Crystallogr Sect F Struct Biol Cryst Commun 66(Pt 4):418–420

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(3):831–839

Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL (1998) Structure and function in GroEL-mediated protein folding. Annu Rev Biochem 67:581–608. doi:10.1146/annurev.biochem.67.1.581

Sonnenberg MG, Belisle JT (1997) Definition of Mycobacterium tuberculosis culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun 65(11):4515–4524

Stapleton MR, Smith LJ, Hunt DM, Buxton RS, Green J (2012) Mycobacterium tuberculosis WhiB1 represses transcription of the essential chaperonin GroEL2. Tuberculosis (Edinb) 92(4):328–332

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(Pt 10):3129–3138

Taneja B, Mande SC (2001) Three-dimensional structure of Mycobacterium Tuberculosis chaperonin-10 reveals a partially stable conformation for its mobile loop. Curr Sci 81:87–91

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(5):903–914

Vinod Kumar K, Lall C, Vimal Raj R, Vedhagiri K, Kartick C, Surya P, Natarajaseenivasan K, Vijayachari P (2016) Overexpression of heat shock GroEL stress protein in leptospiral biofilm. Microb Pathog 102:8–11

Weaver J, Rye HS (2014) The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. J Biol Chem 289(33):23219–23232

Weber F, Keppel F, Georgopoulos C, Hayer-Hartl MK, Hartl FU (1998) The oligomeric structure of GroEL/GroES is required for biologically significant chaperonin function in protein folding. Nat Struct Biol 5(11):977–985

Williams DL, Pittman TL, Deshotel M, Oby-Robinson S, Smith I, Husson R (2007) Molecular basis of the defective heat stress response in Mycobacterium Leprae. J Bacteriol 189(24):8818–8827

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 388(6644):741–750

Chapter 6
Multiple Chaperonins and Their Potential Roles in Rhizobia

Peter A. Lund

Abstract The Rhizobia are a paraphyletic group of gram-negative bacteria which are of enormous importance to agriculture, due to their ability to form nitrogen-fixing nodules on the roots of legumes. These organisms have a remarkable tendency to possess multiple copies of chaperonin genes, with up to seven copies being recorded—a record as far as bacterial genomes are concerned. The regulation of these multiple chaperonin genes is complex, suggesting that they may have roles in addition to protection from stresses that induce protein unfolding, and the proteins that they encode show little tendency to form mixed oligomers, again suggesting at least some specificity of function. Closely related bacteria that lack the ability to fix nitrogen often do not show this high chaperonin gene number. The conclusion is that the presence of the multiple chaperonins is likely to be related at least in part to the complex processes of nodule formation and nitrogen fixation, and some evidence exists to support this hypothesis.

6.1 Nitrogen Fixation and the Biology of the Rhizobia

The concept of nitrogen fixation—the conversion of the somewhat inert atmospheric gas, nitrogen, to ammonia, which can subsequently be used for biosynthesis of nucleotides and amino acids—is introduced to school children when they learn about the nitrogen cycle, befitting its central place in understanding the chemistry of life. Nitrogen fixation can occur in several ways and would have been a prerequisite for the origin and evolution of all life on Earth. Before life emerged, nitrogen fixation must have been an abiotic process, and processes such as photochemical reactions and lightning discharges continue to contribute to the formation of ammonia. But an early step in the evolution of life was the appearance of biological nitrogen

P.A. Lund
School of Biosciences & Institute of Microbiology and Infection, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK
e-mail: p.a.lund@bham.ac.uk

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_6

fixation, catalysed by the enzyme nitrogenase which uses energy from ATP hydrolysis to drive this energetically unfavourable process (16 molecules of ATP are required for the fixation of each molecule of nitrogen, N2). Estimates of when this process evolved vary, but recent evidence suggests it could have been at least 3.2×10^9 years ago (Stüeken et al. 2015). Organisms which can fix nitrogen are referred to as diazotrophs. In the past they have been identified by their occurrence in the environment and study in the laboratory, but genomic methods have more recently been used to look for candidates on the basis of their possession of genes which are essential for formation of the nitrogenase complex. This has identified 13 different bacterial phyla which appear to have the potential to fix nitrogen (Dos Santos et al. 2012), comprising nearly 15% of sequenced bacterial genomes, and showed that nitrogen fixation is limited to prokaryotes (including a few archaea). The fact that this number is greater than had previously been expected suggests that earlier estimates that the activity of nitrogenase is responsible for half of all the fixed nitrogen on Earth (Falkowski 1997) may have underestimated its importance. Eukaryotes are thus in this case, as in so many others, dependent on their prokaryotic cousins for an essential part of their lifestyle.

Of the estimated 153×10^6 tons of nitrogen fixed every year on land, around 25% is fixed by the group of plants known as legumes (Brady and Weil 2007). Legumes between them include many important crop plants (both for human and animal use) and green manures such as peas, beans, soybeans, lentils, clover, and alfalfa, which play an important role in crop rotation, itself a process which among other things helps prevent nitrogen from becoming limiting for growth. An important though not completely universal property of legumes is their symbiotic relationship with bacteria in nodules on their roots which fix nitrogen into ammonia that subsequently becomes available to the plants, and the bacteria which form these nodules are referred to generically as Rhizobia. This relationship is a mutualistic one: the bacteria provide fixed nitrogen in the form of ammonium and amino acids which are used by the plants, while the plant provides reduced carbon compounds to the bacteria which they can use as a source of energy to drive the energetically expensive nitrogen fixation process.

The Rhizobia are not a monophyletic group. Although originally they were classified in terms of their host range (i.e. the crop plant on which they formed root nodules), the use of DNA sequence-based methods to construct bacterial phylogenies led to the realisation that bacteria that nodulate legume roots include members of both the alpha- and beta-proteobacteria and that many Rhizobial species have non-nitrogen-fixing relatives to which they are more closely related than they are to other Rhizobia (Young and Haukka 1996). An example is *Rhizobium leguminosarum*, which can nodulate peas and beans and which is significantly closer in evolutionary terms to the non-nodulating bacterium *Agrobacterium tumefaciens* (causative agent of crown gall disease and widely used in the genetic engineering of plants) than it is to other Rhizobia such as *Bradyrhizobium japonicum.* This can lead to confusion: the term ‘Rhizobia’ is convenient to describe bacteria in terms of their properties, but is not a phylogenetic category. A more phylogenetically correct term is *Rhizobiales*. The order *Rhizobiales* is monophyletic and includes all the root

nodulators of legumes which are alpha-proteobacteria; however, it also includes many other bacteria which have a very broad range of properties and which do not nodulate or fix nitrogen. For the purposes of this chapter, I will mostly refer to the Rhizobia, since it is the biological properties of root nodulation and nitrogen fixation, and their relationship to chaperonin gene number, which is of interest here, but I will also draw comparison between the different Rhizobial and non-Rhizobial members of the *Rhizobiales* and mention briefly other nitrogen-fixing bacteria which are not part of this order.

The steps involved in root nodulation are a fascinating example of a complex bacterial developmental programme and have been studied intensively for many years, both because of their enormous agricultural importance and because of their high biological interest. In order to form root nodules, bacteria must somehow detect the presence of roots, invade them without causing damage to their hosts, set up shop within the root, and create a highly anaerobic environment, as nitrogenase activity is strongly inhibited by oxygen. A detailed description of this process is beyond the scope of this chapter (an excellent recent review can be found in Gage 2004). Briefly, free-living Rhizobia are able to detect compounds which are exuded by plant roots, and this leads to the induction of a number of bacterial genes called *nod* genes that leads to the synthesis and export of a compound that in turn induces alterations in root development. These include a typical curling of root hairs and the induction of cell division in the root cortex. It is these newly diving root cells that will eventually comprise the root nodule tissue. Bacteria now enter the plant through a structure called an infection thread and will eventually enter the newly dividing cells in the root cortex, where they become surrounded by membrane derived from the plant cell and differentiate into structures called bacteroids, which are the nitrogen-fixing structures. Nitrogen fixation itself requires expression and assembly of nitrogenase, plus host of other genes whose products are needed for different aspects of the process. Consequently, it is not surprising to find using transcriptomic analysis that a considerable amount of novel gene expression takes place in root nodules; interestingly, many of the highly expressed genes are not common between different Rhizobia (Karunakaran et al. 2009), although around 30% are, including the *nif* and *fix* genes which have been studied for many years and which are often found on large plasmids in different Rhizobia. This induction of gene expression that is associated with major changes in the properties of the bacteria, including losing their ability to be isolated as free-living organisms, may result in high requirements for chaperone function, as will be discussed below.

6.2 Chaperonin Number and Genome Context in Rhizobia and Related Species

In 1993, two groups working on different genera of Rhizobia noticed that they possessed multiple chaperonin genes. One group in Zurich, Switzerland, identified five *cpn60* homologues in *Bradyrhizobium japonicum*, the bacterium that nodulates the

roots of soybeans (Fischer et al. 1993); genome sequencing subsequently revised this number up to seven in a closely related species, *B. japonicum* USDA 110 (now renamed *B. diazoefficiens* USDA 110), the highest number of chaperonin gene copies known in any bacterium. A separate group working in Canada found three *cpn60* homologues in a strain of *Rhizobium* (now *Sinorhizobium*) *meliloti* (Rusanganwa and Gupta 1993). The following year it was shown that a strain of *Rhizobium leguminosarum* also contained three *cpn60* genes (Wallington and Lund 1994), and later genome sequencing showed that some *R. leguminosarum* strains had four *cpn60* genes (Young et al. 2006). Multiple copies have also been found in *Mesorhizobium japonicum* and in *R. etli* (Kaneko et al. 2000; González et al. 2006).

With the advent of routine genome sequencing, the number of available genomes of Rhizobia and many other nitrogen-fixing bacteria has rapidly increased over recent years. A complete bioinformatics analysis of their chaperonin genes has not been conducted, but a quick examination of representative genomes on the useful comparative site rhizobase (http://genome.microbedb.jp/rhizobase/) yields the numbers shown in Table 6.1, which I have based on annotations supported with BLASTP searches. (Care has to be taken in using annotations alone as the *cpn60* genes are often annotated in multiple different ways, sometimes even in the same genome.) For comparison, Table 6.1 also includes some other bacteria. These are:

(1) Examples of root-nodulating nitrogen-fixing beta-proteobacteria, which are thus Rhizobia by our functional definition, but not members of the *Rhizobiales*.
(2) *Frankia*, which fix nitrogen in association with Actinorhizal plants (and are thus distinguished from Rhizobia which are associated only with legumes) and are gram-positive bacteria.
(3) Some examples of free-living nitrogen fixers which do not form root nodules but can associate endophytically with plants (including wheat, rice, and sugar cane) and enhance plant growth. All of these organisms are alpha-proteobacteria but do not fall within the order *Rhizobiales*.
(4) Examples of *Rhizobiales* which are closely related to the organisms in the first group, but which are not diazotrophs. These include bacteria with a wide range of properties, including one species (*Rhodopseudomonas palustris*) which can grow photoautotrophically, photoheterotrophically, chemoautotrophically, or chemoheterotrophically, a quite remarkable metabolic diversity, and some animal and human pathogens.

Two things emerge clearly from this table. First, the ability to form root nodules and fix nitrogen tends to be associated with a large number of the chaperonin gene family within that organism, particularly in the *Rhizobiales*. The only exception in this table is the organism *Cupriavidus taiwanensis*, which forms nodules on the roots of mimosa. It is particularly intriguing that the *Frankia* spp., which as gram positives are phylogenetically very distant from the other bacteria shown here, also have high numbers of chaperonin genes. However, it is also clear that multiple chaperonins are common, though not universal, in some *Rhizobiales* that are not nitrogen fixers. Thus the correlation between root nodulation, nitrogen fixation, and the

Table 6.1 Numbers of chaperonin and co-chaperonin genes in different Rhizobia and other nitrogen-fixing bacteria

Organism name	Number of *cpn60* genes	Number of *cpn60* genes without adjacent *cpn10*	Number of *cpn10* genes without an adjacent *cpn60*	Genome size (Mb)
Root-nodulating *Rhizobiales*				
Mesorhizobium loti MAFF303099	5	0	0	7.6
Bradyrhizobium japonicum USDA 110	7	2	0	9.1
Sinorhizobium meliloti 1021	5	2	1	6.7
Rhizobium etli CFN42	4	0	0	6.5
Rhizobium leguminosarum bv. viciae 3841	4	1	0	7.8
Bradyrhizobium sp. BTAi1	4	1	0	8.5
Bradyrhizobium sp. ORS278	4	1	0	7.4
Azorhizobium caulinodans ORS571	2	0	0	5.4
Root-nodulating beta-proteobacteria				
Cupriavidus taiwanensis LMG 19424	1	0	0	6.5
Burkholderia phenoliruptrix BR3459a	3	0	0	4.2
***Frankia* species**				
Frankia alni ACN14a	4	3	1	7.5
Frankia sp. CcI3	4	3	1	5.4
Frankia EAN1pec	4	3	1	9.0
Non-nodulating nitrogen-fixing bacteria				
Azospirillum sp. B510	3	1	0	7.6
Gluconacetobacter diazotrophicus PAl 5	2	0	0	3.8
Klebsiella pneumoniae Kp342	1	0	0	5.9
Other non-nodulating, non-nitrogen-fixing *Rhizobiales*				
Bartonella quintana	1	0	0	1.6
Bartonella henselae	1	0	0	1.9
Bartonella bacilloformis	1	0	0	1.4
Brucella melitensis	1	0	0	3.3
Ochrobactrum anthropi	1	0	0	4.8
Agrobacterium tumefaciens	1	0	0	5.7
Nitrobacter hamburgensis	3	0	0	5.0
Rhodopseudomonas palustris	2	0	0	5.5
Xanthobacter autotrophicus	2	0	0	5.3
Parvibaculum lavamentivorans	1	0	0	3.9

possession of multiple chaperonins is significant but not absolute. Several endophytes that also fix nitrogen are shown in Table 6.1, and again, although the correlation is not complete, there is a tendency for them to possess multiple chaperonins. Second, although there is a link between genome size and number of chaperonin genes, increasing chaperonin gene number is not simply a consequence of genome expansion: several organisms with quite large genomes have only single chaperonin genes. However, reduction in genome size which seems to have occurred in the pathogenic *Bartonella* and *Brucella* strains also leads to a loss of the multiple chaperonin genes.

To what extent does genome context also support the hypothesis that the chaperonin genes are playing a role in the Rhizobial lifestyle? They are certainly sometimes found close to genes with a role in symbiosis. The *cpn60.3* operon in *B. japonicum*, for example, is found within in gene cluster involved in nitrogen fixation (Fischer et al. 1993), and one of the *cpn* operons in *Mesorhizobium loti* is in a 'symbiotic island', which contains many genes for nodulation and nitrogen fixation (Kaneko et al. 2000). One of the *cpn60* genes in *R. leguminosarum* A3 is on a large plasmid that also carries some of the genes needed for nitrogen fixation (Young et al. 2006), and three of the five *cpn60* homologues in *S. meliloti* map to the pSym plasmids that carry genes for nitrogen fixation and root nodulation (Galibert et al. 2001). The phylogenetic analysis of the multiple chaperonin genes is rather limited to date, but in general they are very similar to each other, with no clear evidence for specialisation of function. This is very different from the situation that exists in some other bacteria with multiple chaperonin genes such as the Actinobacteria, where it is clear that gene duplication and sequence divergence have happened long before speciation to the current day strains (Goyal et al. 2006; Gould et al. 2007a; Lund 2009). A more detailed bioinformatics analysis of *cpn60* gene position relative to genes involved in different aspects of bacterial growth, and their phylogenetical relationships, both in the diazotrophs and in the other species in Table 6.1, would be of considerable interest.

6.3 Genetic Analysis of Chaperonin Requirement in Root Symbiosis

The simplest hypothesis for the role of chaperonin genes in the nodulation lifestyle would be that one or more of the genes has become specialised for some specific aspect of this lifestyle, such as the assembly of nitrogenase. An alternative hypothesis would be that no specialisation of function has occurred at all, and the high number of chaperonin genes is simply a reflection of the increased requirement for protein folding capacity that may exist during the complex processes of nodulation and differentiation into bacteriods. Sequence analysis alone suggests that the first hypothesis is unlikely to be fully correct, as there is little sign of significant functional divergence in gene sequences. However, the only way to formally test these

hypotheses is to construct mutations of individual genes and examine the effect of this on the bacteria concerned. This has been done only in a few cases to date, and the results suggest that the true story is somewhere between the two alternative hypotheses above.

The first studies on this were done in *B. japonicum*. Knockout mutations were constructed in all five of the *cpn60* genes that had been cloned in the original strain (*B. japonicum* 110spc4), and all were found to be still capable of forming root nodules and fixing nitrogen (Fischer et al. 1993). Subsequently, it was found that a double knockout of two of the cpn operons, called *groESL3* and *groESL4* in the original study, reduced nitrogen fixation activity by 95% (Fischer et al. 1999) and that this was due to a significant reduction in the levels of the NifH and NifDK proteins (components of the nitrogenase proteins) in this strain. However, this defect could be partially suppressed by overexpression of two of the other *cpn* operons, though for full suppression, expression of one of the two deleted *cpn* operons was required. This fits a model where there is some partial specificity of function of some of the chaperonins, but not so high that other chaperonins cannot substitute if expressed at sufficiently high levels. If this is correct, it can be predicted that even chaperonins from other organisms (particularly those with only a single copy where no specialisation of function will have evolved) could suppress this defect. This was tested by overexpression of the *E. coli groESL* operon, and it was indeed found that this could also partially suppress the nodulation and nitrogen fixation defect.

Data from *S. meliloti* supports this model. In this organism, a random Tn5 insertion in one of the *cpn60* homologues was found to lead to delayed formation of nodules and loss of their ability to fix nitrogen (Ogawa and Long 1995). Further study showed that this defect arose from reduced expression of three regulatory activating proteins (NodD1, NodD3, and SyrM) in the mutant, and co-immunoprecipitation experiments showed that this particular Cpn60 protein is associated with the NodD proteins. Similar to the case with *B. japonicum*, overexpression of another *S. meliloti cpn* operon could suppress the defects of the Tn5 mutant, although *E. coli groESL* could not do this (Ogawa and Long 1995), which suggests a higher degree of chaperonin specificity in this case.

Subsequent work on *S. meliloti* showed that each of the five *cpn* operons can be deleted without loss of growth, but either *groESL1* or *groESL2* must be expressed for growth (Bittner et al. 2007). The *groESL1* operon alone appears to be necessary and sufficient for nodule formation and nitrogen fixation, and this can be complemented by overexpression of *groESL2* but not *groESL3* (the other operons have not been tested; Bittner and Oke 2006).

R. leguminosarum shows similar evidence for overlap of chaperonin function. In strain A34 (which has three complete *cpn* operons), only one of the *cpn60* genes (*cpn60.1*) is essential (Rodríguez-Quiñones et al. 2005). However, overexpression of the *cpn60.3* gene can partially but not fully restore function (Gould et al. 2007b). Deletion of both the *cpn60.2* and *cpn60.3* genes led to a 50% reduction in nodule formation and nitrogen fixation (P.A. Lund and J.A. Downie, unpublished data).

6.4 The Complex Regulation and Properties of the Rhizobial Chaperonins

The study of the regulation of multiple chaperonin genes in Rhizobia provides further evidence that links their expression to root nodulation and nitrogen fixation. Chaperonin genes are typically heat shock proteins and in most bacteria are regulated by a repressor-based system—the HrcA protein—binding to a consensus sequence referred to as a CIRCE sequence which is found upstream of heat shock gene-specific promoters. Some alpha-proteobacteria have a separate system which relies on raised activity of an alternative sigma factor (encoded by the *rpoH* gene) for RNA polymerase following heat shock, which directs the RNA polymerase to the promoters of heat shock-specific genes. This system is very well studied, in part because it operates in *E. coli*. Rhizobia frequently possess both systems, an interesting observation as it implies that subtle differences may exist in how the two systems are deployed to raise gene expression after stress, a feature that would enable these organisms to fine-tune their gene expression to particular stress conditions. An added level of complexity is suggested by the fact that some Rhizobia encode multiple *rpoH* genes.

In *B. japonicum*, studies on the five operons originally identified showed that one is expressed constitutively, two are regulated by an HrcA-CIRCE system, and one is regulated by (and in the genome maps next to) one of three *rpoH* homologues that are found in this organism (Fischer et al. 1993; Babst et al. 1996; Minder et al. 2000). The fifth is highly expressed in bacteroids, a result that has subsequently been confirmed in a whole transcriptome analysis (Pessi et al. 2007) and is regulated by the NifA activator and another alternative sigma factor, σ54. This operon is the *groESL3* operon identified in mutational studies as being important for nitrogen fixation. The particular significance of this finding is that genes regulated by these proteins include the genes required for nitrogen fixation (Fischer 1994), strengthening the link between expression of this operon and aspects of a successful symbiosis.

In *S. meliloti*, two of the *cpn* operons are heat shock induced, one apparently being regulated by one of the two RpoH proteins encoded by this organism and the other by HrcA (Mitsui et al. 2004; Bittner and Oke 2006). One of the operons was identified using a genome-wide promoter trap screen as being expressed only in nodules (Oke and Long 1999). Interestingly in *S. meliloti*, mutation of one of the *rpoH* genes leads to defects in nodule formation, and double mutants in both genes cannot form nodules at all. However, overexpression of chaperonin proteins alone cannot suppress these defects, so there must be additional targets of *rpoH* regulation that are required for effective nodulation; these have not been identified to date (Bittner and Oke 2006; Barnett et al. 2012).

In *R. leguminosarum* A34, the essential *cpn* operon (*cpn1*) is regulated by HrcA, while the two dispensable operons are regulated by RpoH (*cpn2*) and NifA (*cpn3*); *cpn3* expression is seen only under anaerobic conditions, which are also found in root nodules because of the high oxygen sensitivity of the nitrogenase enzyme (Rodríguez-Quiñones et al. 2005; Gould et al. 2007b).

A further line of evidence that supports a degree of functional specialisation comes from biochemical studies on the three Cpn60 proteins of *R. leguminosarum* A34. All three of these proteins have been expressed and purified from *E. coli* and compared for a range of biophysical and biochemical properties. The most striking difference was that whereas Cpn60.1 and Cpn60.2 were equally competent at refolding the protein lactate dehydrogenase from a denatured state in vitro (a widely used client to study chaperonin activity), Cpn60.3 was unable to refold this client. It was also noted that if Cpn60.1 and Cpn60.3 were co-expressed in *E. coli*, they showed a high preference for assembling into homo-oligomeric complexes (George et al. 2004).

6.5 Conclusions and Perspectives

The evidence on the biological roles of chaperonins in the crucial process of nitrogen fixation in legumes is incomplete but clearly points to the hypothesis that they do play some role and that there may be a degree of specificity in their association with some key proteins involved in the steps of nodulation and nitrogen fixation.

Much work remains to be done to fill out the details of this association. The availability of large amounts of complete genome sequence information means that there is plenty of scope for further bioinformatics analysis into the occurrence of the multiple chaperonin genes and in particular their genomic context: whether at least some of them are always found in gene clusters either on the chromosome or on the mega-plasmids that are involved with aspects of nodulation and nitrogen fixation. In addition, now that there are many more sequences available, a detailed phylogenetic and covariance analysis has the potential to reveal features of the protein sequences that may be associated with their specialised role.

However, elucidating the mechanistic details of their function will require a combination of more in vivo, in vitro, and in planta studies. Proteomics methods have moved on since most of the studies above were completed, and it is to be hoped that the recognition of the likely role of chaperonins in this crucially important biological process may spur on some new studies in the future that will shed more light on the mechanism of their action in Rhizobia.

References

Babst M, Hennecke H, Fischer HM (1996) Two different mechanisms are involved in the heat-shock regulation of chaperonin gene expression in *Bradyrhizobium japonicum*. Mol Microbiol 19:827–839

Barnett MJ, Bittner AN, Toman CJ, Oke V, Long SR (2012) Dual RpoH sigma factors and transcriptional plasticity in a symbiotic bacterium. J Bacteriol 194:4983–4994

Bittner AN, Oke V (2006) Multiple *groESL* operons are not key targets of RpoH1 and RpoH2 in *Sinorhizobium meliloti*. J Bacteriol 188:3507–3515

Bittner AN, Foltz A, Oke V (2007) Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. J Bacteriol 189:1884–1889

Brady NC, Weil RR (2007) The nature and properties of soils, 14th edn. Prentice Hall

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics 13:162

Falkowski PG (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. Nature 387:272–275

Fischer HM (1994) Genetic regulation of nitrogen fixation in rhizobia. Microbiol Rev 58: 352–386

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

Fischer HM, Schneider K, Babst M, Hennecke H (1999) GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum*. Arch Microbiol 171:279–289

Gage DJ (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiol Mol Biol Rev 68:280–300

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672

George R, Kelly SM, Price NC, Erbse A, Fisher M, Lund PA (2004) Three GroEL homologues from Rhizobium leguminosarum have distinct in vitro properties. Biochem Biophys Res Commun 324:822–828

González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci U S A 103:3834–3839

Gould PS, Burgar HR, Lund PA (2007a) Homologous cpn60 genes in *Rhizobium leguminosarum* are not functionally equivalent. Cell Stress Chaperones 12:123–131

Gould P, Maguire M, Lund PA (2007b) Distinct mechanisms regulate expression of the two major *groEL* homologues in *Rhizobium leguminosarum*. Arch Microbiol 187:41640

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

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res 7:331–338

Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS (2009) Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J Bacteriol 191: 4002–4014

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

Minder AC, Fischer HM, Hennecke H, Narberhaus F (2000) Role of HrcA and CIRCE in the heat shock regulatory network of *Bradyrhizobium japonicum*. J Bacteriol 182:14–22

Mitsui H, Sato T, Sato Y, Ito N, Minamisawa K (2004) *Sinorhizobium meliloti* RpoH1 is required for effective nitrogen-fixing symbiosis with alfalfa. Mol Genet Genomics 271:416–425

Ogawa J, Long SR (1995) The *Rhizobium meliloti* groELc locus is required for regulation of early nod genes by the transcription activator NodD. Genes Dev 9:714–729

Oke V, Long SR (1999) Bacterial genes induced within the nodule during the Rhizobium–legume symbiosis. Mol Microbiol 32:837–849

Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H (2007) Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. Mol Plant Microbe Interact 20:1353–1363

Rodríguez-Quiñones F, Maguire M, Wallington EJ, Gould PS, Yerko V, Downie JA, Lund PA (2005) Two of the three *groEL* homologues in *Rhizobium leguminosarum* are dispensable for normal growth. Arch Microbiol 184:253–265

Rusanganwa E, Gupta RS (1993) Cloning and characterization of multiple *groEL* chaperonin-encoding genes in *Rhizobium meliloti*. Gene 126:67–75

Stüeken EE, Buick R, Guy BM, Koehler MC (2015) Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr. Nature 520:666–669

Wallington EJ, Lund PA (1994) Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes. Microbiology 140:113–122

Young JPY, Haukka KE (1996) Diversity and properties of rhizobia. New Phytol 133:87–94

Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J (2006) The genome of Rhizobium leguminosarum has recognizable core and accessory components. Genome Biol 7:R34

Chapter 7
Multiple Chaperonins in Cyanobacteria: Why One Is Not Enough!

Hema Rajaram, Akhilesh K. Chaurasia, and Akhilesh A. Potnis

Abstract The changing environmental conditions in diverse habitats of cyanobacteria result in adaptive evolution involving both genetic and physiological modulation in order to cope up with extreme environments. One of the major common factors observed under different extreme conditions is denaturation of proteins, and accumulation of denatured proteins in the cells would result in lethality. Cells prevent accumulation of denatured proteins by synthesizing heat shock proteins, which function as chaperones and/or proteases. In cyanobacteria, the chaperonins or the Hsp60 family of proteins are the most predominant HSPs, unlike DnaK or the Hsp70 family in most other bacteria. Most cyanobacterial species have two *groEL* genes, one present as a part of a bicistronic *groESL* operon and the other as a monocistronic *groEL2*/*cpn60* gene. Both the genes have been found to be essential for cyanobacterial growth, with *groEL2* being dispensable in a few cyanobacterial species under ambient growth temperatures. The synthesis of the two proteins was induced in response to not only heat stress but

H. Rajaram (✉)
Molecular Biology Division, Bhabha Atomic Research Centre,
Trombay, Mumbai 400085, India

Homi Bhabha National Institute,
Training School Complex, Anushakti Nagar, Mumbai 400094, India
e-mail: hemaraj@barc.gov.in

A.K. Chaurasia
Molecular Biology Division, Bhabha Atomic Research Centre,
Trombay, Mumbai 400085, India

Samsung Biomedical Research Institute, School of Medicine, Sungkyunkwan University,
Suwon 440-746, Republic of Korea

A.A. Potnis
Molecular Biology Division, Bhabha Atomic Research Centre,
Trombay, Mumbai 400085, India

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_7

several other abiotic stresses, suggesting a more global role for the chaperonins in cyanobacteria. Their expression was regulated mainly through the repressor protein HrcA which bound to the CIRCE element overlapping with the promoter region, with other regulatory elements, such as the K-box and H-box also playing a role during heat and/or light stresses. The two chaperonins exhibited differences in their biochemical activities, which possibly reflected on their distinct physiological roles. The cyanobacterial GroEL were distinct from most other bacterial GroEL in being able to function optimally in a GroES- and ATP-independent manner. Taken together, the multiple cyanobacterial Hsp60 proteins and their *cis* and *trans* regulation upon varying physiological conditions provide an in-depth insight into how the ancient chaperonins might have evolved, functioned, and contributed toward adaptive evolution and diversification of the more modern bacteria.

7.1 Introduction

The photoautotrophic cyanobacteria inhabited the earth nearly 3.5 billion years ago (Brock 1973) and accounted for the initial oxygenation of Earth's atmosphere (Schopf 1975). They exhibit survival under extreme temperatures, demonstrated by the presence of *Synechococcus* sp. (>90 °C) and *Oscillatoria teribriformis* (~54 °C) in hot springs (Castenholz 1968) and *Calothrix parietina*, *Nostoc* sp., *Synechococcus* sp., and *Phormidium frigidum* in the frozen lakes of Antarctica (Wynn-Williams 1990). Cyanobacteria are photoautotrophic diazotrophs as their nitrogen fixation abilities depend entirely on photosynthesis (Stewart 1980). However, due to the adverse effect of oxygen on nitrogen fixation due to the presence of oxygen-labile nitrogenase (Robson and Postgate 1980), photosynthesis and nitrogen fixation need to be separated either in time or space. In the unicellular forms, the temporal separation was achieved by performing photosynthesis in light and nitrogen fixation in dark (Mitsui et al. 1986). In the filamentous cyanobacteria, such as *Anabaena* and *Nostoc*, nitrogen fixation is restricted to terminally differentiated specialized cells, namely, the heterocysts, while photosynthesis takes place in vegetative cells (Haselkorn 1978). A structural linkage is observed between photosynthesis, respiration, and nitrogen fixation due to the sharing of certain components in the electron transport chain of photosynthesis and respiration. The disaccharides resulting from the dark reaction of photosynthesis in the vegetative cells of cyanobacteria are supplied to the heterocysts, the seat of nitrogen fixation, wherein they are metabolized to generate NADPH, ATP, and glutamate required for nitrogen fixation. The interdependence of these activities is schematically depicted in Fig. 7.1.

Heat stress adversely affected both photosynthesis and nitrogen fixation. In the unicellular cyanobacterium, *Synechococcus* PCC7942, exposure to temperatures

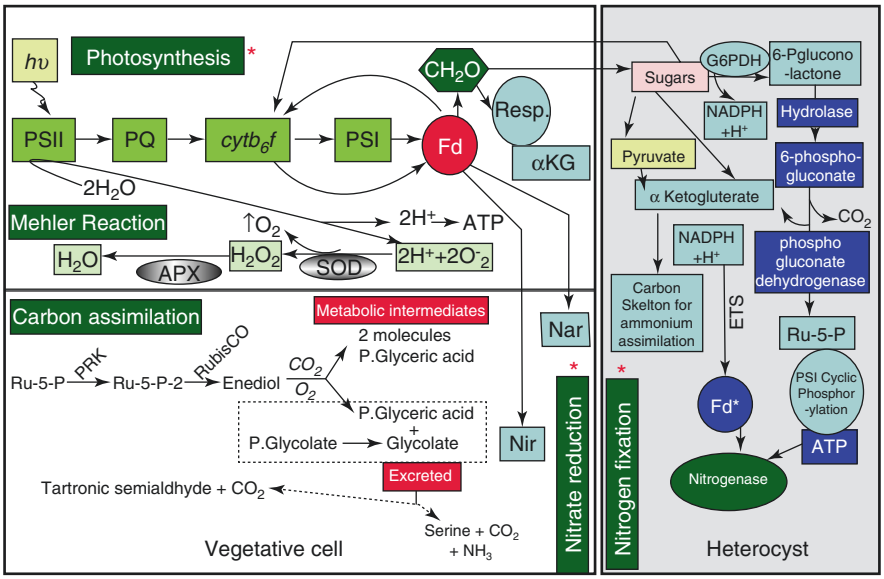


Fig. 7.1 Schematic diagram depicting the interrelation and dependence of photosynthesis, respiration, and nitrogen metabolism and division of labor between vegetative cells and heterocysts of *Anabaena*. The photolysis of the water generates protons ($4H^+$), electrons ($4e^-$), and oxygen free radicals. The oxygen free radicals and two protons enter into the Mehler reaction to produce water and oxygen. The electron transport chain of cyanobacterial respiration and photosynthesis shares some of the common components. The source of reduced ferredoxin, which is required for nitrate assimilation in vegetative cells, is photosynthesis. Reduced ferredoxin for nitrogen fixation in heterocysts is generated by G-6-PDH and Fd-$NADP^+$ oxidoreductase. The ATP generated by the light reaction is used for the carbon fixation through Calvin's cycle in vegetative cells. The cyanobacterial respiration and PS-I-based cyclic photophosphorylation generate ATP that is used for nitrogen fixation and assimilation. The end product of interrupted TCA cycle, α-ketoglutarate, is used as carbon skeleton for the GS-GOGAT/GDH pathway(s), for C/N ratio, and for signal transduction. Photosynthesis and Mehler reaction are redrawn based on (Berman-Frank et al. 2001)

above 48 °C resulted in the inactivation of the photosystem II (Eriksson and Clarke 1996). Heat stress affected the photosynthetic electron transport and phosphorylation reactions resulting in reduced photosynthetic activity in unicellular *Synechocystis* sp. PCC6803 (Mamedov et al. 1993). In the filamentous cyanobacteria, heat stress inhibited nitrogen fixation in *Anabaena cylindrica* and *Mastigocladus laminosus* (Pederson et al. 1986) and photosynthesis as well as nitrogen fixation and assimilation in *Anabaena* sp. (Rajaram and Apte 2003, 2008; Chaurasia and Apte 2009). Effect of heat stress on major photosynthetic pigments and its subsequent recovery upon shift back to normal growth temperatures in *Anabaena* is shown in Fig. 7.2. The adverse effect of heat on such vital activities emphasizes the importance of gaining insight into the heat shock response in cyanobacteria.

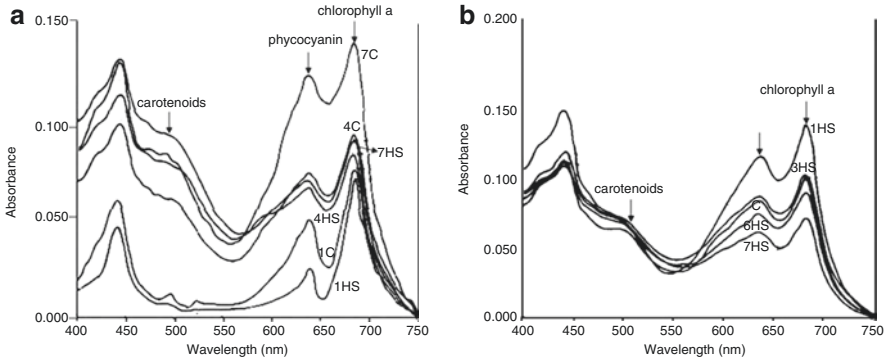


Fig. 7.2 The major photosynthetic pigments, chlorophyll a (A680), phycocyanin (A620), and carotenoids (A500), during (**a**) heat stress and (**b**) recovery from heat stress in *Anabaena* L-31. (**a**) Three-day-old *Anabaena* L-31 culture was inoculated in fresh BG-11, N^- medium at a density of 1 μg chlorophyll *a* mL^{-1} and grown at 27 °C (control) or subjected to heat stress at 42 °C, for 7 days. (**b**) *Anabaena* L-31 cells heat stressed for 0, 1, 3, 5, and 7 days were inoculated in fresh BG-11, (N-) medium at 1 μg chlorophyll *a* mL^{-1} and allowed to recover at 27 °C for 7 days. Spectra of dilute suspensions of whole filaments of cultures exposed to heat stress (HS) for 0, 1, 3, 5, and 7 days were taken at the end of 7 days of growth under control growth conditions. "C" indicates spectrum of control grown cells. Spectra of dilute suspensions of whole filaments were recorded on different days in a Hitachi (Model 150-20) Double Beam UV/Visible Spectrophotometer after adjusting turbidity at 750 nm. The spectra were recorded on day 1, day 4, and day 7 of control (C) or heat stress (HS) condition

7.2 Heat Shock Response in Cyanobacteria

The heat shock response observed upon temperature upshift is transcriptionally activated in cyanobacteria (Borbely et al. 1985; Rajaram and Apte 2010; Webb et al. 1990). The magnitude of induction depended on the growth temperature prior to heat stress and the rise in temperature during stress (Lehel et al. 1993a). Transcriptome analysis revealed induction of several heat shock genes in the unicellular cyanobacteria *Synechocystis* sp. PCC6803 (Suzuki et al. 2006), and *Synechococcus* sp. PCC7002 (Ludwig and Bryant 2012), and a concomitant decrease in the levels of transcripts associated with the major metabolic pathways (Ludwig and Bryant 2012). This was reflected in the increased synthesis and accumulation of host of heat shock proteins, such as HspA1, GroEL1, GroEL2, GroES, HtpG, DnaK2, and ClpB, in *Synechocystis* sp. PCC6803 (Slabas et al. 2006) and in the filamentous nitrogen-fixing cyanobacterium, *Anabaena* sp., in a time-dependent manner (Rajaram and Apte 2003). The cyanobacterial heat shock proteins can be classified into five major families, namely, Hsp100, Hsp90, Hsp70, Hsp60, and sHsps, which contribute toward innate and acquired thermotolerance, protection of photosynthetic and nitrogen-fixing apparatus, and membrane fluidity (Rajaram et al. 2014).

The heat shock proteins were synthesized not only in response to heat stress but also to several other stresses, such as osmotic and salt stresses in *Anabaena* (Apte et al. 1998) and heat, salt, and metal stresses in *Synechocystis* sp. PCC6803 (Castielli

et al. 2009). The common factor in all these stresses is denaturation of proteins, and accumulation of denatured proteins is known to evoke heat shock response (Kanemori et al. 1994). This could account for the observed decreased UV-B toxicity in the filamentous cyanobacterium *Anabaena doliolum* preexposed to heat stress (Mishra et al. 2009). However, contrary to this, in *Synechocystis* sp. PCC6803, preexposure to heat stress did not provide enhanced tolerance to subsequent salt stress (Nikkinen et al. 2012). Of the several Hsps, the Hsp60 proteins are the most predominant Hsps in cyanobacteria and have been discussed in the subsequent sections.

7.3 Hsp60 Family or Chaperonins of Cyanobacteria

Cyanobacteria are among the increasing group of bacteria having multiple *hsp60* or chaperonin genes (Lund 2009). In cyanobacteria, of the two *hsp60* genes, only one exists as part of a bicistronic operon, while the other exists as an individual gene. The monocistronic *hsp60* referred to as either *groEL-2* or *cpn60* are phylogenetically classified into Group A, while the bicistronic *hsp60*, referred to as either *groEL-1* or *groEL*, come under Group B, and a very small group of *hsp60* genes of cyanobacteria constitute the Group C in the phylogenetic tree depicting the *hsp60* genes of cyanobacteria (Rajaram et al. 2014).

7.3.1 Identification of groEL Genes in Different Cyanobacterial Species

An ~8.9 kb region containing cluster of seven genes coding for subunits of ATP synthase in the cyanobacterium *Synechococcus* sp. PCC6301 (Cozens and Walker 1987) includes partial sequences of a gene which bore significant homology with the heat shock genes in *E. coli* and genes coding for 10 kDa and 65 kDa antigen of *Mycobacterium tuberculosis* (Cookson et al. 1989). Thereafter, the complete *groESL* operon of *Synechococcus* sp. PCC7942 was cloned for the first time and revealed a ~120-fold increase in the transcript level of *groEL* and a tenfold increase in the corresponding protein level upon exposure to heat stress at 45 °C (Webb et al. 1990). The transcript corresponding to the *cpn60* gene of *Synechocystis* PCC6803 was found to be monocistronic and exhibit increase in levels upon exposure to heat shock, ultraviolet exposure, and oxidative stress (Chitnis and Nelson 1991). The first instance of a purified cyanobacterial heat shock protein was reported for the *cpn60*-encoded protein of *Synechocystis* PCC6803, which cross-reacted with anti-*E. coli* GroEL antibody (Lehel et al. 1992). The second *groEL* gene of *Synechocystis* PCC6803 was found to be part of a *groESL* operon, and the corresponding transcript showed a 100-fold increase within 15 min of heat shock (Lehel et al. 1993b). In the thermophilic cyanobacterium, *Synechococcus vulcans*, the transcript levels of both

groEL-1 and *groEL-2*, showed significant increase upon heat stress, but only *groEL-1*, which existed as part of *groESL* operon, was able to complement the heat sensitivity of *groEL* defective mutant of *E. coli* (Furuki et al. 1996; Tanaka et al. 1997). In *Anabaena* L-31, the *groESL* operon was identified as the major heat shock inducible *hsp60* gene (Rajaram et al. 2001), while the second gene *cpn60* was found to be monocistronic (Rajaram and Apte 2008).

7.3.2 Abiotic Stress Induction of Chaperonin Genes/Proteins

The chaperonin genes of *Synechocystis* sp. PCC6803 exhibited a differential expression during heat stress as a function of light-dark transition. Both *groEL* and *cpn60* exhibited increased transcription in response to heat stress during illumination, but not under dark conditions (Glatz et al. 1997). The effect of light was observed in terms of modulation of the transcription of the heat shock genes including *groEL*, but not the stability of the corresponding mRNA (Asadulghani et al. 2003). The expression of the GroEL protein was found to be higher than that of Cpn60 during heat stress in *Synechocystis* sp. PCC6803 (Kovács et al. 2001). Transcriptomic and proteomic analysis of the heat shock response in *Synechocystis* sp. PCC6803 further confirmed the induction of the chaperonins (Suzuki et al. 2006). Induction of the chaperonin transcripts in response to heat stress was also observed in *S. vulcans* (Tanaka et al. 1997). In *Thermosynechococcus elongatus*, the *groEL-1* was induced in response to heat stress, while *groEL-2* was induced in response to cold stress (Sato et al. 2008). In the marine cyanobacterium, *Oscillatoria* sp., the GroEL protein was found to be transcriptionally induced in response to UV-stress (Yamazawa et al. 1999).

Expression of GroEL in response to different abiotic stresses has been widely studied in the filamentous cyanobacteria, *Anabaena* sp. The Hsp60 proteins were synthesized throughout the heat stress in *Anabaena* sp. L-31 (Rajaram and Apte 2003). However, the relative expression of the GroEL and Cpn60 proteins was dependent on the availability of fixed nitrogen (N-status) in *Anabaena*. While the heat stress-induced expression of the 59 kDa GroEL was observed irrespective of the N-status (Rajaram and Apte 2003, 2008), that of the 61 kDa Cpn60 was observed only under nitrogen-fixing conditions (Rajaram and Apte 2008). In fact, the Cpn60 was expressed under normal growth conditions in N-supplemented cultures and was repressed upon exposure of these cultures to heat stress (Rajaram and Apte 2008). Enhanced expression of GroEL in *Anabaena* in response to different abiotic stresses, as analyzed by Western blotting and immunodetection, is shown in Fig. 7.3. Proteomic analysis in different *Anabaena* sp. confirmed increased abundance of the GroEL chaperonin protein in response to methyl viologen stress (Panda et al. 2014, 2015), salt stress (Rai et al. 2014), UV-irradiation (Shrivastava et al. 2015), Cd stress (Singh et al. 2015), and exposure to butachlor (Agrawal et al. 2014). This indicates that the cyanobacterial chaperonins are not specifically heat shock proteins but are general stress proteins.

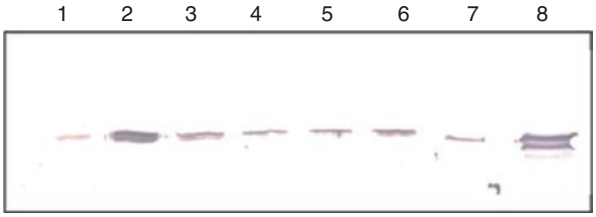


Fig. 7.3 Expression of GroEL in response to different abiotic stresses in *Anabaena* PCC7120. Western blots showing immunodetection of GroEL levels in *Anabaena* 7120 exposed to various stresses. The different lanes correspond to unstressed control (lane 1), heat stress at 42 °C for 4 h (lane 2), salt stress with 150 mM NaCl for 3 days (lane 3), acid stress at pH 5 for 4 h (lane 4), oxidative stress with 5 μM methyl viologen for 2 h (lane 5), heavy metal stress in the presence of 5 mM cadmium (lane 6), dark stress for 24 h (lane 7), and osmotic stress with 200 mM sucrose (lane 8)

7.3.3 Regulation of Expression of Chaperonin Genes

The bacterial heat shock genes, including the *groEL* chaperonin genes, are regulated primarily by either σ^{32} (RpoH) or HrcA, wherein RpoH acts a positive regulator by binding with RNA polymerase, while HrcA inhibits transcription by RNA polymerase (Yura and Nakahigashi 1999). The genome database of cyanobacteria (http://genome.microbedb.jp/cyanobase) revealed the absence of RpoH homolog and presence of HrcA-like protein across cyanobacterial species. The negative regulation by HrcA repressor is achieved by the binding of the HrcA dimer to a 9 bp inverted repeat element [TTAGCACTC-N_9-GAGTGCTAA], also known as CIRCE (*C*ontrolling *I*nverted *R*epeat of *C*haperone *E*xpression) at normal growth temperatures. Denaturation of HrcA during heat stress results in derepression of the genes regulated by it (Zuber and Schumann 1994). Regulation of *groEL* genes by HrcA has been confirmed for *Synechocystis* sp. PCC6803 (Nakamoto et al. 2003) and *Anabaena* sp. (Rajaram and Apte 2010). Cyanobacterial heat shock genes, with the exception of the *groEL* genes, have also been found to be regulated by histidine kinase-34 in *Synechocystis* sp. PCC6803 (Červený et al. 2015) and through SOS response in *Oscillatoria* sp. (Yamazawa et al. 1999).

The CIRCE element was detected upstream of both the chaperonin genes in several cyanobacterial sp. (Nakamoto et al. 2003; Rajaram and Apte 2010) but was absent upstream of *groEL-2* in *S. vulcanus* (Furuki et al. 1996), *Synechococcus* sp. PCC7942, and *T. elongatus* (Kojima and Nakamoto 2007). The CIRCE element partially overlapped (last two nucleotides) with the promoter sequence of *groEL-1*/*groEL* and completely overlapped with the promoter sequence of *groEL-2*/*cpn60* in *Synechocystis* sp. PCC6803 and *Anabaena* sp. L-31. The promoter sequence was similar to *E. coli* σ^{70} promoter (Nakamoto et al. 2003; Rajaram and Apte 2010). Unlike in other bacteria possessing the *hrcA* gene, CIRCE element was absent in the vicinity of the promoter of cyanobacterial *hrcA*, suggesting possible lack of autoregulation. The HrcA protein is characterized by the presence of three boxes, BoxA, BoxB, and BoxC (Schulz and

Schumann 1996), which were also found in cyanobacterial HrcA proteins (Nakamoto et al. 2003), but the overall homology with bacterial *hrcA* was poor. EMSA and supershift studies using *Anabaena* HrcA and the corresponding *groESL* and *cpn60* promoters confirmed that the cyanobacterial HrcA, in spite of sequence differences, is capable of binding to the CIRCE element (Rajaram and Apte 2010). Higher expression of the two Hsp60 chaperonins at normal growth temperatures in the *hrcA* mutants of *Synechocystis* and *Anabaena* reiterated that the cyanobacterial HrcA is active as a repressor (Nakamoto et al. 2003; Rajaram and Apte 2010). Though no further enhancement in the levels of Cpn60 protein was observed upon heat stress in *Anabaena hrcA* mutant, levels of GroEL increased further, indicating multiple regulations for *groESL* operon in *Anabaena* (Rajaram and Apte 2003). In *Synechocystis* PCC6803, two regulatory regions, namely, the K-box and the N-box, were detected upstream of the CIRCE element in the *groESL1* operon and were found to be essential for basal level transcription of the *groESL1* operon, with the K-box also exhibiting regulation in response to both heat and light (Kojima and Nakamoto 2007). The K-box was detected upstream of several cyanobacterial *groESL* operons (Kojima and Nakamoto 2007); however, in *Anabaena* PCC7120, the K-box was involved in only regulation by light and not heat stress (Rajaram and Apte 2010). The *groESL* operon of *Anabaena* PCC7120 has an additional regulatory element, designated as H-box, which comprised of an 11 base pair inverted repeat element. The K-box possibly required a trans-acting protein as suggested for *Synechocystis* (Kojima and Nakamoto 2007) and acted as appositive regulatory element (Kojima and Nakamoto 2007; Rajaram and Apte 2010); the H-box was a *cis*-acting negative regulatory element, contributing to enhance synthesis of GroEL during heat stress only (Rajaram and Apte 2010). The K-box was detected upstream of *groEL2* in *Synechocystis* PCC6803, but not *Anabaena* PCC7120. The unicellular cyanobacterial *groEL2* also possessed a consensus L-box, while the marine cyanobacteria have M-box; however, the exact roles of these boxes are yet to be elucidated (Kojima and Nakamoto 2007).

7.3.4 Physiological Role of Chaperonins

The studies on physiological role of chaperonins in cyanobacteria are mainly restricted to that in *Anabaena*. Due to the presence of two *groEL* genes in cyanobacteria, it was thought that the two chaperonins could compensate for each other, which would allow construction of viable *groEL* mutants. However, on the contrary, in *Anabaena* PCC7120, both *groEL* and *cpn60* were found to be indispensable (Rajaram and Apte 2008), while *groEL2* could be dispensed with in *S. elongatus* PCC7942 (Sato et al. 2007) and *T. elongatus* (Sato et al. 2008) under normal growth conditions but exhibited defective growth both at high and low temperatures in *S. elongatus* (Sato et al. 2008). Mutation in *groESL1* was not viable across cyanobacterial species, indicating it to be an essential gene as in other bacteria, such as *E. coli*. In *Anabaena*, recombinant strains overexpressing either GroEL (with GroES) or Cpn60 were found to be viable and exhibited higher growth under normal growth temperatures (Chaurasia and Apte 2009; Rajaram and Apte 2008). The *hrcA* mutant

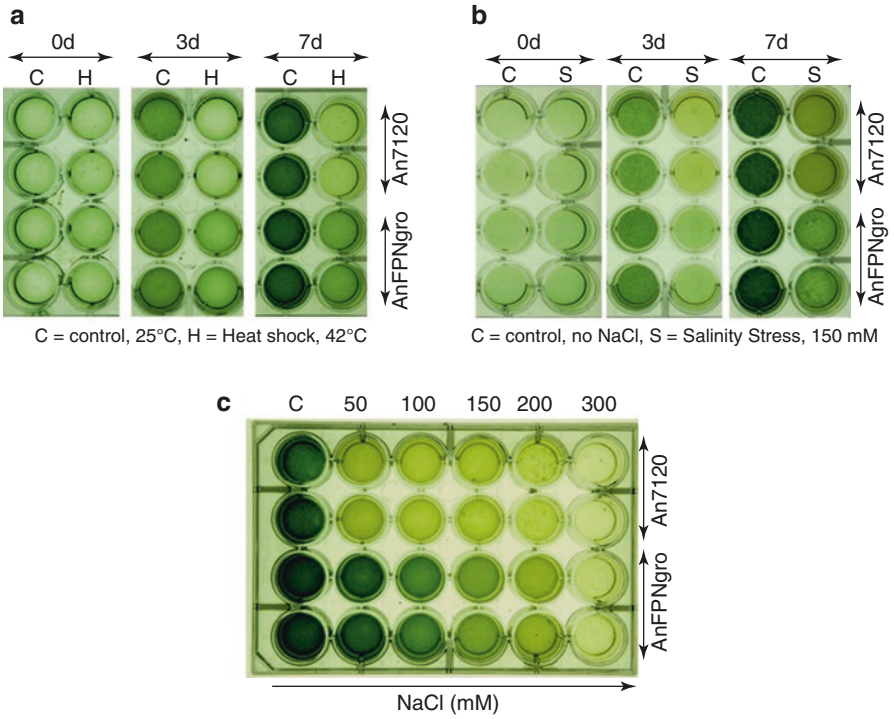


Fig. 7.4 Effect of GroEL overexpression on thermotolerance and salinity tolerance of diazotrophically grown *Anabaena* strains. Microtiter plates showing growth and pigment phenotype of strains during prolonged exposure to (**a**) heat (42 °C) or (**b** and **c**) NaCl

of *Synechocystis*, expressing GroEL1 and GroEL2 at higher levels, also exhibited higher thermotolerance compared to the wild-type cells (Nakamoto et al. 2003).

The 59 kDa GroEL protein of *Anabaena* was expressed during heat stress irrespective of N-status, with the expression under nitrogen-fixing conditions being higher (Rajaram and Apte 2008). The recombinant *Anabaena* strain, AnFPN*gro*, constitutively expressing the *groESL* operon from the *psbA1* promoter exhibited robust growth under normal growth condition, with higher photosynthetic pigments (Chaurasia and Apte 2009) and more blue-green phenotype (Fig. 7.4) both under normal growth conditions and upon exposure to heat and salt stresses. This was reflected in higher photosynthetic activity measured as photoevolution of oxygen during stress in the recombinant cells. GroES-GroEL overexpression resulted in cells retaining 87% and 39% photosynthetic activity after 3 days of continuous salt and heat stress, respectively, as against 54% and 30% in wild-type *Anabaena* PCC7120 under similar conditions (Chaurasia and Apte 2009). The nitrogenase activity, which is more sensitive to heat stress than salt stress and decreased to less than 10% after 4 h at 42 °C in *Anabaena* PCC7120, increased to 27% of corresponding unstressed cells upon overexpression of GroEL (Chaurasia and Apte 2009). The higher stability of the two vital processes along with lowered protein aggregation during stress resulted in increased thermotolerance and salt tolerance of AnFPN*gro* cells under nitrogen-fixing conditions (Chaurasia and Apte 2009).

Unlike the GroEL protein expression during heat stress, that of the 61 kDa protein was dependent on the N-status of growth (Rajaram and Apte 2008). Heat stress induced the expression of Cpn60 under nitrogen-fixing conditions but repressed its expression when grown in combined nitrogen-supplemented media. In addition to the inhibition of transcription of the *cpn60* gene during heat stress under N-supplemented conditions, the Cpn60 protein was also proteolytically degraded (Rajaram and Apte 2008), unlike the observed high stability of both the GroEL and Cpn60 proteins through the heat stress under nitrogen-fixing conditions (Rajaram and Apte 2003). This was reflected in the lower thermotolerance of N-supplemented cultures compared to the nitrogen-fixing *Anabaena* cultures, with both the photosynthetic and nitrate reductase activities decreasing to less than 20% within 6 h of heat stress. However, constitutive overexpression of Cpn60 in *Anabaena*, as in AnFPN*cpn*, resulted in the cells retaining over 20% photosynthetic and nitrate reductase activities even after 4 days of heat stress. This suggested a greater role for Cpn60 under N-supplemented conditions (Rajaram and Apte 2008). *Anabaena hrcA* mutant wherein both GroEL and Cpn60 are constitutively expressed showed growth comparable to AnFPN*gro* under nitrogen-fixing conditions and to AnFPN*cpn* under N-supplemented conditions under heat as well as salt stress (Fig. 7.5). This is

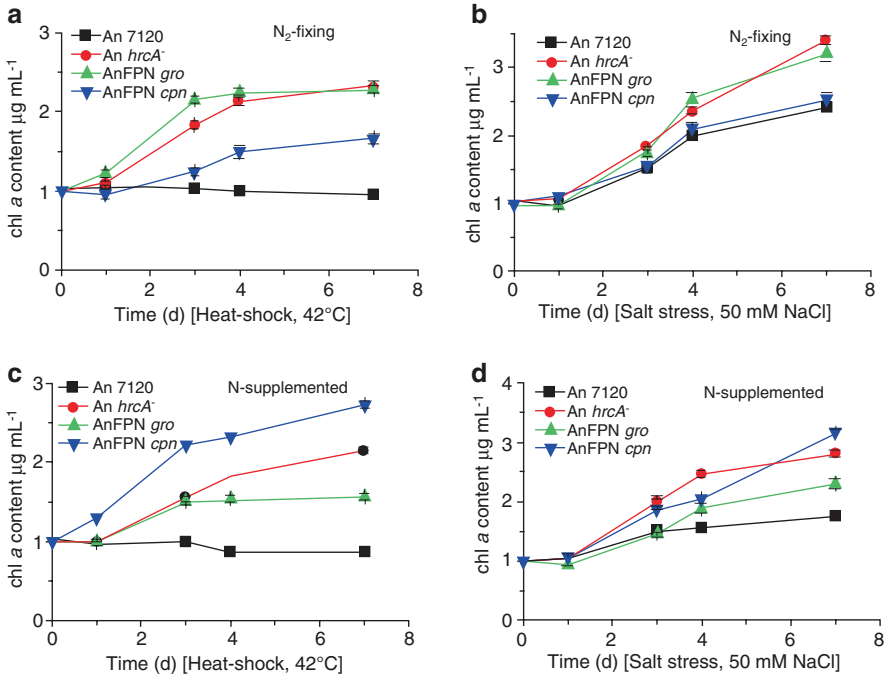


Fig. 7.5 Effect of constitutively higher expression of chaperonin proteins on thermotolerance and salt tolerance. Wild-type and recombinant *Anabaena* strains grown under (**a** and **b**) nitrogen-fixing or (**c** and **d**) nitrogen-supplemented conditions were subjected either to (**a** and **c**) heat stress at 42 °C of (**b** and **d**) salt stress with 50 mM NaCl for 7 days. Growth was measured in terms of chlorophyll *a* content. The different strains used were wild-type An7120: *Anabaena* PCC7120, and recombinant strains An*hrcA*⁻, *hrcA* gene mutant; AnFPN*gro*, GroES and GroEL overexpressing strain; and AnFPN*cpn*, Cpn60 overexpressing strain

suggestive of distinct roles for the two cyanobacterial chaperonins, with the GroEL being primarily engaged in the protection of nitrogenase and to some extent the photosynthetic apparatus, along with other denatured cytosolic proteins, while the Cpn60 is required for stabilization of the nitrate reductase and photosynthetic apparatus in addition to other proteins. A schematic representation of the division of labor between the two Hsp60 proteins has been shown earlier (Rajaram et al. 2014). This division of labor may account for the inability to generate either the *groEL* or *cpn60* mutants of *Anabaena.*

7.4 Cyanobacterial Hsp60 Proteins as Chaperones: A Biochemical Analysis

The above differences in expression suggest possible differences in their chaperonin activity, which was analyzed biochemically for both *S. elongatus* and *Anabaena* sp. PCC7120 chaperonins. The 59 kDa GroEL (GroEL-1) and the 61 kDa Cpn60 (GroEL-2) proteins of cyanobacteria possess the signature sequence "GPKGRN" and exhibit about 60% similarity. However, a major difference is the presence of the "GGM" repeat, typical of bacterial GroEL proteins, at the C-terminal tail of Cpn60, but not GroEL. The *E. coli* GroEL protein exists as a 14-mer, arranged as two stacks of heptameric rings (Hendrix 1979), with the heptameric GroEL acting as a lid and aiding in the ATPase activity of GroEL (Morimoto et al. 1994). Of the two Hsp60 proteins, the *Synechococcus* GroEL1 existed in three oligomeric forms, i.e., 14-mer, heptamer, and a dimer, with the 14-mer being highly unstable, while the GroEL-2 primarily existed in heptameric and dimeric forms, with the higher oligomeric state being unstable (Huq et al. 2010). On the other hand, the *Anabaena* GroEL existed in only two forms, a high oligomeric form (>12-mer) and a dimer, and the Cpn60 protein solely in monomeric form (Potnis et al. 2016). The stability of the higher oligomeric form of GroEL was dependent on the buffer conditions used (Potnis et al. 2016), which suggested that the oligomeric status of GroEL may also vary *in vivo* depending on the intracellular conditions. Such distinct differences in the oligomeric status of the cyanobacterial Hsp60 proteins with respect to *E. coli* GroEL protein had a bearing on their ATPase as well as refolding activities.

The ATPase activity of *Synechococcus* GroEL1 was found to be only 17% of that of *E. coli* GroEL (Huq et al. 2010), while that of *Anabaena* GroEL was about 60% (Potnis et al. 2016). The ATPase activity of GroEL2 was negligible. The higher ATPase activity of *Anabaena* GroEL was due to the higher stability of its higher oligomeric form, suggesting the importance of the higher oligomer in the activity of cyanobacterial Hsp60 proteins. Unlike *E. coli* GroEL, wherein the ATPase activity of GroEL is inhibited in the presence of GroES and absence of a denatured substrate (Horwich et al. 2006), that of both *Anabaena* and *Synechococcus* GroEL showed an increase in the presence of only GroES (Huq et al. 2010; Potnis et al. 2016). These differences were reflected in their refolding activity as well, which was found to be low for both the Hsp60 proteins of *Synechococcus* as well

as *Anabaena* compared to *E. coli*, when the substrate was almost completely denatured (Huq et al. 2010; Potnis et al. 2016). However, a partially denatured substrate, such as malate dehydrogenase (MDH) retaining about 50% activity upon denaturation, could regain almost complete activity in the presence of *Anabaena* GroEL (Potnis et al. 2016). Unlike the increase in chaperonin activity of *E. coli* GroEL in the presence of GroES and ATP (Horwich et al. 2006), GroES and ATP did not aid in the refolding activity of cyanobacterial GroEL (Huq et al. 2010; Potnis et al. 2016). In fact, the refolding activity of *Anabaena* GroEL decreased in the presence of GroES and/or ATP when a substrate native to *Anabaena* was used, but no change was observed with heterologous substrates (Potnis et al. 2016). This suggests that in cyanobacteria, though the *groES* and *groEL* are transcribed from a single operon, they may not be interacting partners, as shown for other bacterial Hsp60 proteins, and this could also be due to the absence of the possible GroES interacting sites on the GroEL proteins of cyanobacteria. The possible non-requirement of the interaction between GroES and GroEL proteins was exemplified by the observation that the overexpression of only GroEL resulted in similar thermotolerance to that observed when the entire *groESL* operon was constitutively expressed in *Anabaena* (Potnis et al. 2016).

The Cpn60/GroEL2 chaperonins exhibited lower ability to refold thermally denatured protein substrates (Huq et al. 2010; Potnis et al. 2016), but Cpn60 could efficiently refold chemically denatured substrate at par with GroEL (Fig. 7.6). The "GGM" repeat of *E. coli* GroEL has been shown to be involved in controlling the volume of the folding cavity (Tang et al. 2008), but with *Anabaena* Cpn60 being a monomer, it was expected to play no role in the refolding abilities of Cpn60. However, the deletion of the tail resulted in complete loss of the refolding activity of *Anabaena* Cpn60, which could, however, be regained in the presence of GroES and/or ATP (Potnis et al. 2016). This suggests possible role of GroES in aiding the activity of Cpn60, which was also suggested for the GroEL2 protein of *Synechococcus*. The low *in vitro* activity of Cpn60 chaperonin does not coordinate well with the observed efficiency in contributing to thermotolerance in vivo, indicating the possible presence of yet to be identified co-chaperonin of Cpn60/GroEL2 required for its complete activity. Thus, the biochemical analysis of the two chaperonins reveals distinct activities for the two proteins, which can further explain the inability for the two genes to complement each other in cyanobacteria, in case of mutation in one of them.

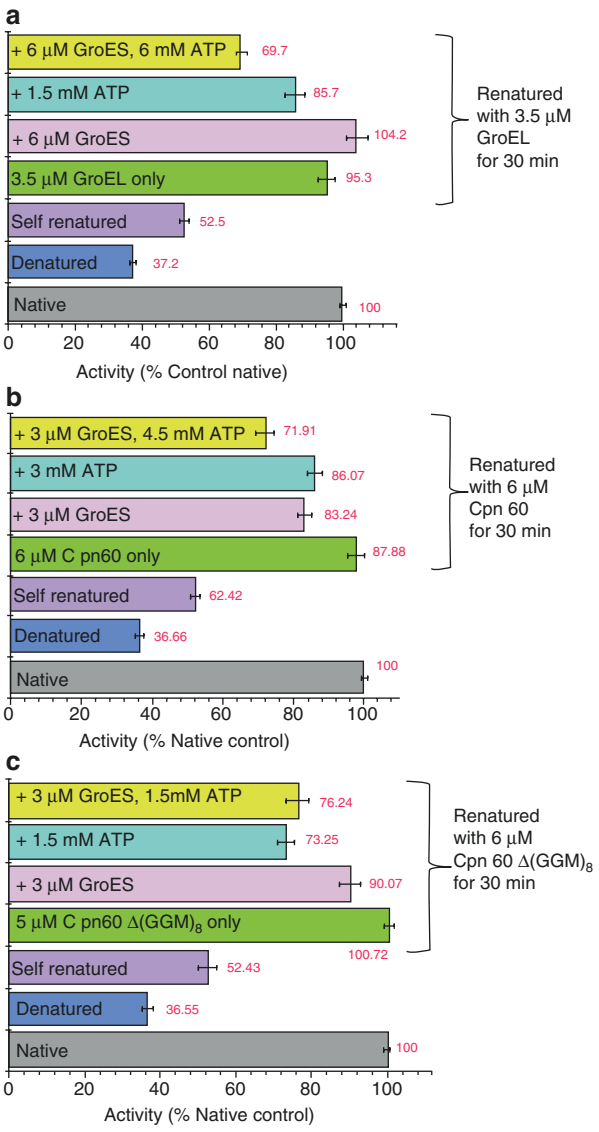


Fig. 7.6 Refolding of urea denatured malate dehydrogenase (MDH) in the presence of Hsp60, GroES, and ATP. MDH was exposed to 2 M urea at 37 °C for 30 min followed by tenfold dilution in refolding buffer. Refolding assays contained chemically denatured MDH and either (**a**) 3.5 μM GroEL, (**b**) 6 μM Cpn60, or (**c**) 5 μM Cpn60Δ$(GGM)_6$ in the presence/absence of specified concentrations of GroES and/or ATP and incubated at 37 °C for 30 min. MDH activity was measured as NADH oxidation in MDH reaction buffer containing 10 mM DTT, 0.5 mM oxaloacetate, 0.28 mM NADH, and 0.15 M potassium phosphate buffer, pH 7.4 in 2 mL final volume at 25 °C. Activity was expressed as % of native enzyme, which was taken as 100%

7.5 Conclusions

The cyanobacterial chaperonins (GroEL/GroEL1 and Cpn60/GroEL2) perform important functions of maintaining the integrity of the enzyme complexes involved in vital metabolic processes. As a result, they contribute significantly to the thermotolerance and salt tolerance of the recombinant strains constitutively overexpressing these proteins. A division of labor between the two Hsp60 proteins renders both the

genes essential. Multiple regulatory elements for the two chaperonin genes ensure a well-coordinated action during different abiotic stress conditions as well as availability of light. Biochemical analysis of the chaperonins indicates a different mode of functioning compared to *E. coli* GroEL. The oligomeric *Anabaena* GroEL functioned in a GroES- and ATP-independent manner, while the Cpn60 functioned as a monomer. These observed differences in biochemical activity and physiological activities may have resulted in cyanobacteria to maintain two chaperonin genes unlike *E. coli* which has only one chaperonin gene.

References

Agrawal C, Sen S, Singh S, Rai S, Singh PK, Singh VK, Rai LC (2014) Comparative proteomics reveals association of early accumulated proteins in conferring butachlor tolerance in three N_2-fixing *Anabaena* spp. J Proteome 96:271–290

Apte SK, Fernandes T, Badran H, Ballal A (1998) Expression and possible role of stress-responsive proteins in *Anabaena*. J Biosci 23:399–406

Asadulghani, Suzuki Y, Nakamoto H (2003) Light plays a key role in the modulation of heat shock response in the cyanobacterium *Synechocystis* sp PCC 6803. Biochem Biophys Res Commun 306:872–879

Berman-Frank I, Lundgren P, Chen YB, Kupper H, Kolber Z, Bergman B, Falkowski P (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. Science 294:1534–1537

Borbely G, Suranyi G, Korcz A, Palfi Z (1985) Effect of heat shock on protein synthesis in the cyanobacterium *Synechococcus* sp. strain PCC 6301. J Bacteriol 161:1125–1130

Brock TD (1973) Evolutionary and ecological aspects of the cyanophytes. In: Carr NG, Whitton BA (eds) The biology of blue-green algae. Blackwell, Oxford, pp 487–500

Castenholz RW (1968) The behavior of *Oscillatoria terebriformis* in hot springs. J Phycol 4:132–139

Castielli O, De la Cerda B, Navarro JA, Hervas M, De la Rosa MA (2009) Proteomic analyses of the response of cyanobacteria to different stress conditions. FEBS Lett 583:1753–1758

Červený J, Sinetova MA, Zavřel T, Los DA (2015) Mechanisms of high temperature resistance of *Synechocystis* sp. PCC 6803: an impact of histidine kinase 34. Life (Basel) 5:676–699

Chaurasia AK, Apte SK (2009) Overexpression of the *groESL* operon enhances the heat and salinity stress tolerance of the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. Appl Environ Microbiol 75:6008–6012

Chitnis PR, Nelson N (1991) Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC 6803. J Biol Chem 266:58–65

Cookson MJ, Baird PN, Hall LM, Coates AR (1989) Identification of two unknown reading frames in *Synechococcus* 6301 as homologues of the 10k and 65k antigen genes of *Mycobacterium tuberculosis* and related heat shock genes in *E. coli* and *Coxiella burnetii*. Nucleic Acids Res 17:6392

Cozens AL, Walker JE (1987) The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. Support for an endosymbiotic origin of chloroplasts. J Mol Biol 194:359–383

Eriksson MJ, Clark AK (1996) The heat shock protein ClpB mediates the development of thermotolerance in the cyanobacterium *Synechococcus* sp. strain PCC 7942. J Bacteriol 178:4839–4846

Furuki M, Tanaka N, Hiyama T, Nakamoto H (1996) Cloning, characterization and functional analysis of *groEL*-like gene from thermophilic cyanobacterium *Synechococcus vulcanus*, which does not form an operon with *groES*. Biochim Biophys Acta 1294:106–110

Glatz A, Horváth I, Varvasovszki V, Kovács E, Török Z, Vigh L (1997) Chaperonin genes of the *Synechocystis* PCC 6803 are differentially regulated under light-dark transition during heat stress. Biochem Biophys Res Commun 239:291–297

Haselkorn R (1978) Heterocysts. Annu Rev Plant Physiol 29:319–344

Hendrix RW (1979) Purification and properties of GroE, a host protein involved in bacteriophage assembly. J Mol Biol 129:375–392

Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. Chem Rev 106:1917–1930

Huq S, Sueoka K, Narumi S, Arisaka F, Nakamoto H (2010) Comparative biochemical characterization of two GroEL homologs from the cyanobacterium *Synechococcus elongatus* PCC 7942. Biosci Biotechnol Biochem 74:2273–2280

Kanemori M, Mori H, Yura T (1994) Induction of heat shock proteins by abnormal proteins results from stabilization and not increased synthesis of sigma-32 in *Escherichia coli*. J Bacteriol 176:5648–5653

Kojima K, Nakamoto H (2007) A novel light- and heat-responsive regulation of the *groE* transcription in the absence of HrcA or CIRCE in cyanobacteria. FEBS Lett 581:1871–1880

Kovács E, van der Vies SM, Glatz A, Török Z, Varvasovszki V, Horváth I, Vígh L (2001) The chaperonins of *Synechocystis* PCC 6803 differ in heat inducibility and chaperone activity. Biochem Biophys Res Commun 289:908–915

Lehel C, Gombos Z, Torok Z, Vigh L (1993a) Growth temperature modulates thermotolerance and heat shock response of cyanobacterium *Synechocystis* PCC 6803. Plant Physiol Biochem 31:81–88

Lehel C, Los D, Wada H, Györgyei J, Horváth I, Kovács E, Murata N, Vigh L (1993b) 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

Lehel C, Wada H, Kovács E, Török Z, Gombos Z, Horváth I, Murata N, Vigh L (1992) Heat shock protein synthesis of the cyanobacterium *Synechocystis* PCC 6803: purification of the GroEL-related chaperonin. Plant Mol Biol 18:327–336

Ludwig M, Bryant DA (2012) *Synechococcus* sp. strain PCC 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and Mixotrophic growth conditions. Front Microbiol 3:354. doi:10.3389/fmicb.2012.00354

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

Mamedov MD, Hayashi H, Murata N (1993) Effects of glycinebetaine and unsaturation of membrane lipids on heat stability of photosynthetic electron transport and phosphorylation reactions in *Synechocystis* PCC 6803. Biochim Biophys Acta 1142:1–5

Mishra Y, Chaurasia N, Rai LC (2009) Heat pretreatment alleviates UV-B toxicity in the cyanobacterium *Anabaena doliolum*: a proteomic analysis of cross tolerance. Photochem Photobiol 85:824–833

Mitsui A, Kumazawa S, Takahashi A, Ikemoto H, Cao S, Arai T (1986) Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. Nature 323:720–722

Morimoto RI, Jurivich DA, Kroeger PE, Mathur SK, Murphy SP, Nakai A, Sarge K, Abravaya K, Sistonen LT (1994) Regulation of heat shock gene transcription by a family of heat shock factors. In: Morimoto RI, Tissiers A, Georgopoulos C (eds) The biology of heat shock proteins and molecular chaperone. Cold Spring Harbor Lab Press, New York, pp 417–455

Nakamoto H, Suzuki M, Kojima K (2003) Targeted inactivation of the *hrcA* repressor gene in cyanobacteria. FEBS Lett 549:57–62

Nikkinen HL, Hakkila K, Gunnelius L, Huokko T, Pollari M, Tyystjarvi T (2012) The SigB σ factor regulates multiple salt acclimation responses of the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Physiol 158:514–523

Panda B, Basu B, Rajaram H, Apte SK (2014) Methyl viologen responsive proteome dynamics of *Anabaena* sp. strain PCC 7120. Proteomics 14:1895–1904

Panda B, Basu B, Rajaram H, Apte SK (2015) Comparative proteomics of oxidative stress response in three cyanobacterial strains native to Indian paddy fields. J Proteome 127:152–160

Pederson DM, Daday A, Smith GD (1986) The use of nickel to probe the role of hydrogen metabolism in cyanobacterial nitrogen fixation. Biochimie 68:113–120

Potnis AA, Rajaram H, Apte SK (2016) GroEL of the nitrogen-fixing cyanobacterium *Anabaena* sp. strain L-31 exhibits GroES and ATP-independent refolding activity. J Biochem 159:295–304

Rai S, Agrawal C, Shrivastava AK, Singh PK, Rai LC (2014) Comparative proteomics unveils cross species variations in *Anabaena* under salt stress. J Proteome 98:254–270

Rajaram H, Apte SK (2003) Heat-shock response and its contribution to thermotolerance of the nitrogen-fixing cyanobacterium *Anabaena* sp. strain L-31. Arch Microbiol 179:423–429

Rajaram H, Apte SK (2008) Nitrogen status and heat-stress-dependent differential expression of the *cpn60* chaperonin gene influences thermotolerance in the cyanobacterium *Anabaena*. Microbiology 154:317–325

Rajaram H, Apte SK (2010) Differential regulation of *groESL* operon expression in response to heat and light in *Anabaena*. Arch Microbiol 192:729–738

Rajaram H, Ballal AD, Apte SK, Wiegert T, Schumann W (2001) Cloning and characterization of the major *groESL* operon from a nitrogen-fixing cyanobacterium *Anabaena* sp. strain L-31. Biochim Biophys Acta 1519:143–146

Rajaram H, Chaurasia AK, Apte SK (2014) Cyanobacterial heat-shock response: role and regulation of molecular chaperones. Microbiology 160:647–658

Robson RL, Postgate JR (1980) Oxygen and hydrogen in biological nitrogen fixation. Annu Rev Microbiol 34:183–207

Sato S, Ikeuchi M, Nakamoto H (2008) Expression and function of a *groEL* paralog in the thermophilic cyanobacterium *Thermosynechococcus elongatus* under heat and cold stress. FEBS Lett 582:3389–3395

Sato M, Nimura-Matsune K, Watanabe S, Chibazakura T, Yoshikawa H (2007) Expression analysis of multiple *dnaK* genes in the cyanobacterium *Synechococcus elongatus* PCC 7942. J Bacteriol 189:3751–3758

Schopf JW (1975) Precambrian paleobiology: problems and perspectives. Annu Rev Earth Planet Sci 3:213–249

Schulz A, Schumann W (1996) *hrcA*, the first gene of the *Bacillus subtilis dnaK* operon encodes a negative regulator of class I heat shock genes. J Bacteriol 178:1088–1093

Shrivastava AK, Chatterjee A, Yadav S, Singh PK, Singh S, Rai LC (2015) UV-B stress induced metabolic rearrangements explored with comparative proteomics in three *Anabaena* species. J Proteome 127:122–133

Singh PK, Shrivastava AK, Chatterjee A, Pandey S, Rai S, Singh S, Rai LC (2015) Cadmium toxicity in diazotrophic *Anabaena* spp. adjudged by hasty up-accumulation of transporter and signaling and severe down-accumulation of nitrogen metabolism proteins. J Proteome 127:134–146

Slabas AR, Suzuki I, Murata N, Simon WJ, Hall JJ (2006) Proteomic analysis of the heat shock response in *Synechocystis* PCC 6803 and a thermally tolerant knockout strain lacking the histidine kinase 34 gene. Proteomics 6:845–864

Stewart WDP (1980) Some aspects of structure and function in N_2-fixing cyanobacteria. Annu Rev Microbiol 34:497–536

Suzuki I, Simon WJ, Slabas AR (2006) The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics. J Exp Bot 57:1573–1578

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

Tang YC, Chang HC, Chakraborty K, Hartl UF, Hayer-Hartl M (2008) Essential role of the chaperonin folding compartment *in vivo*. EMBO J 27:1458–1468

Webb R, Reddy KJ, Sherman LA (1990) Regulation and sequence of the *Synechococcus* sp. strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. J Bacteriol 172:5079–5088

Wynn-Williams DD (1990) Ecological aspects of Antarctic microbiology. Adv Microbiol Ecol 11:71–146

Yamazawa A, Takeyama H, Takeda D, Matsunaga T (1999) UV-A-induced expression of GroEL in the UV-A-resistant marine cyanobacterium *Oscillatoria* sp. NKBG 091600. Microbiology 145:949–954

Yura T, Nakahigashi K (1999) Regulation of the heat-shock response. Curr Opin Microbiol 2:153–158

Zuber U, Schumann W (1994) CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *Bacillus subtilis*. J Bacteriol 176:1359–1363

Part III
Multiple Chaperonins of Archaeal System

Chapter 8
Functional Distribution of Archaeal Chaperonins

Le Gao and Shinsuke Fujiwara

Abstract Chaperonin, also known as heat shock protein 60 (Hsp60), belongs to an evolutionarily conserved protein family that enables cells to survive under stressful conditions, including elevated or reduced temperature, high or low salinity, acidic/alkaline environments, and high or low osmotic pressure. Archaeal genomes generally contain multiple genes encoding chaperonins. For example, the hyperthermophilic archaeon *Thermococcus kodakarensis*, which grows optimally at 85 °C, has both a cold-inducible (CpkA) and a heat-inducible (CpkB) chaperonin, which are involved in adaptation to low and high temperatures, respectively. These two chaperonins share high sequence identity (77%), except in their carboxy-terminal regions. Furthermore, depletion of *cpkA* or *cpkB* results in growth defects under cold stress (60 °C) or heat stress (93 °C), respectively, but not at the optimal temperature (85 °C). These observations indicate that CpkA and CpkB are necessary for cell growth at lower and higher stressed temperatures, respectively. Immunoprecipitation studies using specific antisera revealed that CpkA and CpkB recognize different types of proteins, i.e., that they have distinct substrate spectra. Likewise, several extremophilic archaea encode paralogous chaperonins that are differentially regulated during stresses such as heat, cold, high salt, pH, pressure, and nutrient deprivation, suggesting that these chaperonins might encounter different substrates depending on the type of stress confronting the cell. Extra chaperonin genes may have arisen to assist in protein folding under

L. Gao
Research Center for Environmental Bioscience, Graduate School of Science and Technology, Kwansei-Gakuin University, Sanda, Hyogo 669-1337, Japan

S. Fujiwara (✉)
Research Center for Environmental Bioscience, Graduate School of Science and Technology, Kwansei-Gakuin University, Sanda, Hyogo 669-1337, Japan

Department of Bioscience, School of Science and Technology, Kwansei-Gakuin University, Sanda, Hyogo 669-1337, Japan
e-mail: fujiwara-s@kwansei.ac.jp

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_8

types of selective pressure in which the constitutively expressed chaperonin is unable to function, and here we speculate about the significance of multiple chaperonin genes in archaea.

8.1 Multiplicity in Chaperonin Genes in Archaea

Thermophilic microorganisms grow optimally above 45 °C and inhabit environments with higher temperatures such as hot springs, terrestrial solfatara, deep-sea hydrothermal vents, and composting organic matter. Extreme thermophiles, also known as hyperthermophiles, grow optimally above 80 °C (Stetter 1996); these species are distributed throughout the archaeal and bacterial domains and are positioned near the root of the microbial phylogenetic tree (Fig. 8.1). Such placement has led to the speculation that most ancient life forms dwelt in hotter environments and that these ancestors gradually evolved into modern-day microorganisms, which subsequently adapted to cooler environments.

Archaeal chaperonins are orthologous to the bacterial chaperonins. Bacteria have genes encoding the chaperonin GroEL and co-chaperonin GroES, which form an oligomeric complex. By contrast, archaea lack a GroES homologue; therefore, their

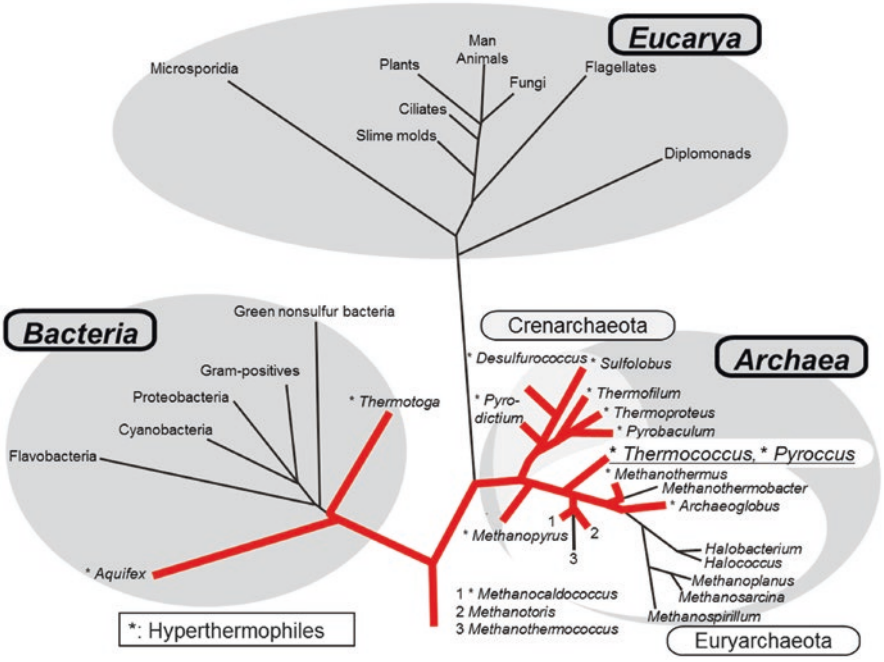


Fig. 8.1 Phylogenetic tree based on sequences of 16S (18S) rRNA genes. *Asterisks* indicate hyperthermophiles. *Bold red lines* indicate speculated evolutionary routes for hyperthermophiles

chaperonins function as oligomeric complexes of 60 kDa group II chaperonin subunits in the absence of a co-chaperonin (Ditzel 1998). According to currently available archaeal genomic data, hosted at http://archaea.ucsc.edu/genomes/archaea/, all the Crenarchaeotal archaea have multiple copies of chaperonin genes, which encode chaperonin homologues (paralogs) with almost identical amino acid sequences, except for a few regions such as the carboxy terminus. Interestingly, about two-thirds of the Crenarchaeotal genomes have two copies of chaperonin genes, whereas the remaining one-third have three; accordingly, the encoded chaperonin protomers assemble into hetero-oligomers with two or three different types of subunits. In the *Euryarchaeota*, the situation is slightly different: approximately 22% of genomes have a single chaperonin gene, and the encoded chaperonin must therefore form a homo-oligomeric complex. About half of *Euryarchaeota* have two chaperonin genes, whereas 22 and 5% have three and four genes, respectively. Accordingly, the encoded chaperonins form hetero-oligomers with two, three, or four types of subunits. About 30% of bacteria have multiple chaperonins that are believed to serve different functions and have different substrate specificities (Lund 2009; Kumar et al. 2015). Compared with bacteria, archaea have higher variation in copy number. In particular, the copy number of the chaperonin genes in archaea is strongly correlated with the number of stresses confronting a given species, as we describe below. *Crenarchaeota*, whose optimal growth temperatures range from 55 to 100 °C, generally have two copies of chaperonin genes, and the encoded subunits are thought to participate in the low- and high-temperature growth phases of the organism. Interestingly, acidophilic archaea of genera *Acidianus* and *Sulfolobus* have three copies of chaperonin genes, suggesting that the third copy is involved in adaptation to acidic conditions. Consistent with this overall principle, hyperthermophiles with a narrow optimal range of growth temperatures, such as the methanogens and genus *Pyrococcus*, have only one chaperonin gene. Most of the *Euryarchaeota*, which have relatively wide optimal temperature ranges (50–100 °C) or tolerate acidic conditions (albeit within a narrow pH range), have two types of chaperonin genes. Species with three types of chaperonin genes are generally thermophiles, halophiles, or alkaliphiles that can grow at wider ranges of temperature, salinity, or pH. *Haloarcula*, *Halopiger*, *Halorhabdus*, and *Natronomonas*, which possess four kinds of chaperonin genes, are alkaliphilic halophiles that grow in environments with relatively wide ranges of salinity and alkaline pH.

8.2 Multiple Chaperonin Subunits in Thermophiles

Most studies have focused on the chaperonins of thermophiles. Species within *Pyrococcus*, a genus of extremophiles that grow at high temperatures within a relatively narrow range (70–103 °C), have homo-oligomeric chaperonins (Okochi et al. 2005; Chen et al. 2007; Garrity 2001). On the other hand, members of *Thermococcus*, which like *Pyrococcus* is a member of the order *Thermococcales*, have two chaperonin genes. *Thermococcus kodakarensis*, which has a relatively

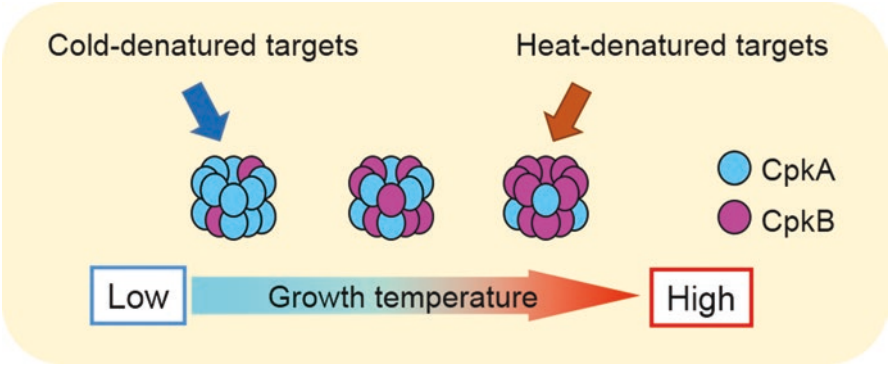


Fig. 8.2 Temperature-dependent molecular chaperonin complex in *T. kodakarensis*. The ratio of subunits CpkA (*blue circle*) and CpkB (*red circle*) changes with growth temperature. Meanwhile, the substrate subsets for each oligomer change with the subunit ratio. *Blue circles*, CpkA; *red circles*, CpkB

wide temperature range (55–100 °C) (Yan et al. 1997; Izumi et al. 1999), has two chaperonin subunits, CpkA and CpkB. Expression of these proteins responds to cold and heat stresses *in vivo* (Izumi et al. 2001). Co-immunoprecipitation (co-IP) experiments revealed that CpkA and CpkB are predominantly expressed below 70 °C and above 90 °C, respectively. In addition, the co-IPs showed that CpkA and CpkB form a hetero-oligomeric complex *in vivo* (~90%) and that the composition of subunits changes depending on temperature (Fig. 8.2). When *cpkA* or *cpkB* was deleted from the *T. kodakarensis* chromosome, the Δ*cpkA* cells grew poorly at 60 °C, whereas Δ*cpkB* cells had a growth defect at 93 °C, indicating that homo-oligomeric chaperonins are functional at the corresponding optimal temperatures, as well as suggesting that the two proteins cannot cross complement each other at either low or high temperatures (Fujiwara et al. 2008). A similar evaluation was performed for *Thermococcus* strain KS-1, which also has orthologs of CpkA and CpkB, and this strain behaved similarly *in vivo*. Expression of the beta subunit (CpkB ortholog) increased with temperature, whereas that of the alpha subunit did not (Yoshida et al. 2001). Thus, the alpha and beta subunits respond to low- and high-temperature stresses, respectively, and consequently the CpkA/CpkB ratio changes with temperature (Fig. 8.2). Chaperonin complexes with different subunit compositions may have different functional properties, as we discuss in more detail in the following subsections.

8.2.1 Multiple Subunits and Thermostability

Further studies at the molecular level revealed that the two chaperonin subunits of *Thermococcus* strain KS-1 have different thermostabilities (Yoshida et al. 2002). Thermal denaturation experiments showed that the beta subunit is more stable than

the alpha subunit. Hence, in the hetero-oligomeric chaperonin, higher levels of the beta subunit improve the thermostability of the complex. The C-terminal region of the beta subunit is believed to be responsible for this protein's higher thermostability (Yoshida et al. 2006). Replacement of the C-terminal region of the beta subunit with the corresponding sequence from the alpha subunit decreased the thermostability of the beta subunit to that of the alpha subunit, confirming that the C-terminal region plays an important role in chaperonin stability. The *Pyrococcus furiosus* chaperonin (Pfcpn) is very similar to the beta subunit (CpkB) of *Thermococcus* (Luo and Robb 2011); mutation of its C-terminal region revealed that an EK-rich motif (528-EKEKEKEGEK-537) is important for high thermostability. In a chaperonin from the psychrophilic methanogenic archaeon *Methanococcoides burtonii*, replacement of this motif with the sequence from a thermophile increased thermostability and thermoactivity. The N-terminal region of Pfcpn is involved in assembly of a double ring, and a mutational study revealed that a salt bridge between Arg22 and Glu37 stabilizes the complex (Luo et al. 2011). Using cryo-electron microscopy, another group showed that the N- and C-terminal regions of a chaperonin from *Acidianus tengchongensis* interweave with each other in the central cavity of the complex (Zhang et al. 2013). Together, these results demonstrate that both the N- and C-terminal regions are important for the stability of the chaperonin complex.

8.2.2 Relation Between ATPase Activity and Chaperonin Function

The ATP hydrolysis cycle is important for the chaperonin reaction cycle, i.e., chaperonin activity is determined by ATPase activity (Iizuka et al. 2003; Sekiguchi et al. 2013). Several studies focused on the temperature dependencies of archaeal chaperonins. The cold-inducible CpkA from *T. kodakarensis*, which grows optimally at 85 °C, exhibited maximal ATPase activity around 65 °C, whereas CpkB remained active up to 93 °C (Gao et al. 2015). These two chaperonins share high sequence identity (77%), except in their C-terminal 20 amino acids. Alignment of the final 20 amino acid residues of chaperonins across prokaryotes with different optimal growth temperatures revealed three types of chaperonin: CpkA-type (with a GGM repeat motif), CpkB-type (rich in charged amino acids), and C-type (unlike the other two types) (Fig. 8.3). Notably, the copy number of CpkA-type chaperonin genes increased as the optimal cell growth temperature decreased (Fig. 8.4). Based on this pattern, we speculate that, in *T. kodakarensis*, CpkA-type chaperonin is involved in handling cold-denatured proteins under low-temperature stress. To explore this idea, we constructed a unique mutant strain, DB3 (*ΔpyrF, ΔcpkB, ΔcpkA1-524*::*cpkB1-524*). In DB3, amino acid residues 1–524 of CpkA were replaced with the corresponding residues of CpkB to construct a chimeric *cpkBA* gene [*cpkA*(1–524)::*cpkB*(1–524)], consisting of amino acid residues 1–524 from CpkB and 525–548 from CpkA; the latter sequence contains the C-terminal region

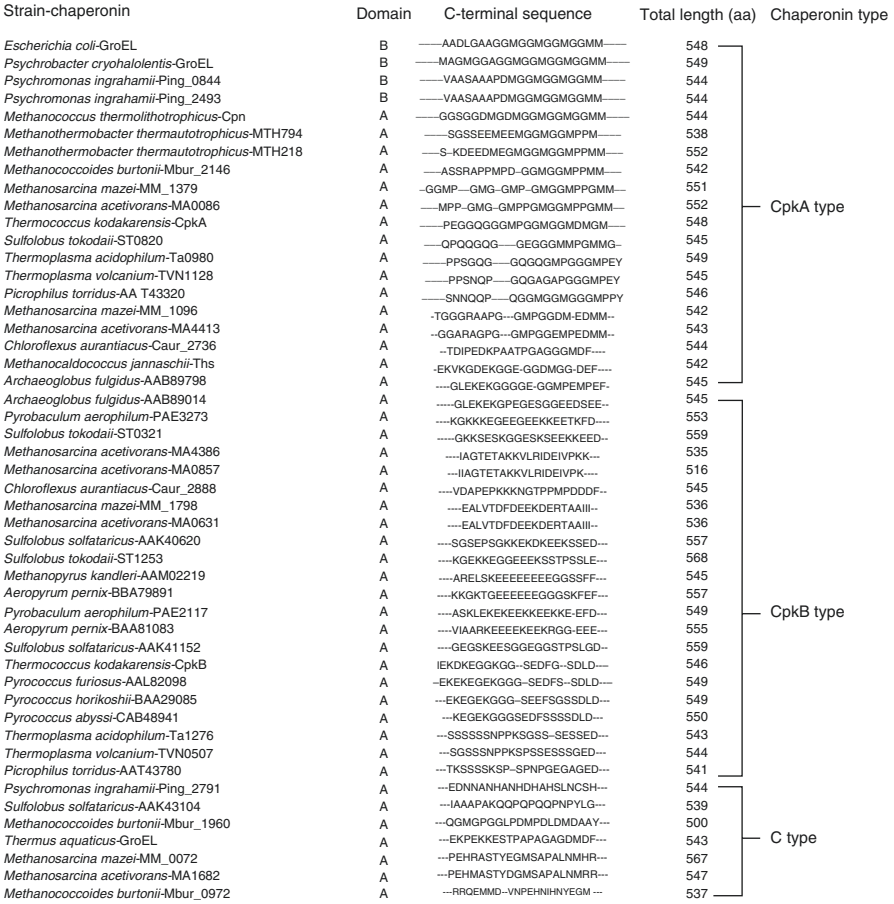


Fig. 8.3 Alignment of C-terminal regions (last 20 amino acids) of chaperonins from archaea and bacteria. Three types of chaperonin are clustered: CpkA-type (A-type), which consists of a "GGM" motif; CpkB-type (B-type), which is rich in charged and polar amino acids; and chimeric (C-type). A and B in the domain section refer to Archaea and Bacteria, respectively. Protein sequences were retrieved from the public National Center for Biotechnology Information (NCBI) database (Coordinators 2016)

(GGM motif). The *cpkBA* chimeric gene was placed under the control of the *cpkA* promoter, making it cold-inducible. DB3 exhibited recovery of growth at 60 °C (Fig. 8.5), indicating that the C-terminal region of CpkA plays a key role in growth at this temperature. Next, we asked whether, if CpkA could be made to function at even lower temperatures, it would help cells grow under these conditions. To explore this possibility, we constructed mutant CpkA that is more active at 50 °C. Specifically, replacement of Glu530 with Gly (CpkA-E530G) increased ATPase and *in vitro* refolding activities at 50 °C relative to those of wild-type

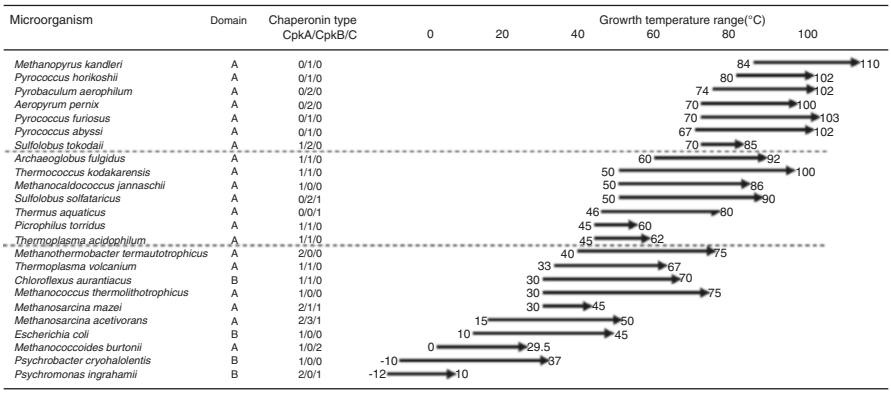


Fig. 8.4 Distribution of chaperonins according to growth temperature range. A and B in the domain section refer to archaea and bacteria, respectively. Growth conditions for each microorganism were obtained from the following sources: *Methanopyrus kandleri* (Garrity 2001), *Pyrococcus horikoshii* (Gonzalez et al. 1998), *Pyrobaculum aerophilum* (Volkl et al. 1993), *Aeropyrum pernix* (Sako et al. 1996), *Pyrococcus furiosus* (Garrity 2001), *Pyrococcus abyssi* (Garrity 2001), *Sulfolobus tokodaii* (Toshio Iwasaki Group Homepage), *Archaeoglobus fulgidus* (Stetter et al. 1987), *Thermococcus kodakarensis* (Izumi et al. 1999), *Methanocaldococcus jannaschii* (Jeanthon et al. 1999), *Sulfolobus solfataricus* (Garrity 2001), *Thermus aquaticus* (Brock and Freeze 1969), *Picrophilus torridus* (Garrity 2001), *Thermoplasma acidophilum* (Darland et al. 1970), *Methanothermobacter thermautotrophicus* (Zeikus and Wolfe 1972), *Thermoplasma volcanium* (Garrity 2001), *Chloroflexus aurantiacus* (Pierson and Castenholz 1974), *Methanococcus thermolithotrophicus* (Zeikus and Wolfe 1972), *Methanosarcina mazei* (Maestrojuan and Boone 1991), *Methanosarcina acetivorans* (Maestrojuan and Boone 1991), *Escherichia coli* (Jones et al. 1987), *Methanococcoides burtonii* (Garrity 2001), *Psychrobacter cryohalolentis* (Bakermans et al. 2006), and *Psychromonas ingrahamii* (Auman et al. 2006)

CpkA. Consistent with this, the mutant strain DA4, which carries the *cpkA-E530G* mutation, grew faster than the wild type at 50 °C (Gao et al. 2015). These results indicated that the C-terminal region of CpkA is important for ATPase and chaperonin activities. Notably, a single-residue mutation (*cpkA-E530G*) enabled cells to adapt to a cooler environment. Thus, chaperonins may play an important role in surviving evolutionary pressures. GroEL overexpression causes changes in substrate specificity, increasing genetic variation and promoting enzyme evolution (Tokuriki and Tawfik 2009). GroEL, or a GroEL-like chaperonin such as CpkA, may specifically recognize these kinds of intermediates and participate in altering their characteristics. Some mutations of the C-terminal region would confer additional evolvability at lower temperatures on *T. kodakarensis*. We speculate that *cpkA* evolved from a parental chaperonin, probably *cpkB*, to adapt to cooler environments over the course of evolution (Fig. 8.6). Thus, *cpkB* might have been present as the original chaperonin gene in ancestral *Thermococcus*, and prototype of *cpkA* would thus have evolved as a *cpkB* paralog, probably via gene duplication. As cells were confronted by cooler environments, *cpkA* have arisen to assist in refolding of cold-denatured cytoplasmic proteins. The C-terminal regions are the most

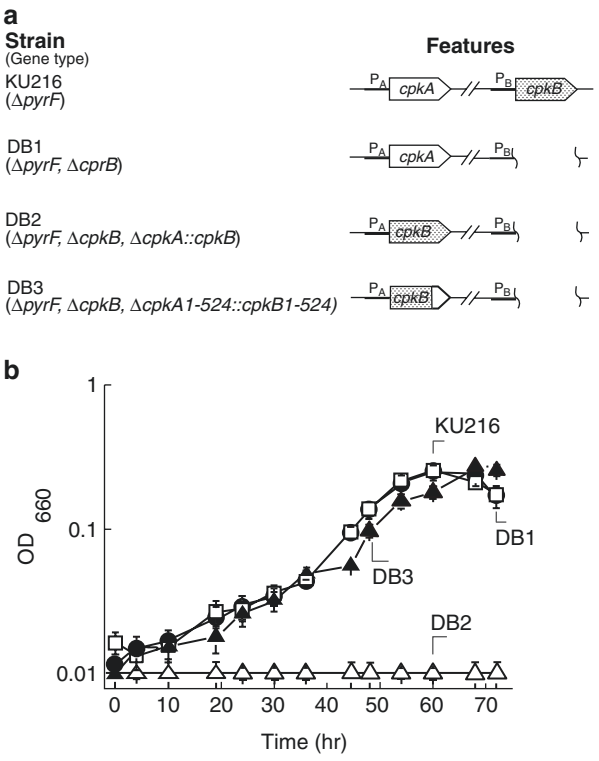


Fig. 8.5 Effect of various *cpk* gene disruptions on cell growth. (**a**) Architectures of disruptants. *White* and *dotted arrows* indicate *cpkA* and *cpkB* genes, respectively. *Double slash* indicates gene disruption, and P_A and P_B indicate the promoter regions of respective genes. (**b**) Growth profiles of various *cpk* disruptants at 60 °C. *Closed circles*, parental-type KU216 (*ΔpyrF*); *open box*, DB1 (*ΔpyrF, ΔcpkB*); *open triangles*, DB2 (*ΔpyrF, ΔcpkA*::*cpkB, ΔcpkB*); *closed triangles*, DB3 (*ΔpyrF, ΔcpkB, cpkA1-524*::*cpkB1-524*). Growth monitoring was conducted in triplicate, and the *plotted data* represent the mean values

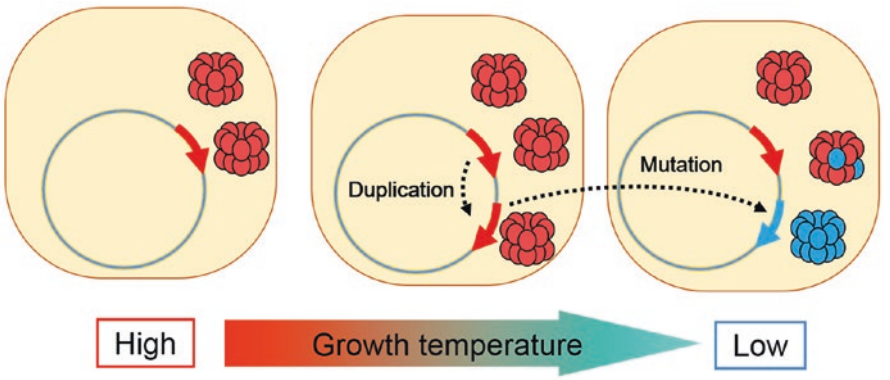


Fig. 8.6 Tentative model of cold adaptation by chaperonin gene duplication and mutation in hyperthermophiles. Original thermophilic chaperonin and cold-adapted chaperonin are indicated in *red* and *blue*, respectively

variable portions of archaeal chaperonins. Chaperonins might recognize target proteins depending on properties of their C-termini. We discuss this possibility in the following section.

8.3 Multiple Subunits and Their Specific Substrates

Deletion mutants DA1 (Δ*cpkA*) and DB1 (Δ*cpkB*) of *T. kodakarensis* exhibited normal cell growth at 85–93 °C and 60–85 °C, respectively, indicating that homo-oligomeric CpkA and CpkB are functional *in vivo* in the corresponding temperature ranges (Fujiwara et al. 2008). Homo-oligomeric CpkA and CpkB are also functional and capable of recognizing cold-denatured and heat-denatured proteins, respectively. Furthermore, a unique mutant strain DB2 (Δ*cpkA*::*cpkB*, Δ*cpkB*), in which *cpkB* was expressed under the control of the *cpkA* promoter, did not grow at 60 °C (Fig. 8.5). The lack of cross complementation indicates that CpkB cannot functionally replace CpkA and that multiple chaperonins in *T. kodakarensis* recognized different types of denatured proteins (stringent substrates) at different temperatures, suggesting that CpkA developed to adapt to cold environments over the course of evolution (Fig. 8.6). To identify stringent substrates of CpkA, we performed co-IP experiments (Gao et al. 2012) and used LC-MS analysis to identify 15 proteins that co-immunoprecipitated with anti-CpkA; all of these proteins had high Mascot scores. There are some classifications, class, fold, superfamily, and family in the SCOP (http://scop.berkeley.edu). The top two proteins had a SCOP fold c.1 (TIM beta/alpha-barrel), and four of the remaining proteins had a SCOP fold c.2 (NAD(P)-binding Rossmann fold domain). Nine of the 15 proteins were of the alpha and beta protein (class c), which consists mainly of parallel beta-sheets. This kind of fold (especially c.1, which has a barrel consisting of eight parallel beta-sheets) is highly prone to be trapped in misfolding pathways during arrangement of beta-sheets, which stringently requires the assistance of a chaperonin (Wu et al. 2007). The highest-scoring candidate CpkA substrate, indole-3-glycerol-phosphate synthase ($TrpC_{Tk}$), was studied further *in vitro*. For this purpose, $TrpC_{Tk}$ was obtained as a recombinant protein and then chemically denatured with urea. Denatured $TrpC_{Tk}$ was refolded by dialysis in the presence or absence of CpkA (or CpkB). An increase in $TrpC_{Tk}$ activity was observed only in the presence of CpkA, suggesting that homo-oligomeric CpkA assisted in refolding of specific denatured states of $TrpC_{Tk}$, whereas homo-oligomeric CpkB did not. Each subunit of chaperonin distinguishes target molecules by an unknown mechanism. Structure-specific activity was reported for a CpkA-type chaperonin of *Methanococcus maripaludis* (Sergeeva et al. 2014), which specifically recognized certain partially folded intermediates of human γD-crystallin, which also contains a beta barrel. The features of the proteins identified in *M. maripaludis* were similar to those of CpkA substrates identified in

T. kodakarensis. Proteins trapped by heat-inducible CpkB in *T. kodakarensis* were also examined by LC-MS, revealing that the substrates had relatively low Mascot scores. Fewer CpkB-trapped proteins had alpha and beta folds than in the case of CpkA-trapped substrates. Two of these proteins were aminopeptidases that also co-immunoprecipitate with anti-chaperonin (anti-SSO0282) of *Sulfolobus solfataricus* (Condo et al. 1998). Moreover, the chaperonin (SSO0282) of *S. solfataricus* is of the CpkB-type, i.e., rich in negatively charged amino acids in the C-terminal region. Positively charged ribosomal proteins of *T. kodakarensis* are trapped by CpkB (data unpublished). *S. solfataricus* chaperonin also interacts with 16S rRNA *in vivo* and thereby participates in RNA processing (Ruggero et al. 1998), suggesting that ribosomal protein-bound 16S rRNA is recognized by the C-terminal region of the chaperonin.

Another group tried to distinguish the substrate subsets of chaperonins from *Methanosarcina mazei* (Hirtreiter et al. 2009). The genome of *M. mazei* encodes four kinds of chaperonin genes (*Mm*GroEL, MM1379; *Mm*Ths, MM1096; MM1798; and MM0072). Co-IP experiments revealed three subsets of chaperonin-captured proteins: *Mm*GroEL-selective interactors, *Mm*Ths-selective interactors, and substrates trapped by both *Mm*GroEL and *Mm*Ths. Compared with *Mm*GroEL substrates, *Mm*Ths substrates are more negatively charged; moreover, *Mm*Ths traps a wider range of domain folds, whereas *Mm*GroEL favors alpha and beta folds. These observations suggest that the multiple chaperonins of *Methanosarcina* respond to different stresses and evolved to favor specific targets *in vivo*, resulting in distinct target specificities.

8.4 Multiple Subunits in Halophiles

Halophiles, including halotolerant species that can grow at high concentrations (>1 M) of sodium chloride (NaCl), also possess several types of chaperonin genes in their genomes. The NaCl concentration ranges and numbers of chaperonin paralogs in the genome are summarized in Fig. 8.7. In comparison with moderate and slight halophiles, all extreme halophiles have two CpkA-type chaperonins with C-terminal mildly hydrophobic "GGM" motifs longer than those of bacterial GroEL. Moreover, the bacterial extreme halophile *Salinibacter ruber* also has two CpkA-type chaperonins.

The halophile *Haloferax volcanii* has three chaperonin genes [*cct1* (HVO0133), *cct2* (HVO0455), and *cct3* (HVO0778)]. All three genes were knocked out, and the growth of the disruptants was monitored (Kapatai et al. 2006). Each gene was individually dispensable. Both CCT1 and CCT2 have CpkA-type sequences, whereas CCT3 has more charged amino acids in the C-terminal region (CCT1, GGMGGGMGGMGGMGGMGGAM; CCT2, GGAPGGMGGMGGMGGMGGAM; CCT3, ESATEAATMIVRIDDVIAAK).

CCT3 may have evolved to respond to a new stress, such as higher salinity. In genus *Halobacterium*, whose members are extreme halophiles, multiple CpkA-type chaperonins are likely to play important roles in salt tolerance. Halophiles use two strategies to cope with a high-salt environment: accumulation of potassium chloride (KCl) to maintain an internal osmotic pressure equivalent to the external pressure (the so-called high salt–in strategy) and production of excess organic compatible solutes to balance osmotic pressure (the "compatible solute–in strategy") (Oren 2008). Under both strategies, halophiles consume more energy for a given growth rate and metabolism than non-halophilic microorganisms (Oren 1999).

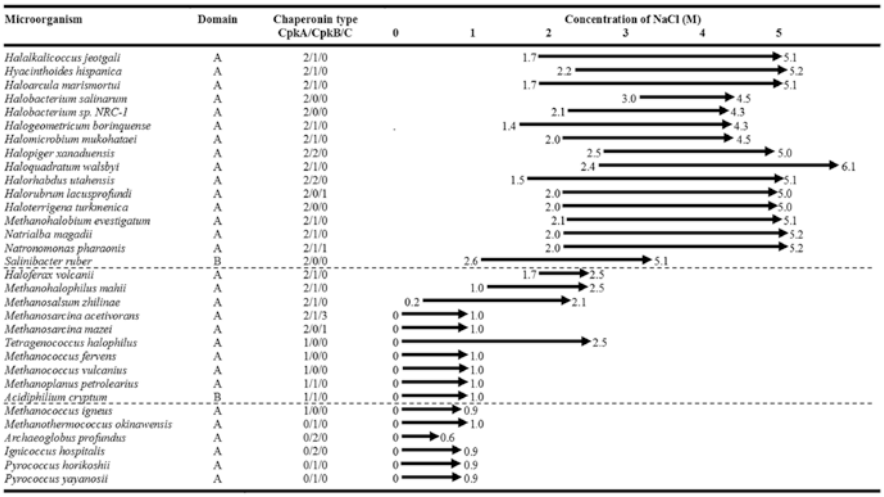


Fig. 8.7 Distribution of chaperonins according to NaCl concentration range. A and B in the domain section refer to Archaea and Bacteria, respectively. Growth conditions for each microorganism were obtained from the following sources: *Halalkalicoccus jeotgali* (Roh et al. 2007), *Hyacinthoides hispanica* (Garrity 2001), *Haloarcula marismortui* (Oren et al. 1990), *Halobacterium salinarum* (Garrity 2001), *Halobacterium sp. NRC-1* (Gruber et al. 2004), *Halogeometricum borinquense* (Montalvo-Rodriguez et al. 1998), *Halomicrobium mukohataei* (Ihara et al. 1997), *Halopiger xanaduensis* (Gutierrez et al. 2007), *Haloquadratum walsbyi* (Burns et al. 2007), *Halorhabdus utahensis* (Waino et al. 2000), *Halorubrum lacusprofundi* (Bowers and Wiegel 2011), *Haloterrigena turkmenica* (Bowers and Wiegel 2011), *Methanohalobium evestigatum* (Garrity 2001), *Natrialba magadii* (Garrity 2001), *Natronomonas pharaonis* (Garrity 2001), *Salinibacter ruber* (Anton et al. 2002), *Haloferax volcanii* (Mullakhanbhai and Larsen 1975), *Methanohalophilus mahii* (Garrity 2001), *Methanosalsum zhilinae* (Garrity 2001), *Methanosarcina acetivorans* (Sowers et al. 1984), *Methanosarcina mazei* (Maestrojuan and Boone 1991), *Tetragenococcus halophilus* (Garrity 2001), *Methanococcus fervens* (Jeanthon et al. 1999), *Methanococcus vulcanius* (Jeanthon et al. 1999), *Methanoplanus petrolearius* (Ollivier et al. 1997), *Acidiphilium cryptum* (Harrison 1981), *Methanococcus igneus* (Garrity 2001), *Methanothermococcus okinawensis* (Takai et al. 2002), *Archaeoglobus profundus* (Burggraf et al. 1990), *Ignicoccus hospitalis* (Jahn et al. 2008), *Pyrococcus horikoshii* (Gonzalez et al. 1998), and *Pyrococcus yayanosii* (Birrien et al. 2011)

8.5 Multiple Subunits in Other Extremophiles

In addition to the extremophiles mentioned above, some other species with multiple chaperonins have also been studied. The piezophile *Thermococcus barophilus*, which has two chaperonin subunits, was studied under various pressure conditions by proteomic and transcriptomic methods (Marteinsson et al. 1999a; Marteinsson et al. 1999b; Vannier et al. 2015). At in situ pressure (40 MPa), no induction of chaperonins occurred, indicating that *T. barophilus* is already adapted to this condition. However, at suboptimal growth conditions [lower (0.1 MPa) or higher (70 MPa) pressure than the optimal 40 MPa or higher temperature (98 °C) than the optimal 80 °C], a CpkB-type chaperonin gene (TERMP_02050) was induced. Another chaperonin (TERMP_00583) of the CpkA-type was consistently expressed under all conditions, indicating that it functions as a housekeeping chaperonin.

Chaperonins have been studied in only a few other extremophiles. Methanogens have greatest diversity in sequence features and subunit numbers. *Methanosarcina acetivorans* has the largest complement of chaperonin genes (six) among Archaea. *M. acetivorans* is a mesophilic methanogen, as well as a moderate halophile and alkaliphile. A comparison of chaperonins from *M. acetivorans* (six genes) and *M. maize* (four subunits) revealed that the amino acid sequences of the two extra genes in *M. acetivorans* have a highly conserved C-terminal region, IAGTETAKKVLRIDEIVPKK(R), which resembles the C-terminal regions of chaperonins from the *Halobacteriaceae* family but has more basic amino acids (Lys and Arg). Because basic amino acids could function as a buffer for acetate ions, these extra chaperonins with basic C-terminal regions may contribute to the unique ability of *M. acetivorans* to metabolize acetate. The mesophile *Methanocella paludicola* has three chaperonin genes (CpkA-type, CpkB-type, and C-type). It is unclear what kinds of stresses are involved in induction of these three chaperonin genes and how they respond to environmental change. In addition, the properties of homo-oligomeric or hetero-oligomeric complexes are likely to differ.

Study of archaea with multiple chaperonins, and the significance of the diversity among these proteins, has just begun. Further genetic studies are required to confirm the speculations described above. We hypothesize that extra chaperonins play important roles in surviving various evolutionary pressures. Extra chaperonin would accept unusual protein structure and promote changes of substrate specificity of enzyme, resulting in enzyme evolution.

References

Anton J, Oren A, Benlloch S, Rodriguez-Valera F, Amann R, Rossello-Mora R (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the bacteria from saltern crystallizer ponds. Int J Syst Evol Microbiol 52(Pt 2):485–491. doi:10.1099/00207713-52-2-485

Auman AJ, Breezee JL, Gosink JJ, Kampfer P, Staley JT (2006) *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. Int J Syst Evol Microbiol 56(Pt 5):1001–1007. doi:10.1099/ijs.0.64068-0

Bakermans C, Ayala-del-Rio HL, Ponder MA, Vishnivetskaya T, Gilichinsky D, Thomashow MF, Tiedje JM (2006) *Psychrobacter cryohalolentis* sp. nov. and *Psychrobacter arcticus* sp. nov., isolated from Siberian permafrost. Int J Syst Evol Microbiol 56(Pt 6):1285–1291. doi:10.1099/ijs.0.64043-0

Birrien JL, Zeng X, Jebbar M, Cambon-Bonavita MA, Querellou J, Oger P, Bienvenu N, Xiao X, Prieur D (2011) *Pyrococcus yayanosii* sp. nov., an obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol 61(Pt 12):2827–2831. doi:10.1099/ijs.0.024653-0

Bowers KJ, Wiegel J (2011) Temperature and pH optima of extremely halophilic archaea: a mini-review. Extremophiles 15(2):119–128. doi:10.1007/s00792-010-0347-y

Brock TD, Freeze H (1969) *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. J Bacteriol 98(1):289–297

Burggraf S, Jannasch HW, Nicolaus B, Stetter KO (1990) *Archaeoglobus-Profundus* Sp-Nov, represents a new species within the sulfate-reducing archaebacteria. Syst Appl Microbiol 13(1):24–28

Burns DG, Janssen PH, Itoh T, Kamekura M, Li Z, Jensen G, Rodriguez-Valera F, Bolhuis H, Dyall-Smith ML (2007) *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. Int J Syst Evol Microbiol 57(Pt 2):387–392. doi:10.1099/ijs.0.64690-0

Chen HY, Chu ZM, Ma YH, Zhang Y, Yang SL (2007) Expression and characterization of the chaperonin molecular machine from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Basic Microbiol 47(2):132–137. doi:10.1002/jobm.200610215

Condo I, Ruggero D, Reinhardt R, Londei P (1998) A novel aminopeptidase associated with the 60 kDa chaperonin in the thermophilic archaeon *Sulfolobus solfataricus*. Mol Microbiol 29(3):775–785

Coordinators NR (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44(DI):D7–19. doi:10.1093/nar/gkv1290

Darland G, Brock TD, Samsonoff W, Conti SF (1970) A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. Science 170(3965):1416–1418

Fujiwara S, Aki R, Yoshida M, Higashibata H, Imanaka T, Fukuda W (2008) Expression profiles and physiological roles of two types of molecular chaperonins from the hyperthermophilic archaeon *Thermococcus kodakarensis*. Appl Environ Microbiol 74(23):7306–7312. doi:10.1128/AEM.01245-08

Gao L, Danno A, Fujii S, Fukuda W, Imanaka T, Fujiwara S (2012) Indole-3-glycerol-phosphate synthase is recognized by a cold-inducible group II chaperonin in *Thermococcus kodakarensis*. Appl Environ Microbiol 78(11):3806–3815. doi:10.1128/AEM.07996-11

Gao L, Imanaka T, Fujiwara S (2015) A mutant Chaperonin that is functional at lower temperatures enables hyperthermophilic archaea to grow under cold-stress conditions. J Bacteriol 197(16):2642–2652. doi:10.1128/JB.00279-15

Garrity GM (2001) Bergey's manual of systematic bacteriology. The archaea and the deeply branching and phototrophic bacteria, 2nd edn. Springer, New York

Gonzalez JM, Masuchi Y, Robb FT, Ammerman JW, Maeder DL, Yanagibayashi M, Tamaoka J, Kato C (1998) *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa trough. Extremophiles 2(2):123–130

Gruber C, Legat A, Pfaffenhuemer M, Radax C, Weidler G, Busse HJ, Stan-Lotter H (2004) *Halobacterium noricense* sp. nov., an archaeal isolate from a bore core of an alpine Permian salt deposit, classification of *Halobacterium* sp. NRC-1 as a strain of *H. salinarum* and emended description of *H. salinarum*. Extremophiles 8(6):431–439. doi:10.1007/s00792-004-0403-6

Gutierrez MC, Castillo AM, Kamekura M, Xue Y, Ma Y, Cowan DA, Jones BE, Grant WD, Ventosa A (2007) *Halopiger xanaduensis* gen. nov., sp. nov., an extremely halophilic archaeon isolated from saline lake Shangmatala in Inner Mongolia, China. Int J Syst Evol Microbiol 57(Pt 7):1402–1407. doi:10.1099/ijs.0.65001-0

Harrison AP (1981) *Acidiphilium-Cryptum* Gen-Nov, Sp-Nov, heterotrophic bacterium from acidic mineral environments. Int J Syst Bacteriol 31(3):327–332

Hirtreiter AM, Calloni G, Forner F, Scheibe B, Puype M, Vandekerckhove J, Mann M, Hartl FU, Hayer-Hartl M (2009) Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. Mol Microbiol 74(5):1152–1168. doi:10.1111/j.1365-2958.2009.06924.x

Ihara K, Watanabe S, Tamura T (1997) *Haloarcula argentinensis* sp. nov. and *Haloarcula mukohataei* sp. nov., two new extremely halophilic archaea collected in Argentina. Int J Syst Bacteriol 47(1):73–77. doi:10.1099/00207713-47-1-73

Iizuka R, Yoshida T, Shomura Y, Miki K, Maruyama T, Odaka M, Yohda M (2003) ATP binding is critical for the conformational change from an open to closed state in archaeal group II chaperonin. J Biol Chem 278(45):44959–44965. doi:10.1074/jbc.M305484200

Izumi M, Fujiwara S, Takagi M, Fukui K, Imanaka T (2001) Two kinds of archaeal chaperonin with different temperature dependency from a hyperthermophile. Biochem Biophys Res Commun 280(2):581–587. doi:10.1006/bbrc.2000.4154

Izumi M, Fujiwara S, Takagi M, Kanaya S, Imanaka T (1999) Isolation and characterization of a second subunit of molecular chaperonin from *Pyrococcus kodakaraensis* KOD1: analysis of an ATPase-deficient mutant enzyme. Appl Environ Microbiol 65(4):1801–1805

Jahn U, Gallenberger M, Paper W, Junglas B, Eisenreich W, Stetter KO, Rachel R, Huber H (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. J Bacteriol 190(5):1743–1750. doi:10.1128/JB.01731-07

Jeanthon C, L'Haridon S, Reysenbach AL, Corre E, Vernet M, Messner P, Sleytr UB, Prieur D (1999) *Methanococcus vulcanius* sp. nov., a novel hyperthermophilic methanogen isolated from East Pacific Rise, and identification of *Methanococcus* sp. DSM 4213T as *Methanococcus fervens* sp. nov. Int J Syst Bacteriol 49(Pt 2):583–589. doi:10.1099/00207713-49-2-583

Jones PG, VanBogelen RA, Neidhardt FC (1987) Induction of proteins in response to low temperature in *Escherichia coli*. J Bacteriol 169(5):2092–2095

Kapatai G, Large A, Benesch JL, Robinson CV, Carrascosa JL, Valpuesta JM, Gowrinathan P, Lund PA (2006) All three chaperonin genes in the archaeon *Haloferax volcanii* are individually dispensable. Mol Microbiol 61(6):1583–1597. doi:10.1111/j.1365-2958.2006.05324.x

Kumar CM, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. Cell Stress Chaperones 20(4):555–74. doi:10.1007/s12192-015-0598-8

Lars Ditzel, Jan Löwe, Daniela Stock, Karl-Otto Stetter, Harald Huber, Robert Huber, Stefan Steinbacher, (1998) Crystal Structure of the Thermosome, the Archaeal Chaperonin and Homolog of CCT. Cell 93(1):125–138

Lund PA (2009) Multiple chaperonins in bacteria – why so many? FEMS Microbiol Rev 33(4):785–800. doi:10.1111/j.1574-6976.2009.00178.x

Luo H, Robb FT (2011) A modulator domain controlling thermal stability in the group II chaperonins of archaea. Arch Biochem Biophys 512(1):111–118. doi:10.1016/j.abb.2011.04.017

Luo H, Zhang P, Robb FT (2011) Oligomerization of an archaeal group II chaperonin is mediated by N-terminal salt bridges. Biochem Biophys Res Commun 413(2):389–394. doi:10.1016/j.bbrc.2011.08.112

Maestrojuan GM, Boone DR (1991) Characterization of Methanosarcina-Barkeri Mst and 227, *Methanosarcina-Mazei* S-6t, and *Methanosarcina-Vacuolata* Z-761t. Int J Syst Bacteriol 41(2):267–274

Marteinsson VT, Birrien JL, Reysenbach AL, Vernet M, Marie D, Gambacorta A, Messner P, Sleytr UB, Prieur D (1999a) *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. Int J Syst Bacteriol 49(Pt 2):351–359. doi:10.1099/00207713-49-2-351

Marteinsson VT, Reysenbach AL, Birrien JL, Prieur D (1999b) A stress protein is induced in the deep-sea barophilic hyperthermophile *Thermococcus barophilus* when grown under atmospheric pressure. Extremophiles 3(4):277–282

Montalvo-Rodriguez R, Vreeland RH, Oren A, Kessel M, Betancourt C, Lopez-Garriga J (1998) *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. Int J Syst Bacteriol 48(Pt 4):1305–1312. doi:10.1099/00207713-48-4-1305

Mullakhanbhai MF, Larsen H (1975) *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol 104(3):207–214
Okochi M, Matsuzaki H, Nomura T, Ishii N, Yohda M (2005) Molecular characterization of the group II chaperonin from the hyperthermophilic archaeum *Pyrococcus horikoshii* OT3. Extremophiles 9(2):127–134. doi:10.1007/s00792-004-0427-y
Ollivier B, Cayol JL, Patel BK, Magot M, Fardeau ML, Garcia JL (1997) *Methanoplanus petrolearius* sp. nov., a novel methanogenic bacterium from an oil-producing well. FEMS Microbiol Lett 147(1):51–56
Oren A (1999) Bioenergetic aspects of halophilism. Microbiol Mol Biol Rev 63(2):334–348
Oren A (2008) Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Syst 4:2–14
Oren A, Ginzburg M, Ginzburg BZ, Hochstein LI, Volcani BE (1990) *Haloarcula marismortui* (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. Int J Syst Bacteriol 40(2):209–210. doi:10.1099/00207713-40-2-209
Pierson BK, Castenholz RW (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch Microbiol 100(1):5–24
Roh SW, Nam YD, Chang HW, Sung Y, Kim KH, HM O, Bae JW (2007) *Halalkalicoccus jeotgali* sp. nov., a halophilic archaeon from shrimp jeotgal, a traditional Korean fermented seafood. Int J Syst Evol Microbiol 57(Pt 10):2296–2298. doi:10.1099/ijs.0.65121-0
Ruggero D, Ciammaruconi A, Londei P (1998) The chaperonin of the archaeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing. EMBO J 17(12):3471–3477. doi:10.1093/emboj/17.12.3471
Sako Y, Nomura N, Uchida A, Ishida Y, Morii H, Koga Y, Hoaki T, Maruyama T (1996) *Aeropyrum pernix* gen nov, sp nov, a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100 degrees C. Int J Syst Bacteriol 46(4):1070–1077
Sekiguchi H, Nakagawa A, Moriya K, Makabe K, Ichiyanagi K, Nozawa S, Sato T, Adachi S, Kuwajima K, Yohda M, Sasaki YC (2013) ATP dependent rotational motion of group II chaperonin observed by X-ray single molecule tracking. PLoS One 8(5):e64176. doi:10.1371/journal.pone.0064176
Sergeeva OA, Yang J, King JA, Knee KM (2014) Group II archaeal chaperonin recognition of partially folded human gammaD-crystallin mutants. Protein Sci 23(6):693–702. doi:10.1002/pro.2452
Sowers KR, Baron SF, Ferry JG (1984) *Methanosarcina acetivorans* sp. nov., an Acetotrophic methane-producing bacterium isolated from marine sediments. Appl Environ Microbiol 47(5):971–978
Stetter KO (1996) Hyperthermophilic procaryotes. FEMS Microb Rev 18:149–158
Stetter KO, Lauerer G, Thomm M, Neuner A (1987) Isolation of extremely thermophilic sulfate reducers: evidence for a novel branch of archaebacteria. Science 236(4803):822–824. doi:10.1126/science.236.4803.822
Takai K, Inoue A, Horikoshi K (2002) *Methanothermococcus okinawensis* sp. nov., a thermophilic, methane-producing archaeon isolated from a Western Pacific deep-sea hydrothermal vent system. Int J Syst Evol Microbiol 52(Pt 4):1089–1095. doi:10.1099/00207713-52-4-1089
Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668–673
Vannier P, Michoud G, Oger P, Marteinsson V, Jebbar M (2015) Genome expression of *Thermococcus barophilus* and *Thermococcus kodakarensis* in response to different hydrostatic pressure conditions. Res Microbiol 166(9):717–725. doi:10.1016/j.resmic.2015.07.006
Volkl P, Huber R, Drobner E, Rachel R, Burggraf S, Trincone A, Stetter KO (1993) *Pyrobaculum aerophilum* sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. Appl Environ Microbiol 59(9):2918–2926
Waino M, Tindall BJ, Ingvorsen K (2000) *Halorhabdus utahensis* gen. nov., sp. nov., an aerobic, extremely halophilic member of the archaea from great salt lake, Utah. Int J Syst Evol Microbiol 50(Pt 1):183–190. doi:10.1099/00207713-50-1-183
Wu Y, Vadrevu R, Kathuria S, Yang X, Matthews CR (2007) A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha sub-

unit of tryptophan synthase, a TIM barrel protein. J Mol Biol 366(5):1624–1638. doi:10.1016/j.jmb.2006.12.005

Yan Z, Fujiwara S, Kohda K, Takagi M, Imanaka T (1997) *In vitro* stabilization and *in vivo* solubilization of foreign proteins by the beta subunit of a chaperonin from the hyperthermophilic archaeon *Pyrococcus* sp. strain KOD1. Appl Environ Microbiol 63(2):785–789

Yoshida T, Ideno A, Hiyamuta S, Yohda M, Maruyama T (2001) Natural chaperonin of the hyperthermophilic archaeum, *Thermococcus* strain KS-1: a hetero-oligomeric chaperonin with variable subunit composition. Mol Microbiol 39(5):1406–1413

Yoshida T, Ideno A, Suzuki R, Yohda M, Maruyama T (2002) Two kinds of archaeal group II chaperonin subunits with different thermostability in *Thermococcus* strain KS-1. Mol Microbiol 44(3):761–769

Yoshida T, Kanzaki T, Iizuka R, Komada T, Zako T, Suzuki R, Maruyama T, Yohda M (2006) Contribution of the C-terminal region to the thermostability of the archaeal group II chaperonin from *Thermococcus* sp. strain KS-1. Extremophiles 10(5):451–459. doi:10.1007/s00792-006-0519-y

Zeikus JG, Wolfe RS (1972) *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. J Bacteriol 109(2):707–715

Zhang K, Wang L, Liu Y, Chan KY, Pang X, Schulten K, Dong Z, Sun F (2013) Flexible interwoven termini determine the thermal stability of thermosomes. Protein Cell 4(6):432–444. doi:10.1007/s13238-013-3026-9

Chapter 9
Archaeal Chaperonins: A Cornucopia of Information and Tools to Understand the Human Chaperoning System and Its Diseases

Alberto J.L. Macario and Everly Conway de Macario

Abstract The chaperoning system in organisms across Archaea, Bacteria, and Eukarya encompasses a wide range of complexities. This range extends from the very simple, consisting of a few genes-proteins in the most primitive archaea to the extremely complex like the one in humans (made of many chaperones, co-chaperones, and chaperone cofactors and close interactors/receptors) and including various degrees of complexity in between. Diseases caused by defects in the chaperoning system, named chaperonopathies, are important because many are widespread and frequent and can be life threatening. Chaperonopathies are genetic or acquired. The former are caused by mutations of chaperone genes, whereas the acquired chaperonopathies are typically caused by posttranslational modifications of the chaperone proteins. In both instances, the affected chaperones are nonfunctional, or function incorrectly (e.g., gain of function), or their levels/functioning is increased or decreased. Little is known on the impact of mutations or posttranslational modifications on the properties of the chaperone molecule or on its chap-

A.J.L. Macario (✉)
Department of Microbiology and Immunology, School of Medicine, University of Maryland at Baltimore, Baltimore, MD, USA

Institute of Marine and Environmental Technology (IMET), Columbus Center, Baltimore, MD, USA

Euro-Mediterranean Institute of Science and Technology (IEMEST), Palermo, Italy
e-mail: Ajlmacario@som.umaryland.edu

E. Conway de Macario
Department of Microbiology and Immunology, School of Medicine, University of Maryland at Baltimore, Baltimore, MD, USA

Institute of Marine and Environmental Technology (IMET), Columbus Center, Baltimore, MD, USA

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_9

eroning functions. Studies are hindered by the extreme complexity of the chaperoning machines and networks in humans and by the scarcity of experimental models. Here, we report on some archaeal chaperoning systems, focusing on the chaperonins only, which are suitable for standardizing experimental models mimicking the human situations observed in chaperonopathies. We discuss archaeal chaperonins that are similar to those of humans and present an illustrative example of the use of one of these archaeal chaperonins to elucidate the molecular abnormalities generated by a pathogenic mutation in a human chaperonin subunit that causes neuropathy.

9.1 Introduction

This is not a comprehensive review on archaeal chaperonins but a brief account on some of them that are potentially useful as experimental models to study molecular aspects of human chaperonopathies. These diseases are caused by abnormal chaperones and have recently been assembled in a coherent nosological group: they constitute a new chapter of Medicine. Their clinical and pathological features, and inheritance mode for those that are due to chaperone gene mutations, are reasonably well characterized. However, the molecular mechanisms that cause the lesions observed in the cells, tissues, and organs of the patients are largely unknown. Likewise, the impact of the mutations in the chaperone genes on the intrinsic properties and chaperoning functions of the mutant proteins is for the most part poorly understood. The same can be said in what pertains to the effect of aberrant post-translational modifications on pathogenic chaperones that cause acquired chaperonopathies. This lack of knowledge ought to be remedied not only because some chaperonopathies are life threatening but also because many of them are very frequent: chaperonopathies are not rare or orphan diseases as perhaps some people believe.

Some of the reasons for the scarcity of knowledge on the molecular aspects of the chaperonopathies are as follows: (1) many physicians, scientists, and the society at large are unaware of their existence mostly because they have been systematically grouped only recently and are not yet included in the medical curricula; (2) the chaperones function mostly as multimolecular machines that are very difficult to dissect for analyzing their interactions, changes, and movements when at work; and (3) there are too few experimental models that mimic the human situations with accuracy but simpler, amenable to close examination at the molecular level.

Archaea offer unique possibilities for developing experimental models pertinent to chaperonopathies because they have chaperones similar in various ways to those from humans. The objective of this article is to describe the chaperonins found in archaea that are promising tools/targets for analyzing molecular aspects of human

chaperonopathies. All archaeal species have chaperonins of Group II, and some have also chaperonins of Group I. Therefore, both Groups are available for study in a single cell.

9.2 The Human Chaperoning System and the Chaperonopathies

The identification of the chaperonopathies, i.e., diseases caused by abnormal or defective chaperones, as a coherent nosological group of importance in Medicine, has boosted interest in elucidating the molecular mechanisms underlying the pathological lesions observed in the patients (Macario and Conway de Macario 2005, 2007). Likewise, the concept that all chaperones, co-chaperones, and chaperone cofactors and specific interactors/receptors constitute a physiological system has provided a novel, cohesive platform to examine many human and animal diseases and understand their pathogenic mechanisms (Macario and Conway de Macario 2009; Macario et al. 2010).

The chaperones in humans are present in all cells, tissues, and organs and in the intercellular space and in circulation, as illustrated in Fig. 9.1. Thus, one may say that the chaperoning system (CS) is ubiquitous, widespread in the body, and interacts with other physiological systems, particularly the immune system (IS), and when one of these two systems is disturbed, the other also suffers and disease develops (Macario et al. 2010).

The CS is a key component of the cellular anti-stress mechanisms and, probably, started its own progression toward the complexity we see today in humans very early in evolution to protect cells from the harsh environmental conditions we imagine that predominated in the primitive Earth. The IS also started early, but most likely after the CS, to protect cells and organisms against infectious agents, and today both the CS and the IS are inextricably related functionally in health and disease.

Chaperones form teams to exercise their canonical functions, pertaining to protein homeostasis, and also form networks (Fig. 9.2) (Macario and Conway de Macario 2005; Macario et al. 2013; Zako et al. 2016). Thus, the correct functioning of the CS requires associations and interactions between molecules of various kinds to integrate chaperoning machines and networks that must act in coordination to maintain health. If one of the components of a chaperoning team is altered because, for instance, of an inheritable mutation or an aberrant posttranslational modification, the conditions are given for a pathological disorder to occur. These disorders caused by abnormal, defective, malfunctioning chaperones are the chaperonopathies (Macario and Conway de Macario 2005). They can affect all types of cells, tissues, and organs and, therefore, are of interest to practically all medical specialties. Furthermore, it should be emphasized again that chaperonopathies are not rare; on the contrary, they are quite frequent and must not be ignored by physicians and other healthcare personnel in their daily service and research activities.

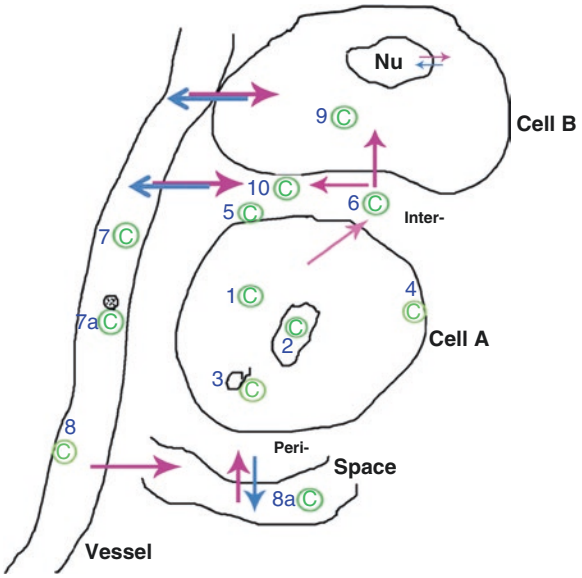


Fig. 9.1 The chaperoning system: schematic diagram showing the locations and migrations of Hsp60 (HSP60 or HSPD1, *Homo sapiens*), or Hsp70 family members, or both. Circled C (*green*), molecular chaperone; **1**, mobile chaperone in the cytosol; **2**, chaperone inside an organelle, such as the nucleus or a mitochondrion; **3**, sessile chaperone anchored to a particle (e.g., ribosome) in the cytoplasm; **4** and **5**, sessile chaperone anchored to the plasma-cell membrane on the cytoplasmic side (4) or on the outside in the extracellular space; (5), chaperones can also be located, at least transitorily, in the plasma-cell membrane (i.e., transmembrane location); **6**, mobile chaperone in the intercellular space; **7**, mobile chaperone in circulation inside a vessel (blood or lymph) in suspension or, **7a**, in microvesicles (e.g., exosomes) and on circulating erythrocytes, lymphocytes, granulocytes, or platelets; **8**, sessile chaperone anchored to the vessel wall on the inside; **8a**, chaperone inside a biological space, such as the intersynovial space in the cavity of the joints and the space between the pia and the arachnoid maters in the nervous system; **9**, mobile chaperone in the cytosol like that shown in 1, but imported from another cell. Molecular chaperones can be found also in other locations such as cerebrospinal fluid (8a) and secretions (e.g., saliva and urine), the latter two not shown in this figure; **10,** mobile or sessile chaperone that originated in the blood or in a nearby cell (same as 5 if sessile, or same as 6 if mobile) and is now in the intercellular space. *Arrows* indicate the various directions of movement of a chaperone molecule from inside a cell to the extracellular space or vessel's lumen and vice versa or from inside an organelle, such as the nucleus as shown, or a mitochondrion (not shown), to the cytoplasm and vice versa. A chaperone can gain the extracellular space (intercellular or pericellular) from inside a cell or from inside a vessel, and it can go into a vessel directly from a cell or from the extracellular space. *Nu* nucleus, *Space* biological space, e.g., intersynovial space and space between pia and arachnoid maters (cerebrum ventricles, cisterns, sulci, and spinal cord central canal); *Inter-*, intercellular space; *Peri-*, pericellular space, e.g., between the pia mater and a superficial brain cortex cell. Source: Macario and Conway de Macario (2009) and Macario et al. (2010)

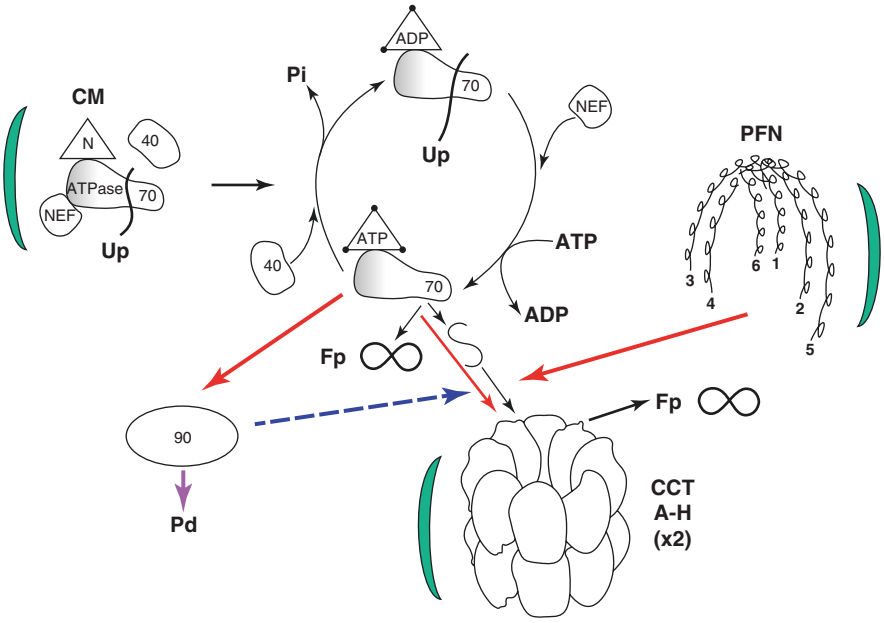


Fig. 9.2 Chaperoning teams and networks. **CM.** The Hsp70-Hsp40-nucleotide exchange factor (NEF) chaperone machine. Hsp70 (70) binds the unfolded client polypeptide (Up), via its peptide-binding domain near the middle of the molecule, when ADP is bound to its ATPase domain and Hsp40 (40) is bound to its C-terminal domain. The nucleotide exchange factor (e.g., BAG-1 or GrpE, shown as NEF) promotes exchange of ADP for ATP. When Hsp70 is bound to ATP, its affinity for the folding polypeptide decreases, and the folded polypeptide (Fp) is released. The nucleotide exchange factor is replaced by Hsp40, ATP hydrolysis occurs with release of pyrophosphate (Pi), and a new cycle of peptide binding, folding, and release begins. *N* nucleotide, e.g., ATP or ADP. Other alternatives are as follows: the partially folded polypeptide is handed over to the chaperonin-containing TCP-1 (CCT) complex, directly or with participation of Hsp90 (90) and/or prefoldin (PFN). The polypeptide is then folded inside the CCT-folding chamber and released. CM and Hsp90 can also be involved in a pathway leading to protein degradation (Pd) in the proteasome, for instance. While CM occurs in the three life domains, CCT and PFN exist in eukaryotes (typically the cytosol) and archaea but not in bacteria with rare exceptions. CM also exists in the eukaryotic cell compartments: nucleus, mitochondria, endoplasmic reticulum, and chloroplasts. **PFN**, prefoldin. This chaperone complex in eukaryotes is composed of several (six in humans) different but evolutionarily related subunits (named 1–6) arranged in a medusa type of structure. Typically PFN exists in the eukaryotic cell cytosol and in the cytoplasm of archaea. The latter have typically two instead of six different by evolutionarily related subunits. **CCT**, Group II chaperonin. This large structure is built like a barrel and consists of two stacked rings each made of eight different but evolutionarily related subunits (1–8, or A-H, or alpha-theta) arranged around a central cavity, the polypeptide-folding chamber. The *green* thick arches indicate that there is movement (allosteric changes) during assembly and functioning of the teams. Source: Macario et al. (2013)

9.3 Chaperoning Teams and the Chaperonins

The main chaperoning teams in humans are depicted on the left side of Fig. 9.3 (Macario et al. 2013). The chaperonins are those whose molecular weight ranges around 60 kDa. As it can be seen in Fig. 9.3, they form functional oligomers, typically rings of 7 or 8 subunits. The rings in turn associate in pairs in such a way as to form a barrel with a central cavity. This structure characterizes both the chaperonins of Group I and Group II. In the former, the rings have 7 subunits and occur in

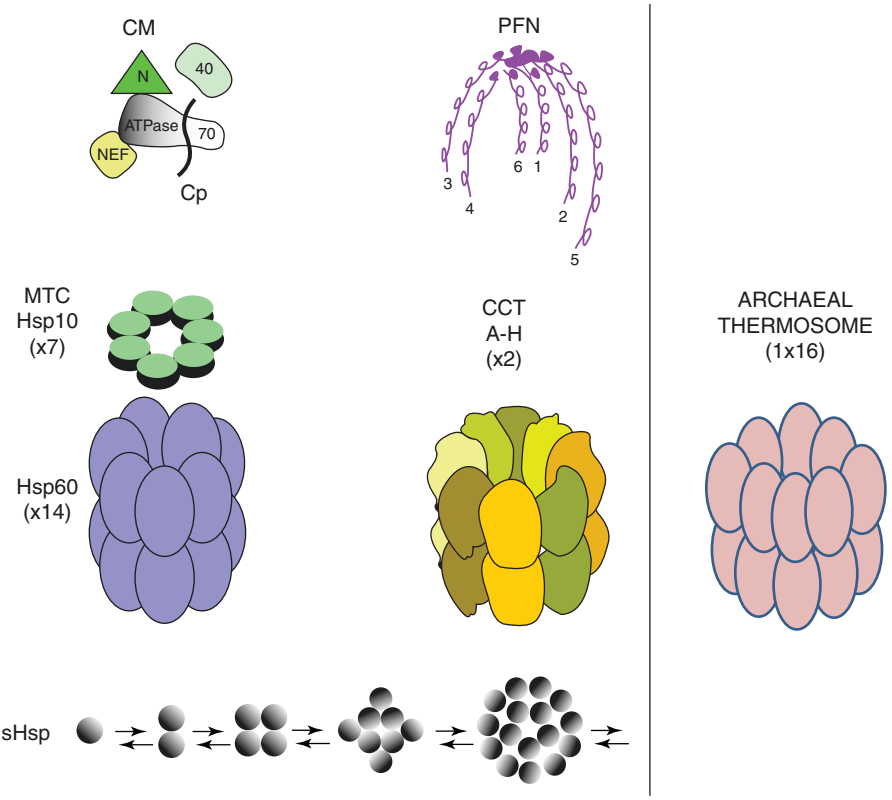


Fig. 9.3 The main chaperoning teams in humans to the *left* and an example of an archaeal Group II chaperonin hexadecamer to the *right*. **CM**, **PFN**, and **CCT**: same as in Fig. 9.2. **MTC**: mitochondrial chaperonin complex or Group I chaperonin. This is also a large structure, resembling CCT. However, e.g., in the bacterial GroEL complex [Hsp60 (×14)], each ring has seven identical subunits (Hsp60 or Cpn60). In addition, a third ring, the GroES complex [Hsp10 (×7)], formed by seven smaller subunits (Hsp10 or Cpn10) joins the team to integrate the chaperoning machine GroEL/GroES. The smaller ring (Hsp10) forms a sort of lid that closes the folding chamber while the folding of a client polypeptide proceeds inside. **sHsp**: small heat shock proteins. These comparatively small chaperones, e.g., the alpha crystallins, are usually 30 kDa or less in MW and may occur as monomers but form multimers of various sizes in response to stress and to participate in the protection of unfolded polypeptides. The hexadecamer to the right is an example of chaperoning machine seen in some archaeal species that possess a single chaperonin gene (see Table 9.1 and Fig. 9.4). Source: Macario et al. (2013)

bacteria and in the eukaryotic organelles that are considered to be of bacterial origin, i.e., mitochondria and chloroplasts. The chaperonins of Group II are constituted of 8 subunits and occur in archaea and in the cytosol of eukaryotic cells. The number of different subunits varies with organisms (Table 9.1), but as far as we know at the present time, they all can build hexadecamers, as illustrated in Fig. 9.4.

Table 9.1 Chaperonin subunits in archaea and eukaryotes: examples[a]

Archaea	Subunits	
Group	Number	Name
Euryarchaeota	1, 2, or 3	a, b; α, β; 1, 2, 3
Halophiles	2, 3, or 4	1, 2, 3, 4
Crenarchaeota	2, or 3	α, β, γ
Nanoarchaeota	1	1
Methanosarcina acetivorans	5	Hsp60–1, 2, 3, 4, 5
Eucarya	9	1–8 (6A and 6B); α, β, γ, δ, ε, ζ1, ζ2, η, θ

[a]Source: Macario et al. (1999, 2004), Conway de Macario et al. (2003), Maeder et al. (2005), and Large and Lund (2009)

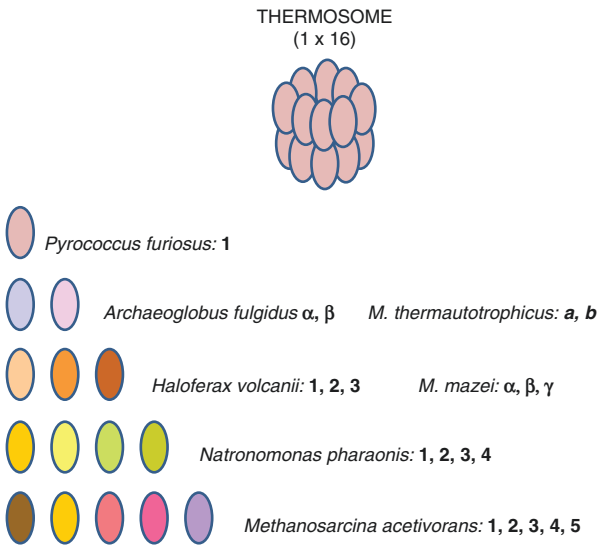


Fig. 9.4 Archaeal chaperonins of Group II. Archaeal species vary in their content of Group II chaperonin genes-proteins from only one through a maximum (at least from what we know at the present time) of five; some examples are mentioned in the figure. These subunits are variously named with Arabic numbers, English letters, or Greek letters. As far as we know, they all form hexadecamers, the thermosome, of the type shown on *top*, which is an example of a hexadecamer found in archaea with only one chaperonin subunit, e.g., *P. furiosus*. The composition of hexadecamers in all archaea that have two or more subunits is not yet fully elucidated. *M. thermautotrophicus*, *Methanothermobacter thermautotrophicus* ΔH, previously known as *Methanobacterium thermoautotrophicum* ΔH. *M. mazei*, *Methanosarcina mazei*. Source: Macario et al. (1999, 2004), Conway de Macario et al. (2003), and Large and Lund (2009)

9.4 Chaperonins and Associated Chaperonopathies

The chaperoning system, including the chaperonins, is present in all organisms with various degrees of complexity, from the simplest in primitive prokaryotes to the very complex such as that occurring in humans. Interestingly, while many organisms have been found to lack the Hsp70 team, which was considered essential for life (Macario and Conway de Macario 1999), no organism has been discovered which does not have chaperonins of Group I, or of Group II, or both (Tables 9.2 and 9.3) (Macario et al. 1999, 2004; Large and Lund 2009; Techtmann and Robb 2010). Furthermore, many species have more than one chaperonin of Group I (Kumar et al. 2015; see other Chapters in this volume). It may be inferred that chaperonins of one Group or another are essential for life; they are key to cell physiology under normal conditions and under stress. Therefore, it is important to understand their roles in humans, in health, and in the diseases in which they play a determinant pathogenic role, namely, the chaperonopathies associated with abnormal chaperonins. Examples of genetic chaperonopathies involving chaperonins are listed in Table 9.4. There are also a large variety of noninheritable diseases in which abnormal chaperonins participate in the pathogenic mechanism although the primary amino acid sequence is not changed: these are the acquired chaperonopathies (Macario et al. 2013), but they will not be discussed here.

Table 9.2 Chaperones and chaperonins in the three life Domains[a]

Bacteria	Archaea	Eucarya
DnaK(Hsp70)	No, yes[b]	Yes
DnaJ(Hsp40)	No, yes	Yes
GrpE	No, yes	mt, chl (and NEFs in cytosol)
GroEL/S (Group I; Hsp60) and Hsp10	No, yes	mt, chl (Cpn, Hsp60) and Hsp10
No	Prefoldin (2)[c]	Yes (human: 6)
No	Chaperonins Group II (1–3; 4, 5)	Chaperonins Group II (human CCT: 9 + 5)

chl choloroplast, *mt* mitochondria, *NEF* nucleotide exchange factor
[a]Source: Macario et al. (1999, 2004)
[b]Some species do not have it (No) and others do have it (Yes)
[c]Figure(s) within parentheses, number of subunits, i.e., genes

Table 9.3 Examples of archaea with both Group I and II chaperonin genes[a]

	Chaperonin genes of group	
Organism	I	II (subunits)
Methanococcus vannielii SB	1	1
Methanospirillum hungatei JF-1	1	2
Methanosarcina barkeri str. Fusaro	1	3
Methanosarcina mazei Go1	1	3
Methanosarcina acetivorans C2A	1	5

[a]Source: Galagan et al. (2002), Deppenmeier et al. (2002), Conway de Macario et al. (2003), Maeder et al. (2005), and Large and Lund (2009)

Table 9.4 Examples of structural hereditary chaperonopathies caused by mutant chaperonins[a]

Gene/protein affected	Diseases/syndromes
Chaperonin group I, Hsp60 (HSPD1)	
Mitochondrial Hsp60 (Cpn60)	Hereditary spastic paraplegia (SPG13); MitCHAP-60 (Pelizaeus-Merzbacher-like)
Chaperonin Group II, CCT subunits (canonical and noncanonical)	
CCT4, CCT5	Hereditary sensory neuropathy
MKKS, BBS10, and BBS12	McKusick-Kaufman, and Bardet-Biedl

[a]Source: Macario et al. (2013)

The impact of disease-associated mutations on the chaperonin itself is poorly understood. It is now necessary to elucidate the effect of the pathogenic mutations on the intrinsic properties of the chaperonin molecule and on its chaperoning functions in vitro, and if possible in vivo, in real life. This type of studies is, as mentioned in the Introduction, hindered by various factors, among which are, on one side, the complexity of the human teams and networks very difficult to examine under controlled experimental conditions and, on the other side, the scarcity of simple standardized experimental models amenable to systematic dissection of molecular structure-function parameters.

9.5 The Potential of Archaea to Study Human Chaperonopathies

With regard to the difficulties to study human chaperonopathies mentioned above, it has to be realized that some archaeal species offer promising alternatives because they possess chaperonins of Group II like those in the eukaryotic cell cytosol and some possess both Group II and the bacterial-type Group I like that in the eukaryotic cell mitochondria and chloroplasts (Table 9.3) (Galagan et al. 2002; Conway de Macario et al. 2003). *Methanosarcina acetivorans* possesses 5 subunits of Group II (Galagan et al. 2002; Maeder et al. 2005), which is comparatively close to nine, i.e., the total number of ortholog subunits in humans. Another species, *Methanosarcina mazei*, has 3 subunits (Deppenmeier et al. 2002; Klunker et al. 2003), while other archaea have only one or two subunits (Macario et al. 2004; Large and Lund 2009).

As far as we know, the subunits in all these organisms, even in those that have only one, form functional hexadecamers (Fig. 9.3, right side). The homo-oligomers formed by the single chaperonin in the archaea that possess a single chaperonin gene offer a much simpler system for study than those species with two or more genes (Fig. 9.4). As we will discuss in more detail later, we have standardized an experimental model using the archaeon *Pyrococcus furiosus* (Pf), which has only one subunit (1, Pf-Cpn) that forms functional hexadecamers of the type shown in Fig. 9.3 (right side) and Fig. 9.4 (Min et al. 2014).

9.6 Group II Chaperonins in *Methanosarcina acetivorans*

Methanosarcina acetivorans offers the possibility of examining the role of each of its 5 subunits and determining, for instance, if they associate in various combinations to form hexadecamers with specificity for distinct substrates or set of substrates.

A comparative analysis of the *M. acetivorans* 5 subunits shows a clustering of subunits 4 and 5 (Fig. 9.5), subunits that are unique to this archaeal species. Likewise, a comparative analysis of the *M. acetivorans* subunits with the human counterparts (Fig. 9.6) shows that these two subunits, 4 and 5, are closer to the human 6A and 6B subunits than to any other, and 6A and 6B also cluster together when all human subunits are analyzed as a separate group (Mukherjee et al. 2010). It is noteworthy that *M. acetivorans* subunits 4 and 5 are distinctive of this organism as compared with other closely related *Methanosarcina* species. Interestingly, subunit 3 from *M. acetivorans* clusters together with subunit 3 from *M. mazei* and both are closer to the human CCT8 than to any other. The other two subunits, 1 and 2, from *M. acetivorans* and from *M. mazei* form a separate cluster. This latter cluster was also observed when comparing all subunits of *M. acetivorans*, *M. mazei*, and *Methanosarcina barkeri* (Fig. 9.7) (Maeder et al. 2005).

It can be hypothesized, for the purpose of developing strategies to study human chaperonopathies, that there are at least three subpopulations of Group II chaperonin subunits in *M. acetivorans* with counterparts (descendants?) among the eukaryotic (e.g., human) CCT subunits. If this were the case, one may wonder what are the structural and functional differences between these three subpopulations; what would be the distinctive pathogenetic role of the human counterparts if anyone of these counterparts were abnormal (mutation), or damaged (posttranslation modification); or if its production was dysregulated (dysregulatory chaperonopathies).

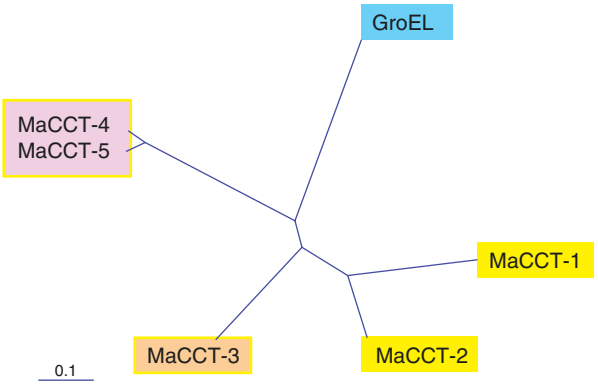


Fig. 9.5 Clustering of the chaperonin subunits from *Methanosarcina acetivorans*. Neighbor-joining Poisson-corrected tree of GroEL and the five Hsp60 (chaperonin of Group II subunits or CCT orthologs) proteins from *M. acetivorans*. Corresponding GenBank accession numbers are GroEL, GI:20089519; Hsp60-1, GI:20088985; Hsp60-2, GI:20093200; Hsp60-3, GI:20090534; Hsp60-4, GI:20093173; and Hsp60-5, GI:20089741. Bootstrap values were derived from 1000 interactions. Analyses were conducted using the MEGA version 2.1

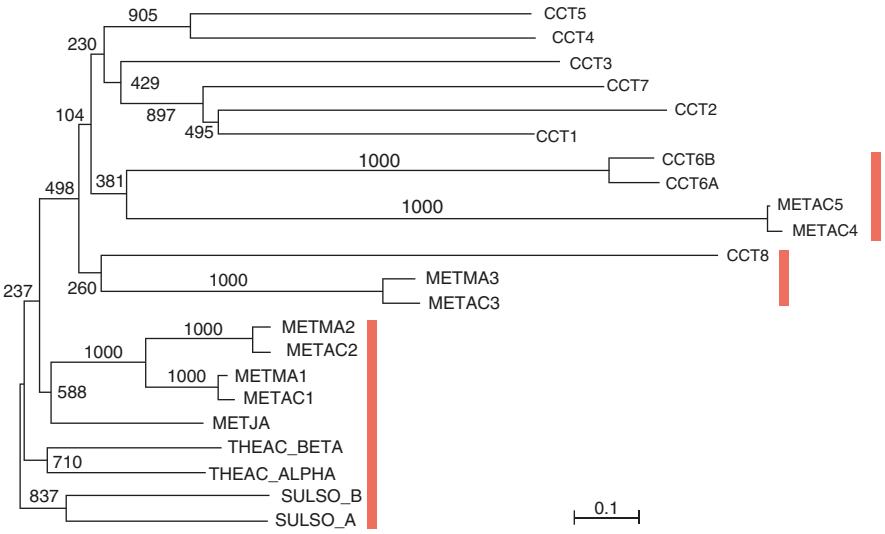


Fig. 9.6 Comparative tree of archaeal and human chaperonin Group II subunits. Preferential clustering of archaeal and human subunits is apparent (see text). METAC, *Methanosarcina acetivorans*; METMA, *Methanosarcina mazei*; METJA, *Methanococcus jannaschii*; THEAC, *Thermoplasma acidophilum*; SULSO, *Sulfolobus solfataricus*. The scale bar represents the indicated number of substitutions per position for a unit branch length

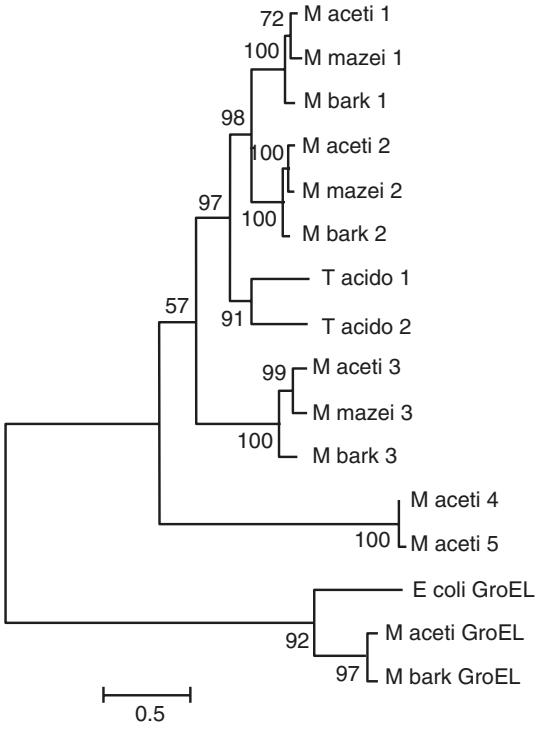


Fig. 9.7 Comparative tree of chaperonin subunits. Preferential clustering of archaeal subunits is apparent (see text and Fig. 9.6). M. aceti, *Methannosarcina acetivorans*; M. mazei, *Methanosarcina mazei*; M. bark, *Methanosarcina barkeri*; T. acido, *Thermoplasma acidophilum*; E. coli, *Escherichia coli*. Fully resolved quartet puzzling tree with maximum likelihood (branch lengths showing support levels). Source: Maeder et al. (2005)

Table 9.5 *Methanosarcina acetivorans* GroEL and GroES compared with bacterial and human counterparts

	Bacillus subtilis		*Streptococcus albus*		*Homo sapiens*	
	Identity[a]	Similarity	Identity	Similarity	Identity	Similarity
GroEL						
M. acetivorans	56	67	51	62	44	66
B. subtilis			63	75	n.d.	n.d.
GroES						
M. acetivorans	34	52	36	55	30	49
B. subtilis			57	74	n.d.	n.d.

[a]Identity and similarity (identical amino acids plus conservative substitutions) percent; *n.d.* not done

9.7 The Chaperonin *groES/groEL* Operon of *Methanosarcina acetivorans*

As we can see from data in Table 9.3, *M. acetivorans* also possesses genes encoding the so-called bacterial-type, Group I chaperonins GroES and GroEL. The proteins are quite similar to those of bacteria (Hsp10 and Hsp60, also called Cpn10 and Cpn60, respectively) and humans (Hsp10 or HSPE1 and HSP60 or HSPD1, also named mitochondrial Cpn10 and mitochondrial Cpn60, respectively), particularly GroEL (Table 9.5). The human HSPD1 is closely related to bacterial GroEL and considerably less related to human CCT subunits (Mukherjee et al. 2010).

9.8 Functional Studies in *Methanosarcina mazei* on the Cooperativity Between Group I and II Chaperonins

In *M. mazei*, functional studies were carried out on its GroEL/GroES team in comparison with that from *Escherichia coli* (Figueiredo et al. 2004). The *groES/groEL* operon from *M. mazei* was unable to replace functionally its counterpart in *E. coli*. The GroES protein from *M. mazei* did complement in vivo a mutant *GroES* in *E. coli*. The ATPase rate of the archaeal GroEL was very low by comparison with that of the *E. coli* GroEL, and the dissociation rate of the archaeal GroES from GroEL was 15-fold slower than the dissociation rate of the *E. coli* GroEL/GroES complex. This slow ATPase cycle caused a prolonged enclosure time of the substrate as compared with that in *E. coli*. These results were interpreted to

indicate that the substrate encapsulation mechanism of the GroEL/GroES machine in the archaeon *M. mazei* is basically the same as that of *E. coli* but proceeds at a slower rate to match the slower growth rate of the archaeal organism by comparison with *E. coli*.

A comparative study was performed to determine if GroEL, i.e., the bacterial-type Group I chaperonin and the Group II chaperonin, both present in archaeon *M. mazei*, can distinguish substrates in such a way that one chaperonin will ignore substrates preferred by the other and vice versa (Hirtreiter et al. 2009). Interactors for the two chaperonins were identified by proteome-wide analysis, and it was found that at least 13% of soluble *M. mazei* proteins interact with chaperonins, with the archaeal Group II chaperonin and the bacterial-type GroEL having overlapping subsets of substrates. Interestingly, substrate preference was independent of the phylogenetic origin of the substrate but was rather determined by its structural and biochemical features. GroEL preferred conserved complex (a/b domains) proteins, while the substrates of the archaeal Group II chaperonin included a wide range of protein folds, including small all-a and all-b topologies and many multidomain proteins. It was concluded that the Group II chaperonins must have evolved to allow the evolution of the more complex proteomes characteristic of eukaryotes.

9.9 Interchangeability Between an Archaeal Chaperonin of Group II and the Bacterial-Type Chaperonin of Group I GroEL

Demonstration of interchangeability of roles between the chaperonins of Group II from archaea and the chaperonins of Group I from bacteria was achieved by using the chaperonin from the archaeon *Methanococcus maripaludis* to complement *E. coli* GroEl, in vivo (Shah et al. 2016). Indeed, the archaeal chaperonin of Group II was able to partially replace GroEL functionally in *E. coli*. Two single-point mutations were identified that made the archaeal chaperonin even more efficient than the wild type to replace functionally the *E. coli* GroEL. One of these mutations was able to sustain growth of *E. coli* even in the absence of the *E. coli groEL* gene. Based on these observations, the authors argue that despite the fact that the two chaperonin Groups diverged approximately 3.7 billion years ago, the molecules of the two Groups still have very similar quaternary structures and assist protein folding in a similar way. However, the two chaperonin Groups differ in a number of details, such as structural features pertaining to interaction with substrate, cofactor requirements, and reaction cycles. The authors conclude that despite the long time since they separated in evolution, the chaperonins of Group I and II still have overlapping, easily recognizable, structural and functional properties.

9.10 Application of an Archaeal Model System to the Elucidation of the Effect of Pathogenic Mutations on a Human Chaperonin

As we said earlier, we have used the archaeon *Pyrococcus furiosus* (Pf) to standardize an experimental model for dissecting the molecular impact of mutation in chaperonin genes (Min et al. 2014). *Pyrococcus furiosus* has a single chaperonin gene encoding Pf-Cpn (subunit 1), which builds hexadecamers like those of humans (Fig. 9.3 (right side) and Fig. 9.4).

We have begun to characterize the properties of the Pf-Cpn molecule (Pf-CD1) carrying the mutation Ile138Arg (Pf-R), which matches the pathogenic mutation in the human Group II chaperonin subunit CCT5 that causes distal neuropathy (Bouhouche et al. 2006). We investigated the impact of the mutation on the intrinsic properties of the protein and on its chaperoning functions and we will discuss, here, only a few illustrative results.

The pathogenic mutation, Arg instead of Ile, was introduced in the Pf-Cpn at position 138, which is the equivalent to 147 in the human CCT5 as determined by extensive comparative analyses using complementary methods (Min et al. 2014). In order to determine the impact of the mutation on the intrinsic properties of the model chaperonin, several tests were carried out to measure, for example, protein stability when confronted with the stressor heat, i.e., a temperature elevation. Ongoing experiments are showing that Pf-CD1 has greater stability as compared with Pf-H (a nonpathogenic mutation) and Pf-R (the pathogenic mutation), which is the least resistant to heat stress.

Other series of tests were applied to assess the impact of the mutation on chaperoning capabilities, for example, protection of an enzyme from heat denaturation and dissociation of amyloid fibrils and bundles. Illustrative protection results pertaining to the enzymes malate dehydrogenase (MDH) and shrimp alkaline phosphatase (SAP) are shown in Fig. 9.8. Mixed oligomers of Pf-CD1, or Pf-H, or Pf-R were tested for protection of MDH at 37 and 42 °C, and purified hexadecamers were tested for protection of MDH at 37 °C and of SAP at 50 °C. It can be seen from the results that Pf-R-mixed oligomers had a diminished protective capability as compared with Pf-CD1 and Pf-H. However, when purified hexadecamers were tested, it became clear that Pf-R had the same protective capability as the other two molecules. In conclusion, these tests with the archaeal model went a long way to demonstrate that the pathogenic mutation interferes with hexadecamer formation and stability and, thereby, impairs considerably the chaperoning abilities of the chaperonin.

Another important function of chaperones is the dissolution of protein precipitates and there are a variety of assays to measure it. We applied a test that uses insulin fibrils that can be dispersed by active, efficient chaperones and found that Pf-R was deficient (Fig. 9.9) (Min et al. 2014). While fibers were dispersed rapidly by Pf-CD1 and a little less so by Pf-H, the pathogenic mutant failed to do so entirely.

The results discussed above are encouraging in as much as the archaeal experimental model we used, which reproduces closely the situation with the human pathogenic mutant CCT5, did provide information on the impact of the mutation on the chaperonin molecule. Likewise, the same can be expected from testing other mutations in the same chaperonin or in others, also present in archaeal species as discussed in the preceding sections of this chapter.

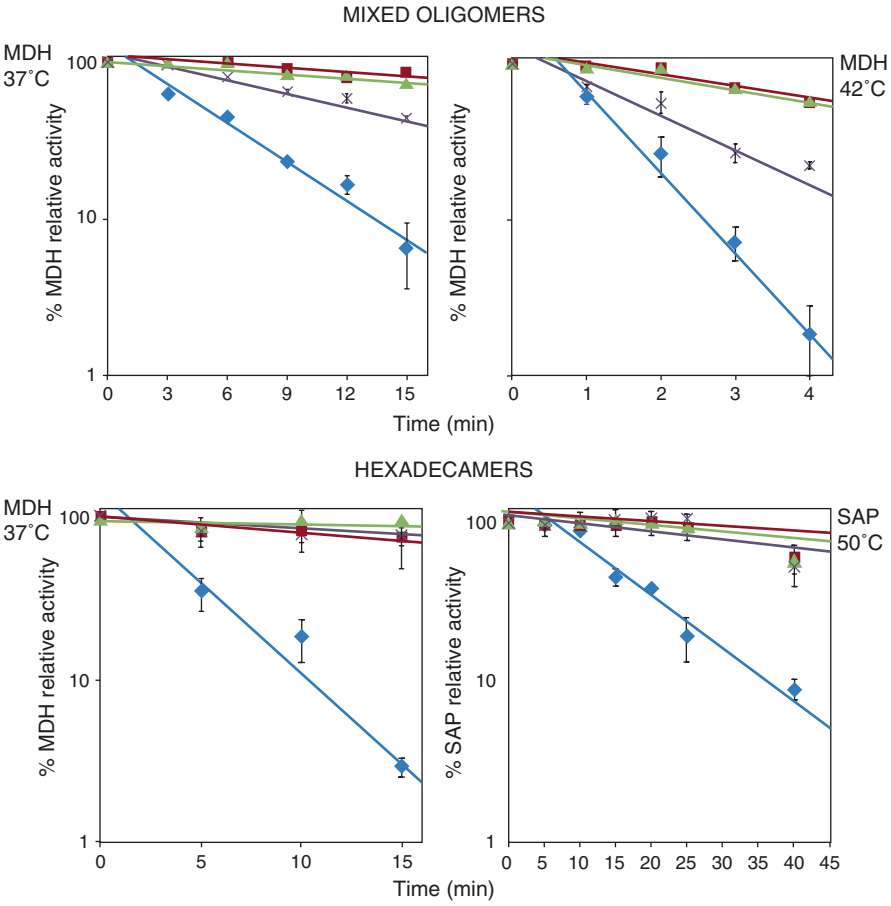


Fig. 9.8 Comparative analyses of protective capacity of mutant and wild-type chaperonins. Malate dehydrogenase (MDH) heat-protection activity profiles of mixed oligomers of Pf-CD1 (*red square*), Pf-H (*green triangle*), and Pf-R (*purple star*) at 37 °C (*top left panel*) and 42 °C (*top right panel*); and MDH heat-protection activity profiles at 37 °C (*bottom left panel*) and shrimp alkaline phosphatase (SAP) heat-protection activity profiles at 50 °C (*bottom right panel*) of pure hexadecamers. Negative control (no chaperonin added), i.e., MDH or SAP (in *bottom right panel*) alone: *blue* diamond. The results shown are mean values (±SD) of triplicate experiments. Source: Min et al. (2014)

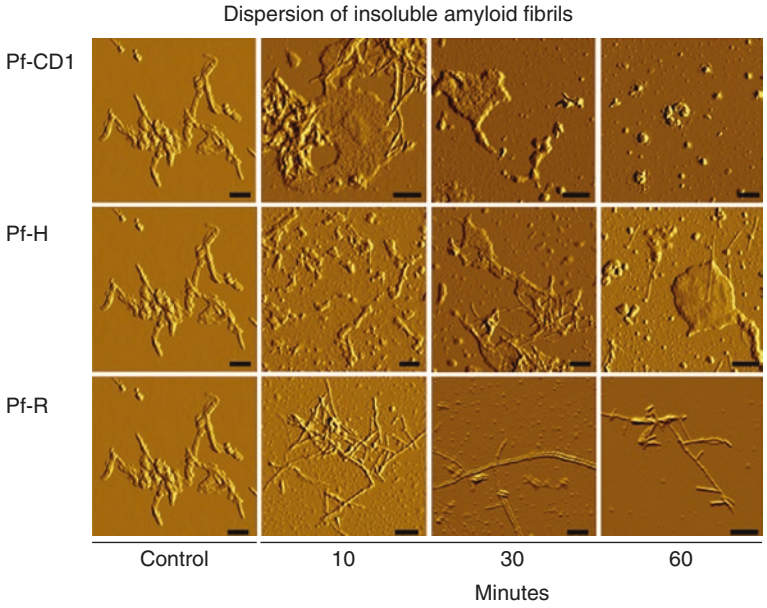


Fig. 9.9 The pathogenic mutant fails to disperse amyloid fibrils. Dispersion of amyloid fibrils by archaeal Pf-CD1 (*top row* of panels), partial dispersion by Pf-H (*middle row* of panels), and no dispersion by Pf-R (*bottom row* of panels). Atomic force microscopy (AFM) of bovine insulin amyloid fibrils treated with Cpn and Mg++ and ATP. Scale bar: 250 nm. Fibril preparations were incubated for the specified times (minutes) with the chaperonins, as indicated. Control: the fibril preparation was incubated for 60 min with no chaperonin added. Source: Min et al. (2014)

9.11 Conclusion and Perspectives

Archaea offer promising tools for analyzing the molecular features and functions of chaperonins, particularly those of Group II, when normal and when bearing a pathological structural alteration, e.g., a mutation or posttranslational modification that in humans causes disease. The archaeal models mimic the human Group II, cytosolic chaperonins with adequate accuracy and are simpler and, therefore, are more easily amenable to manipulations and analysis in vitro and in vivo. In what pertains to simplicity, there are archaea, e.g., *Pyrococcus furiosus*, that have a single chaperonin gene, but its product forms double-ring hexadecamers just like the human CCT. On the other side of the spectrum, some *Methanosarcina* possess three Group II chaperonin genes, and *Methanosarcina acetivorans* has five, encoding related but distinct subunits that also form typical hexadecamers. These methanosarcinas can be considered candidates to examine the formation of various types of hexadecamers, each with a different set of subunits, and to dissect specific interactions between subunits and substrate preferences, as suggested by studies with *Sulfolobus solfataricus* (Chaston et al. 2016). The archaeal organisms with two or only one chaperonin can be used to amplify subtle defects caused by nonlethal but pathogenic mutations, as we have done with *P. furiosus* in our studies of the chaperonopathy caused by a mutation in the human CCT5 subunit.

Notably, some archaeal species have both chaperonins of Group II and chaperonins of Group I. Therefore, both Groups are available for study in a single cell,

within the same compartment: this is a unique situation since in human cells, which also possess the two Groups, one (Group II) is in the cytosol whereas the other (Group I) is inside the mitochondria. This is the canonical picture of human cells. However, now we know that human chaperonins of Group I are also present in the cytosol (and other places in the body, Fig. 9.1), which makes those archaea with the two groups in the cytoplasm very convenient models to examine interactions between them as they might occur in the human cell cytosol.

Last but not least, in view of the progress made with chaperonotherapy to treat various chaperonopathies at the experimental stage, one can foresee an application of archaeal chaperones in therapeutics. Positive chaperonotherapy consists of the use of normal chaperones, or its genes, to replace defective chaperones causing disease. It also includes the use of compounds that would interact with a defective chaperone and boost its capacity to function correctly. These are examples of positive chaperonotherapy in which archaeal models could be applied in preclinical tests. There is also negative chaperonotherapy, in which the objective is to inhibit the action of pathogenic chaperones like those known to favor cancer growth and dissemination. In this case, archaeal chaperones could be used as targets to screen compounds and find which ones have the greatest inhibitory capacity.

Acknowledgments The authors were partially supported by the Institute of Marine and Environmental Technology (IMET), Baltimore, MD, USA; and by the Euro-Mediterranean Institute of Science and Technology (IEMEST), Palermo, Italy. This is IMET contribution number IMET 16-193.

References

Bouhouche A, Benomar A, Bouslam N, Chkili T, Yahyaoui M (2006) Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia. J Med Genet 43(5):441–443. Epub 2006 Jan 6. PMID: 16399879. http://jmg.bmj.com/content/43/5/441.long; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2564519/

Chaston JJ, Smits C, Aragão D, Wong AS, Ahsan B, Sandin S, Molugu SK, Molugu SK, Bernal RA, Stock D, Stewart AG (2016) Structural and functional insights into the evolution and stress adaptation of Type II chaperonins. Structure 24(3):364–374. doi:10.1016/j.str.2015.12.016. Epub 2016 Feb 4. PMID: 26853941

Conway de Macario E, Maeder DL, Macario AJL (2003) Breaking the mould: archaea with all four chaperoning systems. Biochem Biophys Res Commun 301(4):811–812. No abstract available. PMID: 12589784

Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Bäumer S, Jacobi C, Brüggemann H, Lienard T, Christmann A, Bömeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol 4(4):453–461. PMID: 12125824

Figueiredo L, Klunker D, Ang D, Naylor DJ, Kerner MJ, Georgopoulos C, Hartl FU, Hayer-Hartl M (2004) Functional characterization of an archaeal GroEL/GroES chaperonin system: significance of substrate encapsulation. J Biol Chem 279(2):1090–1099. Epub 2003 Oct 23. PMID: 14576149

Galagan JE, Nusbaum C, Roy A, Endrizzi MG, Macdonald P, FitzHugh W, Calvo S, Engels R, Smirnov S, Atnoor D, Brown A, Allen N, Naylor J, Stange-Thomann N, DeArellano K, Johnson R, Linton L, McEwan P, McKernan K, Talamas J, Tirrell A, Ye W, Zimmer A, Barber RD, Cann I, Graham DE, Grahame DA, Guss AM, Hedderich R, Ingram-Smith C, Kuettner HC, Krzycki JA, Leigh JA, Li W, Liu J, Mukhopadhyay B, Reeve JN, Smith K, Springer TA, Umayam LA,

White O, White RH, Conway de Macario E, Ferry JG, Jarrell KF, Jing H, Macario AJL, Paulsen I, Pritchett M, Sowers KR, Swanson RV, Zinder SH, Lander E, Metcalf WW, Birren B (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. Genome Res 12(4):532–542. PMID: 11932238

Hirtreiter AM, Calloni G, Forner F, Scheibe B, Puype M, Vandekerckhove J, Mann M, Hartl FU, Hayer-Hartl M (2009) Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. Mol Microbiol 74(5):1152–1168. doi: 10.1111/j.1365-2958.2009.06924.x. Epub 2009 Oct 15. PMID: 19843217 http://www.nature.com/srep/2014/141027/srep06688/full/srep06688.html; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4209464/

Klunker D, Haas B, Hirtreiter A, Figueiredo L, Naylor DJ, Pfeifer G, Müller V, Deppenmeier U, Gottschalk G, Hartl FU, Hayer-Hartl M (2003) Coexistence of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. J Biol Chem 278(35):33256–33267. Epub 2003 Jun 9. PMID: 12796498

Kumar CM, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. Cell Stress Chaperones 20(4):555–574. doi:10.1007/s12192-015-0598-8. Epub 2015 May 20. Review. PMID: 25986150

Large AT, Lund PA (2009) Archaeal chaperonins. Front Biosci (Landmark Ed) 14:1304–1324. Review. PMID: 19273132

Macario AJL, Cappello F, Zummo G, Conway de Macario E (2010) Chaperonopathies of senescence and the scrambling of interactions between the chaperoning and the immune systems. Ann N Y Acad Sci 1197:85–93. doi:10.1111/j.1749-6632.2010.05187.x. Review. PMID: 20536837

Macario AJL, Conway de Macario E (1999) The archaeal molecular chaperone machine: peculiarities and paradoxes. Genetics 152(4):1277–1283. Review. PMID: 10430558

Macario AJL, Conway de Macario E (2005) Sick chaperones, cellular stress, and disease. N Engl J Med 353(14):1489–1501. Review. No abstract available. PMID: 16207851

Macario AJL, Conway de Macario E (2007) Chaperonopathies by defect, excess, or mistake. Ann N Y Acad Sci 1113:178–191. Epub 2007 May 4. Review. PMID: 17483209

Macario AJL, Conway de Macario E (2009) The chaperoning system: physiology and pathology. Exp Med Rev 2-3:9–21

Macario AJL, Conway de Macario E, Cappello F (2013) The chaperonopathies. In: Diseases with defective molecular chaperones. Springer, Dordrecht. http://www.springer.com/biomed/book/978-94-007-4666-4

Macario AJL, Lange M, Ahring BK, Conway de Macario E (1999) Stress genes and proteins in the archaea. Microbiol Mol Biol Rev 63(4):923–967. table of contents. Review. PMID: 10585970

Macario AJL, Malz M, Conway de Macario E (2004) Evolution of assisted protein folding: the distribution of the main chaperoning systems within the phylogenetic domain archaea. Front Biosci 9:1318–1332. Review. PMID: 14977547

Maeder DL, Macario AJL, Conway de Macario E (2005) Novel chaperonins in a prokaryote. J Mol Evol 60(3):409–416. PMID: 15871051

Min W, Angileri F, Luo H, Lauria A, Shanmugasundaram M, Almerico AM, Cappello F, Conway de Macario E, Lednev IK, Macario AJL, Robb FT (2014) A human CCT5 gene mutation causing distal neuropathy impairs hexadecamer assembly in an archaeal model. Sci Rep 4:6688. doi:10.1038/srep06688. PMID: 25345891

Mukherjee K, Conway de Macario E, Macario AJL, Brocchieri L (2010) Chaperonin genes on the rise: new divergent classes and intense duplication in human and other vertebrate genomes. BMC Evol Biol 10:64. doi:10.1186/1471-2148-10-64. PMID: 20193073

Shah R, Large AT, Ursinus A, Lin B, Gowrinathan P, Martin J, Lund PA (2016) Replacement of GroEL in *Escherichia coli* by the group II chaperonin from the archaeon *Methanococcus maripaludis*. J Bacteriol. pii: JB.00317-16. [Epub ahead of print] PMID: 27432832

Techtmann SM, Robb FT (2010) Archaeal-like chaperonins in bacteria. Proc Natl Acad Sci U S A 107(47):20269–20274. doi:10.1073/pnas.1004783107. Epub 2010 Nov 5. PMID: 21057109

Zako T, Sahlan M, Fujii S, Yamamoto YY, Tai PT, Sakai K, Maeda M, Yohda M (2016) Contribution of the C-terminal region of a group II chaperonin to its interaction with prefoldin and substrate transfer. J Mol Biol 428(11):2405–2417. doi:10.1016/j.jmb.2016.04.006. Epub 2016 Apr 11. PMID: 27079363

Part IV
Evolution of Multiple Copies of Chaperonins

Chapter 10
Evolution of Multiple Chaperonins: Innovation of Evolutionary Capacitors

Mario A. Fares

Abstract Molecular chaperones and chaperonins are part of large and evolutionarily conserved protein families that are involved in a large variety of functions in the cell. Chief among these functions is their ability to assist other slow-folding proteins in acquiring their native conformation. Because of their roles in the cell, chaperonins have enthralled scientists for decades and are considered the most important quality control mechanisms of the cell. In this chapter, I present evidence that supports a remarkable expansion of chaperonin protein families through gene duplication. Because of their ability in modulating phenotype through genotype, chaperonins are potent capacitors of evolution, as they allow the survival of innovative mutations despite their destabilizing effects for protein structures. In this sense, chaperonins increase the resistance of proteins to mutations and fuel evolvability by enabling proteins for a wider exploration of genotypic network. The complexity of the range of functions in which chaperonins are involved and the latest studies magnify the importance of these molecules as moonlighting proteins involved in a wide variety of independent, however equally important, functions. Future studies will aim at understanding how these proteins give origin to novel functions and using these capacitors as a mean to define the functional landscape of the cell.

10.1 Two Families of the Chaperonins

The primary amino acid sequence of synthesized proteins contains the required information to drive folding through a tortuous energetic landscape (Fig. 10.1) toward the minimum energy and most stable native conformation (Anfinsen 1973).

M.A. Fares
Department of Abiotic Stress, Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Valencia, Spain

Department of Genetics, Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland
e-mail: mfares@ibmcp.upv.es

C. M. Santosh Kumar, S.C. Mande (eds.), *Prokaryotic Chaperonins*, Heat Shock Proteins 11, DOI 10.1007/978-981-10-4651-3_10

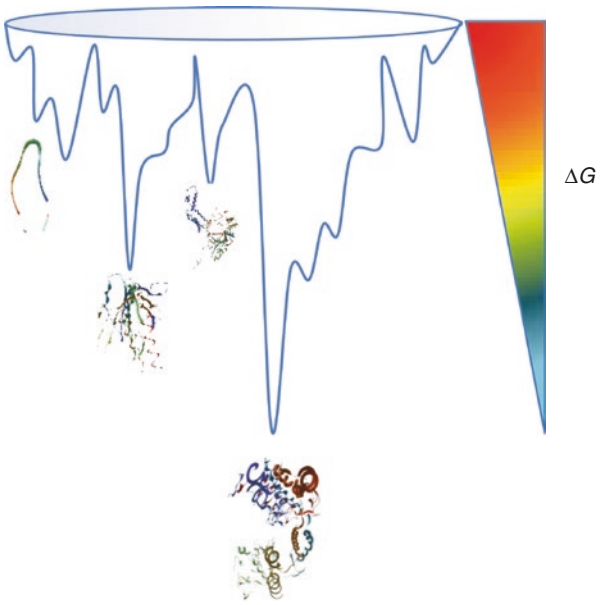


Fig. 10.1 Energy landscape for protein folding. A protein is synthesized as a linear polypeptide chain, which starts a complex pathway of folding to achieve the minimum energy (ΔG) state. At the beginning of this pathway, the ΔG is high (*red section of the triangle*), and the protein is unstable. The structure of the protein will stabilize as it approaches the lowest energy section of the landscape, where it acquires the native, functional, conformation

However, in a crowded environment such as the cytosol of a cell or organelle, where a 100 nm cube contains roughly 30,000 small molecules and 50,000 ions (Goodsell 1991), nonspecific aggregates are often formed through interactions between hydrophobic-exposed surfaces of partially unfolded proteins (Ellis and Minton 2006; Gershenson and Gierasch 2011). This problem is even bigger for slow-folding proteins, which constitute about 30% of all the proteins in the bacterium *Escherichia coli* (Ellis and Minton 2006), where the assistance of folding machines, such as chaperones, becomes absolutely essential. Finally, most non-synonymous nucleotide mutations (i.e., mutations that involve a change in the codon and encoded amino acid) occurring in protein-coding genes are deleterious owing to their stochastic nature leading to destabilized protein structures and eventually resulting in the formation of nonspecific protein aggregates (Tokuriki and Tawfik 2009b). Chaperones bind exposed hydrophobic surfaces of non-native proteins in a nonspecific way, thus forestalling misfolding and nonspecific aggregation and allowing new opportunities for native folding to occur.

Chaperonins are a subset of the folding molecules that constitute megadaltonic-sized machines structurally organized around an empty crowding-safe cage that constitutes an independent dimension within the crowded cytosolic space. Chaperonins are generally organized as multiple subunits, each of which comprises a number of different domains (three in general) that are intertwined into a mechanical relationship, where the movement of one domain impinges conformational changes in the other domains. Each subunit presents two main domains, the apical that contains the substrate binding regions and the equatorial that contains the ATP-binding sites as well as the substrate-interaction sites.

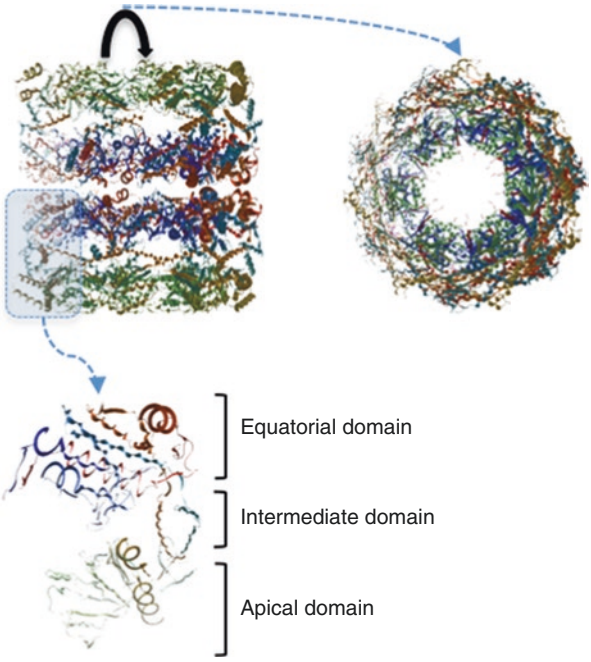


Fig. 10.2 General structure of chaperonins. This molecular machine is organized in two rings back-to-back oriented, with a folding chamber, and each formed by several subunits. Subunits comprise three domains: apical, intermediate, and equatorial

Both these domains move in a relative manner to one another through complex and substantial conformational changes mediated by the third flexible domain, the intermediate domain (Saibil and Ranson 2002). These subunits are organized into one or two rings oriented back to back and that each forms a chamber that works in trans with its sister ring (Fig. 10.2). In an ATPase-dependent manner, these chaperonins undergo conformational changes that allow transferring the unfolded protein from the hydrophobic lining in the opening of the ring to the cage, completing the cycle of protein folding in the cage and releasing the fully or partially folded protein to the cell environment (Ranson et al. 1998; Sigler et al. 1998). Effectively, therefore, these amazing machines are able to lead unfolded proteins through the most efficient paths in the energy landscape and rescue trapped intermediates from high-energy non-native folding forms (Dill and Chan 1997; Wolynes et al. 1995).

There are two main families of chaperonins that occupy all three domains of life, attesting to their absolute essentiality for the cell functioning. Chaperonin family I, or group I chaperonins, resides in the bacterial cytoplasm (also known as GroEL) or in the eukaryotic organelles of bacterial endosymbiotic origin, including mitochondria (Hsp60) or chloroplasts and hydrogenosomes (Brocchieri and Karlin 2000). Type II chaperonins, on the other hand, are present in the archaebacterial cytoplasm (generally termed thermosome) and in the eukaryotic cytosol (CCT/TRiC) (Frydman et al. 1992; Gutsche et al. 1999; Kubota et al. 1995; Trent et al. 1991). Despite the fact that these chaperonins are ubiquitous in all three domains of life, it remains intriguing the observation that some organisms lack chaperonins that are considered absolutely essential in all tested physiological conditions. For example, many species of

Mycoplasma and *Ureaplasma* are unique in that they lack a chaperonin homolog of any kind (Lund 2009). It was hypothesized that the loss of GroEL in these bacteria must have been accompanied by the loss in these bacteria of proteins requiring GroEL for folding, or alternatively, protein GroEL clients have evolved in these bacteria to fold independently of GroEL (Williams and Fares 2010). Indeed, there is some evidence (Fujiwara et al. 2010) that some homologs of *E. coli* GroEL clients fold in a GroEL-independent manner in these bacteria. Evolutionary analyses support this idea, as they show that *Mycoplasma* GroEL client proteins have undergone relaxed selective constraints and accumulation of mutations in these client proteins that have possibly favored the evolution of a GroEL-independent protein folding (Williams and Fares 2010). Finally, mutagenesis experiments using the MetK protein, an obligate client of *E. coli* GroEL but which homolog in *Ureaplasma* folds in a GroEL-independent manner, have shown that even a single mutation can convert MetK from a GroEL-independent into a GroEL-dependent protein, suggesting that GroEL independence is marginal (Ishimoto et al. 2014). There are differences in the structural properties between GroEL-independent and GroEL-dependent proteins, with GroEL being able to accelerate the rate of TIM-barrel domain folding (Georgescauld et al. 2014).

Groups I and group II chaperonins are different in a number of respects. Firstly, group I chaperonins includes homo-oligomeric complexes, whereas group II chaperonins comprises, with the exception of few archaeal chaperonins, hetero-oligomeric proteins. This hetero-oligomerism likely evolved multiple times independently (Archibald et al. 1999) through a neutral rather than selective evolutionary process (Ruano-Rubio and Fares 2007). In mouse, the hetero-oligomeric nature of the type II chaperonin CCT is extensive and includes eight subunits highly divergent at the sequence level and with precise arrangements in the structural octameric ring (Kubota et al. 1994, 1995; Liou et al. 1998). The identification of orthologs for each of the eight mouse CCT subunits in yeast suggests that these subunits are the result of ancient duplication events that were concomitant with the origin of eukaryotes (Archibald et al. 2000). Secondly, the ability of group I chaperonins to fold proteins is conditioned by the mechanical action of a co-chaperone that acts as a cap for the protein folding cavity, while group II chaperonins use an additional helical protrusion in the apical domain of the subunits that provides this function (Pereira et al. 2012; Reissmann et al. 2007). Thirdly, the allosteric behavior of the two chaperonin groups is different owing to the different subunits organization in the double-ringed structure (Braig et al. 1994; Ditzel et al. 1998; Lopez et al. 2015; Sewell et al. 2004). Fourthly, the inner cavity presents different distribution of surface charges between the two chaperonin groups, with group I being dominated by negatively charged amino acids (Cong et al. 2010). GroEL, the best representative of type I chaperonins, works assisted by the cofactor GroES, a 10 kDa protein that binds the apical domain of the protein inducing the required conformational changes for the transfer of the unfolded substrate to the center of the heptameric ring. In contrast to GroEL, type II chaperonins possess a protrusion stemming from the apical domain that makes the role of a lid pushing substrates into the central cavity (Horwich and Saibil 1998; Klumpp et al. 1997; Llorca et al. 1999). Type II chaperonin have been found, nevertheless, to require assistance during the folding cycle from GroES-unrelated co-chaperonins (Gebauer et al. 1998; Geissler et al. 1998; Siegers et al. 1999;

Vainberg et al. 1998). Substrate specificity is also unique to each of the chaperonin types as revealed by extensive interactome studies (Dekker et al. 2008; Fujiwara et al. 2010; Houry et al. 1999; Kerner et al. 2005; Yam et al. 2008). Indeed, many eukaryotic proteins can be folded by type II, but not type I, chaperonins (Tian et al. 1995), and vice versa, many proteins are folded by type I chaperonins only but not by type II chaperonins (Hirtreiter et al. 2009). In a recent study, however, Shah and colleagues were able to replace *E. coli groEL* with a thermosome from *M. maripaludis* and improve the function of this thermosome in folding GroEL client proteins by introducing a number of mutations at two sites in the apical domain, suggesting that the differences in the specificity of the two groups of chaperonins for their protein clients are small and can be overcome with only nuanced changes in key amino acid positions (Shah et al. 2016).

10.2 Chaperonins Are Evolutionary Capacitors

Chaperones and chaperonins assist slow-folding proteins in reaching their native conformation, prevent protein aggregation, and refold misfolded proteins (Hartl et al. 2011; Hartl and Hayer-Hartl 2009; Young et al. 2004). Because of these functions, molecular chaperones and chaperonins can restore the native conformation of proteins destabilized by environmental or genetic (i.e., mutations) perturbations. For instance, denatured proteins by heat stress can recover their native conformation assisted by molecular chaperones, which usually are highly abundant under a number of stresses, including high temperatures. Protein structures destabilized by mutations can also reach a native conformation aided by molecular chaperones. This ability to buffer the effects of destabilizing mutations allows some chaperones and chaperonins providing protein structures with resistance against destabilizing mutations. This alteration of the mapping of protein genotypes to their phenotypes (i.e., their structures) is known as mutational or genetic robustness, a property underlying most biological systems (de Visser et al. 2003; Fares 2015; Wagner 2005).

Genetic robustness increases the capacity of biological systems to evolve by incrementing the genotypic space that can be explored in a neutral manner and the eventual finding of new phenotypes by subsequent mutations (Wagner 2012). It follows then that increasing the robustness of an organism enhances its ability to generate heritable genetic variation, a property known as evolvability (Fares 2015). However, a trade-off exists between the size of the neutral genotypic space and the evolvability of a system. As the genotypic network increases, the set of possible phenotypes accessible through subsequent mutations from different genotypes of the same genotypic network start to overlap, eventually leading to lower evolvability with higher robustness of the system (Fig. 10.3) (Draghi et al. 2010). Take, for instance, two genotypic networks generated by a number of genotypic backgrounds. Genotypic backgrounds within the same network are connected by a single mutation and mutations that link the genotypes within the same network have no phenotypic effect. In Fig. 10.3, for instance, there is one network containing three genotypes (nodes), all of which encode a single phenotype (color black), and these

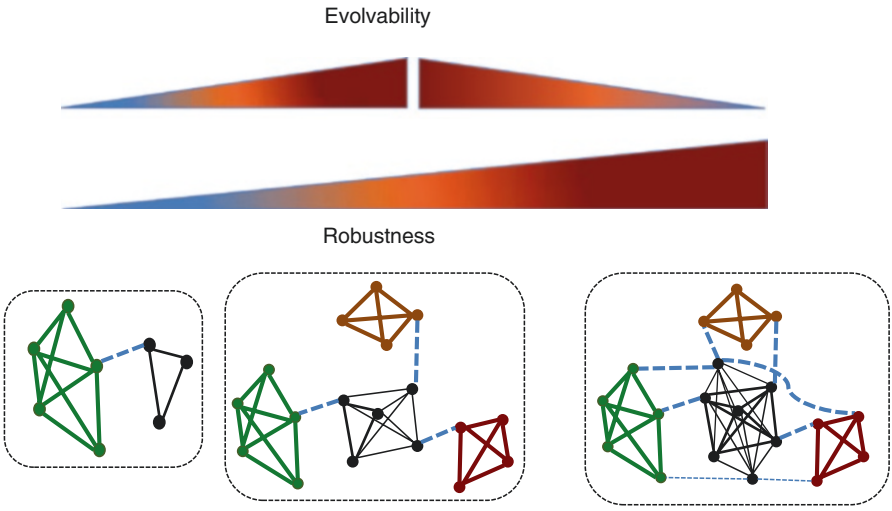


Fig. 10.3 Robustness and evolvability are correlated, with intermediate levels of robustness exhibiting maximum evolvability. In this figure the evolvability of a genotypic network, whose genotypes (nodes) all encode the same phenotypes (*black color*), is measured. The different phenotypes are color-coded

genotypes are linked through single mutations. On the other hand, there is the network of five genotypes that encode color green also connected through single mutations. In the example, the two networks are accessible through a mutation in one of the genotypes. Potentially, therefore, the black network contains three genotypes, but two potential phenotypes (black and green) are accessible through single mutations (evolvability of the black genotypic network in this case could be computed as the number of accessible phenotypes divided by the number of genotypes in the network: $E = 2/3 = 0.66$). Let us assume that one is able to increase the neutral genotypic network that encodes the black color by molecular means, for instance, using molecular chaperones, adding two additional genotypes that also code for black color but which allow accessing two additional genotypic networks encoding red and brown colors. In this case, increasing the robustness of the black network by two genotypes has increased the number of accessible phenotypes from two to four, hence yielding an evolvability higher than the previous, less robust network (evolvability = 4 phenotypes/5 genotypes = 0.8). Additional increases of the size of the black network lead to an overlap between the phenotypic spaces accessible through the genotypes of this network, reducing its evolvability.

Overwhelming evidence supports the role of chaperones and chaperonins in increasing the genetic robustness of organisms, hence their evolvability. For instance, the heat-shock protein 90 kDa, Hsp90, has been proposed to be a capacitor of evolution by increasing the resistance of phenotypes to mutations—that is, keeping the mutations cryptic in the populations. Hsp90 folds proteins involved in the signal transduction pathway, and thus impairment of its functions leads to dramatic consequences for the organism's development. In agreement with this asser-

tion, impairment or disruption of Hsp90 function through heat stress or pharmacological inhibition using drugs such as radicicol or geldanamycin reveals cryptic genetic variation in populations of a variety of organisms and wide range of distant species (Queitsch et al. 2002; Rutherford and Lindquist 1998; Sangster et al. 2004). This cryptic genetic variation is heritable and leads to novel adaptations in populations. For example, a sharp decline of light in an environment, in which natural surface populations of the cavefish *Astyanax mexicanus* live, can lead to the manifestation of phenotypic variants in the form of eyeless fishes that were silent in the previous conditions, in which light was abundant, and this new phenotypes can now be better fit to living in cave environments, where eye development is costly and useless (Rohner et al. 2013). Independent evidence also supports the role of Hsp90 as an evolutionary capacitor, as the gene copies of the duplicated kinases that interact with Hsp90 exhibit faster rates of evolution (i.e., greater number of mutations per site in the gene sequence when compared to orthologs in other organisms) than their sister gene copies that do not require or interact with Hsp90 (Lachowiec et al. 2015).

The chaperonin GroEL, which is unrelated to Hsp90, folds many proteins in *E. coli* (a minimum of 250 proteins) and was also shown to increase the tolerance of populations of *E. coli* to protein-destabilizing mutations. To show this increased tolerance, Fares and colleagues experimentally evolved a hyper-mutagenic version of the bacterium *E. coli* under conditions of strong genetic drift (Fares et al. 2002b): a single colony was transferred daily to a new plate with fresh growth media (Fig. 10.4). Under these conditions, the effective population size of the bacterium is 1 because the transmission from generation to generation is clonal. Under these clonal conditions, mutations accumulate in the genome in an irreversible manner, such that mutations at time t_1 accumulate those produced at time t_0 as well. After roughly 3200 generations of evolution, the biological fitness of the evolved *E. coli* populations was compared to their ancestral parental populations, yielding a sharp decline in fitness that resulted from the irreversible fixation of genome-wide deleterious mutations, a well-known symptom of genetic drift (Moran 1996).

Overexpression of the chaperonin GroEL in the evolved *E. coli* cells with declined fitness allowed an almost full recovery of the fitness of these strains, strongly supporting the capacity of GroEL to buffer the deleterious effects of mutations (Fares et al. 2002b). The role of GroEL in buffering the effects of mutations is also supported by the fixation of folding-enhancing mutations in GroEL from endosymbiotic bacteria of insects (Fares et al. 2002a, 2005), known to undergo strong genetic drift effects (Moran 1996), and by the fact that proteins that require GroEL for folding evolve faster than proteins that fold in a GroEL-independent manner (Bogumil and Dagan 2010, 2012; Sabater-Munoz et al. 2015; Williams and Fares 2010). The capacity of GroEL of buffering the effects of mutations in *E. coli* and other bacteria (Maisnier-Patin et al. 2005) was used in another study to evolve four enzymes with weak specificities for alternative substrates in the presence of high concentrations of GroEL and under mutational drift. Enhancing GroEL concentrations allowed the folding of enzyme variants with destabilizing core mutations and the emergence of enzymes with strong specificities for alternative substrates (Tokuriki and Tawfik 2009a).

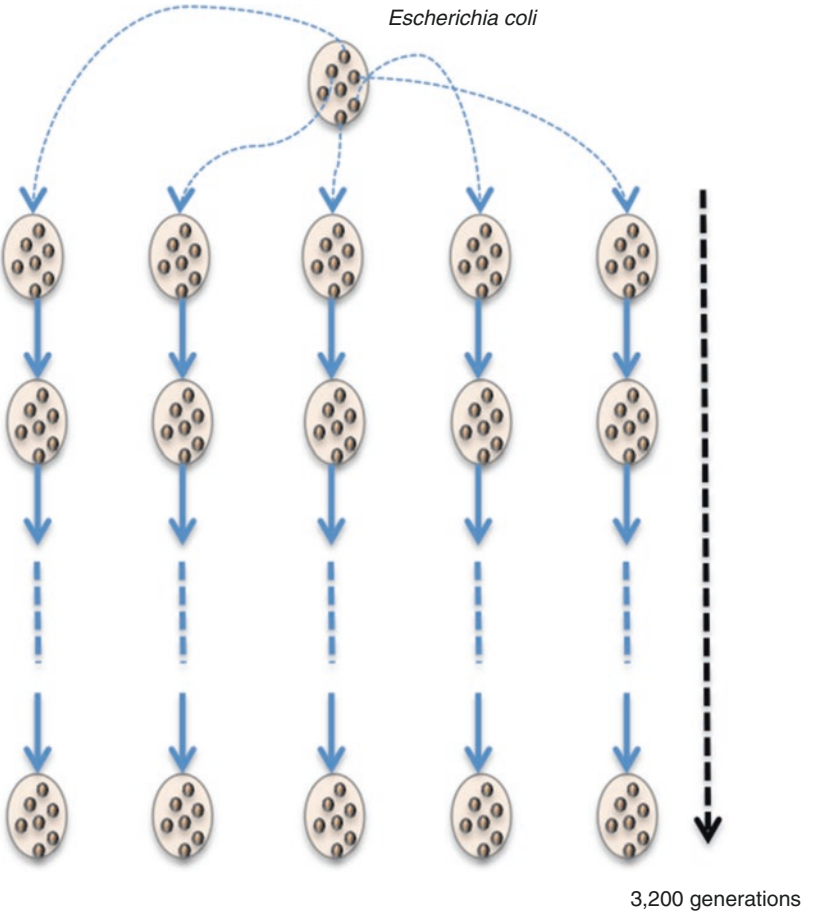


Fig. 10.4 Experimental evolution of the bacterium *Escherichia coli* under strong genetic drift effects. A single ancestral population was daily transferred by picking one colony and restriking it in a new plate containing fresh growth medium. The population was evolved for 3200 generations of the bacterium

The mitigating effect of chaperones does not seem to be the only property of Hsp90 and GroEL. Recently, it has been shown in an evolution experiment similar to that conducted with GroEL that the chaperone DnaK (Hsp70) is also able to buffer the effects of deleterious mutations and allow the rapid evolution of proteins requiring DnaK for folding. Two studies support the role of DnaK in accelerating the evolution of its protein clients. Firstly, a theoretical study in which authors studied the evolutionary parameters derived from the comparison of *E. coli* proteome to that of 1149 other bacteria showed that proteins with high binding affinity to DnaK evolve on average 4.3-fold faster than proteins in the lowest binding affinity class (Kadibalban et al. 2016). Secondly, a theoretical and experimental study in which populations of *E. coli* were subjected to string genetic drift effects showed that proteins with higher affinity for DnaK withstand more deleterious mutations than proteins with lower DnaK affinities and that such result is observable at long

and short evolutionary distances, implying that many of the mitigated mutations are eventually fixed in the populations and perhaps lead to adaptive phenotypes (Aguilar-Rodriguez et al. 2016).

10.3 Gene Duplication and the Expansion of Chaperonin Functions

Gene duplication involves the doubling of a gene creating two identical copies of that gene and of the encoded functions. Gene duplications are errors that have a stochastic origin as by-products of a number of fundamental cellular processes, including unequal crossing between homologous genome regions, retrotransposition, capture of DNA fragments through breaks in the double helix, and ectopic exchange of DNA fragments (Fig. 10.5). The duplication of genes that encode

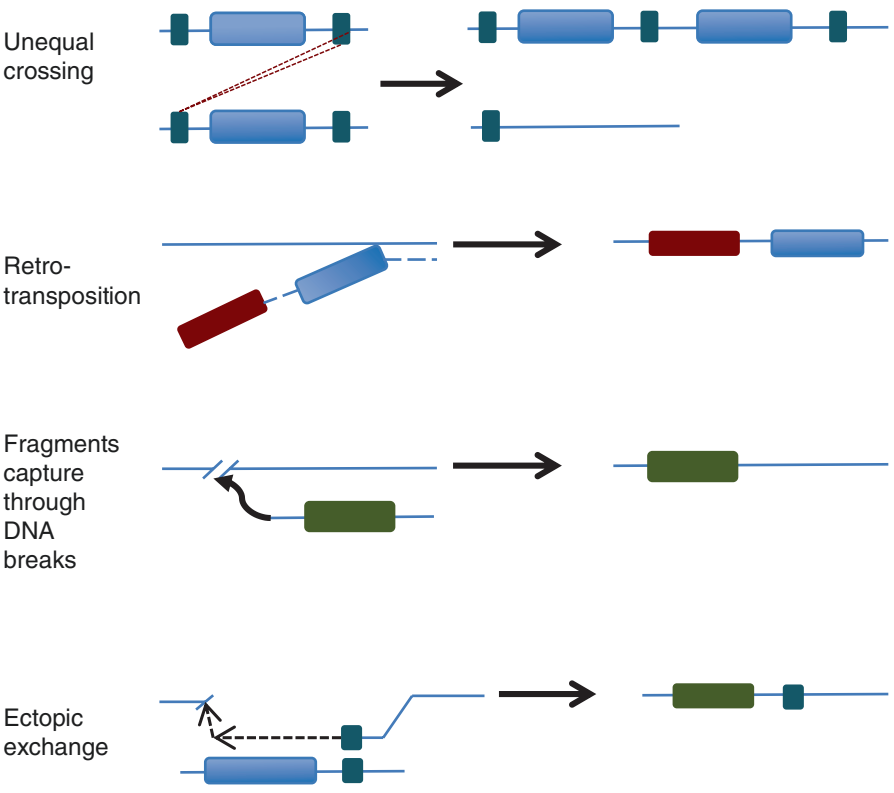


Fig. 10.5 Gene duplication is a phenomenon that takes place as a by-product of a number of cellular events, including unequal crossing between homologous DNA fragments, retrotransposition, capturing fragments of DNA through DNA breaks, and ectopic DNA exchange. Large rectangular shapes represent genes, while short ones represent neighboring regions with homologs elsewhere in the genome

functions with important contribution to the fitness of organisms has important evolutionary consequences that can lead to the origin of novel functions. Indeed, early evolutionary theory predicts that, as one gene copy is performing the ancestral needed function, natural selection relaxes against the sister copy, allowing it explore a wider genotypic space and eventually find novel functions (Fig. 10.6) (Ohno 1970).

Population genetics theory predicts that most duplicated genes return to single-copy genes shortly after duplication because an entirely redundant gene will accumulate most of the mutations that occur regardless of their deleterious effect due to relaxed selective constraints, being this followed by gene nonfunctionalization and

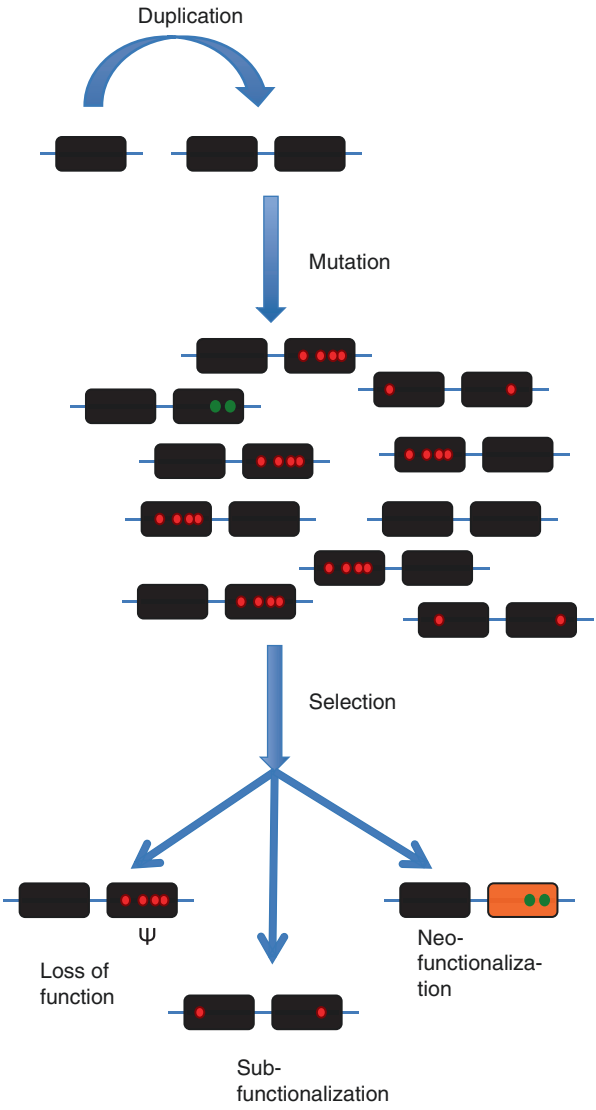


Fig. 10.6 Gene duplication, mutation, and selection. After gene duplication, one or two gene copies (*black rectangles*) can undergo deleterious mutations (*red dots*), which can lead to the nonfunctionalization of the affected gene copy or the subfunctionalization of each gene copy. In rare occasions, mutations may be beneficial (*green dots*) allowing the origin of novel functions in a process known as neo-functionalization

eventual genome erosion (Innan and Kondrashov 2010; Lynch and Conery 2000; Moore et al. 2005; Ohno 1970) (Fig. 10.6). Selection can, nevertheless, favor the fixation of a second gene copy if this copy either endues organisms with mutational robustness (Fares et al. 2013; Gu et al. 2003; Keane et al. 2014), increases of gene dosage are selectively favorable (Conant and Wolfe 2008), gene copies have diverged in their functions through a process of subfunctionalization (i.e., the partitioning of ancestral functions) (Barkman and Zhang 2009; Des Marais and Rausher 2008; Force et al. 1999) or neo-functionalization (i.e., the finding of new functions) (He and Zhang 2005), or copies have diverged in their expression levels (Francino 2005).

Gene duplication is generally followed by the return of the duplicated gene to single-copy genes, as supported by the death of the second gene copy in 92% of all duplicates originated from a whole-genome duplication in the yeast *Saccharomyces* roughly 100 MYA (Marcet-Houben and Gabaldon 2015; Wolfe 2015; Wolfe and Shields 1997). Against this prediction, the yeast *S. cerevisiae* preserves 32% of its genome in duplicate (Fares et al. 2013), and plants exhibit more than 50% of their genes in duplicate (Blanc and Wolfe 2004; Cui et al. 2006).

This preservation is likely the result of complex events of sub- and neo-functionalization, as evidenced by the concomitant occurrence of genome duplication events and the emergence of major morphological, metabolic, and functional innovations in plants and, however, much rarer, in animals (Blanc and Wolfe 2004; Hoegg et al. 2004; Holub 2001; Kim et al. 2004; Lespinet et al. 2002; Maere et al. 2005). Indeed, gene duplication is responsible for the expansion of many gene families, which have increased the ability of organisms to respond to environmental and cellular perturbations. Examples abound in the scientific literature, but the most noticeable include protein families from plants such as pepsin- and subtilisin-like proteases (Lespinet et al. 2002), metacaspases (Holub 2001), regulatory genes (Maere et al. 2005), and developmentally important MADS-box genes (Aoki et al. 2004; Kim et al. 2004; Kramer et al. 1998; Purugganan et al. 1995; Roque et al. 2016).

In microbes, gene duplication has also led to the expansion of the chaperonins, with some microbes possessing up to nine chaperonins. For example, *E. coli* requires only one chaperonin gene (*cpn60*, also known as *groEL*), which is essential under all known physiological conditions. However, analysis of 669 complete bacterial genomes has revealed that about 30% of bacteria have more than 1 *cpn60*; however, only 1 of these seems essential for their viability (Lund 2009). The presence of different copy numbers of chaperonins in microbes does not appear to obey a random process. For example, while some mycoplasmas lack *cpn60*, many do require *cpn60* for survival (Clark and Tillier 2010). Some bacterial groups, such as Spirochaetes, only require one cpn60 gene, while almost entire phylogenetic groups of others (e.g., Actinobacteria) possess multiple *cpn60* copies (Henderson et al. 2013).

Three explanations have been given for the presence of multiple chaperonins in these bacteria: (1) the gene products have similar functions but act as part of a complex cell cycle or regulatory system, (2) there has been a co-evolution between the duplication and divergence of *cpn60* gene copies and the evolution of their specificities for

those proteins they fold, and (3) the extra genes encode proteins with functions that are alternative to the original folding function (i.e., they moonlight) (Henderson et al. 2013). There is evidence in support of one of the three alternative hypotheses mentioned above or of more than one such hypothesis. For example, the Alphaproteobacteria group contains microorganisms with the greatest number of *cpn60* genes, with some, such as *Bradyrhizobium japonicum*, having seven cpn60 genes (Kaneko et al. 2002). *B. japonicum* group contains many microorganisms that associate symbiotically with plants and contribute to plants capacity in the fixation of nitrogen. In some of these microorganisms, such as *Sinorhizobium meliloti*, the expression of *groEL*1 is essential for the expression of the *nod* gene protein, required for nodule formation in the plant where nitrogen is fixed, and this gene cannot be replaced by *E. coli groEL* (Ogawa and Long 1995). Of the four *groE* operons that *S. meliloti* possess, only two are essential for viability and growth (Bittner et al. 2007). *Rhizobium leguminosarum*, another bacterium form the same group, contains three *cpn60* genes, named *cpn60.1*, *cpn60.2*, and *cpn60.3* with 60–80% sequence identity between one another resulting from these genes sharing a common gene ancestor (Rodriguez-Quinones et al. 2005). These chaperonins show clear evidence of divergence in their functions since their evolution from the last gene common ancestor. In support of this claim, in *R. leguminosarum*, the protein encoded by *cpn60.1* is present at higher abundances than its paralog encoded by *cpn60.2*. By contrast, there is no evidence for the expression of *cpn60.3* under normal growth conditions. Of the three chaperonins, only *cpn60.1* was shown to be essential for cell viability, while lacking *cpn60.2*, *cpn60.3*, or both of these chaperonins does not seem to influence the viability of the cells (Rodriguez-Quinones et al. 2005). Their folding activity also provides evidence of their functional divergence: *cpn60.1* and *cpn60.2* encode proteins able to fold the lactate dehydrogenase in *R. leguminosarum*, while their paralog encoded by *cpn60.3* does not fold this protein (George et al. 2004).

Another attractive example of chaperonins duplications is that of the *Chlamydiae*, a group of bacteria with all its members sharing three *cpn60* paralogous genes that resulted from two events of gene duplication (Karunakaran et al. 2003; Lund 2009). The presence of all three cpn60 in the entire group supports as plausible hypothesis that the duplication events giving rise to all three *cpn60* genes predated the origin of the different *Chlamydiae* (Fig. 10.7) (McNally and Fares 2007). The peculiarity of these chaperonins is that the sequence identity between the three paralogs (~20%) is substantially below that reported for cpn60 paralogs in other bacterial groups. This low sequence identity is accompanied by a number of differences in the expression and functional dynamics of the three chaperonins. For instance, *cpn60.1* are the one most expressed in HeLa cells and are responsive to high temperatures by increasing their expression levels (Karunakaran et al. 2003). *cpn60.2* or *cpn60.3* are unable to rescue a *groEL*-lacking *E. coli* from lethality (Karunakaran et al. 2003). In *Chlamydia*-infected joins, isolated *Chlamydia* exhibit no expression of *cpn60*.3 but high concentrations of proteins encoded by *cpn60*.1 and *cpn60*.2, with the latter being more expressed than the former (Gerard et al. 2004).

The functional divergence between the three paralogous cpn60 genes was further demonstrated using bioinformatics analyses using a formal mathematical test in the

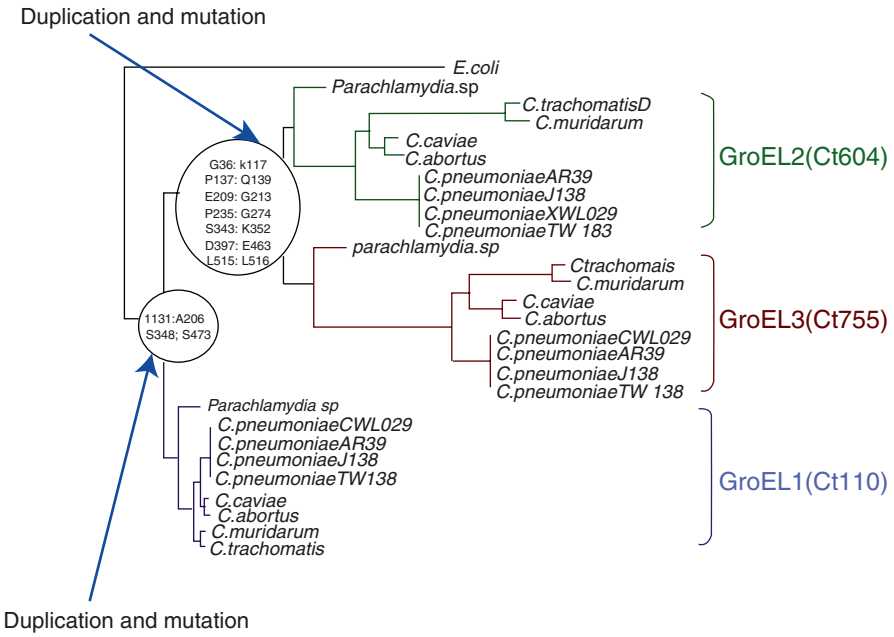


Fig. 10.7 Evolution of multiple chaperonins in the bacterial group *Chlamydiae*. Two events of gene duplication took place during *Chlamydiae* evolution, yielding three *groEL* gene copies. After each duplication event, many mutations have allowed the divergence of *groEL* functions

maximum likelihood framework of functional divergence (McNally and Fares 2007). Briefly, the method tests two competing hypothesis: (a) the null hypothesis considers that there is no evidence in support of the observed patterns of sequence divergence in the phylogenetic tree of paralogous genes, while (b) the alternative hypothesis assumes that such differential evolution between the groups of paralogous sequences is the result of functional divergence (Gu 1999). To compare these competing hypotheses, the method estimates the value of the functional divergence parameter (θ), with values of this parameter higher than 0 supporting the alternative hypothesis, whereas values of this parameter of 0 support the null hypothesis.

The two models (i.e., the null model assuming $\theta = 0$, hence no functional divergence, and the alternative model assuming $\theta > 0$, hence functional divergence) are fitted to the data assuming a given phylogenetic tree, and a log-likelihood value is computed to measure the likelihood of the model (l_0 is the likelihood of the null model and l_1 is the likelihood of the alternative model). The log-likelihood values of both of the models are then compared to one another by the likelihood-ratio test (LRT = $2\Delta l = 2(l_0 - l_1)$), this LRT value is approximated to a χ^2 with 1 degree of freedom. Posterior Bayesian probabilities are also calculated for each of the amino acid sites in the alignment of the paralogous sequences, which allow identifying those amino acid changes responsible for the divergence in the function between

paralogous genes. Applying this method to the *Chlamydiae* group allowed authors in concluding that the three groups of *cpn60* paralogs have diverged in their functions significantly since their lost common gene ancestor and that most of this divergence has taken place between *cpn60*.1 and *cpn60*.2. Other examples of functional divergence between multiple chaperonin copies exist, details of which can be found elsewhere (Henderson et al. 2013).

10.4 Moonlighting in the Chaperonins

The existence of enzymes with multiple functions was first discovered by Joram Piatigorsky, who reported that the lens crystallin protein in the duck was the metabolic enzyme, argininosuccinate lyase—he called this phenomenon gene sharing (Piatigorsky et al. 1988). Later, the structural biologist Connie Jeffery transformed the term gene sharing into protein moonlighting, a term that stems from the exercise of a second job, at night, in addition to a daytime occupation (Jeffery 1999, 2009). Since then, defining protein moonlighting clashed with a number of other terminologies with nuanced differences that made it difficult to reach a consensus for a formal definition of protein moonlighting. Today, proteins generated by gene fusions, homologous but nonidentical proteins, splice variants, protein decoration variants, protein fragments, and proteins with identical functions performed in different cellular locations are not considered examples of protein moonlighting (Jeffery 2009). Therefore, moonlighting proteins are those proteins that have one or more functions, independent biological activities, in addition to the initial activity by which the protein was first known.

Chaperonin 60 (Cpn60) has been shown to exhibit a large number of independent functions since the discovery that the mucus-binding proteins of the enteric bacteria, *Salmonella typhimurium*, was the Cpn60 of this organism (Ensgraber and Loos 1992). This report was surprising owing to two main reasons. First, Cpn60 has as main activity folding of partially folded proteins in the cell. Second, Cpn60 is an eminently intracellular protein, so it was surprising finding it bound to the surface of a cell. In fact, later reports showed that Cpn60 can be found bound to the surface of the cell in 22 different bacteria, where it is thought to function as an adhesin, binding to components of the host (see (Henderson et al. 2013), including humans (De Bruyn et al. 2000). Gene duplication and functional divergence is likely to be one of the causes for the functional promiscuity of the chaperonins. Cpn60.2, the paralog gene of *cpn60.1*, from *Mycobacterium tuberculosis* stimulates human monocytes to synthesize pro-inflammatory cytokines (Friedland et al. 1993). Many other functions have been reported for the chaperonins, which can be summarized in the following:

1. In eukaryotes, Cpn60 has been found in the cytosol of the cell, outside the mitochondrion, hinting at a possible folding-independent function for this chaperonin.
2. Cpn60 has been found in the surface of cells, perhaps acting as receptors.

3. Ligand binding.
4. Intracellular signaling,

A complete list of known independent functions for Cpn60 can be found elsewhere (Henderson et al. 2013).

10.5 Molecular Basis of Functional Innovation in the Chaperonin GroEL

As mentioned earlier in this chapter, *groEL* is part of an operon, *groE*, in which the cofactor GroES is also encoded and the expression of both of the genes, *groE* and *groS*, allows the construction of a homotetradecamer formed by two back-to-back oriented rings, each formed by seven identical subunits of GroEL and GroES. Each of the subunits is formed by three domains that are structurally distinguishable and functionally different: the apical domain, which contains the amino acid regions that bind unfolded proteins and peptides and the cofactor GroES; the equatorial domain, which contains the ATP-binding sites; and the intermediate domain that allows the flexible movement of the apical domain and facilitates the transition of the subunits between *cis* and *trans* conformations that are required for the protein folding cycle of GroESL (Braig et al. 1994; Hunt et al. 1996; Xu et al. 1997). The functions performed by each of the subunits have been assigned to specific amino acid positions of *E. coli* GroEL. However, since GroEL has undergone many duplications in bacteria (Lund 2009), events of adaptive evolution (Fares et al. 2002a), and functional divergence (McNally and Fares 2007), one would expect certain flexibility in the functional code of this protein. In fact, specific amino acid evolutionary replacements have been linked to important structural changes and functional shifts in GroEL. In agreement with this, changes in the amino acid composition of the co-chaperonin GroES can determine GroEL functioning as either a double- or single-ring complex (Liu et al. 2009).

Given the wide range of different and unrelated functions that GroEL can perform, the high sequence conservation of GroEL across long phylogenetic distances is striking. In particular, it is not known how amino acid changes in this protein map to functional changes despite a large body of directed mutagenesis experiments. Having said this, it has been hypothesized that, perhaps, the multifunctional nature of GroEL results from the existence of a reservoir of functionalities encoded through specific interactions between structurally linked amino acid regions within each of the domains of GroEL. This picture is even more complicated considering that such functionalities may be kept as cryptic alternative weak functions under certain physiological conditions but become strong primary functions under other conditions (Ruiz-Gonzalez and Fares 2013).

To demonstrate the functional and structural links between amino acid sites within GroEL domains, authors conducted coevolutionary analyses under the assumption that such links may be under selective constraints. Indeed, using a formal

test of coevolution in which the co-variability of amino acid sites in a GroEL alignment for hundreds of bacteria is calculated (Fares and McNally 2006; Fares and Travers 2006), authors found strong coevolution signatures between functionally important sites. In such coevolutionary networks, authors found that GroES amino acids Leu27 and Gly29, known for their essentiality in the interaction of GroES with GroEL, exhibited strong coevolutionary signatures. Other amino acids structurally close to these ones were also identified to coevolve with Leu27 and Gly29, forming what authors called functional sector—amino acids located in structurally continuous regions. Such functional sectors were also found in each of the domains of GroEL subunits. Remarkably, different sectors seem to include amino acid sites that have individually been involved in functions different to protein folding. These results revealed an enormous evolutionary plasticity for GroEL across the entire bacterial phylogeny, illuminating the molecular basis of its functional diversity.

10.6 Future Perspectives in the Understanding of Chaperonins' Evolution

While reading throughout this chapter, and I believe other chapters in this book, the reader will quickly realize that chaperonins offer an unprecedented potential to understand many fundamental questions in molecular evolution, including the mapping of genotypes to phenotypes, the link between the sequence of a protein and its function, and the functional plasticity hidden in the amino acid sequence and structural code of this protein. Chaperonins have been implicated in a large number of different functions, and a challenge for the future is to identify all possible functions that remain hidden in these proteins.

Perhaps, nature has already provided hints in a way to identify these functions, as the duplication of chaperonins in bacteria has led to the discovery of a wide range of previously unsuspected functions for these proteins. Identifying these functions is also exciting from the biotechnological and medical perspectives, particularly given the role that these proteins have in promoting the evolution of other proteins in the cell and in signal transduction. Future studies should aim at experimentally evolving this protein under controlled laboratory conditions and identifying the parameters that define its functional landscape. In the same way, domesticating the buffering capacity of the chaperonins can aid advances in biotechnology by fueling the evolution of alternative functions in other enzymes that require chaperonins for folding. This domestication could make use of the discovery made by Lund's lab on the interchangeability of protein clientele between chaperonins of different groups despite their long evolutionary distance by nuanced amino acid replacements. Breaking the chaperonin code to identify its functions and evolutionary plasticity will make its domestication a reality in the foreseeable future.

References

Aguilar-Rodriguez J, Sabater-Munoz B, Montagud-Martinez R, Berlanga V, Alvarez-Ponce D, Wagner A, Fares MA (2016) The molecular chaperone DnaK is a source of mutational robustness. Genome Biol Evol

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M (2004) Phylogeny and divergence of basal angiosperms inferred from APETALA3- and PISTILLATA-like MADS-box genes. J Plant Res 117:229–244

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

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

Barkman T, Zhang J (2009) Evidence for escape from adaptive conflict? Nature 462:E1; discussion E2–3

Bittner AN, Foltz A, Oke V (2007) Only one of five groEL genes is required for viability and successful symbiosis in Sinorhizobium meliloti. J Bacteriol 189:1884–1889

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16:1679–1691

Bogumil D, Dagan T (2010) Chaperonin-dependent accelerated substitution rates in prokaryotes. Genome Biol Evol 2:602–608

Bogumil D, Dagan T (2012) Cumulative impact of chaperone-mediated folding on genome evolution. Biochemistry 51:9941–9953

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature 371:578–586

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

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

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950

Cong Y, Baker ML, Jakana J, Woolford D, Miller EJ, Reissmann S, Kumar RN, Redding-Johanson AM, Batth TS, Mukhopadhyay A, Ludtke SJ, Frydman J, Chiu W (2010) 4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. Proc Natl Acad Sci U S A 107:4967–4972

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res 16:738–749

De Bruyn J, Soetaert K, Buyssens P, Calonne I, De Coene JL, Gallet X, Brasseur R, Wattiez R, Falmagne P, Montrozier H, Laneelle MA, Daffe M (2000) Evidence for specific and non-covalent binding of lipids to natural and recombinant Mycobacterium bovis BCG hsp60 proteins, and to the *Escherichia coli* homologue GroEL. Microbiology 146(Pt 7):1513–1524

de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC (2003) Perspective: evolution and detection of genetic robustness. Evolution 57:1959–1972

Dekker C, Stirling PC, McCormack EA, Filmore H, Paul A, Brost RL, Costanzo M, Boone C, Leroux MR, Willison KR (2008) The interaction network of the chaperonin CCT. EMBO J 27:1827–1839

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454:762–765

Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4:10–19
Ditzel L, Lowe J, Stock D, Stetter KO, Huber H, Huber R, Steinbacher S (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. Cell 93:125–138
Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010) Mutational robustness can facilitate adaptation. Nature 463:353–355
Ellis RJ, Minton AP (2006) Protein aggregation in crowded environments. Biol Chem 387:485–497
Ensgraber M, Loos M (1992) A 66-kilodalton heat shock protein of Salmonella typhimurium is responsible for binding of the bacterium to intestinal mucus. Infect Immun 60:3072–3078
Fares MA (2015) The origins of mutational robustness. Trends Genet 31:373–381
Fares MA, McNally D (2006) CAPS: coevolution analysis using protein sequences. Bioinformatics 22:2821–2822
Fares MA, Travers SA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. Genetics 173:9–23
Fares MA, Barrio E, Sabater-Munoz B, Moya A (2002a) 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, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E (2002b) Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature 417:398
Fares MA, Moya A, Barrio E (2005) Adaptive evolution in GroEL from distantly related endosymbiotic bacteria of insects. J Evol Biol 18:651–660
Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW (2013) The roles of whole-genome and small-scale duplications in the functional specialization of *Saccharomyces cerevisiae* genes. PLoS Genet 9:e1003176
Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545
Francino MP (2005) An adaptive radiation model for the origin of new gene functions. Nat Genet 37:573–577
Friedland JS, Shattock RJ, Griffin GE (1993) Phagocytosis of Mycobacterium tuberculosis or particulate stimuli by human monocytic cells induces equivalent monocyte chemotactic protein-1 gene expression. Cytokine 5:150–156
Frydman J, Nimmesgern E, Erdjument-Bromage H, Wall JS, Tempst P, Hartl FU (1992) Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. EMBO J 11:4767–4778
Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J 29:1552–1564
Gebauer M, Zeiner M, Gehring U (1998) Interference between proteins Hap46 and Hop/p60, which bind to different domains of the molecular chaperone hsp70/hsc70. Mol Cell Biol 18:6238–6244
Geissler S, Siegers K, Schiebel E (1998) A novel protein complex promoting formation of functional alpha- and gamma-tubulin. EMBO J 17:952–966
George R, Kelly SM, Price NC, Erbse A, Fisher M, Lund PA (2004) Three GroEL homologues from Rhizobium leguminosarum have distinct in vitro properties. Biochem Biophys Res Commun 324:822–828
Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, Hayer-Hartl M, Hartl FU (2014) GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. Cell 157:922–934
Gerard HC, Whittum-Hudson JA, Schumacher HR, Hudson AP (2004) Differential expression of three Chlamydia trachomatis hsp60-encoding genes in active vs. persistent infections. Microb Pathog 36:35–39
Gershenson A, Gierasch LM (2011) Protein folding in the cell: challenges and progress. Curr Opin Struct Biol 21:32–41
Goodsell DS (1991) Inside a living cell. Trends Biochem Sci 16:203–206

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16:1664–1674
Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421:63–66
Gutsche I, Essen LO, Baumeister W (1999) Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. J Mol Biol 293:295–312
Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 16:574–581
Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332
He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164
Henderson B, Fares MA, Lund PA (2013) Chaperonin 60: a paradoxical, evolutionarily conserved protein family with multiple moonlighting functions. Biol Rev Camb Philos Soc 88:955–987
Hirtreiter AM, Calloni G, Forner F, Scheibe B, Puype M, Vandekerckhove J, Mann M, Hartl FU, Hayer-Hartl M (2009) Differential substrate specificity of group I and group II chaperonins in the archaeon Methanosarcina mazei. Mol Microbiol 74:1152–1168
Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59:190–203
Holub EB (2001) The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev Genet 2:516–527
Horwich AL, Saibil HR (1998) The thermosome: chaperonin with a built-in lid. Nat Struct Biol 5:333–336
Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of in vivo substrates of the chaperonin GroEL. Nature 402:147–154
Hunt JF, Weaver AJ, Landry SJ, Gierasch L, Deisenhofer J (1996) The crystal structure of the GroES co-chaperonin at 2.8 A resolution. Nature 379:37–45
Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97–108
Ishimoto T, Fujiwara K, Niwa T, Taguchi H (2014) Conversion of a chaperonin GroEL-independent protein into an obligate substrate. J Biol Chem 289:32073–32080
Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci 24:8–11
Jeffery CJ (2009) Moonlighting proteins—an update. Mol Biosyst 5:345–350
Kadibalban AS, Bogumil D, Landan G, Dagan T (2016) DnaK-dependent accelerated evolutionary rate in prokaryotes. Genome Biol Evol 8:1590–1599
Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K, Kohara M, Matsumoto M, Shimpo S, Tsuruoka H, Wada T, Yamada M, Tabata S (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement). DNA Res 9:225–256
Karunakaran KP, Noguchi Y, Read TD, Cherkasov A, Kwee J, Shen C, Nelson CC, Brunham RC (2003) Molecular analysis of the multiple GroEL proteins of Chlamydiae. J Bacteriol 185:1958–1966
Keane OM, Toft C, Carretero-Paulet L, Jones GW, Fares MA (2014) Preservation of genetic and regulatory robustness in ancient gene duplicates of *Saccharomyces cerevisiae*. Genome Res 24:1830–1841
Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122:209–220
Kim S, Yoo MJ, Albert VA, Farris JS, Soltis PS, Soltis DE (2004) Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. Am J Bot 91:2102–2118
Klumpp M, Baumeister W, Essen LO (1997) Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. Cell 91:263–270

Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. Genetics 149:765–783

Kubota H, Hynes G, Carne A, Ashworth A, Willison K (1994) Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. Curr Biol 4:89–99

Kubota H, Hynes G, Willison K (1995) The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. Eur J Biochem 230:3–16

Lachowiec J, Lemus T, Borenstein E, Queitsch C (2015) Hsp90 promotes kinase evolution. Mol Biol Evol 32:91–99

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res 12:1048–1059

Liou AK, McCormack EA, Willison KR (1998) The chaperonin containing TCP-1 (CCT) displays a single-ring mediated disassembly and reassembly cycle. Biol Chem 379:311–319

Liu H, Kovacs E, Lund PA (2009) Characterisation of mutations in GroES that allow GroEL to function as a single ring. FEBS Lett 583:2365–2371

Llorca O, Smyth MG, Carrascosa JL, Willison KR, Radermacher M, Steinbacher S, Valpuesta JM (1999) 3D reconstruction of the ATP-bound form of CCT reveals the asymmetric folding conformation of a type II chaperonin. Nat Struct Biol 6:639–642

Lopez T, Dalton K, Frydman J (2015) The mechanism and function of group II chaperonins. J Mol Biol 427:2919–2930

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

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A 102:5454–5459

Maisnier-Patin S, Roth JR, Fredriksson A, Nystrom T, Berg OG, Andersson DI (2005) Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet 37:1376–1379

Marcet-Houben M, Gabaldon T (2015) Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the Baker's yeast lineage. PLoS Biol 13:e1002220

McNally D, Fares MA (2007) In silico identification of functional divergence between the multiple groEL gene paralogs in Chlamydiae. BMC Evol Biol 7:81

Moore RC, Grant SR, Purugganan MD (2005) Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Mol Biol Evol 22:91–103

Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A 93:2873–2878

Ogawa J, Long SR (1995) The Rhizobium meliloti groELc locus is required for regulation of early nod genes by the transcription activator NodD. Genes Dev 9:714–729

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin

Pereira JH, Ralston CY, Douglas NR, Kumar R, Lopez T, McAndrew RP, Knee KM, King JA, Frydman J, Adams PD (2012) Mechanism of nucleotide sensing in group II chaperonins. EMBO J 31:731–740

Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, Borras T, Nickerson JM, Wawrousek EF (1988) Gene sharing by delta-crystallin and argininosuccinate lyase. Proc Natl Acad Sci U S A 85:3479–3483

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 140:345–356

Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417:618–624

Ranson NA, White HE, Saibil HR (1998) Chaperonins. Biochem J 333(Pt 2):233–242

Reissmann S, Parnot C, Booth CR, Chiu W, Frydman J (2007) Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. Nat Struct Mol Biol 14:432–440

Rodriguez-Quinones F, Maguire M, Wallington EJ, Gould PS, Yerko V, Downie JA, Lund PA (2005) Two of the three groEL homologues in Rhizobium leguminosarum are dispensable for normal growth. Arch Microbiol 183:253–265

Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, Lindquist S, Tabin CJ (2013) Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. Science 342:1372–1375

Roque E, Fares MA, Yenush L, Rochina MC, Wen J, Mysore KS, Gomez-Mena C, Beltran JP, Canas LA (2016) Evolution by gene duplication of *Medicago truncatula* PISTILLATA-like transcription factors. J Exp Bot 67:1805–1817

Ruano-Rubio V, Fares MA (2007) Testing the neutral fixation of hetero-oligomerism in the archaeal chaperonin CCT. Mol Biol Evol 24:1384–1396

Ruiz-Gonzalez MX, Fares MA (2013) Coevolution analyses illuminate the dependencies between amino acid sites in the chaperonin system GroES-L. BMC Evol Biol 13:156

Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396:336–342

Sabater-Munoz B, Prats-Escriche M, Montagud-Martinez R, Lopez-Cerdan A, Toft C, Aguilar-Rodriguez J, Wagner A, Fares MA (2015) Fitness trade-offs determine the role of the molecular chaperonin GroEL in buffering mutations. Mol Biol Evol 32:2681–2693

Saibil HR, Ranson NA (2002) The chaperonin folding machine. Trends Biochem Sci 27:627–632

Sangster TA, Lindquist S, Queitsch C (2004) Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. Bioessays 26:348–362

Sewell BT, Best RB, Chen S, Roseman AM, Farr GW, Horwich AL, Saibil HR (2004) A mutant chaperonin with rearranged inter-ring electrostatic contacts and temperature-sensitive dissociation. Nat Struct Mol Biol 11:1128–1133

Shah R, Large AT, Ursinus A, Lin B, Gowrinathan P, Martin J, Lund PA (2016) Replacement of GroEL in *Escherichia coli* by the group II chaperonin from the archaeon Methanococcus maripaludis. J Bacteriol

Siegers K, Waldmann T, Leroux MR, Grein K, Shevchenko A, Schiebel E, Hartl FU (1999) Compartmentation of protein folding in vivo: sequestration of non-native polypeptide by the chaperonin-GimC system. EMBO J 18:75–84

Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL (1998) Structure and function in GroEL-mediated protein folding. Annu Rev Biochem 67:581–608

Tian G, Vainberg IE, Tap WD, Lewis SA, Cowan NJ (1995) Specificity in chaperonin-mediated protein folding. Nature 375:250–253

Tokuriki N, Tawfik DS (2009a) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668–673

Tokuriki N, Tawfik DS (2009b) Stability effects of mutations and protein evolvability. Curr Opin Struct Biol 19:596–604

Trent JD, Nimmesgern E, Wall JS, Hartl FU, Horwich AL (1991) A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. Nature 354:490–493

Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 93:863–873

Wagner A (2005) Robustness and evolvability of living systems. Princeton University Press, Princeton

Wagner A (2012) The role of robustness in phenotypic adaptation and innovation. Proc Biol Sci 279:1249–1258

Williams TA, Fares MA (2010) The effect of chaperonin buffering on protein evolution. Genome Biol Evol 2:609–619

Wolfe KH (2015) Origin of the yeast whole-genome duplication. PLoS Biol 13:e1002221
Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713
Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. Science 267:1619–1620
Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 388:741–750
Yam AY, Xia Y, Lin HT, Burlingame A, Gerstein M, Frydman J (2008) Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. Nat Struct Mol Biol 15:1255–1262
Young JC, Agashe VR, Siegers K, Hartl FU (2004) Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol 5:781–791