Functional Analysis of

M. tuberculosis Chaperones

C. M. Santosh Kumar

Laboratory of Structural Biology,

Centre for DNA Fingerprinting and Diagnostics (CDFD),

Nampally, Hyderabad

Functional Analysis of

M. tuberculosis Chaperones

Thesis submitted to

Manipal University, Manipal

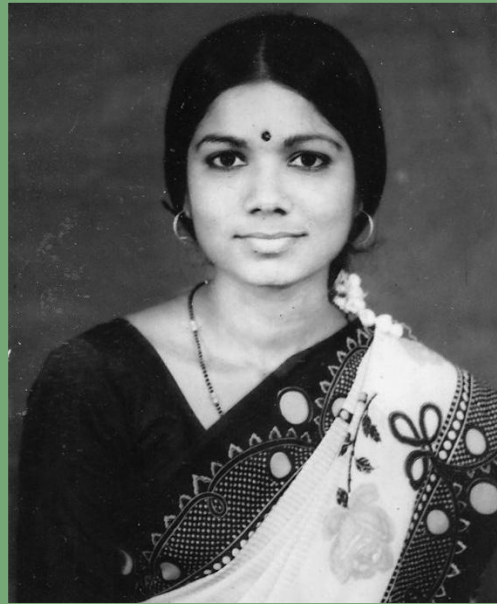For the Degree of

Doctor of Philosophy

By

C. M. Santosh Kumar

(Registration Number: 050100015)

Laboratory of Structural Biology,

Centre for DNA Fingerprinting and Diagnostics (CDFD),

Nampally, Hyderabad

2009

Dedicated to my Mother

DECLARATION

I hereby declare that the work towards this thesis entitled "Functional analysis of *M. tuberculosis* Chaperones" has been carried out by me under the supervision of Dr. Shekhar C. Mande, at the Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad. The work is original and has not been submitted in part or full for any other degree or diploma of any other Institute or University, earlier.

C. M. Santosh Kumar,

Laboratory of Structural Biology,

Centre for DNA Fingerprinting

and Diagnostics (CDFD),

Hyderabad.

Certificate

This is to certify that this thesis entitled **"Functional Analysis of *M. tuberculosis* chaperones"**, submitted by Mr. **C. M. Santosh Kumar**, for the degree of Doctor of Philosophy to the Manipal University, is based on the work carried out by him at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad. This work is original and has not been submitted in part or full for any other degree or diploma of any other university or institution.

Dr. Shekhar C. Mande,
Thesis Supervisor and
Dean, Academic Affairs,
Centre for DNA Fingerprinting
and Diagnostics (CDFD), Hyderabad.

Contents

Chapter I

Review of Literature

Chapter II

In vivo Complementation Studies on Mycobacterial GroEL Homologues

Chapter III

Gene Shuffling and Domain Swapping of Mycobacterial GroEL

Chapter IV

Biochemical Characterization of GroEL variants

Chapter V

Isolation and Characterization of GroES Independent GroEL Mutants

Chapter VI

Conclusions

Appendix I

Functional Analysis of *Mycobacterium tuberculosis* Hsp70

Appendix II

Oligonucleotide Primers and Plasmid Vectors

Acknowledgements

It is in fact a pleasant opportunity to thank everyone who has helped me in accomplishing this feat. My first thanks to my mentor Dr. Shekhar C. Mande for his everlasting encouragement and boundless independence, which have resulted in developing freedom of thought in me. I have joined CDFD after an exceptionally bad phase at my earlier institute. Dr. Shekhar kindly accepted me as his student. During this wonderful course of my association with him, I understand that he is a very gentle person and a nice human being. I am really fortunate to be associated with him. Words are not enough to thank him for the support and encouragement he had shown on me.

I thank the Director, CDFD, Dr. J. Gowrishankar for providing the facilities and encouragement to carry out research work. My sincere regards to the former Director, CDFD, Dr. Syed E Hasnain, for allowing me to join CDFD. I thank Vice Chancellor, Registrar and Head, Dept. of Biotechnology at Manipal University for allowing me to register as a Ph. D. student. I thank the faculty at CDFD, Dr. Ranjan, Dr. Sanjeev, Dr. Abhijit, Dr. Hari, Dr. Rupinder, Dr. Rashna, Dr. shubhadeep for helpful discussions and support.

I express my sincere thanks to Prof. Anil Tyagi for allowing me to work in his lab and for elaborate discussions we used to have. During my stay at UDSC, Delhi, I found that Prof. Tyagi as a systematic and caring person and this association certainly boosted my scientific temper. I also take this opportunity to thank Garima, Vineel for all the patience and support during my stay at UDSC and after.

It gives me immense pleasure in thanking my seniors Srikanth, Jagpreeth, Naresh and Radha for their support whenever I was in need. Many cheerful thanks

ABBREVIATIONS

°C	:	Degree centigrade
ADP	:	Adenosine-5'-diphosphate
ATP	:	Adenosine-5'-triphosphate
BCG	:	Bacillus of Calmette Guerin
BSA	:	bovine serum albumin
bp	:	base pair
cfu	:	colony forming unit
Ci	:	Curies
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxynucleotide triphosphate
DNAse	:	deoxyribonuclease
DTT	:	1, 4-Dithiothreitol
dTTP	:	2'-deoxythymidine-5'-triphosphate
E. coli	:	*Escherichia coli*
Mtb	:	*Mycobacterium tuberculosis*
EDTA	:	Ethylene diamine tetra acetic acid (disodium salt)
EtBr	:	Ethidium Bromide
HEPES	:	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
IPTG	:	Isopropyl-b-D-thiogalactopyranoside
kb	:	Kilo base pair
kDa	:	Kilo Dalton(s)

KOH	:	potassium hydroxide
MCS	:	Multiple cloning site
RBS	:	Ribosome binding site
mg	:	Milli gram (10^{-3} gram)
min	:	Minute(s)
ml	:	Millilitres (10^{-3} litres)
mM	:	Millimolar
mmol	:	Millimoles (10^{-3} moles)
ng	:	Nano gram ($10-9$ gram)
ADC	:	Oleic Albumin Dextrose Catalase growth supplement
OD	:	Optical density
ORF	:	Open reading frame
ori	:	origin of replication
PAGE	:	polyacrylamide gel electrophoresis
PAS	:	Para-aminosalisylic acid
PCR	:	Polymerase chain reaction
PMSF	:	phenylmethanesulphonylfluoride
rpm	:	rotations per minute
SDS	:	Sodium dodecyl sulfate
V	:	Volts
µg	:	Micro gram (10^{-6} grams)
µl	:	Micro litre (10^{-6} litre)
µM	:	Micro molar (10^{-6} Molar)

Preface

Mycobacterium tuberculosis (Mtb) is a dreadful pathogen causing about 2 lakh deaths a year. Several proteins, including the heat shock proteins (Hsps), of this organism are reported to be antigenic. Heat shock proteins of Mtb have been reported to be secreted and hence were shown to elicit immune response. This phenomenon is supposed to be advantageous for the pathogen in evading the host defences. Heat shock proteins, as the name suggests, are the proteins which are present in bulk amounts when the cells are shifted to higher temperature. Although initially identified to be up regulated during heat shock, the HSPs are over-expressed under a variety of stresses, such as nutrient deprivation, phagocytosis, oxidative stress etc. HSPs are ubiquitous and highly conserved across the taxons. Functionally, HSPs bind to a wide array of substrate proteins by virtue of stretches of hydrophobic residues that are otherwise buried in the properly folded native protein and help their folding, in most cases, by ATP driven cycles of binding and release.

Two of the well-characterized Heat shock proteins belong to the Hsp70 and the Hsp60 families. The current understanding of these proteins was derived from the studies on their *E. coli* counterparts DnaK and GroEL, respectively. These studies have established that the two proteins are the principal chaperones functioning in the cell for the protein quality control and are structurally and functionally distinct. Hsp70, working in concert with its co-chaperones, DnaJ and GrpE, binds to native and extended polypeptides ranging from 30 kDa to 150 kDa. Structural analysis of full length DnaK is not known, although crystal structures of different individual domains have been characterized. GroEL forms large oligomeric structures with two isologues heptameric rings enclosing two cavities for the substrate protein binding. GroEL binds to proteins that are partially un-folded or kinetically trapped intermediates, which are less than 50 kDa in mass. The genes encoding these proteins, in *E. coli* constitute the σ^{32} regulon. Various regulatory mechanisms for the same are instrumental in other organisms, details of which are beginning to be understood.

During the tenure of my Ph. D. thesis I have carried out functional analysis of the Hsp60 and Hsp70 homologues of Mtb. The genome of Mtb harbors two copies of genes encoding GroEL homologues, *groEL1* (Cpn60.1) and *groEL2* (Cpn60.2). Earlier

biochemical and structural studies have shown that these proteins are unusual. Mtb GroELs have been shown to exist as dimers, unlike the canonical tetradecamers form and the intermolecular interactions are mediated by apical domains, in contrast to the equatorial domains as shown in the *E. coli* GroEL. Moreover, they have been shown to be inefficient in prevention of aggregation, ATP hydrolysis and refolding of the substrate polypeptides.

Work done during my Ph. D. is divided into six chapters. I wish to present here a brief summary of the chapter wise contents of my thesis. **Chapter-I** reviews the current literature and attempts to present a comprehensive understanding of the concept of molecular chaperones. Various topics concerning the biology of molecular chaperones, including the structural and functional understanding on the two principal chaperone molecules, Hsp70 and Hsp60, are covered in this chapter. Moreover, notes on the discovery of the heat shock response and heat shock proteins, classification of these proteins and different mechanisms operating in heat shock response are included in this chapter.

The unusual behavior of Mtb GroELs in vitro led us to study their behavior in vivo. We tested if the two Mtb GroEL could complement the loss of GroEL function in *E. coli.* The **Chapter-II** describes the methods followed and the results obtained during the complementation studies. For this, we have employed two strains of *E. coli,* SV2, which harbors as TS *groEL44* allele and LG6, in which expression of chromosomal *groES/L* operon is under the control of P_{lac} promoter. Neither of Mtb GroELs could rescue loss of GroEL function in *E. coli* (Fig. 1A). This led us to probe for the differences in the molecular features of the GroELs from Mtb and *E. coli*.

In order to investigate the molecular features which lead to the differences in the behavior of *E. coli* and Mtb GroELs, we generated a pool of *groEL* variants via DNA shuffling, starting with ORFs encoding Mtb GroEL1 and GroEL2 as template DNA. ORFs capable of encoding active versions of GroEL were selected in *E. coli* SV2. Sequence analysis of the mutants obtained showed that the substrate interacting apical domains can tolerate variations but the oligomerization equatorial domain is conserved among the variants and is homologus to *E. coli* GroEL. We therefore hypothesized that due to the presence of an "*E. coli* GroEL-like" equatorial domain, the GroEL variants could

function in vivo. To check this possibility, we have constructed *E. coli* and Mtb GroEL chimeras by exchanging the equatorial domains. The chimera derived from Mtb GroEL1 with the equatorial domain exchanged from *E. coli* GroEL, GroELMEF, was capable of complementing the loss of GroEL function when tested in different strains of *E. coli*. On the other hand the chimera derived from *E. coli* GroEL with the equatorial domain exchanged from Mtb GroEL1, GroELMER, turned inactive. Furthermore, GroELMEF was shown to function with *E. coli* GroES and could support bacteriophage morphogenesis. The **Chapter-III** describes the methods followed and the results obtained during this part of the project.

Encouraged by the results of the complementation with GroELMEF and GroELMER, I have attempted to study the biochemical characteristics of these proteins, including their oligomerization properties. In the **Chapter-IV**, I describe the detailed biochemical characterization of these proteins. Purified GroELMEF was capable of existing in a higher oligomeric state, similar to that seen for *E. coli* GroEL, whereas GroELMER displayed a lower oligomeric character. Moreover, GroELMEF was capable of exhibiting chaperonin characteristics similar to *E. coli* GroEL. However, GroELMEF was poor in substrate binding whereas GroELMER acts similar to *E. coli* GroEL, suggesting roles for the parental domains and that the variations in apical domain, to some extent can be absorbed without impairing chaperonin function as long as the molecule retains its ability to encapsulate the substrate proteins. These results therefore confirm that mere substrate recognition is not sufficient for the chaperonin function. On the contrary oligomerization is an important attribute of GroEL. Having established that oligomerization is the principal attribute of GroEL function we attempted to explore the oligomeric status of the Mtb GroEL in native conditions. Surprisingly, Mtb GroEL1 exhibited multiple oligomeric forms, monomer, dimer, heptamer and tetradecamer. Immunochemical studies followed by mass spectrometric analysis showed that the conversion from heptameric (single ring) form to the teteradecameric form is mediated by phosphorylation on a serine residue, viz. Ser-393. This is the first ever observation on a post-translational modification for a bacterial chaperone.

GroEL and GroES were shown to be essential for *E. coli* under all growth conditions tested. Precise reason for the essentiality of GroEL came from proteomic studies which showed that a fraction of the obligate GroEL substrates are essential for *E.*

coli growth. Moreover, eukaryotic GroEL homologues exhibit co-chaperone independent behaviour, owing to their built in lid. However the reasons for the essentiality of GroES are not clear. We have isolated two mutants of GroEL which could function in the absence of GroES. The **Chapter-V** describes the genetic and biochemical studies on these variants.

The **Chapter-VI** is the summary of the complete results. Also a discussion, on the possible roles of GroELs in Mtb is presented in this chapter.

Hsp70 of Mtb is encoded by *dnaK*, which exists as an operon with the genes *dnaJ*, *grpE* and *hspR*. DnaJ and GrpE are the co-chaperones and HspR acts as the co-repressor for this operon, as a complex with DnaK. On the other hand, Hsp70 has been shown to interact with the extra-cytopalsmic domain of human CD40 and thereby elicit chemokine response. Considering these two facets of Mtb Hsp70, we have attempted to study its interactions with HspR and CD40, at the molecular detail. The proteins were purified employing several expression systems, including bacterial, yeast and insect cell systems. Purified proteins were subjected to crystallizations for structural studies. Experimental details including the results obtained are presented as **Appendix-I**.

Appendix-II constitutes the lists of oligonucleotide primers and plasmids vectors from this study.

CHAPTER I

Review of Literature

1.1 Introduction

One of the central processes in biology concerns with the conversion of genetic information encoded by DNA into proteins that finally carry out the genetic program. Ribosomes play an important role in translating the genetic information from nucleotide bases into amino acid residues and consequently bringing about the formation of polypeptide chains. Subsequently, the naive polypeptides, in order to perform the preordained function, are required to be folded to form functional proteins.

The intracellular milieu being complex and extremely crowded, favors intermolecular interactions, which potentially lead to non-productive aggregation of unfolded species. As a result the translating polypeptides have an intrinsic propensity to populate non-native conformations awaiting sufficient structural information to be available, enhancing the problem of aggregation. The phenomenon of aggregation has many toxic consequences as in several neurodegenerative disorders, Parkinson's disease and Huntington's disease. To overcome such phenomena, cells have developed a sophisticated system of molecular chaperones, which are able to prevent protein aggregation and catalyze protein folding.

Molecular chaperones are a broad family of proteins that are ubiquitous and highly conserved (Lindquist and Craig, 1988; Georgopoulos and Welch, 1993). Molecular chaperones sequester the unfolded substrate proteins by binding to their otherwise buried hydrophobic patches and thereby, with the cycles of binding and release, assist the proteins fold in a sequestered environment. Apart from their role as folding catalysts, these proteins are involved in a multitude of biological processes, including disaggregation of protein aggregates, assembling multi-subunit proteins, polypeptide transport across biological membranes, and proteolysis (Bukau and Horwich, 1998; Hartl and Hayer-Hartl, 2002). Many of these proteins were initially identified as the abundant proteins during heat shock in Drosophila and later in *E. coli* and were hence termed heat shock proteins.

1.2 Discovery of Heat Shock Proteins - the Time Line

All life forms, including bacteria, animals and plants, respond to elevated temperatures and to chemical and physiological stress by a rapid and transient increase in the synthesis of a special class of proteins known as the Heat Shock Proteins (HSPs) and the response is called Heat Shock Response (HSR). Although these proteins are particularly important in stress conditions, they have been shown to be essential for normal growth, as well. For many years the Hsp genes were of academic interest till 1982 when a meeting on *Heat Shock: From Bacteria to Man*, held at the Cold Spring Harbour Laboratory revealed the heat shock field as a major study area in experimental biology (Ashburner, 1982). Although the "homeostatic activity" and the association with heat shock and other stresses of these proteins was understood, the functions of the HSPs remained mysterious at that time and the details of regulation of *hsp* gene expression were only beginning to emerge.

Initially, the discovery of heat shock response and the chaperone function of HSPs suffered from lack of acceptability. The central dogma proposed by Francis Crick in 1957 and the experiments of Christian B. Anfinsen in the 1950s and 1960s supported the notion of spontaneous folding of proteins, according to which, the information required for functional folding resides in the primary sequence of the protein. Despite the fact that these experiments were done with purified proteins and hence the observations were difficult to extrapolate to the situations in the crowded cellular milieu, the idea of assisted folding and the discovery of function of chaperones in protein folding were not considered relevant by the scientific community. Here, I aim to present a brief note on the discoveries that led to the discovery and understanding of heat shock response and Hsps.

1.2.1 Discovery of Heat Shock Response

Beginning of the discovery of heat shock response and the chaperone function too, suffered from wide criticism although both were accidental. Ferruccio Ritossa, working on nucleic acid synthesis in the salivary glands of *Drosophila melanogastor*, found in early 1960s that a transient increase in temperature would induce differences in the puffing pattern of special sections of Drosophila polytene chromosomes and these

differences were later demonstrated to facilitate a rapid increase in the RNA synthesis (Ritossa, 1962; 1963; 1964). Although this was one of the important demonstrations on the direct effect of environment on gene expression, the relevance of these observations was not clear (Ritossa, 1996). The increased RNA was later demonstrated to encode a special class of proteins, Heat Shock Proteins. However, it was to take another 10–15 years before the first *Drosophila* HSP mRNA was isolated (Ashburner, 1982). Around this time the HSPs were discovered in various organisms like in mammalian tissue culture cells, in *E. coli*, in yeast, and in plants (Kelley and Schlesinger, 1978; Lemaux et al., 1978; Bouche et al., 1979; Miller et al., 1979; Barnett et al., 1980; Hightower and White, 1981). In the following years, it was demonstrated that this response is conserved from bacteria to mammals, as are the different families of induced proteins (heat shock proteins, HSPs).

Several observations from late 1970s attempted to decipher the role of heat shock proteins: over expression of heat shock proteins under variety of stress and metabolic settings such as, glucose starvation, cancer and during particular steps of differentiation and development and the association of some of these proteins, which later were found to be members of Hsp90 family, with oncogenic protein kinases as well as steroid hormone receptors (SHRs) (Godowski and Picard, 1989; Morange, 2005). Moreover, the roles of these proteins were hypothesised in metabolism and in the control of cytoskeletal structure (Morange, 2005). Thus, although the biological processes in which the HSPs participated became known, the actual biochemical functions of these proteins remained largely unknown.

1.2.2 Discovery of Molecular Chaperones

Coincidentally, the members of two major families of chaperones, Hsp60 and Hsp70 and their physiological functions, were discovered at about the same time. The 70 kDa heat shock proteins from Drosophila were demonstrated to possess high sequence similarity with the *E. coli* DnaK (Bardwell and Craig, 1984). Although the Hsp70 along with its co-chaperones was first discovered in *E. coli* as host factors that, when mutated, block the bacteriophage propagation (Georgopoulos, 1977; Saito and Uchida 1977; Sunshine et al., 1977). It was Hugh Pelham in 1985, who reported the first demonstration of a chaperone function. Working on the expression of BiP, a member of

Hsp70 family from Drosophila, in mouse and monkey cell lines, he showed that the over expression of BiP rescues the morphology of the nucleolus and export of ribosomes, following heat stress (Pelham, 1984; 1986; Lewis and Pelham, 1985; Munro and Pelham, 1986). Following this, in the next five years, the notion of chaperones emerged successfully. Hsp70 was later shown to be required for import of proteins into sub-cellular organelles by its close association with the imported proteins on both sides of the organellar membranes (Deshaies et al, 1988; Chirico et al, 1988). Two years later, Hsp70 was shown to bind transiently with nascent polypeptides (Beckmann et al., 1990).

Based on the land mark observations on the assembly of ribulose 1–5 diphosphate carboxylase (Rubisco), the enzyme responsible for assimilation of CO_2 in chloroplasts, John Ellis in 1987 was able to name the function (Ellis, 1987). During the radio labelling studies with intact chloroplasts, the large subunit of rubisco was found not to assemble into holoenzyme, instead, co-migrating, on a 5% native PAGE, with a prominently staining protein with an apparent mass of about 700 kda, which erroneously was concluded to be different oligomeric form of Rubisco large subunit (Blair and Ellis, 1973). However, Roger Barraclough later demonstrated that the staining protein was not an oligomeric form of rubisco large subunit; on the contrary it was a different protein (Barraclough and Ellis, 1980). This was the first demonstration of a protein binding to a newly synthesized form of another protein (Ellis, 1996). Immuno-chemical studies with antibodies specific to the binding protein and rubisco large subunit showed that principal part of the large subunit is buried in the complex. Time course experiments on rubisco assembly and reconstitution studies ascribed the role of a chaperone to the binding protein. These observations were supported by Harry Roy and colleagues and extended their support by showing that the dissociation of large subunit from the binding protein, thereby its assembly into holoenzyme in the presence of small subunit is ATP dependent (Bloom et al., 1983). Later studies revealed the subunit composition of the binding protein (Hemmingsen and Ellis, 1986; Musgrove et al., 1987) and its homologues in the bacterial extracts, named GroEL (Hemmingsen et al., 1988). In analogy to *E. coli* DnaK, the role of GroEL and a few other Hsps was initially thought to be responsible in the morphogenesis of coliphages by mutational analysis, however its role in uninfected cells was not studied (Takanao and Kakefuda,

1972; Georgopoulos et al., 1973). GroEL was purified from bacterial extracts (Hendrix, 1979; Hohn et al., 1979) and oligomers of GroEL were demonstrated to associate transiently and non-covalently with protein B, followed by detection of the complex on density gradients and hence were supposed to be required for the assembly of phage lambda protein B into pre-connector (Kochan and Murialdo, 1983). GroEL and its co-chaperonin were purified and the chaperone function of GroEL and its association with GroES was demonstrated with the purified proteins a few years later (Goloubinoff et al., 1989 a & b). By protein cross linking, GroEL was shown to bind nascent polypeptides in *E. coli* based cell free translation system. Successive observations then followed to establish the importance of chaperone function.

1.2.3 Origin of the Term Molecular Chaperone

The term Chaperone was first used for the neurotoxin purified from the venom of the Australian taipan snake. Active form of the neurotoxin is composed of three subunits with homology at the amino terminus. One of the three subunits displays toxicity to an extent lesser than the holotoxin and the other two were proposed to act as chaperones that increase the specificity of the toxin and protect it against degradation (Fohlman et al., 1976). The authors however, did not attempt to extend the notion of chaperone to other systems.

The famous discovery of the chaperone action of nucleoplasmin, purified from Xenopus egg homogenates, on histones, allowing their correct assembly into nucleosomes in vitro followed similar path (Laskey et al., 1978). The role of nucleoplasmin was fit to be termed as a "molecular chaperone", i. e., its interaction with the substrates is transient, it is not part of the assembled nucleosome, it functions to prevent premature, improper interactions between oppositely charged histones and DNA by decreasing the rate of reaction. Hence, Ron Laskey and colleagues termed Nucleoplasmin as a Molecular Chaperone (Laskey et al., 1978).

Later, this notion was extended to the rubisco binding protein (Musgrove and Ellis, 1987) and to Hsp70 and Hsp90 from Drosophila (Pelham, 1986). The concept of Molecular chaperones was however generalized by Ellis (Ellis, 1987) and this was later extended and generalized to various systems. In addition, the term "Chaperonin", which

is used to refer the Hsp60 family of chaperones, was coined by Sean Hemmingsen in 1988 (Hemmingsen et al., 1988). The wave of research on the structural and functional aspects that followed with the discovery of various heat shock proteins from different forms of life enhanced the understanding of the chaperone function and the regulation of heat shock response thereafter. Increase in the number of heat shock proteins further led to the convenient classification of these proteins, based on their molecular masses.

1.3 Classification of Molecular Chaperones

Molecular chaperones are the key components of proteome contributing to cellular homeostasis and are responsible for protein folding, assembly, translocation and degradation in a broad array of normal cellular processes and the stabilization of proteins and membranes and protein refolding under stress conditions (Lindquist, 1986; Lindquist and Craig, 1988). A wide range of proteins has been reported to have chaperone activity and are classified according to their molecular masses. Since majority of the molecular chaperones were originally identified as heat-shock proteins (Hsps), the names of these molecular chaperones follow their early nomenclatures and are referred as Hsps/chaperones. Typical heat shock response includes five major families of heat shock proteins, Hsp100 (Clp) family, Hsp90 family, Hsp70 (DnaK) family, Hsp60 (GroEL) family the small Hsp (sHsp) Family (Bukau and Horwich, 1998; Narberhaus, 2002; Chang et al., 2007; Tang et al., 2007).

1.3.1 Hsp100 Family

Hsp100 family of molecular chaperones are the members of large ATPase family known as ATPases associated with various cellular activities (AAA Super family), owing to the nucleotide binding domains (NBD1 and NBD2) present within their primary sequences (Schirmer et al., 1996; Patel and Latterich, 1998; Neuwald et al., 1999; Agarwal et al., 2001). The prokaryotic homologue is known as ClpB. These molecular chaperones exist as hexameric ring like structures of 80-100 kD protomers (Lee et al., 2003). Apart from the regular chaperone function, in conjunction with ClpP, the protease, the Hsp100 family functions in protein disaggregation and/or protein degradation and therefore is important for the maintenance of cellular homeostasis (Glover and Lindquist, 1998; Golobionoff et al., 1999; Beuron et al., 1998; Weber-Ban et al., 1999; Horwich et al., 2001).

1.3.2 Hsp90 Family

Hsp90 family of molecular chaperones is the major species which requires ATP for its functions and is among the most abundant proteins in cells comprising 1–2% of total cellular protein (Buchner, 1999; Frydman, 2001). The Hsp90 family of molecular chaperones is distinct from other families of molecular chaperones because of a variety

of functions that have been assigned to this family, ranging from a role in morphological evolution and stress adaptation in Drosophila and Arabidopsis, to the principal role in assembly and maintenance of 26S proteasome (Pratt et al., 2001; Rutherford and Lindquist, 1998; Queitsch et al., 2002; Imai et al., 2003). The major role attributed to the Hsp90 family is to control protein folding in the cell. Since majority of its substrates are signal transduction proteins such as steroid hormone receptors and signaling kinases, it is therefore believed to play a key role in the signal-transduction networks, cell-cycle control, protein degradation and protein trafficking (Young et al., 2001; Ritcher and Buchner, 2001). Surprisingly, the families comprising Hsp100 and Hsp90 are absent from archeal branch of life, which constitutes many extremophilic organism (Laksanalamai et al., 2004).

1.3.3 Hsp70 Family

The family of Hsp70 chaperones comprises 70 kDa molecular chaperones that are ubiquitous and highly conserved (Bukau and Horwich, 1998). Hsp70 functions, together with their co-chaperones, Hsp40 and GrpE, in preventing aggregation, assisting refolding of non-native proteins under both normal and stress conditions, in protein import and translocation processes, in maintaining the cellular homeostasis by targeting the non-native and non-functional proteins to ClpP, lysosomes or proteasomes and in regulating other stress-associated gene expression (Hartl, 1996; Bimston et al., 1998; Frydman, 2001). Constitutively expressed members of this family are referred to as Hsc70 (70-kDa heat-shock cognate), which are often involved in assisting the folding of de novo synthesized polypeptides and the import/translocation of precursor proteins.

1.3.4 Hsp60 Family

The Hsp60 family comprises a distinct family of molecular chaperones, called chaperonins, found in the cytoplasm of prokaryotes and in the mitochondria and plastids but absent from endoplasmic reticulum of eukaryotes and form large oligomeric structures of 60 kDa protomers, comprising two isologus cavities for encapsulation of substrates (Bukau and Horwich, 1998; Hartl, 1996; Horwich et al., 2001). Chaperonins are further classified into two subfamilies, based on their cellular

localization and requirement of co-chaperonins for their action. The Group I chaperonins require the assistance from their co-chaperonin, Hsp10, which acts as a lid, and are found in cytoplasm of bacteria, mitochondria and chloroplasts of eukaryotes. The Group II chaperonins on the other hand contain an built-in lid within their primary sequence therefore are independent of co-chaperonins and are found in archaea and in the cytosol of eukaryotes (Ranson et al., 1998).
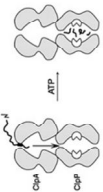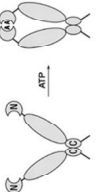
1.3.5 Small Hsp Family

The small Heat shock proteins (sHsps) family is the largest among the chaperone families comprising Hsps with molecular masses ranging from 12–40 kDa and share a conserved 90-residue C-terminal Alpha-crystalline Domain (ACD) (Waters et al., 1995; Boston et al., 1996; Vierling, 1991) The sHsps are ubiquitous, although predominant fraction is found in plants, and are synthesized in response to heat and other stresses and during certain developmental stages (Vierling, 1991; Scharf et al., 2001). Although, sHsps are not able to refold non-native proteins, they are capable of binding to the unfolded substrates and thereby stabilize and prevent aggregation (Ehrnsperger et al., 1997; Lee et al., 1997; Reddy et al., 2000). These substrate proteins are reported to be transferred to the ATP-dependent chaperones such as the Hsp70 system or Hsp100/Hsp70 complexes, for subsequent refolding (Veinger et al., 1998; Lee and Vierling, 2000; Mogk et al., 2003 a & b).
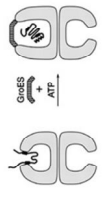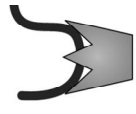
1.3.5 Other Speciallized Chaperones
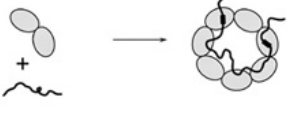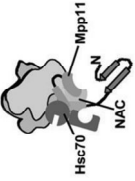
Several groups of proteins are considered as molecular chaperones such as Hsp47/SPARC, which assist the assembly of collagen triple helix (Dafforn et al., 2001); ribosomal associated chaperones, which assist nascent chain folding (Young et al., 2004; Bukau, 2005); periplasmic chaperones, which assist the folding of exported proteins (Baneyx and Mujacic, 2004; Nishiyama et al., 2005; Riuz et al., 2006); Hsp33, which is responsible for redox homeostasis in the cell (Sitia and Molterni, 2004); calnexin/calreticulin, which assists the folding of glucosylated proteins in the endoplasmic reticulum (Horwich et al., 2001). A representative classification of the heat shock proteins is presented in Figure 1.01.
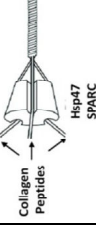
Figure1.01: Families of Heat Shock Proteins.

Chaperone Family	Topology of Binding	Chaperone	Systematic Occurrence	Oligomeric status	Sub-cellular localization	Co-chaperone	Known Function
Hsp100 Family		ClpA	E. coli	Hexamer	Cytoplasm	ClpP, SspB	Works with ClpP protease in ATP-dependent unfolding and proteolysis.
		ClpB	E. coli	Hexamer	Cytoplasm	DnaK, DnaJ, GrpE	ATP-dependent protein disaggregase. Works with DnaK.
		Hsp104	S. cerevisiae	Hexamer	Cytosol	Hsp70, Hsp40	Reactivates heat-damaged proteins, establishes and maintains the yeast prion phenotype.
		Hsp78	S. cerevisiae	Trimer & Hexamer	Mitochondria	Ssc1, Pim1	Prevents aggregation, degradation and turnover of unassembled mitochondrial proteins. Works with mtHsp70 as bichaperone system in restoring mitochndria DNA synthesis following heat shock.
		ClpB, ClpA/C, ClpD	Plantae	Hexamer	Cytosol, Mitochondria	ClpP	Disaggregation of substrate protein aggregates, unfolding
		ClpM, ClpN, ClpX, ClpY	Plantae	Hexamer	Chloroplast	ClpP	Works with the ClpP protease to promote proteolysis.
		HtpG	E. coli	Dimer	Cytoplasm		Protein refolding in stressed cells. Probable secretory chaperone.
Hsp90 Family		Hsp82	S. cerevisiae	Dimer	Cytosol	Sti1, Aha1, Cdc37, p23/Sba1, Ppt1, CPR6	Folding and conformational regulation of signalling proteins, reactivation of stress-denatured protans and acts as a capacitor of phenotypic variation.
		Hsp90 (Hsp83, Hsp89)	Mammalia and Insecta	Dimer	Cytosol	Hop, Hip, Hsp70, p50, p23, CHIP, Sgt1, TPR2, Immunophilins (Cyp40, FKBP38, FKBP51, FKBP52)	Major cytosolic chaperone. Protein folding peptide chaperone, cytoprotection, intracellular signaling (e.g: steroid hormone receptors and many kinases), cell-cycle control and buffering of harmful mutations.
		Grp94 (ERgp99)	Mammalia	Dimer	ER and Cell surface	Grp78	Controls protein homeostasis in the ER and folding and assembly of some secretory proteins. Implicated in the activation of dendritic cells and chaperoning of antigenic peptides in the process of antigen presentation. Tumor-derived gp96 tested in phase III trials against melanoma and renal cell carcinoma.
		AtHsp90-1 – 4 (Hsp81.1-4)	A. thaliana	Dimer	Cytosol	Hop, Sgt1, RAR1 Immunophilin	Plant Immunity via integration of diverse disease resistance mechanisms. Cytokinin signalling and plant development.
		AtHsp90-5 (Hsp88.1)	A. thaliana	Dimer	Chloroplast	-	Function unknown.
		AtHsp90-6 (Hsp89.1)	A. thaliana	Dimer	Mitochondria	-	Function unknown.
		AtHsp90-7	A. thaliana	Dimer	ER	-	Function unknown, proposed to function similar to Grp94.

Chaperone Family	Topology of Binding	Chaperone	Systematic Occurrence	Oligomeric status	Sub-cellular Localization	Co-chaperone	Known Function
Hsp70 Family		DnaK	*E. coli*	Monomer	Cytoplasm	DnaJ, GrpE, ClpB	Stabilizes newly synthesised polypeptides and preserves folding competence. Reactivates heat-denatured proteins and controls heat-shock response. Works with ClpB as disaggregase.
		HscA	*E. coli*	Monomer	Cytoplasm	HscB	Iron sulphur cluster protein assembly.
		HscC	*E. coli*	Monomer	Cytoplasm	YbeV, YbeS	σ 70 regulation.
		DnaK	Methanosarcinae	Monomer	Cytoplasm	DnaJ, GrpE	Principal folding chaperone. Works with ClpB as disaggregase.
		Ssa1-4	Eukaryota	Monomer	Cytosol	Ydj1, Sis1, Stl1, Snl1, Fes1, Sse1/2, Cns1	Protein transport across organelle membranes into ER and Mitochondria. Folding of newly synthesized proteins. Dissociating clathrin from vesicles and promoting lysosomal degradation of cytosolic proteins.
		Ssb1, 2	*S. cerevisiae*	Monomer	Cytosol	Zuotin, Sse1/2, Ssz1	Folding of ribosome-bound nascent polypeptide chains. Dissociation of clathrin from coated vesicles, promoting of lysosomal degradation of cytosolic proteins.
		Pdr13p	*S. cerevisiae*	Hetero-dimer	Cytosol	Zuotin	Zuotin cofactor, modifies Zuotin-Ssb interaction, folding of nascent chains on ribosomes.
		Kar2	*S. cerevisiae*	Monomer	ER	Sec63, Scj1, Jem1, Lhs1, Sil1/Sls1	Protein translocation into ER, binds to unassembled/misfolded ER protein subunits, regulates unfolded protein response.
		Ssc1	*S. cerevisiae*	Monomer	Mitochondria	Tim14, Tim16, Tim44, Mdj1, Mge1	Protein translocation into and folding in mitochondria.
		Ssc3	*S. cerevisiae*	Monomer	Mitochondria	Mge1	Assists some mitochondrial protein complex assembly, reported to mediate folding of imported proteins.
		Ssq1	*S. cerevisiae*	Monomer	Mitochondria	Jac1, Mge1	Protein complex assembly, Fe/S cluster insertion.
		Hsc70, Hsp70	Mammalia	Monomer	Cytosol	Hsp40, Hop, Bag1-5, Hip, HspBP1, CHIP, SGT, TPR1, Tom70, Hsp110 homologs	Mediates nascent polypeptide binding and folding, protein transport into ER, mitochondria and nucleus, promotes lysosomal degradation of cytosolic proteins, clathrin uncoating, inhibits polyglutamine fibril formation. Leukemia-derived Hsc70 is now being tested as a leukemia vaccine.
		Hsp70L1	Mammalia	Hetero-dimer	Cytosol/nucleus	MPP11	Cytoprotection and antiapoptotic. Folding of nascent chains on ribosomes. Implicated in spermatogenesis.
		BiP	Mammalia and Plantae	Monomer	ER	Grp170, Sil1	Protein translocation into ER, binds to unassembled/misfolded ER protein subunits, regulates unfolded protein response. Multimeric ER protein assembly. Tolerance to water stress in plants.
		Grp78	Mammalia	Monomer	ER	Sls1	Protein (eg, immunoglobulin) folding Protein translocation into ER, binds to unassembled/misfolded ER protein subunits, regulates unfolded protein response. Constitutively expressed and induced under glucose starvation.
		mtHsp70	Mammalia	Monomer	Mitochondria	Tim44	Protein translocation and folding in mitochondria. Involved in antigen presentation and radio-resistance Also oncogenic (overexpression of Mot-2 leads to p53 inactivation and cell transformation). Implicated in in myeloid leukemias and myelodysplasia.
		Hsp68	Plantae	Monomer	Mitochondria	mtDnaJ, mtGrpE	General chaperone for mitochondrial matrix proteins
		ctHsp70	Plantae	Monomer	Chloroplast	Tic40	Folding of imported proteins, insertion of light-harvesting complex protein into thylakoid membranes.

Chaperone Family		Topology of Binding	Chaperone	Systematic Occurrence	Oligomeric Status	Sub-cellular Localization	Co-chaperone	Known Function
Hsp60 Family	Group I		GroEL	*E. coli*	Homo-tetradecamer	Cytoplasm	GroES	Folding of a cytosolic protein subset, stabilizes proteins during heat stress, promotes folding in vivo of over-produced proteins and refolding of many proteins in vitro.
			Cpn60.1 & 2	*M. tuberculosis*	?	Cytoplasm & Secretory	?	Cpn60.1 implicated in biofilm and granuloma formation.
			GroEL	Methanosarcinae	Homo-tetradecamer	Cytoplasm	GroES	Folding of cytosolic proteins, the only archaeal species that has GroEL.
			Hsp60	*S. cerevisiae*	Homo-tetradecamer	Mitochondria	Hsp10	Folding of newly imported proteins, binds to heat-denatured mitochondrial proteins and prevents Aggregation.
			mtHsp60	Mammalia	Homo-tetradecamer	Mitochondria	Hsp10	Folding of newly imported proteins. Cytoprotection, macrophage activator possibly through Toll like receptors. Linked with autoimmunity such as rheumatoid arthritis.
			RUBISCO-Subunit Binding Protein (RBP)	Plantae	Hetero oligomer	Chloroplast	-	Binding to, preventing mis-folding and assisting in correct folding of RUBISCO large subunit.
			Cpn60 (α and β)	*A. thaliana*	Hetero-tetradecamer	Cytosol	Cpn10, Cpn21	General chaperone for folding and assembly of chloroplast proteins.
	Group II		Thermosome (α and β)	*T. acidophilum*	Hetero-hexadecamer	Cytoplasm	Prefoldin/GimC	Stress-inducible, promotes folding of a protein subset, refolding of unfolded polypeptides in vitro.
			TF55 (Rosettasome or Archeasome)	Archea	Hetero octadecamer	Cytoplasm	-	General chaperone for folding cytoplasmic proteins
			TRiC (α - θ) or TCP1	*S. cerevisiae*	Hetero-hexadecamer	Cytosol	GimC	Folding of a cytosolic protein subset, including actin, tubulins, and WD40 domain proteins, downstream of Hsp70 in de novo folding assembly of polyglutamine expansion proteins into nontoxic oligomers.
			CCT (α - θ)	Mammalia	Hetero-hexadecamer	Cytosol	Prefoldin, PhLP	Folding of a cytosolic protein subset, including actin, tubulins, and WD40 domain proteins, downstream of Hsp70 in de novo folding assembly of polyglutamine expansion proteins into nontoxic oligomers.
Prefoldin			MtGimC	*M. thermoautotrophicum*	Hexamer (2α+ 4β)	Cytosol	-	General chaperone, able to prevent aggregation and stabilize the non-native states of a variety of proteins
			GimC	*S. cerevisiae*	Hexamer (α₂, α₃, β₁, β₃, β₄ and β₆)	Cytosol	-	Stabilizes nascent or nonnative chains of actins and tubulins. Cooperates with TRiC in the post-translational folding
			Prefoldin	Bovine	Hexamer (α₂, α₃, β₁, β₃, β₄ and β₆)	Cytosol	-	Stabilizes nascent or nonnative chains of actins and tubulins and can transfer them to TRiC for folding.

Chaperone Family	Topology of Binding	Chaperone	Systematic Occurrence	Oligomeric status	Sub-cellular Localization	Co-chaperone	Known Function
		IbpA	*E. coli*	Monomer Dimer	Cytoplasm	-	Prevents heat-denatured protein aggregation, associates with inclusion bodies, works with DnaK in protein refolding.
		IbpB	*E. coli*	Multimer	Cytoplasm	-	Stabilizes unfolded polypeptides and prevents aggregation.
		Hsp16.5	*M. jannashii*	24mer	Cytoplasm	-	Stabilizes unfolded polypeptides and prevents aggregation.
		Hsp16.3	*M. tuberculosis*	Nonamer	Cytoplasm	-	Prevention of protein aggregation, temperature-dependent dissociation required for efficient non-native substrate binding
		Hsp26	*S. cerevisiae*	24mer	Cytosol	-	General chaperone for protein homeostasis at physiological and stress conditions
		Hsp42	*S. cerevisiae*	Dimer and 12-16 mer	Cytosol	-	Structural protein of eye lens. Prevents heat-denatured protein aggregation, associates with inclusion bodies and works with DnaK in protein refolding. Constitutes 35% of proteins in a mammalian lens. Mutations cause cataract.
Small Heat Shock Proteins (sHsps)		α-A crystallin	Mammalia	32mer	Majorly in eye lens	-	Antiapoptotic, thermoprotection. Increased in many neurological diseases, such as Alzheimer's disease, induced by heat and osmotic shock.
		α-B crystallin	Mammalia	32mer	Cytosol	-	Protection of the NADH: ubiquinone (Complex I) oxidoreductase activity, during salt stress.
		Hsp17.6 (Class I) Hsp17.9 (ClassII)	Plantae	12mer	Cytosol	-	Protection of the NADH: ubiquinone (Complex I) oxidoreductase activity, during salt stress.
		Hsp21 and Hsp26.2 (Class III)	Plantae	24mer	Chloroplast	-	Prevention of Aggregation of substrate proteins
		Hsp22 (Class IV)	Plantae	Multiple oligomers	ER	-	Protection of the NADH: ubiquinone (Complex I) oxidoreductase activity, during salt stress.
		Hsp23 (Class V)	Plantae	Homo-oligomer	Mitochondria	-	Protection of the nascent polypeptide chains from aggregation.
		Hsp22.3 (Class VI)	Plantae	Homo-oligomer	Membrane	-	Assists folding of nascent chains, catalyzes peptidyl-prolyl isomerisation in vitro.
		Trigger factor	*E. coli*	Dimer	Cytoplasm	-	Interacts with ribosomes SRP and translating polypeptide chains, role proposed in protein folding/quality control.
Ribosome Associated Chaperones		NAC (α, β)	*S. cerevisiae*	Hetero-dimer	Cytosol	SRP	Interacts with Ssb, Reported to be involved in folding of nascent chains on ribosome.
		RAC (Ssz1/Pdr2 + Zuotin)	*S. cerevisiae*	Hetero-dimer	Cytosol	Ssb	Folding of nascent chains on ribosomes.
		Hsp70L1	Mammalia	Hetero-dimer	Cytosol	Mpp11	

Chaperone Family	Topology of Binding	Chaperone	Systematic Occurrence	Oligomeric status	Sub-cellular localization	Co-chaperone	Known Function
		FimC	*E. coli*	Monomer	Periplasm	FimD	Folds proteins involved in type 1 Pili biosynthesis and assembly.
		PapD family	*E. coli*	Monomer	Periplasm	PapC	Pili assembly in the chaperone-usher pathway.
		Skp (OmpH)	*E. coli*	Trimer	Outer Membrane	-	Rescues off pathway aggregation prone OMPs, maintain solubility of folding intermediates in the periplasm.
Periplasmic Chaperones		FkpA	*E. coli*	Dimer	Periplasm	-	Maintain solubility of folding intermediates in the periplasm.
		SurA	*E. coli*	Dimer	Periplasm	-	Peptidyl prolyl isomerase activity, aids folding of LamB, prevents thermal aggregation of substrates. Escorts OMPs across the periplasm.
		LolA	*E. coli*	Monomer	Periplasm	LolB, LolCDE	Folding and escorting outer membrane lipo-proteins.
		Hsp33	*E. coli*	Dimer (active form)	Cytoplasm	-	Redox-regulated holding chaperone, prevents aggregation of thermally unfolded and oxidatively damaged proteins.
		Hsp31	*E. coli*	Homo dimer	Cytoplasm	-	Binds early unfolding intermediates in times of stress.
		SecB	*E. coli*	Tetramer	Cytoplasm	SecA, SecYEG	Stabilizes some secretory proteins in an unfolded state for export. Can overcome TF and DnaK deficiency.
		Calnexin	Mammalia	Monomer	ER	ERp57, EDEM	Works with glycosyltransferase to fold ER glycoproteins. Interacts with some non-native proteins independent of glycosylation. Plays a role in generation of MHC Iβ2microglobulin peptide complex.
Other Chaperones		Calreticulin	Mammalia	Monomer	ER Lumen and Cell Surface	ERp57, EDEM	Folding of glycoproteins, facilitates peptide loading to the class I molecule of the MHC. Implicated in the clearance of apoptotic bodies, and as an autoantigen in Systemic Lupus Erythematosus.
		Hsp47	Mammalia	Trimer	ER	P4H	Binds to collagen, chaperone in the biosynthetic pathway of various collagens. Member of serpin (serine protease inhibitor) superfamily. Involved in tumor migration and metastasis
		SPARC (Osteonectin)	Invertibrata to mammalia	Trimer	ER	Hsp27	Implicated in collagen folding and stabilization.

1.4 Regulation of Heat Shock Response

The evolutionary process of life forms, in a large measure, is a reflection of the recurring conflict between the organisms and the surrounding environment. It is assumed therefore, that stress is an unavoidable part of the life of all organisms. In view of the fact that any stressful condition is potentially harmful to the cells, the most primitive organisms also, are equipped with cellular mechanisms, which safeguard against the potential damages due to stress. Temperature is the principal physical parameter under constant surveillance in all life forms and cells respond to a sudden or absolute increase in growth temperature by transient but enhanced expression, termed the Heat Shock Response, of a distinct set of about twenty heat shock genes which encode the Hsps. (Narberhaus et al., 2006). This evolutionarily conserved cellular protection mechanism is primarily regulated at the level of transcription. Several mechanisms were described in bacteria that regulate the heat shock response including the activation of σ^{32} factor in *E. coli* and majority of gram negative bacteria, repression by of the *groE* operon HrcA in *Bacillus subtilis*, and by HspR in *Streptomycetes coelicolor* and many other gram positive bacteria. The regulation of heat shock response in multicellular organisms has been shown to be much more complex wherein a conserved family of transcription factors, termed heat shock transcription factors (HSFs) bind to specific heat shock elements (HSEs) within the promoters of their target genes and regulate the expression of the same. Apart from the molecular chaperones, cells have devised various mechanisms, collectively termed as thermal sensors, which are capable of detecting changes in temperature.

1.4.1 Thermal Sensors in Bacteria

The underlying dogma of the thermal sensors is alteration in the conformations of molecules in response to changes in temperature. Three such sensors, termed either as type I sensors or direct sensors, have been described: a, bending in DNA; b, melting in mRNA loops and c, unfolding of heat labile proteins (Bahl et al., 1987; Parsel and Sauer, 1989; Chaudhuri et al., 2004). Moreover, another class of thermosensors, termed either type II or indirect sensors, have been identified, which act according to the responses from the type I sensors. Two classes of such sensors have been characterized: a, the molecular chaperones, which assist the unfolded proteins during heat shock to fold into

native state and b, the proteases, which clear the terminal protein aggregates (McCarty and Walker, 1991; Wickner et al, 1999; Waldminghaus et al., 2009).

1.4.1.1 DNA Bending in Thermal Sensing

Effect of structural variations in DNA conformation on the information access and thereby the phenotype is becoming evident. Bending in DNA occurs usually at the promoter sites of the genes encoding thermal transcriptional activators. This is due to either innate sequence, such as AT tracts, or due to the binding of a protein, such as H-NS, predominantly at the sites of promoters. This is usually aimed at preventing access of the RNA polymerase due to the three-dimensional configuration and thereby keeping the gene (s) following the promoter in a repressed state (Tanaka et al., 1991; Atlung and Ingmer, 1997). After a shift to higher temperature, bends get melted allowing access of the RNA polymerase to the downstream genes. This principle is used by many pathogenic bacteria such as *Shigella flexneri* and *Yersinia enterocolitica* to induce expression of their virulence genes, such as *virF* (Lambert et al., 1992; Prosseda et al., 1998).

1.4.1.2 Thermal Sensing in mRNA Molecules

In addition to the role as carrier of information, involvement of mRNAs in regulation of gene expression in response to the temperature changes is well described. The ability of mRNAs in detecting changes in environment was demonstrated to be directed by metabolites such as riboswitches, transcription factors or small noncoding RNAs, or temperature (Mandal and Breaker, 2004; Kaempfer, 2003; Storz et al., 2004; Narberhaus et al., 2006). The mRNAs form secondary structure within their 5' untranslated region (UTR) during ambient temperature, for sequestering part or the complete Shine-Dalgarno sequence, thereby preventing translation of the downstream gene (Morita et al., 1999; Nocker et al., 2001). Upon increase in temperature, the secondary structure gets destabilized, resulting in the production of more protein. Examples for RNA thermosensors include, transcripts coding for the bacteriophage λ cIII protein, the heat shock sigma factor σ^{32} of *E. coli* and transcriptional activator PrfA, for virulence genes in *L. monocytogenes* (Morita et al., 1999; Johansson et al., 2002).

1.4.1.3 Temperature Sensing by Proteins

Majority of the heat labile DNA binding proteins are labelled as thermosensors (Schumann, 2007). Temperature perturbations result in major fluctuations in the protein conformation and dynamics, thereby rendering the proteins incapable of recognising their substrates, for example, the target operator sequences upon temperature up shift (Servant and Mazodier 1995; Hurme et al., 1996; 1997). The protein sensors therefore are the most efficient in controlling expression of several genes in response to the changes in environmental temperature.

1.4.2 Heat Shock Response in *E. coli*

The first studies on bacterial heat-shock response were performed in *E. coli* K-12, wherein the expression of the heat shock operons, including the Hsp70 operon that comprises the genes encoding DnaK, GrpE and DnaJ and the Hsp60 operon comprising the encoding GroES and GroEL, are regulated by the stress-inducible subunit of RNA polymerase, σ^{32}. σ^{32}, the product of the *rpoH* gene, is alternative sigma factor, classified as Group II sigma factors, which direct the RNA polymerase to distinct promoters that differ from the constitutive promoters (Grossman et al., 1984; Arsène et al., 2000). Regulation by σ^{32} is controlled at the transcriptional and post-translational levels. Under the normal growth conditions, DnaK-DnaJ-GrpE chaperone machinery sequesters σ^{32} and possibly delivers it to the ATP dependent metallo-protease FtsH, the product of the *hflB* gene for degradation. Consequently, the reduced rate of translation of *rpoH* mRNA due to the sequestration of Shine-Dalgarno sequence and the start codon in the secondary structure ensures low levels of σ^{32} and the expression of downstream genes (Tilly et al., 1983; Straus et al., 1990; Tomoyasu et al., 1998; Tatsuta et al., 1998; Morita et al., 1999). Under the heat shock conditions, the secondary structure in *rpoH* mRNA is released, thereby increasing its expression. Concomitantly, DnaK is sequestered by the damaged proteins, thus releasing σ^{32} and thereby initiating a cascade of events including increased synthesis σ^{32}. Induction of GroEL/S system by the stable σ^{32} and the expression of the downstream heat-shock genes is also enhanced (Straus et al., 1987; Yura and Nakahigashi 1999; Arsène et al., 2000). After the normal growth conditions are restored, the amount of active σ^{32} is rapidly brought to low via a negative feedback

mechanism (Figure 1.02). Therefore, the DnaK/DnaJ serves a direct sensor of cellular stress and regulator of heat shock transcription (Tomoyasu et al., 1998).
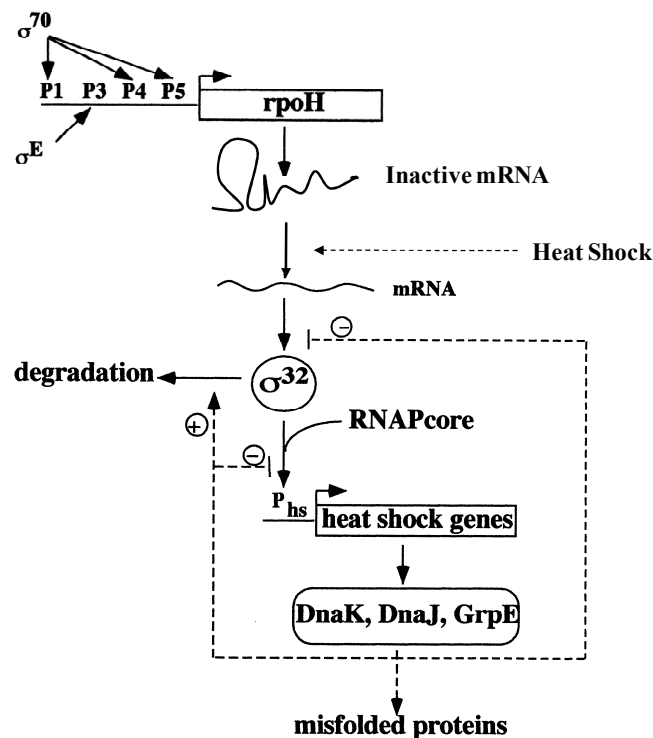
Figure 1.02: Regulatory Mechanism of the σ³² Regulon in E. coli. Heat shock enhances the level of σ³² both by making and activating translation of *rpoH* mRNA. σ³² is stabilized under heat shock by sequestering DnaK/J chaperones by unfolded proteins produced upon heat shock. Stable σ³² binds to RNAP and concomitantly activating heat-shock gene transcription. Subsequent build-up of HSPs brings about negative feedback control by inhibiting translation, stabilization, and activity of σ³². The illustration is reproduced with permission from Arsène et al., 2000.

Recent reports show that the regulation of the heat shock response in *E. coli* is different when compared to that in other bacteria. The major heat shock genes in *E. coli* and other gram-negative bacteria comprise the single σ³² regulon, whereas in other bacteria, especially the Gram-positive bacteria, the heat shock genes are parts of several regulons (Yura and Nagahigashi, 1999). For example, in *B. subtilis,* among the three major classes of heat shock genes, only one is regulated by the σ³² homologue, σ^B. The observation that even in a σ^B mutant in *B. subtilis,* several heat shock genes were upregulated, led to the evidence of presence of additional regulatory mechanisms

(Chang et al., 1994). Other regulatory mechanisms on heat shock genes were also discovered in several bacteria and the corresponding regulatory DNA elements were identified. A brief note on these additional regulatory mechanisms is presented below.

1.4.3 Heat Shock Regulation in Gram-positive Bacteria

Apart from the central role played by σ^{32} in heat shock response in *E. coli*, distinct regulatory mechanisms that direct the expression of specific heat shock genes have been characterized in other bacteria. For example, regulation of *groES/L* operon in Bacillus was shown to be regulated by HrcA repressor and the *dnaK* operon in Streptomycetes by HspR. Several such examples were discovered.

1.4.3.1 HrcA-CIRCE System

In *B. subtilis*, expression of the *dnaK* and *groEL* operons, classified as Class I heat shock genes, are under the negative regulation of the HrcA repressor binding to the CIRCE (**C**ontrolling **I**nverted **R**epeat of **C**haperone **E**xpression; with a consensus sequence of TTAGCACTC-N9-GAGTGCTAA) operator elements (Baird et al., 1989; Zuber and Schumann, 1994; Yura and Nagahigashi, 1999). Activity of HrcA is modulated by the GroES/L system, probably by facilitating its folding (Figure 1.03). In response to increased protein damage, upon heat shock, the GroES/L folding machinery is occupied by the unfolded proteins and folding of the HrcA repressor is stalled thereby releasing the repression; providing a direct sensing mechanism for protein misfolding (Mogk et al., 1997). The CIRCE/HrcA system is one of the best understood operator-repressor pairs in heat shock response and has a widespread occurrence in the bacterial kingdom with presence in more than 40 different species (Hecker et al., 1996; Narberhaus and Bahl, 1992; Narberhaus et al., 1992; Wetzstein et al., 1992).
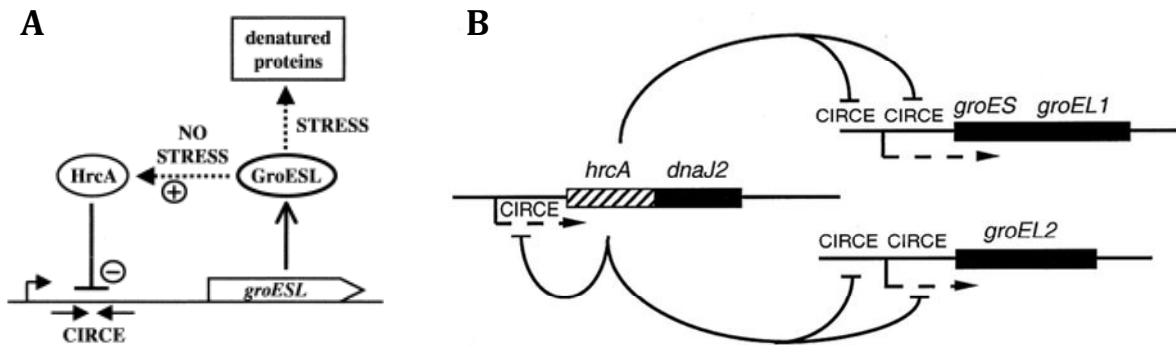
Figure 1.03: Models for Regulation by HrcA. **A.** GroES/L helps folding of HrcA, which in turn binds the CIRCE elements. Upon Heat shock GroEL is titrated by denatured proteins, and HrcA folding is stalled. The model is reproduced with permission from Narberhaus, 1999. **B.** Major Heat shock genes in gram-positive bacteria, *groES/L1*, *groEL2*, *hrcA* and *dnaJ2* are under the regulation of HrcA. The model is reproduced with permission from Servant and Mazodier, 2001.

1.5.3.2 HspR-HAIR System

HspR-HAIR system has been identified in *Streptomyces* and controls the *dnaK* operon and the *clpB* gene in *Streptomyces*. This sytem is also present in other bacterial species, including Mycobacteria (Bucca et al., 1995; Grandvalet et al., 1997, Cole et al., 1998). The HspR repressor is very unstable, where it has been proposed to be stabilised only in complex with DnaK, and the complex in turn binds the inverted repeat elements of the operator sequence with a consensus motif (CTTGAGT-N7-ACTCAAG). The inverted repeat sequence been designated as HAIR for **H**spR **A**ssociated **I**nverted **R**epeat. Upon Heat shock the denatured polypeptides compete for binding with DnaK and therefore the HspR is released, thereby releasing the repression (Figure 1.04).
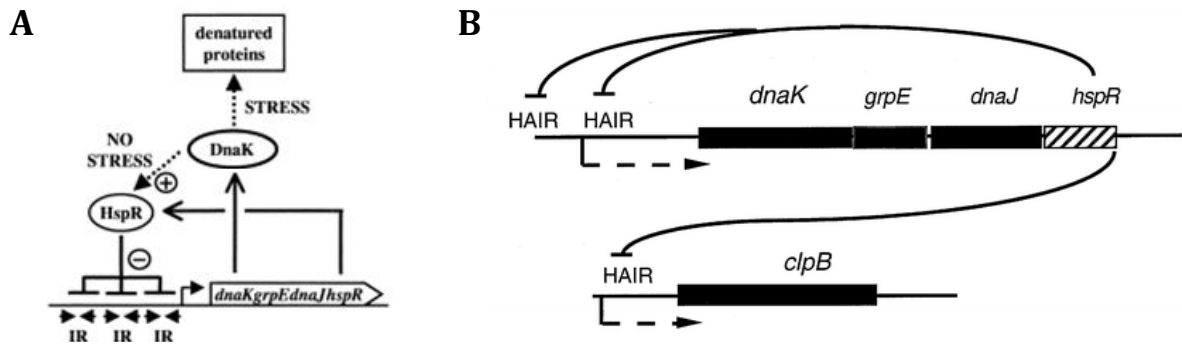
Figure 1.04: Models for Regulation by HspR. A. DnaK complexes with otherwise unstable HspR and the complex in turn binds the HAIR elements. Upon Heat shock DnaK is titrated by denatured proteins, and HspR is released from the promoter. The model is reproduced with permission from Narberhaus, 1999. **B.** The *dnaK* operon and *clpB* in gram-positive bacteria are under the regulation of HspR. The model is reproduced with permission from Servant and Mazodier, 2001.

1.5.3.3 ROSE Regulon

In *Bradyrhizobium japanicum*, a negative *cis*-acting element called ROSE (**R**epression **O**f heat **S**hock-gene **E**xpression) was reported to control the expression of five heat shock operons encoding small heat shock proteins including a σ^{32} homologue and the periplasmic protease DegP (Narberhaus et al., 1998 a; Münchbach et al., 1999) (Figure 1.05). ROSE is a conserved DNA element of approximately 100 bp that is positioned between the transcription and translation start sites of the first gene of each operon and confers temperature regulation to a σ^{70}-type promoter (Narberhaus et al., 1998 a & b; Nocker et al., 2001). Deletions in the distal half, but not the proximal half in this unusually long regulatory element resulted in derepression of the downstream genes at low temperatures. Critical bases and the repressor interacting with ROSE remain elusive.
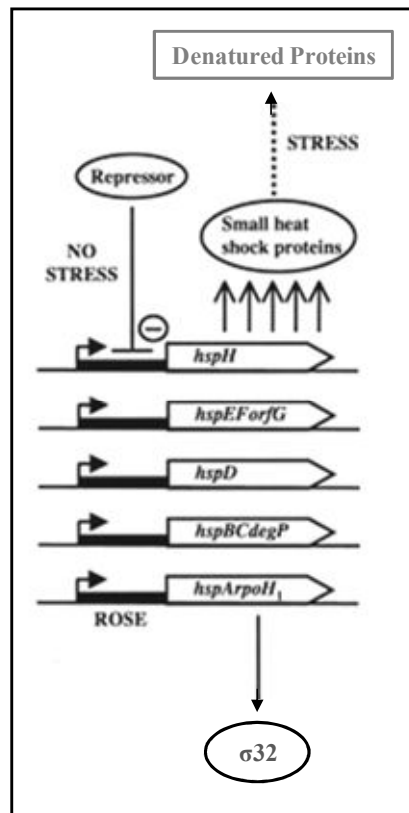
Figure 1.05: Models for Regulation by ROSE. Under normal conditions, ROSE element is bound by the putative repressor element, represses the expression of the downstream genes. Upon heat shock, the repressor is released from ROSE thereby inducing the genes encoding several sHsps, which help folding of the accumulating denatured proteins, and σ^{32}, which induces other major heat shock genes. The model is reproduced with permission from Narberhaus, 1999.

1.5.3.4 RheA

S. albus employs another regulatory mechanism to control expression of *hsp18* which encodes a small HSP protein that plays a role in thermotolerance. Transcription of *hsp18* is strongly induced by heat shock from a streptomyces vegetative promoter (Servant and Mazodier, 1995; 1996). RheA (**R**epressor of *hsp* **E**ighteen) repressor encoded by a gene *orfY*, situated 150 bp upstream and in the opposite orientation to that encoding Hsp18, was shown to contribute to the transcriptional regulation of *hsp18*. The mechanism of RheA-mediated repression is proposed to be via binding to the

inverted-repeat sequence (GTCATC-N5- GATGAC) that overlaps with the -35 region of the *hsp18* promoter. Presence of another similar sequence (GTCGTC-N5-GATGAC) centred 50 bp upstream of the orfY start codon, suggests that RheA appears to function as a negative autoregulator (Servant et al., 1999) (Figure 1.06). Further, the observation that *hsp18* mRNA was not translated under ambient temperature conditions, suggested a role for an unknown post-transcriptional mechanism. However, unlike the traditional feedback mechanisms proposed, induction of *hsp18* after temperature up-shift prolonged, suggesting that RheA follows a different mechanism of regulation.
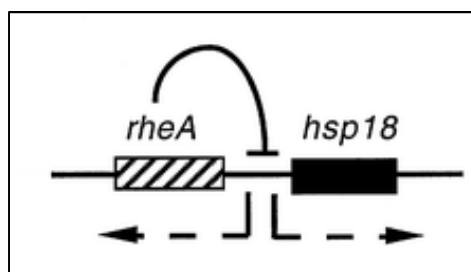
Figure 1.06: Model for Regulation by RheA. RheA functions as auto repressor as it binds to the target sequences located upstream of *hsp18* and *rheA* thereby represses the expression these genes. The model is reproduced with permission from Servant and Mazodier, 2001.

1.5.3.5 CtsR

Class III Heat shock genes in *B. subtillis* (*clpP*, *clpC* operon, *clpE*) and thioredoxin are negatively regulated by CtsR (**C**lass **T**hree **S**tress-gene **R**epressor), the product of the first gene of the *clpC* operon, which is proposed to recognize a directly repeated heptanucleotide operator sequence with its helix–turn–helix motif (Kruger et al., 1996; Derré et al., 1999 a & b). CtsR is composed of three functional domains, an N-terminal dimerization domain, a DNA binding domain which encompasses a helix-turn-helix motif and a proposed heat sensing central Glycine rich domain (Derre et al., 2000). The proteins McsA and McsB, which are encoded by two of the eight genes in CtsR regulon, function as modulators of CtsR (Figure 1.07). McsA, with its zinc finger motif, was proposed to function as DNA-binding protein and thereby switch CtsR into active conformation whereas McsB, with its arginine kinase like activity is proposed to

inactivate CtsR by phosphorylation and consequently to ClpCP-mediated proteolysis (Krüger et al., 2001). Although CtsR regulon is characterized in *B. subtillis*, genome sequences from several gram-positive bacteria, especially those with low G + C contents, displayed conserved CtsR-binding sites upstream of their *clp* genes, suggesting that this regulon might be widely distributed (Derré et al., 1999 b).
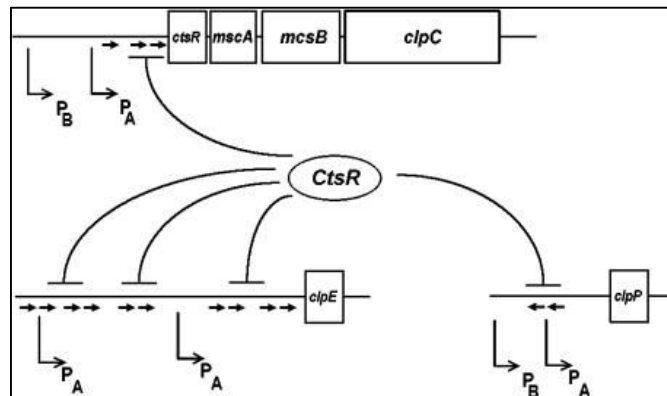
Figure 1.07: The CtsR Regulon. Cartoon showing 3 transcriptional units, the *clpC*, *clpP* and *clpE* operons, which are under the negative control of the CtsR repressor. Arrows indicate the direct repeats for CtsR binding. The model is reproduced from Schumann, 2003.

1.5.4 Conclusions

Work in the past few years has demonstrated the existence of various specific regulatory mechanisms for heat shock response. Presence of several redundant mechanisms raises the possibility of a cross talk between different regulons and existence of a master regulator. The regulons, however, seem to follow independent modes of expression control (Schumann, 2003). The significance of this de-centralization in heat shock response could be discerned in future studies.

1.5 Heat Shock Proteins as Molecular Chaperones

Majority of the heat shock proteins enable the aggregation-prone intracellular proteins to attain their native conformations. These proteins, therefore, are aptly termed as molecular chaperones. The two best-studied families of molecular chaperones are the ATP dependent ubiquitous Hsp70 family and the Hsp60 (Chaperonin) system. These families recognize the exposed hydrophobic surfaces on the extended and collapsed polypeptides, respectively (Bukau and Horwich, 1998). Majority of the newly synthesized polypeptides are assisted by these two chaperone machines, to reach their functional conformation. A simplified cartoon below displays a view of the action of these two principal chaperones. The nascent or extended polypeptides initially interact with Hsp70. In an ATP dependent mechanism, these are then assisted to reach their correctly folded state or alternately are 'handed over' further to other set of molecular chaperones, such as the Hsp60. The substrate proteins that do not get folded by either of these chaperones would generally be destined to the cellular proteases.
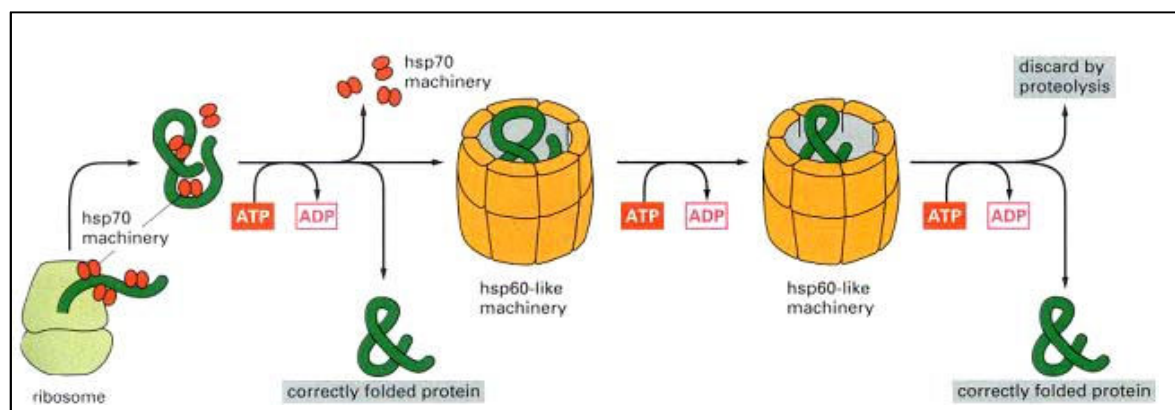
Figure 1.08: Hsp70 and Hsp60 are the Major Chaperones. Nascent polypeptides, as they emerge from Ribosome are sequestered by the Hsp70 machinery and are folded to their correct conformation. Unfolded polypeptides are relayed to the Hsp6o wherein upon encapsulation, the polypeptides are folded to correct conformation but the unfolded ones are cleared by the action of proteolysis. The illustration is reproduced with permission from Alberts et al., 2004.

1.6 Structural and Functional Aspects of Hsp70 mediated Protein Folding

70-kDa heat shock proteins (Hsp70s), along with their co-chaperone molecules comprise abundant cellular machines that assist a wide range of folding processes, including the folding of naive proteins, refolding of misfolded proteins, disengaging aggregated proteins, membrane translocation of organellar and secreted proteins, controlling activity of regulatory proteins and assisting proteolytic degradation of unstable proteins (Hartl, 1996; Toft, 1999; Bukau et al., 2000; Ben-Zvi and Goloubinoff, 2001; Hartl and Hayer-Hartl, 2002; Neupert and Brunner, 2002; Ryan and Pfanner, 2002; Pratt and Toft, 2003; Young et al., 2003). The Hsp70 family includes members with high sequence similarity and similar molecular mass. The observation that many organisms from bacteria to human encode multiple members of the Hsp70 family suggests that the multiple copies have evolved specific cellular roles. Similarly a large number of the co-chaperones, Hsp40 (also known as JDP for J-domain Protein) and GrpE (also known as NEF for Nucleotide Exchange Factor), also suggests that the multiple copies have different functional roles. Noticeably, a few copies are constitutively expressed but many are expressed in response to various environmental stimuli including heat shock or growth conditions (Genevaux et al., 2007). Studies on the prokaryotic counterpart of heat shock inducible Hsp70, known as DnaK, have provided insights into the chaperone function. In *E. coli*, DnaK along with its canonical co-chaperones DnaJ, the Hsp40 homologue and GrpE, the NEF homologue, functions effectively as a chaperone. Two other Hsp70 homologues, HscA and HscC in *E. coli*, were shown to be constitutively expressed and likewise five other JDP like proteins, CbpA, DjlA, DjlB, DjlC and HscB are not heat shock inducible. Genetic studies have indicated that the three Hsp70s share a common substrate pool with the ribosome associated trigger factor (TF) (Deuerling et al., 1999; Deuerling et al., 2003). Therefore, Hsp70s are dispensable in TF expressing *E. coli*, but the mutants are synthetically lethal in the absence of TF (Hesterkamp and Bukau, 1998; Kluck et al., 2002).

1.6.1 Hsp70 Structure

Biochemical and structural studies on the Hsp70 homologues from *E. coli*, *T. thermophilus*, yeast and bovine have delineated the two domain architecture of Hsp70 family of proteins; highly conserved 42-44 kDa N-terminal ATPase domain and the C-terminal domain which is further divided into a 15-18 kDa Substrate binding domain and a less conserved 10 kDa flap domain (Young et al., 2004; Mayer and Bukau, 2005; Jiang et al., 2005; Revington et al., 2005; Swain et al., 2007). Structural information from full length Hsp70 is still lacking. The communication between the ATPase domain and the substrate binding domain is mediated by a 10-12 residue linker that connects these two domains (Figure 1.09).

Crystal structures of the Bovine Hsc70 ATPase domain in complex with several adenosine nucleotides has revealed that the ATPase domain is composed of two globular sub-domains I and II, each with a smaller sub-domains A and B, separated by a deep cleft and connected by two crossed helices forming a hydrophobic binding pocket for the nucleotides (Flaherty et al., 1990). Nucleotide, positioned in complex with one Mg^{++} and two K^+ ions at the binding pocket with the β and γ-phosphate binding loops, interacts with the four sub-domains (Meyer and Bukau, 2005). Solution structure of ATPase domain suggested that the tilting motions of the sub-domains, which result in opening and closing of the nucleotide binding cleft might be directed by the nucleotide, with the highest opening rate in nucleotide free form and the lowest in the ATP bound form (Gassler et al., 2001; Zhang and Zuiderweg, 2004).

Structural studies on the peptide binding domain have revealed the presence of a beta sandwich of 2 four strand sheets with the four loops, connecting the strands, protruding upwards (Zhu et al., 1996; Pellecchia et al., 2000; Cupp-Vickery et al., 2004). Two of the four loops, $L_{1,2}$ and $L_{3,4}$, along with the Helix B form the substrate binding pocket, with a cross section of about 5 × 7 Å (Bukau and Horwich, 1998). Helix A along with helix B functions as a lid in closing the substrate binding pocket, by forming salt bridges with loops $L_{3,4}$ and $L_{5,6}$. Genetic and biochemical analysis have resulted in identifying a conserved segment, DVLLLD, in the linker, which mediates the communication between these two domains (Rist et al., 2006; Vogel et al., 2006).
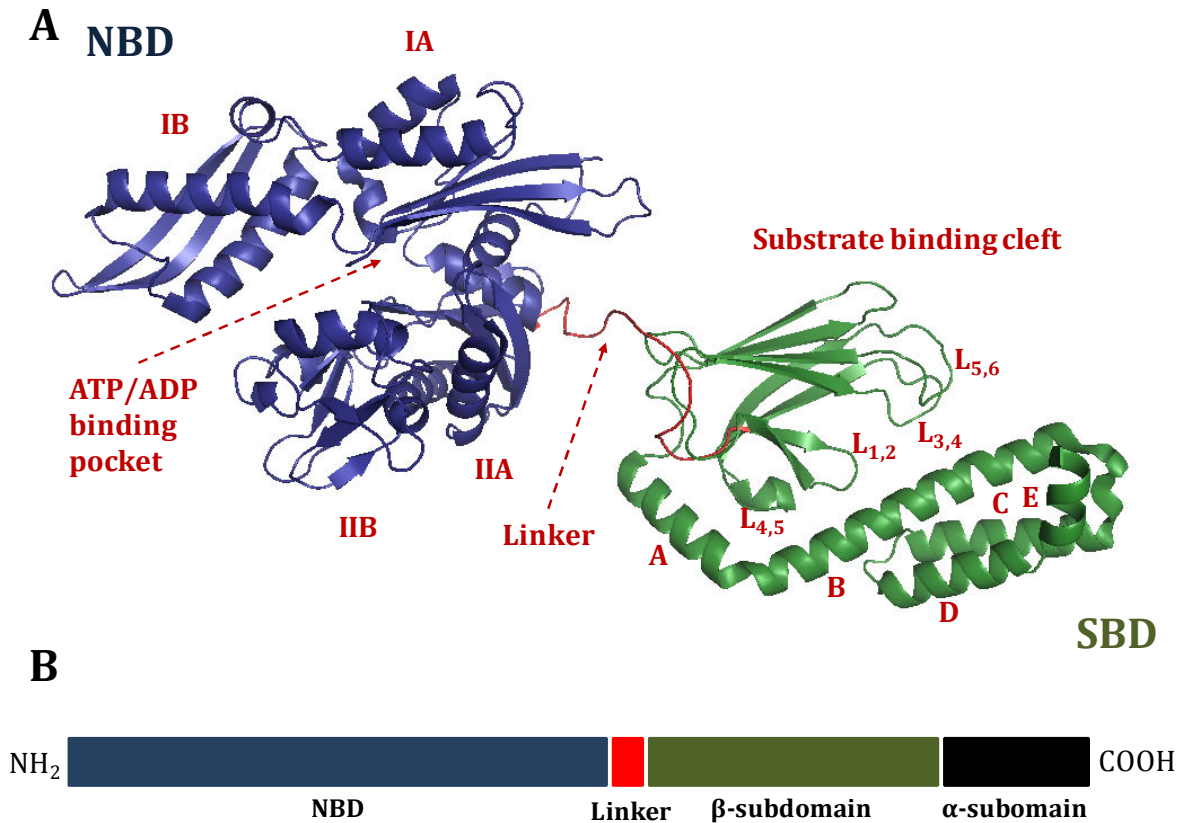
Figure 1.09: Architecture of Hsp70. **A.** Domain architecture of the Hsp70. Cartoon diagram showing different structural features of Hsp70 with individual domains colour coded for distinction. N-terminal Nucleotide Binding Domain (NBD) is shown in blue and the Substrate Binding Domain (SBD) in Green. The linker connecting these two domains is shown in red. Sub-domains of the ATPase domain, IA – IIB along with ATP/ADP binding pocket, Substrate binding cleft with the component loops ($L_{1,2}$ – $L_{5,6}$) and five helices (A - E) are indicated. Cartoon is generated using Pymol 0.99, molecular visualization software from DeLano Scientific LLC, USA. Co-ordinates for the molecule were obtained from the structure deposited in PDB with the ID: 1KHO. **B.** Domain organisation in Hsp70 defining approximate domain boundaries of the N-terminal NDB, Linker, SBD and the C-terminal 10 kDa domain are presented.

1.6.2 Mechanism of Action of Hsp70

In the Hsp70 assisted folding reactions, substrate polypeptides undergo repeated cycles of binding and release, usually at a stoichiometry of a single Hsp70 monomer per substrate molecule (Szabo et al., 1994; Buchberger et al., 1996). Hsp70 generally acts in concert with its cohorts, Hsp40 (DnaJ) and GrpE. The target sequences for binding to Hsp70 are the extended polypeptides that are seven residues long and rich in hydrophobic residues, preferably Leucine and Isoleucine (Schmid et al., 1994; Zhu et al., 1996). Eukaryotic members of Hsp70 family are shown to be associated with ribosomes and bind to the nascent chains at the ribosome exit tunnel. Mechanism of action of the prokaryotic Hsp70 homologue, the *E. coli* DnaK, has been well understood at molecular details through a large number of studies.

In the ATP bound form, DnaK displays low affinity towards the substrates and therefore high binding and release rates, of the order of a few seconds to milliseconds, while in the ADP bound form it displays high affinity and low rates of binding and release, of the order of a few minutes to sometimes a few hours. Prokaryotic Hsp40 homologue, DnaJ has an affinity for unfolded proteins and binds them with its hydrophobic C-terminal domain (CTD). The characteristic 75 residue long J domain in DnaJ recognizes ATP bound DnaK at its ATPase domain and thereby delivers the substrate to DnaK at SBD, which is in an open conformation. DnaJ is involved in accelerating DnaK's ATPase activity. Although DnaK and DnaJ act independently in terms of peptide recognition, the interaction between DnaJ and DnaK seems to be essential for two reasons; (A) Presence of greater number of DnaJ copies in the cell is believed to impart substrate specificity for chaperone action, and (B) DnaJ is shown to be involved in stabilizing the substrate-chaperone complex by accelerating ATPase activity of DnaK. Surprisingly, sequences of DnaK from the gram positive bacteria lack the characteristic 23 residue long DnaJ interacting region located in the ATPase domain of DnaK. The observation that DnaK from these organisms is inactive in vivo and in vitro further strengthens the notion on the essentiality of DnaJ in DnaK mediated protein folding. However, physiological significance of such a deletion in the context of Gram positive bacteria is yet to be discerned.

After ATP is hydrolyzed to ADP, the helix B in the lid domain latches on to the loops $L_{3,4}$ and $L_{5,6}$, thereby locking the substrate binding cleft in a closed conformation. DnaK-ADP-Substrate complex is stable till the other co-chaperone, GrpE gets into action. GrpE is the nucleotide exchange factor for DnaK, which binds the NBD thereby triggering swirling and twisting motion in the NBD sub-domains IA, IB, IIA and IIB. This conformational change opens up the nucleotide binding cavity thereby helping it to release ADP and re-bind ATP. ATP binding triggers opening of the substrate binding cleft thereby releases the substrate proteins in either partially folded or completely folded state. However, rate of nucleotide exchange appears to be controlled at the post-translational level. GrpE, being a thermally unstable protein, gets destabilized upon heat shock and therefore is not able to interact with DnaK and exchanges the ADP with ATP, whereby the DnaK-ADP-Substrate complex is stabilized during the heat shock. Noticeably a few Hsp70 homologues, such as *E. coli* HscC, were shown to be independent of GrpE. The partially folded substrates might either re-bind DnaJ/DnaK for another round of folding, or may interact with other class of chaperones, the GroES/L chaperonins. The cartoon below describes the mechanism of action of Hsp70 in folding of an unfolded substrate protein (Figure 1.10).
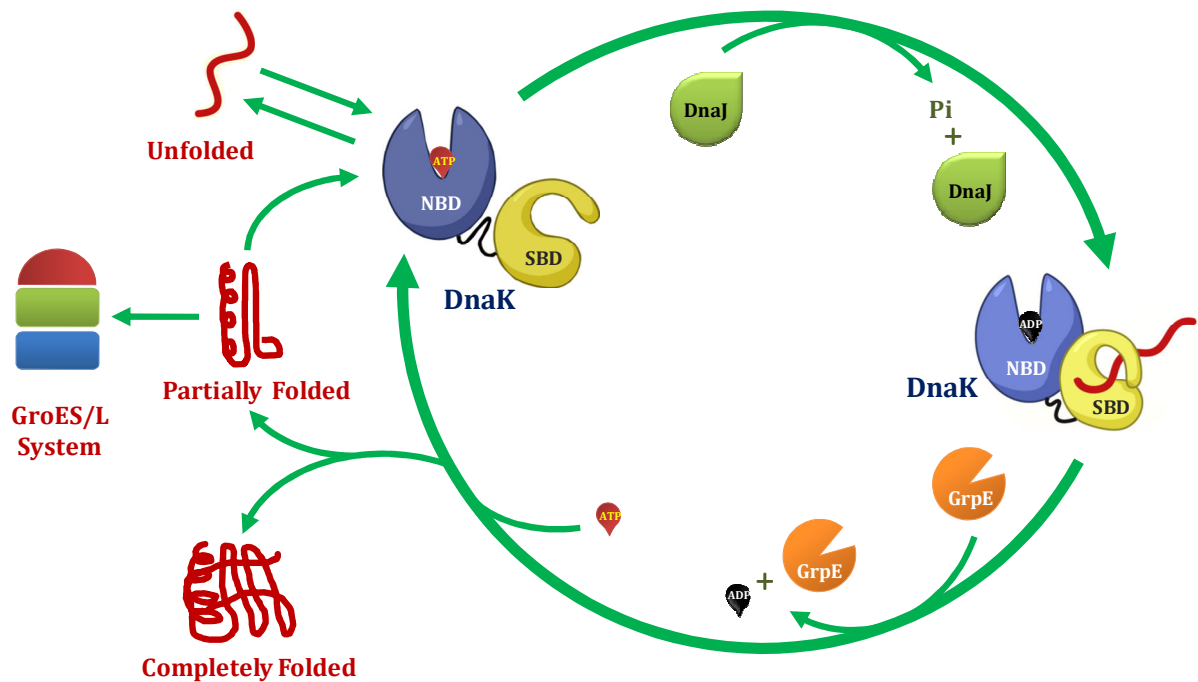
Figure 1.10: Co-operation of Co-chaperones in Hsp70 Mediated Protein Folding. Cartoon diagram showing the mechanism of action of Hsp70 mediated protein folding. ATP or ADP binds to the Nucleotide Binding Domain (NBD) in the DnaK. In the ATP bound form, the Substrate Binding Domain (SBD) exists in open conformation allowing substrate binding. Unfolded or naive polypeptides bind at the SBD in this conformation. Upon ATP Hydrolysis, an event mediated by DnaJ, conformational changes in SBD switches it to closed form and thereby locking the substrate in the cleft. Binding of the other co-chaperone, GrpE exchanges the ATP for ADP, thereby switching the conformation of SBD from closed to open state and releasing the substrate, either in a partially folded or completely folded state. Substrates in the partially folded state might either re-enter DnaK pathway or the chaperonin (GroES/L) mediated pathway. Illustrations of DnaK and the unfolded substrate protein are reproduced with permission from Bukau et al., 2006.

1.6.3 Hsp70 from *Mycobacterium tuberculosis*

Mycobacterium tuberculosis (Mtb), the causative of tuberculosis, is a dreadful pathogen causing about 2 lakh deaths a year. The heat shock proteins of this organism are reported to be antigenic and also are able to elicit immune response, as many of them are secretary. Mtb genome harbors one copy of the *dnaK* gene, designated as Rv0350, which encodes the Hsp70 homologue (Cole et al., 1998). The ORF encoding DnaK (*hsp70*, Rv0350) along with those encoding the co-chaperones, GrpE (Rv0351) and DnaJ1 (hsp40, Rv0352) followed by the co-repressor HspR (Rv0353) constitute an operon. These genes of 1878, 708, 1188 and 381 bp in length encode proteins of 66.8, 24.5, 41.3 and 14.1 kDa molecular mass, respectively. The operon appears to be up-regulated by SigH, encoded by Rv3223c (Raman et al., 2001).

Presence of two inverted repeats, with close relation to HAIR elements, upstream of the *dnaK* operons, suggests an additional HspR mediated regulation (Stewart et al., 2001; 2002). Binding of HspR to HAIR and enhancement of HspR activity in the presence of Mtb GroELs, DnaK and the co-chaperones were demonstrated in vitro using denatured HspR (Stewart et al., 2001; Das Gupta et al., 2008). Moreover, loss of persistence in HspR knockout mutants suggested an indirect role for HspR in virulence possibly via overexpression of highly antigenic DnaK, which might result in a better immune response in the host (Stewart et al., 2003). On the contrary, HspR is proposed to regulate several other virulence-related genes such as *acr2*, suggesting that the mechanism of HspR action might be intricate. Although several biochemical and genetic studies have led to the present understanding on the HspR, precise information at the molecular details, a structural study is lacking.

Hsp70 homologues from several organisms including Mtb were shown to play important roles in humoral and adaptive immune responses. Recent immunological studies showed that Mtb Hsp70 acts as a cytokine and could elicit CD40 and CCR5 mediated cytokine and CC chemokine response in PBMC, THP-1 and HEK 293 cell lines (Wang et al., 2001; Lazarevic et al., 2003; MacAry et al., 2004; Whittal et al., 2006; Floto et al., 2006). CD40 is a member of the growing TNF receptor family triggering various immunological pathways. This multifunctional protein is a 50-kDa type I trans-membrane protein principally expressed on B-cells and a large variety of other cells

together with Antigen Presenting Cells (Banchereau et al., 1994; Young et al., 1998). Though the crystal structure of CD40 is not yet determined, the domain organization was mapped to three domains: a 20-22-kDa cystein rich extracellular domain of four imperfect repeats, a 22-residue long trans-membrane helix and a 42 residue long Cytoplasmic tail (Foy et al., 1996).

Vast lines of investigation have established the importance of CD40 in the development of humoral and cell mediated immunity and other immunological pathways (Kiener et al., 1995; Cella et al., 1996). CD40, with its extra cellular domain, is known to interact with its natural ligand CD154, also known as CD40L and trigger many diverse immunological processes such as B cell proliferation, rescue from apoptosis, imunoglobulin isotype class switching (Hyper IgM syndrome), germinal center formation T cell activation or tolerance etc. Initial steps of CD40-CD40L ligation and the further signaling pathway involve variety of the TNFr Associated Factors such as TRAF 2, TRAF3, TRAF5 and TRAF 6.

Given the importance of CD40 and its interaction with the highly antigenic Mtb Hsp70 in modulating host immune responses, we have attempted to study the molecular details of the interactions by crystallizing the complex made by Mtb Hsp70 and extracellular domain of human CD40 for structural studies. Details of the experimental procedures and the results obtained therein are presented in Appendix I.

1.7 Structural and Functional Aspects of Hsp60 Mediated Protein Folding

The Hsp60 family of chaperones, also known as the chaperonins, are characterized by large ring assemblies that assist misfolded substrate proteins to reach the native state by ATP dependent cycles of binding and release. The substrate proteins are encapsulated into the cavity to fold productively. Based on the phylogenetic distribution and requirement of a co-chaperone, these are classified into two classes:

Group I constitutes the members present in the cytosol of prokaryotes and the endo-symbiotically related membrane bound eukaryotic organelle, mitochondria and chloroplast and require the co-chaperone, GroES or Cpn10. Moreover, these molecules are characterized by the formation of isologus homo-tetradecameric ring enclosing two cavities for the substrate proteins to bind. Examples include the GroEL/GroES system from *E. coli* and several eubacteria (Bukau and Horwich, 1998; Horwich et al., 2001). Group II chaperonins constitute the members localized in archeal and eukaryotic cytosol and possess built-in lid for encapsulation and thus act independent of co-chaperonin. CCT chaperonins and the well studied thermosome from archea are the members of this class (Gutsche et al., 1999).

1.7.1 Structures of Group I and Group II Chaperonins

Although the cellular roles played by the two families of chaperonins are similar, differences in the tertiary and quarternary structures of these groups of chaperonins suggest distinct mechanisms of encapsulation. As mentioned above, group I chaperonins utilize a detachable lid in the form of GroES or Hsp10 that binds in an ATP-dependent fashion. On the other hand, Group II chaperonins are covered by a built-in lid (Klump et al., 1997; Ditzel et al., 1998; Pappenberger et al., 2002). Moreover, whereas the group I chaperonins recognize the substrates directly, chaperonins belonging to group II are shown to be assisted by Prefoldin and Hsp70 homologues for substrate delivery (Iizuka et al., 2004; Cuéllar et al., 2008). Conformational changes induced by the nucleotide binding result in opening and closing of the substrate binding cavity (Meyer et al., 2003). Domain architecture of the two classes of chaperonins is presented in Figure 1.11.
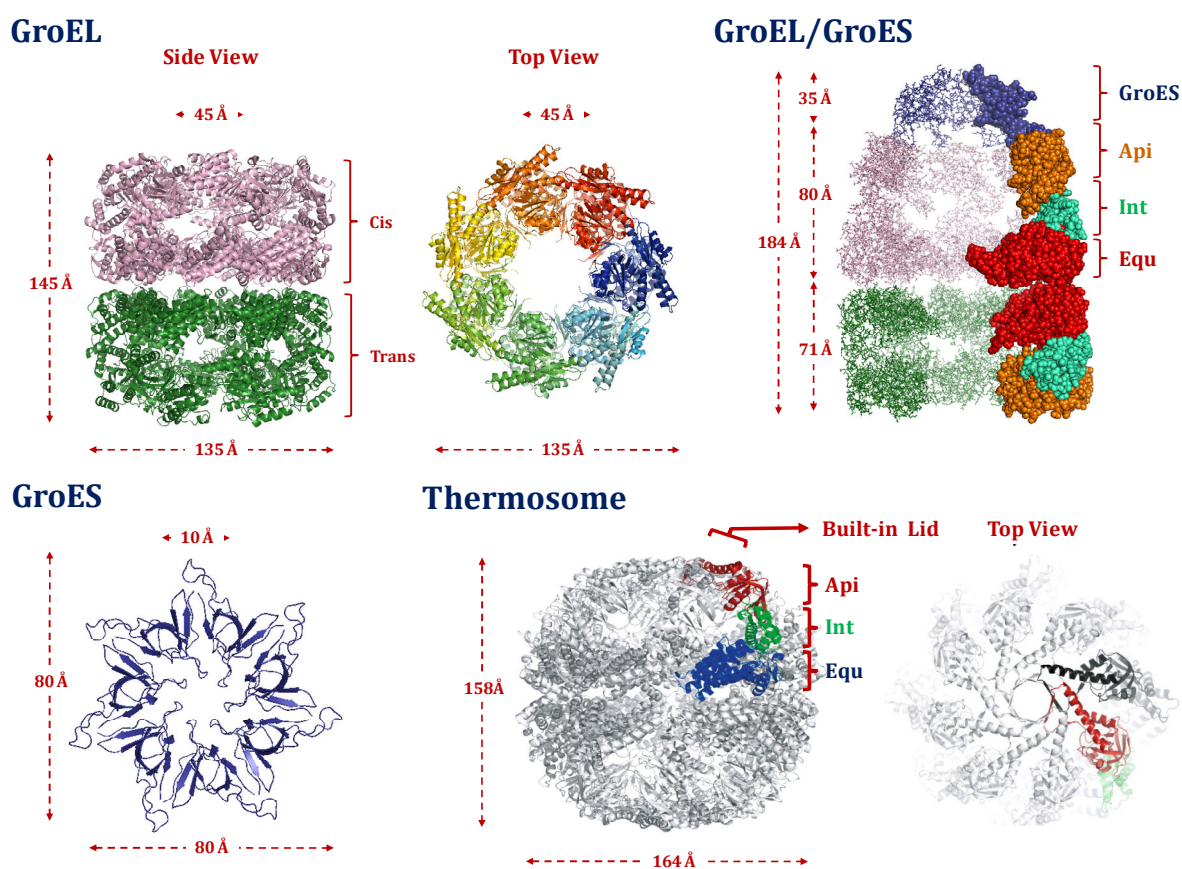
Figure 1.11*: **Architecture of Group I and II Chaperonins.** Crystallographic models of *Escherichia coli* GroEL, GroES and GroEL-GroES representing the Group I chaperonins and the thermosome from *T. acidophilum*, representing the Group II chaperonins, are presented. Individual domains in one subunit of GroEL and thermasome are indicated. Api; apical domain; Int, intermediate domain; Equ, equatorial domain. GroES acting as a lid binds to GroEL asymmetrically, at the *cis* GroEL ring, wherein the general substrate polypeptides are encapsulated. Other open ring is termed the *trans* ring. Thermosome forms a symmetric complex, showing a "closed" cavity. Illustrations for GroEL, GroES and GroEL-GroES are generated using Pymol 0.99, molecular visualization software from DeLano Scientific LLC, USA. Co-ordinates for the molecules were obtained from the structures deposited in PDB with the ID: 1OEL for GroEL, 1AON for GroEL-GroES and GroES structures. Models are of thermosome were reproduced with permission from Horwich et al., 2007.

1.7.2 Inter Domain Communication in GroEL

Genetic, biochemical and structural studies have delineated the three-domain architecture of *E. coli* GroEL monomers and the GroES-GroEL interactions (Braig et al., 1994; Xu et al., 1997). The central region of the GroEL polypeptide, spanning amino acid residues 191-376, constitutes the apical domain that is rich in hydrophobic residues and is a 3-Layer (bba) sandwich which binds the non-native substrates and GroES (Chen et al., 1994; Fenton et al., 1994; Lin et al., 1995). The equatorial ATPase domain spanning two extremities of the GroEL polypeptide, that is, residues 6-133 and 409-523, is responsible for the ATPase activity and the bulk of inter-subunit and inter ring interactions (Mayhew et al., 1996; Roseman et al., 1996). Equatorial domain is structurally an alpha helical orthogonal bundle and harbours a pseudo-Walker motif where the ATP binds. The hinge forming intermediate domain, is structurally a sandwich of two alpha beta layers. Intermediate domain spans two regions on the polypeptide namely, residues 134-190 and 377-408, and connects the equatorial and apical domains (Saibil et al., 1993; Ma et al., 2000). The conformational changes resulting from ATP binding and hydrolysis at the equatorial domain are transmitted to the apical domain via this region (Hayer-Hartl et al., 1995; Weissman et al., 1995; Ueno et al., 2004). ATP binding at the equatorial domain induces a rotation along the hinge region near the apical domain, whereby the apical domain is twisted up releasing the substrate into the cavity and exposing the hydrophibic patches for the GroES to bind (Figure 1.12).

GroEL binds a wide-range of unfolded or partially unfolded proteins via hydrophobic interactions in the parallel α-helical groove formed by H and L helices of the apical domain (Fenton et al., 1994; Braig et al., 1994). Co-operative binding of ATP at the equatorial domain triggers inward movement of the intermediate domain resulting in 60° upward movement and 90° counter clock wise rotation of the apical domain, which moves the substrate binding patches away and upward thereby releasing the substrates and allowing GroES to bind (Xu et al., 1997). Hydrolysis of ATP induces a negative co-operativity for ATP binding the trans ring, which subsequently induces allostery in the cis ring. Conformational changes therein release ADP, GroES and the bound substrate proteins (Fenton et al., 1994; Horwich et al., 2001)
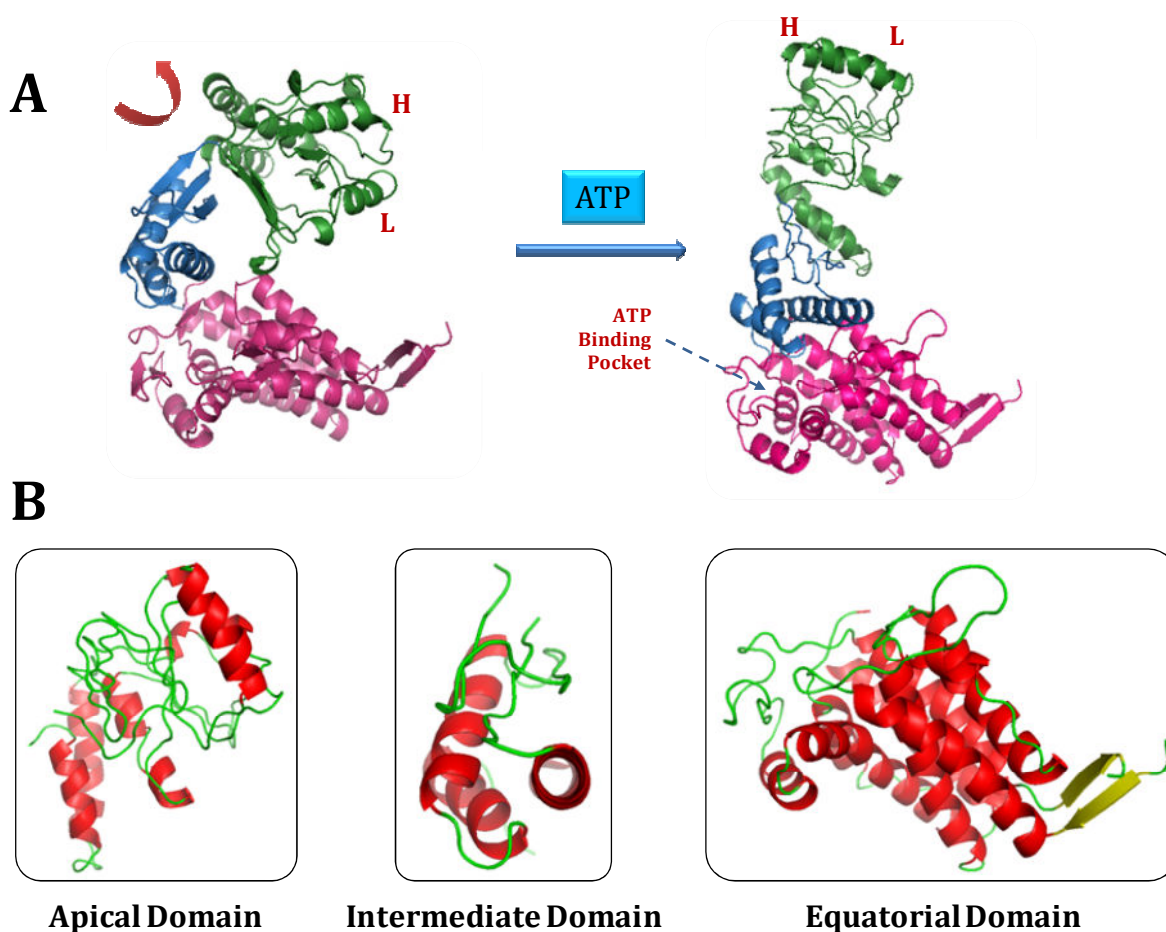
Figure 1.12: ***Structural Features of the GroEL Domains.*** **A.** Crystallographic models of domain motions in *E. coli* GroEL monomer upon ATP binding are presented. GroEL in substrate bound form assumes a constricted conformation, positioning the substrates at the opening of the cavity. Upon ATP binding, rotational movements in the intermediate domain and thereby upward movement of the apical domain result in extended conformation of the molecule. This releases the substrate into the cavity and exposes the Helices H and L for GroES. Co-ordinates for the extended and constrained GroEL monomer were obtained from PDB: 1AON chain A and chain H, respectively. **B.** Individual domains in GroEL are presented. Illustrations for the individual domains were obtained from SCOP data base. Co-ordinates for the illustrations of individual domains were obtained from PDB: 1AON chain A.

1.7.3 Mechanism of Action of GroEL

Much of the present understanding about the mechanism of action of chaperonins came from the biochemical, biophysical and genetic studies on the prokaryotic counterpart, *E. coli* GroEL. GroEL function has been shown to be a complex interplay between its interaction with and encapsulation of substrate proteins, with concomitant conformational changes induced by ATP binding, hydrolysis and GroES binding (Hayer-Hartl et al., 1995; Weissman et al., 1995; Ueno et al., 2004). Current understanding about GroEL–GroES mediated protein folding suggests two mechanisms mechanisms, the so called cis and trans mechanisms, differ in the cavity the substrate binds and co-chaperonin – GroES.

1.7.3.1 The *cis* Mechanism

Majority of the GroEL substrates follow this mechanism. GroES binds to the same side of the GroEL ring as of the substrate polypeptide and hence this mechanism is termed the cis mechanism. GroEL binds to the unfolded or kinetically trapped substrate protein intermediates, by virtue of exposed hydrophobic patches. ATP binding to the pseudo walker motif at the equatorial domain induces a conformational change in the apical domain leading to two consequences; a, the substrate is released into the hydrophilic cavity and b, capping of the substrate bound cavity by GroES upon its interaction with the hydrophobic patches on GroEL, which earlier were bound by the substrate (Figure 1.12). The substrate protein is thus sequestered from the crowded cellular milieu and allowed to fold independently in the hydrophilic cavity (Weissman et al., 1995; Mayhew et al., 1996). The protein stays in the cavity till the ATP is hydrolyzed to ADP, which usually takes about 10-15 seconds. Consequently, binding of ATP to the trans ring, induces a further conformational change which facilitates the release of GroES and subsequently the substrate protein. The substrate protein is released either in a completely folded form or in a kinetically trapped form, which can further bind to the trans ring, which is active (Rye et al., 1999).
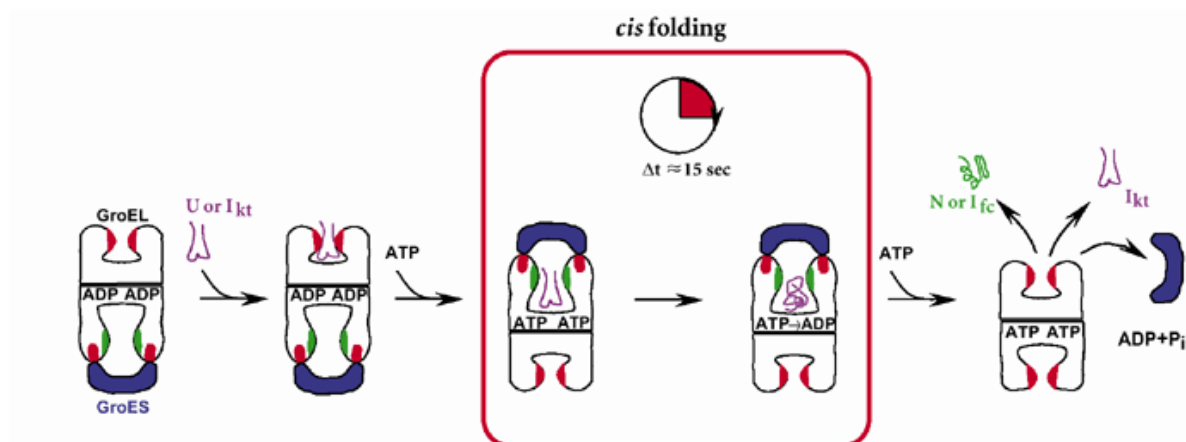
Figure 1.13: *Model for cis Folding Pathway.* Unfolded protein (U) or kinetically trapped intermediates (I_{kt}) binds the asymmetric GroEL by hydrophobic patches (red). ATP binding induces conformational changes which releases the substrate into hydrophilic cavity (green) and facilitate GroES binding (blue). Substrate protein gets folded in a span of about 15 seconds when the ATP is hydrolysed to ADP. ATP binding to the trans cavity releases the substrates in either native (N) or as folding committed intermediates (I_{fc}) or as kinetically trapped intermediates (I_{kt}). The illustration is reproduced with permission from Erbse et al., 2003.

1.7.3.2 The *trans* Mechanism

The usual size limit for the substrate proteins, as shown by both in vitro and in vivo studies, is around 57 kDa although the *cis* cavity is reported to theoretically accommodate larger proteins of the order of 104 kDa (Houry et al., 1999; Sakikawa et al., 1999; Lin and Rye, 2004). Productive *in vivo* folding of the proteins larger than the usual size limit, such as 86 kDa maltose binding protein fusion and 82 kDa mitochondrial aconitase has also been reported (Huang and Chuang, 1999; Chaudhuri et al., 2001). Since such large substrates are difficult to accommodate in the central cavity, it has been suggested that their productive folding might occur outside the *cis* cavity. These studies therefore indicate that the substrate recognition patterns of GroEL may be more diverse than initially thought.

Binding a large substrate protein by GroEL would prevent GroES binding in *cis* cavity. Therefore a reaction cycle with GroES binding to the *trans* cavity for folding large protein substrates such as 82 kDa Iron sulphur protein, yeast mitochondrial aconitase was demonstrated by Arthur Horwich and colleagues (Chaudhuri et al., 2001). Unlike the cis folding, only a part of the substrate protein, the Iron-Sulphur centre, was shown to be folded by the GroEL/GroES machinery, suggesting nucleation step, where from rest of the protein folds independent of the chaperonins.

On the contrary, electron microscopic studies on single ring version of GroEL, SR1, displayed that the *cis* ring can accommodate large substrates, such as 86 kDa mitochondrial branched-chain α-ketoacid dehydrogenase by expansion of the cavity by 80% than shown in crystal structure (Chen et al., 2006). Notably, productive release of the large encapsulated proteins was not demonstrated in *cis* cavity. The cartoon below displays the proposed mechanism for trans folding.
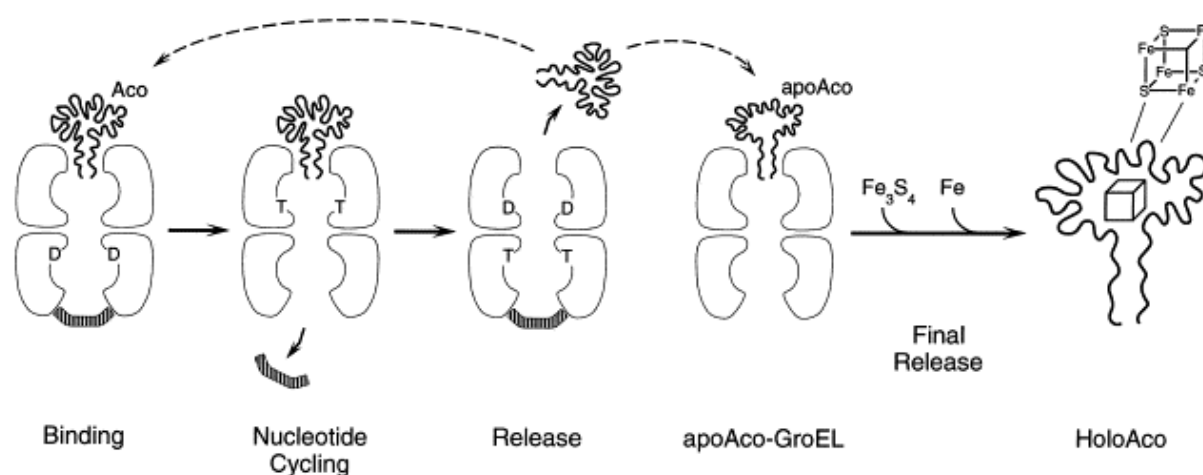
Figure 1.14: *Model for trans Folding Pathway.* Unfolded Aconitase (Aco), was proposed to bind the open cavity of the GroEL/GroES/ADP complex. Followed by ATP binding in the polypeptide-bound ring (*trans*) and cosequent release of GroES from the *cis* ring. Binding of ATP and GroES at the cis ring and releases Aconitase in the bulk solution. During this, Aconitase is folded to apoaconitase. Molecules that are not folded enter the one more round of GroEL/GroES cycle. Aconitase is released during the formation of Iron-Sulphur Cluster to produce holoenzyme. The illustration is reproduced with permission from Chaudhuri et al., 2001.

These two mechanisms of action of GroEL, therefore emphasize the requirement of the tetradecameric ring for the functioning of GroEL as a chaperonin.

1.7.4 GroELs from *Mycobacterium tuberculosis*

Recent genome annotation studies on various bacteria have revealed that a few bacterial genomes possess multiple copies of *groEL* genes (Fischer et al., 1993; Karunakaran et al., 2003; Barreiro et al., 2005). The *Mycobacterium tuberculosis* (Mtb) genome bears two copies of *groEL* genes (*groELs*). One of these, *groEL1,* is arranged in an operon, with the cognate co-chaperonin *groES*, being the first gene, while the second copy, *groEL2* exists separately on the genome (Kong et al., 1993; Cole et al., 1998). Recombinant mycobacterial GroELs were shown to possess biochemical features that deviated significantly from the trademark properties of *E. coli* GroEL. The most striking feature of Mtb GroELs, however, was their oligomeric state, where contrary to expectations, *in vitro* they did not form the canonical tetradecameric assembly when purified from *E. coli* (Figure 1.15). The proteins rather existed as lower oligomers (dimers) irrespective of the presence or absence of cofactors such as the cognate GroES or ATP (Qamra et al., 2004; Qamra and Mande, 2004). Furthermore, they displayed weak ATPase activities and GroES independence in preventing aggregation of the denatured polypeptides.

Evolutionary studies on Mtb *groEL* sequences have suggested rapid evolution of the *groEL1* gene, yet not turning these into pseudogenes (Goyal et al., 2006). The other hypothesis suggests that Mtb, being a slow growing organism, might require GroEL function that does not utilize ATP rapidly, but rather with a slow turnover rate. Alternately, additional mechanisms might exist in Mtb, which could mediate regulated oligomerization of Mtb chaperonins. Such regulation might help in the controlled utilization of ATP in nutrient deprived Mtb, as observed for other chaperones such as small heat shock proteins (Haslbeck et al., 2005).
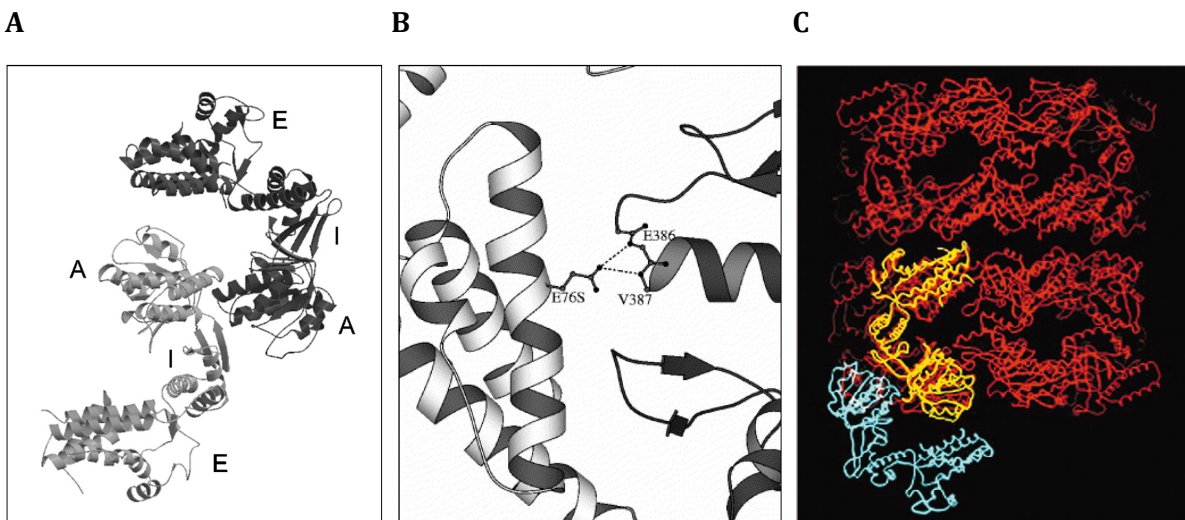
Figure 1.15: *Unusual Oligomeric Nature of Mtb GroEL.* **A.** Crystallographic model showing dimeric form of Mtb GroEL2. The intersuunit interactions are mediated via the apical domains, unlike by the equatorial domains in *E. coli* GroEL. Domain architecture is indicated to show apical (A), intermediate (I) and equatorial (E) domains. The illustration is reproduced with permission from Qamra and Mande, 2004. **B.** Glu76, which is converted to Ser75 in Mtb GroELs, is thought to play key role in intersubunit interactions. Glu76 in one subunit (showin in gray) interacts with Glu386 of the neighbouring subunit (shown in black) and facilitates the oligomerization in *E. coli* GroEL. Therefore, presence of serine at this position, in Mtb GroELs is thought to result in the loss of its oligomeric structure. This Figure was reproduced with permission from Qamra et al., 2004. **C.** Crystallographic alignment of *E. coli* GroEL tetradecamer (shown in red) with Mtb GroEL2 dimer (cyan and yellow). Differences in the intersubunit interactions, between the two GroEL homologues, are revealed. The illustration is reproduced with permission from Qamra and Mande, 2004.

During the course of this thesis, we have exploited the unusual oligomeric status of the recombinant Mtb GroELs to study the significance of oligomer formation in GroEL function as a molecular chaperone. Furthermore we have explored the possibility of existence of regulated oligomerization for native Mtb GroELs in their natural setting. We first show that Mtb *groEL* genes are not capable of complementing a conditional allele of *E. coli groEL* namely, *groEL44* (Chapter II). Phenotypic and biochemical analyses of GroEL variants obtained by gene shuffling and domain swapping studies suggest that the impaired chaperoning ability of recombinant Mtb GroELs is a consequence of their inability to form higher order oligomers in *E. coli* and that oligomerization is the prelude to the formation of an active GroEL chaperonin (Chapter III). Further by immunochemical and mass spectrometric analysis of native mycobacterial GroELs, we show that Mtb GroEL1 exists in multiple oligomeric forms viz. monomeric, dimeric, heptameric (single-ring) and tetradecameric (double-ring) forms, and a switch between single ring and double ring variants is operated by phosphorylation on a Serine residue (Chapter VI). These observations suggest that determinants of oligomerization for Mtb GroEL1 are distinct from its *E. coli* counterpart, and that it is capable to oligomerise in Mtb (its native environment), whereas loses its oligomerization capability when expressed in *E. coli*. It could thus be possible that Mtb GroEL1 requires certain native Mtb protein, probably a eukaryotic-like Ser-Thr protein kinase, to oligomerise properly. Detailed methodologies followed and the results obtained are discussed in the following chapters as mentioned below.

1.8 References

- Agarwal, M., Katiyar-Agarwal, S., Sahi, C., Gallie, D. R. and Grover, A. (2001) *Arabidopsis thaliana* Hsp100 proteins: kith and kin. *Cell Stress Chaperones* **6**, 219–224

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter P. (2007) Molecular Biology of the Cell. 5e, Garland Science, U. K.

- Arsène. F., Tomoyasu, T. and Bukau, B. (2000) The heat shock response of *Escherichia coli. Int. J. Food Microbiol.* **55**, 3-9

- Ashburner, M. (1982) The genetics of a small autosomal region of *Drosophila melanogaster* containing the structural gene for alcohol dehydrogenase. III. Hypomorphic and hypermorphic mutations affecting the expression of hairless. *Genetics* **101**, 447-459

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

- Bahl, H., Echols, H., Straus, D. B., Court, D., Crowl, R. and Georgopoulos, C. P. (1987) Induction of the heat shock response of *E. coli* through stabilization of σ^{32} by the phage cIII protein. *Genes Dev.* **1,** 57-64

- Baird, P. N., Hall, L. M. C. and Coates, A. R. M. (1989) Cloning and sequence analysis of the 10 kDa antigen gene of *Mycobacterium tuberculosis. J. Gen. Microbiol.* **135**, 931–939

- Banchereau, J., Bazan, F., Blanchard, D., Brière, F., Galizzi, J. P., van Kooten, C., Liu, Y. J., Rousset, F. and Saeland, S. (1994) The CD40 and its antigen. *Annu. Rev. Immunol.* **12**, 881-922

- Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli. Nat. Biotechnol.* **22**, 1399-1408

- Bardwell, J. C. and Craig, E. A. (1984) Major heat shock gene of Drosophila and the *Escherichia coli* heat-inducible *dnaK* gene are homologous*. Proc. Natl. Acad. Sci. USA* **81**, 848-852

- Barnett, T. M., Altschuler, C. N., McDaniel, C. N. and Mascarentes, J. P. (1980) Heat shock induced proteins in plant cells. *Dev. Genet.* **1**, 331–340

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

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

- Beckmann, R. P., Mizzen, L. E. and Welch, W. J. (1990) Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. *Science* **248**, 850-854

- Ben-Zvi, A. P. and Goloubinoff, P. (2001) Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J. Struct. Biol.* **135**, 84-93

- Beuron, F., Maurizi, M. R., Belnap, D. M., Kocsis, E., Booy, F. P., Kessel, M. and Steven, A. C. (1998) At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. *J. Struct. Biol.* **123**, 248–259

- Bimston, D., Song, J., Winchester, D., Takayama, S., Reed, J. C. and Morimoto, R. I. (1998) BAG-1, a negative regulator of Hsp70 chaperone activity, uncouples nucleotide hydrolysis from substrate release. *EMBO J.* **17,** 6871–6878

- Blair, G. E. and Ellis, R. J. (1973) Protein synthesis in chloroplasts. I. Light-driven synthesis of the large subunit of fraction I protein by isolated pea chloroplasts. *Biochem. Biophys. Acta.* **319,** 223-234

- Bloom, M. V., Milos, P. and Roy, H. (1983) Light-dependent assembly of ribulose-1,5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* **80**, 1013-1017

- Boston, R. S., Viitanen, P. V. and Vierling, E. (1996) Molecular chaperones and protein folding in plants. *Plant Mol. Biol.* **32**, 191–222

- Bouche, G., Amalric, F., Caizergues-Ferrer, M. and Zalta, J. P. (1979) Effects of heat shock on gene expression and subcellular protein distribution in Chinese hamster ovary cells. *Nucleic Acids Res.* **7**, 1739–1747

- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. and Sigler, P. B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586

- Bucca, G., Ferina, G., Puglia, A. M. and Smith, C. P. (1995) The *dnaK* operon of *Streptomyces coelicolor* encodes a novel heat-shock protein which binds to the promoter region of the operon. *Mol. Microbiol.* **17**, 663–674

- Buchberger, A., Schröder, H., Hesterkamp, T., Schönfeld, H. J. and Bukau, B. (1996) Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. *J. Mol. Biol.* **261**, 328–333

- Buchner, J. (1999) Hsp90 & Co. a holding for folding. *Trends Biochem. Sci.* **24**, 136–141

- Bukau B., Deuerling E., Pfund C. and Craig E. A. (2000) Getting newly synthesized proteins into shape. *Cell* **101:** 119–122

- Bukau, B. (2005) Ribosomes catch Hsp70s, *Nat. Struc. Mol. Biol.* **12**, 472-473

- Bukau, B. and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351–366

- Bukau, B., Weissman, J. and Horwich, A. (2006) Molecular chaperones and protein quality control. *Cell* **125**, 443–451

- Cella, M., Scheidegger, D., Palmer-Lehmann, K., Lane, P., Lanzavecchia, A. and Alber, G. (1996) Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: T-T help via APC activation. *J. Exp. Med.* **184**, 747-752

- Chang, B. Y., Chen, K. Y., Wen, Y. D. and Liao, C. T. (1994) The response of a *Bacillus subtilis* temperature-sensitive sigA mutant to heat stress. *J. Bacteriol.* **176**, 3102-3110

- Chang, H., Tang, Y., Hayer-Hartl, M. and Hartl, F. U. (2007) SnapShot: Molecular Chaperones, Part I. *Cell* **128**, 212-212.e1

- Chaudhuri, S., Jana, B., Sarkar, S. and Basu, T. (2004) Accumulation of the periplasmic protein alkaline phosphatase in cell cytosol induces heat shock response in *E. coli. Curr. Sci.* **87**, 986-990

- Chaudhuri, T. K., Farr, G. W., Fenton, W. A., Rospert, S. and Horwich, A. L. (2001) GroEL/GroES-mediated folding of a protein too large to be encapsulated. *Cell* **107,** 235–246

- Chen, D. H., Song, J. L., Chuang, D. T., Chiu, W. and Ludtke, S. J. (2006) An expanded conformation of single-ring GroEL-GroES complex encapsulates an 86 kDa substrate. *Structure*. **14**, 1711-1722

- Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R. and Saibil, H. R. (1994) Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature* **371**, 261–264

- Chirico, W. J., Waters, M. G. and Blobel, G. (1988) 70K heat shock related proteins stimulate protein translocation into microsomes. *Nature* **332**, 805-810

- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. III., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544

- Cuéllar, J., Martín-Benito, J., Scheres, S. H., Sousa, R., Moro, F., López-Viñas, E., Gómez-Puertas, P., Muga, A., Carrascosa, J. L. and Valpuesta, J. M. (2008) The structure of CCT-Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nat. Struct. Mol. Biol.* **15**, 858-864

- Cupp-Vickery, J. R., Peterson, J. C., Ta, D. T. and Vickery, L. E. (2004) Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. *J. Mol. Biol.* **342**, 1265-1278

- Dafforn, T. R., Della, M. and Miller A. D. (2001) The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J. Biol. Chem.* **276**, 49310-49319

- Das Gupta, T., Bandyopadhyay, B. and Das Gupta, S. K. (2008) Modulation of DNA-binding activity of *Mycobacterium tuberculosis* HspR by chaperones. *Microbiology* **154**, 484-490

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

- Derré, I., Rapoport, G., Devine, K., Rose, M. and Msadek, T. (1999a) ClpE, a novel type of HSP100 ATPase, is part of the CtsR heat shock regulon of *Bacillus subtilis. Mol. Microbiol.* **32**, 581-593

- Derré, I., Rapoport, G. and Msadek, T. (1999b) 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

- Deshaies, R. J., Koch. B. D., Werner-Washburne, M., Craig, E. A. and Schekman, R. (1988) A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* **332**, 800-805

- Deuerling, E., Patzelt, H., Vorderwülbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H. and Bukau, B. (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* **47**, 1317-1328

- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. and Bukau, B. (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* 400, 693-696

- Ditzel, L., Löwe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R. and Steinbacher, S. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93,** 125–138

- Ehrnsperger, M. S., Gräber, S., Gaestel, M. and Buchner, J. (1997) Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J.* **16**, 221–229

- Ellis, R. J. (1996) Discovery of molecular chaperones *Cell Stress Chaperones* **1**, 155-160.

- Ellis. J. (1987) Proteins as molecular chaperones. *Nature (London)* **328,** 378–379

- Erbse, A., Dougan, D. A. and Bukau, B. (2003) A folding machine for many but a master of none. *Nat. Struct. Biol.* **10,** 84-86

- Fenton, W., Kashi, Y., Furtak, K. and Horwich, A. (1994) Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614-619

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

- Flaherty, K. M., DeLuca-Flaherty, C. and McKay, D. B. (1990) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**, 623-628

- Floto, R. A., MacAry, P. A., Boname, J. M., Mien, T. S., Kampmann, B., Hair, J. R., Huey, O. S., Houben, E. N., Pieters, J., Day, C., Oehlmann, W., Singh, M., Smith, K. G. And Lehner, P. J. (2006) Dendritic cell stimulation by mycobacterial Hsp70 is mediated through CCR5. *Science* **314**, 454-458

- Fohlman, J., Eaker, D., Karlsoon, E. and Thesleff, S. (1976) Taipoxin, an extremely potent presynaptic neurotoxin from the venom of the australian snake taipan (*Oxyuranus s. scutellatus*). Isolation, characterization, quaternary structure and pharmacological properties. *Eur. J. Biochem.* **68**, 457-469

- Foy, T. M., Aruffom A., Bajorath, J., Buhlmann, J. E. and Noelle, R. J. (1996) Immune regulation by CD40 and its ligand GP39. *Annu. Rev. Immunol.* **14**, 591-617

- Frydman, J. (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**, 603–647

- Gassler, C. S., Wiederkehr, T., Brehmer, D., Bukau, B. and Mayer, M. P. (2001) Bag-1M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration-dependent as positive and negative cofactor. *J. Biol. Chem.* **276**, 32538-32544

- Genevaux, P., Georgopoulos, C. and Kelley, W. L. (2007) The Hsp70 chaperone machines of *Escherichia coli*: a paradigm for the repartition of chaperone functions. *Mol. Microbiol.* **66**, 840-857

- Georgopoulos, C. and Welch, W. J. (1993) Role of the major heat shock proteins as molecular chaperones. *Annu. Rev. Cell. Biol.* **9,** 601-634

- Georgopoulos, C. P. (1977) A new bacterial gene (*groPC*) which affects lambda DNA replication. *Mol. Gen. Genet.* **151**, 35-39

- Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R. and Kaiser, A. D. (1973) Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76**, 45-60

- Glover, J. R. and Lindquist, S. (1998) Hsp104, Hsp70 and Hsp40: a novel chaperone system that rescues previously aggregated proteins. *Cell* **94**, 73–82

- Godowski, P. J. and Picard, D. (1989) Steroid receptors. How to be both a receptor and a transcription factor. *Biochem. Pharmacol.* **38**, 3135-3143

- Goloubinoff, P., Christeller, J. T., Gatenby, A. A. and Lorimer, G. H. (1989a) Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**, 884-889

- Goloubinoff, P., Gatenby, A. A. and Lorimer, G. H. (1989b) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli. Nature* **337**, 44-47

- Goloubinoff, P., Mogk, A., Zvi, A. P., Tomoyasu, T. and Bukau, B. (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. USA* **96**, 13732–13737

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

- Grandvalet, C., Servant, P., Mazodier, P. (1997) Disruption of *hspR*, the repressor gene of the *dnaK* operon in *Streptomyces albus* G. *Mol. Microbiol.* **23**, 77–84

- Grossman, A. D., Ericksonm J. W. and Gross, C. A. (1984) The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* **38**, 383-390

- Gutsche, I., Essen, L. O. and Baumeister, W. (1999) Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293,** 295-312

- Hartl, F. U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**, 571–580

- Hartl, F. U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295,** 1852–1858

- Haslbeck, M., Franzmann, T., Weinfurtner, D. and Buchner, J. (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12,** 842-846

- Hayer-Hartl, M. K., Martin. J. and Hartl, F. U. (1995) Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. *Science* **269**, 836-841

- Hecker, M., Schumann, W. and Völker, U. (1996) Heat-shock and general stress response in *Bacillus subtilis. Mol. Microbiol.* **19**, 417–428

- Hemmingsen, S. M. and Ellis, R. J. (1986) Purification and Properties of Ribulosebisphosphate Carboxylase Large Subunit Binding Protein. *Plant Physiol.* **80**, 269-276

- Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W. and Ellis, R. J. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333,** 330-334

- Hendrix, R. W. (1979) Purification and properties of *groE*, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* **129**, 375-392

- Hesterkamp, T. and Bukau, B. (1998) Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in *E. coli. EMBO J.* **17**, 4818-4828

- Hightower, L. E., and White, F. P. (1981) Cellular responses to stress: comparison of a family of 71–73-kilodalton proteins rapidly synthesized in rat tissue slices and canavanine treated cells in culture. *J. Cell Physiol.* **108**, 261–275

- Hohn, T., Hohn, B., Engel, A., Wurtz, M. and Smith, P. R. (1979) Isolation and characterization of the host protein *groE* involved in bacteriophage lambda assembly. *J. Mol. Biol.* **129**, 359-373

- Horwich, A. L., Fenton, W. A., Chapman, E. and Farr, G. W. (2007) Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell. Dev. Biol.* **23**, 115-145

- Horwich, A. L., Fenton, W. A. and Rapoport, T. A. (2001) Protein folding taking shape. Workshop on molecular chaperones. *EMBO Rep.* **2**, 1068-1073

- Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F. U. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402,** 147–154

- Huang, Y. S. and Chuang, D. T. (1999) Mechanisms for GroEL/GroES-mediated Folding of a Large 86-kDa Fusion Polypeptide in vitro. *J. Biol. Chem.* **274,** 10405-10412

- Hurme, R., Berndt, K. D., Namok, E. and Rhen, M. (1996) DNA binding exerted by a bacterial gene regulator with an extensive coiled coil domain. *J. Biol. Chem.* **271,** 12626–12631

- Hurme, R., Berndt, K. D., Normark, S. J. and Rhen, M. (1997) A proteinaceous gene regulatory thermometer in Salmonella. *Cell* **90**, 55–64

- Iizuka, R., So, S., Inobe, T., Yoshida, T., Zako, T., Kuwajima, K. and Yohda, M. (2004) Role of the helical protrusion in the conformational change and molecular chaperone activity of the archaeal group II chaperonin. *J. Biol. Chem.* **279**, 18834-18839

- Imai, J., Maruya, M., Yashiroda, H., Yahara, I. and Tanaka, K. (2003) The molecular chaperone Hsp90 plays a role in the assembly and maintenance of the 26S proteasome. *EMBO J.* **22**, 3557–3567

- Jiang, J., Prasad, K., Lafer, E. M. and Sousa, R. (2005) Structural basis of interdomain communication in the Hsc70 chaperone. *Mol. Cell.* **20**, 513-524

- Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M. and Cossart, P. (2002) An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes. Cell* **110,** 551–561

- Kaempfer, R. (2003) RNA sensors: novel regulators of gene expression. *EMBO Rep.* **4,** 1043–1047

- Karunakaran, K. P., Noguchi, Y., Read, T. D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C. C. and Brunham, R. C. (2003) Molecular Analysis of the Multiple GroEL Proteins of Chlamydiae. *J. Bacteriol.* **185**, 1958-1966

- Kelley, P. M. and Schlesinger, M. J. (1978) The effect of amino acid analogues and heat shock on gene expression in chicken embryo fibroblasts. *Cell* **15,** 1277-1286

- Kiener, P. A., Moran-Davis, P., Rankin, B. M., Wahl, A. F., Aruffo, A. and Hollenbaugh, D. (1995) Stimulation of CD40 with purified soluble gp39 induces proinflammatory responses in human monocytes. *J. Immunol.* **155,** 4917-4925

- Kluck, C. J., Patzelt, H., Genevaux, P., Brehmer, D., Rist, W., Schneider-Mergener, J., Bukau, B. and Mayer, M. P. (2002) Structure-function analysis of HscC, the *Escherichia coli* member of a novel subfamily of specialized Hsp70 chaperones. *J. Biol. Chem.* **277**, 41060-41069

- Klump, M., Baumeister, W. and Essen, L. O. (1997) Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* **91**, 263–270

- Kochan, J. and Murialdo, H. (1983) Early intermediates in bacteriophage lambda prohead assembly. II. Identification of biologically active intermediates. *Virology* **131**, 100-115

- Kong, H., Coates, A. R., Butcher, P. D., Hickman, C. J. and Shinnick, T. M. (1993) *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc. Natl. Acad. Sci. USA* **90**, 2608–2612

- Krüger, E., Zühlke, D., Witt, E., Ludwig, H. and Hecker, M. (2001) Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *EMBO J.* **20**, 852-863

- Krüger, E., Msadek, T. and Hecker, M. (1996) Alternate promoters direct stress-induced transcription of the *Bacillus subtilis clpC* operon. *Mol. Microbiol.* **20**, 713–723

- Laksanalamai, P., Whitehead, T. A. and Robb, F. T. (2004) Minimal protein-folding systems in hyperthermophilic archaea. *Nat. Rev. Microb.* **2**, 315-324

- Lambert, de. R. C., Sluiters, C. and Cornelis, G. R. (1992) Role of the transcriptional activator, VirF, and temperature in the expression of the pYV plasmid genes of *Yersinia enterocolitica. Mol. Microbiol*. **6**, 395–409

- Laskey, R. A., Honda, B. M., Mills, A. D. and Finch, J. T. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275**, 416-420

- Lazarevic, V., Myers, A. J., Scanga, C. A. and Flynn, J. L. (2003) CD40, but not CD40L, is required for the optimal priming of T cells and control of aerosol *M. tuberculosis* infection. *Immunity* **19**, 823 - 835

- Lee, G. J. and Vierling, E. (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiol.* **122**, 189–198

- Lee, G. J., Roseman, A. M., Saibil, H. R. and Vierling, E. (1997) A small heat shock protein stably binds heat denatured model substrates and can maintain a substrate in a folding competent state. *EMBO J.* **16**, 659–671

- Lee, S., Sowa, M. E., Watanabe, Y., Sigler, P. B., Chiu, W., Yoshida. M. and Tsai, F. T. S. (2003) The Structure of ClpB: A Molecular Chaperone that Rescues Proteins from an Aggregated State. *Cell* **115**, 229-240

- Lemaux, P. G., Herendeen, S. L., Bloch, P. L. and Neidhardt, F. C. (1978) Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts. *Cell* **13**, 427-434

- Lewis, M. J. and Pelham, H. R. (1985) Involvement of ATP in the nuclear and nucleolar functions of the 70 kd heat shock protein. *EMBO J.* **4**, 3137-3143

- Lin, Z. and Rye, H. S. (2004) Expansion and compression of a protein folding intermediate by GroEL. *Mol. Cell,* **16**, 23–34

- Lin, Z., Schwartz, F. P. and Eisenstein, E. (1995). The hydrophobic nature of GroEL-substrate binding. *J. Biol. Chem.* ***270,*** 1011–1014

- Lindquist, S. (1986) The heat-shock response. *Annu. Rev. Biochem.* **55**, 1151–1191

- Lindquist, S. and Craig, E. A. (1988) The heat-shock proteins. *Annu. Rev. Genet.* **22**, 631–677

- Ma, J., Sigler, P. B., Xu, Z. and Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **302**, 303-313

- MacAry, P. A., Javid, B., Floto, R. A., Smith, K. G., Oehlmann, W., Singh, M. and Lehner, P. J. (2004) HSP70 peptide binding mutants separate antigen delivery from dendritic cell stimulation. *Immunity* **20**, 95-106

- Mandal, M. and Breaker, R. R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463

- Mayer, M. P. and Bukau, B. (2005) Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci.* **62**, 670-684

- Mayhew, M., da Silva, A. C., Martin, J., Erdjument-Bromage, H., Tempst, P. and Hartl, F. U. (1996) Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* **379**, 420-426

- McCarty, J. S. and Walker, G. C. (1991) DnaK as a thermometer: threonine-199 is site of autophosphorylation and is critical for ATPase activity. *Proc. Natl. Acad. Sci. USA* **88**, 9513–9517

- Meyer, A. S., Gillespie, J. R., Walther, D., Millet, I. S., Doniach, S. and Frydman, J. (2003) Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **113**, 369–381

- Miller, M. J., Xuong, N. H. and Geiduschek, E. P. (1979) A response of protein synthesis to temperature shift in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **76**, 1117–1121

- Mogk, A., Deuerling, E., Vorderwülbecke, S., Vierling, E. and Bukau, B. (2003a) Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol. Microbiol.* **50**, 585–595

- Mogk, A., Homuth, G., Scholz, C., Kim. L. X., Schmid, F. and 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

- Mogk, A., Schlieker, C., Friedrich, K. L., Schönfeld, H. J., Vierling, E. and Bukau, B. (2003b) Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/ DnaK. *J. Biol. Chem*. **278**, 31033–31042

- Morange, M. (2005) What history tells us II. The discovery of chaperone function. *J. Biosci.* **30**, 461-464

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

- Münchbach, M., Nocker, A., Narberhaus, F. (1999) Multiple small heat shock proteins in *Rhizobia*. *J. Bacteriol*. **181**, 83-90

- Munro, S. and Pelham, H. R. (1986) An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein. *Cell* **46**, 291-300

- Musgrove, J. E., Johnson, R. A. and Ellis, R. J. (1987) Dissociation of the ribulosebisphosphate-carboxylase large-subunit binding protein into dissimilar subunits. *Eur. J. Biochem.* **163**, 529-534

- Narberhaus F (2002) Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol. Mol. Biol. Rev.* **66**, 64–93

- Narberhaus, F. (1999) Negative regulation of bacterial heat shock genes. *Mol. Microb.***31**, 1-8

- Narberhaus, F., Giebeler, K. and Bahl, H. (1992) Molecular characterization of the *dnaK* gene region of *Clostridium acetobutylicum*, including *grpE*, *dnaJ*, and a new heat shock gene. *J. Bacteriol.* **174**, 3290–3299

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

- Narberhaus, F., Kowarik, M., Beck, C. and Hennecke, H. (1998b) Promoter selectivity of the *Bradyrhizobium japonicum* RpoH transcription factors in vivo and in vitro. *J. Bacteriol.* **180,** 2395–2401

- Narberhaus, F., Waldminghaus, T. and Chowdhury, S. (2006) RNA thermometers. *FEMS Microbiol. Rev.* **30,** 3–16

- Narberhaus, F. and Bahl, H. (1992) Cloning, sequencing, and molecular analysis of the *groESL* operon of *Clostridium acetobutylicum*. *J. Bacteriol.* **174**, 3282–3289

- Neupert, W. and Brunner, M. (2002) The protein import motor of mitochondria. *Nat. Rev. Mol. Cell. Biol.* **3**, 555–565

- Neuwald, A. F., Aravind, L., Spouge, J. L. and Koonin, E. V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation and disassembly of protein complexes. *Genome Res.* **9**, 27–43

- Nishiyama, M., Horst, R., Eidam, O., Herrmann, T., Ignatov, O., Vetsch, M., Bettendorff, P., Jelesarov, I., Grütter, M. G., Wüthrich, K., Glockshuber, R. and Capitani, G. (2005) Structural basis of chaperone-subunit complex recognition by the type 1 pilus assembly platform FimD. *EMBO J.* **24**, 2075-2086

- Nocker, A., Krstulovic, N. P., Perret, X. and Narberhaus, F. (2001) ROSE elements occur in disparate rhizobia and are functionally interchangeable between species. *Arch. Microbiol.* **176**, 44-51

- Pappenberger, G., Wilsher, J. A., Roe, S. M., Counsell, D. J., Willison, K. R. and Pearl, L. H. (2002) Crystal structure of the CCTγ apical domain: implications for substrate binding of the eukaryotic cytosolic chaperonin. *J. Mol. Biol.* **318**, 1367–1379

- Parsell, D. A. and Sauer, R. T. (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

- Patel, S. and Latterich, M. (1998) The AAA team: related ATPases with diverse functions. *Trends. Cell Biol.* **8**, 65-71.

- Pelham, H. R. (1984) Hsp70 accelerates the recovery of nucleolar morphology after heat shock. *EMBO J.* **3,** 3095-3100

- Pelham, H. R. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46**, 959-961

- Pellecchia, M., Montgomery, D. L., Stevens, S. Y., Vander Kooi, C. W., Feng, H. P., Gierasch, L. M. and Zuiderweg, E. R. (2000) Structural insights into substrate binding by the molecular chaperone DnaK. *Nat. Struct. Biol.* **7**, 298-303

- Pratt, W. B. and Toft, D. O. (2003) Regulation of signalling protein function and trafficking by the Hsp90/Hsp70-based chaperone machinery. *Exp. Biol. Med. (Maywood)* **228,** 111–133

- Pratt, W. B., Krishna, P. and Olsen, L. J. (2001) Hsp90-binding immunophilins in plants: the protein movers. *Trends. Plant Sci.* **6**, 54–58

- Prosseda, G., Fradiani, P. A., Di, L. M., Falconi, M., Micheli, G., Casalino, M. Nicoletti, M. and Colonna, B. (1998) A role for H-NS in the regulation of the *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Res. Microbiol.* **149**, 15–25

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

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

- Queitsch, C., Sangster, T. A. and Lindquist S. (2002) Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624

- Raman. S., Song, T., Puyang, X., Bardarov, S., Jacobs, W. R. Jr. and Husson, R. N. (2001) The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*. *J. Bacteriol.* **183**, 6119–6125

- Ranson, N. A., White, H. E. and Saibil, H. R. (1998) Chaperonins. *Biochem. J.* **333**, 233–242

- Reddy, G. B., Das, K. P., Petrash, J. M. and Surewicz, W. K. (2000) Temperature-dependent chaperone activity and structural properties of human αA- and αB-crystallins. *J. Biol. Chem.* **275**, 4565–4570

- Revington, M., Zhang, Y., Yip, G. N., Kurochkin, A. V. and Zuiderweg, E. R. (2005) NMR investigations of allosteric processes in a two-domain *Thermus thermophilus* Hsp70 molecular chaperone. *J. Mol. Biol.* **349,** 163-183

- Richter, K. and Buchner, J. (2001) Hsp90: chaperoning signal transduction. *J. Cell. Physiol.* **188**, 281–290

- Rist, W., Graf, C., Bukau, B. and Mayer, M. P. (2006) Amide hydrogen exchange reveals conformational changes in hsp70 chaperones important for allosteric regulation. *J. Biol. Chem.* **281**, 16493–16501

- Ritossa, F. (1962) A new puffing pattern induced by temperature shock and DNP in *Drosophila Experientia* **18**, 571-573

- Ritossa, F. (1963) New puffs induced by temperature shock, DNP and salicilate in salivary chromosomes of *D. melanogaster. Drosophila Infor. Serv.* **37**, 122-123

- Ritossa, F. (1964) Expermental activation of specific loci in polytene chromosomes of *Drosophila. Exp. Cell. Res.* **35**, 601-607

- Ritossa, F. (1996) Discovery of the heat shock response. *Cell Stress Chaperones* **1**, 97-98

- Roseman, A. M., Chen, S., White, H., Braig, K., and Saibil, H. R. (1996) The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* **87,** 241–251

- Ruiz, N., Daniel Kahne, D. and Silhavy, T. J. (2006) Advances in understanding bacterial outer-membrane biogenesis. *Nat. Rev. Microb.* **4**, 57-66

- Rutherford, S. L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342

- Ryan, M. T. and Pfanner, N. (2002) Hsp70 proteins in protein translocation. *Adv. Protein Chem.* **59**, 223–242

- Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. and Horwich, A. L. (1999) GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97,** 325-338

- Saibil, H. R., Zheng, D., Roseman, A. M., Hunter, A. S., Watson, G. M., Chen, S., Auf Der Mauer, A., O'Hara, B. P., Wood, S. P., Mann, N. H., Barnett, L. K. and Ellis, R. J. (1993). ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Curr. Biol.* **3**, 265-273

- Saito, H. and Uchida, H. (1977) Initiation of the DNA replication of bacteriophage lambda in *Escherichia coli* K12. *J. Mol. Biol.* **113,** 1-25

- Sakikawa, C., Taguchi, H., Makino, Y. and Yoshida, M. (1999) On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J. Biol. Chem.* **274**, 21251–21256

- Scharf, K. D., Siddique, M. and Vierling, E. (2001) The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α-crystallin domains (ACD proteins). *Cell Stress Chaperones* **6**, 225–237

- Schirmer, E. C., Glover, J. R., Singer, M. A. and Lindquist, S. (1996) Hsp100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.* **21**, 289–296

- Schmid, D., Baici, A., Gehring, H. and Christen, P. (1994) Kinetics of molecular chaperone action. *Science* **263**, 971-973

- Schumann, W. (2003) The *Bacillus subtilis* heat shock stimulon. *Cell Stress Chaperones* **8,** 207-217

- Schumann, W. (2007) Thermosensors in eubacteria: role and evolution. *J. Biosci.* **32,** 549–557

- Servant, P. and Mazodier, P. (1996) Heat induction of *hsp18* gene expression in *Streptomyces albus* G: Transcriptional and posttranscriptional regulation. *J. Bacteriol.* **178**, 7031–7036

- Servant, P. and Mazodier, P. (2001) Negative regulation of the heat shock response in Streptomyces. *Arch. Microbiol.* **176,** 237–242

- Servant, P., Rapoport, G., Mazodier, P. (1999) RheA, the repressor of *hsp18* in *Streptomyces albus* G. *Microbiology*, **145**, 2385–2391

- Servant, P. and Mazodier, P. (1995) Characterization of *Streptomyces albus* 18-kilodalton heat shock-responsive protein. *J. Bacteriol.* **177**, 2998–3003

- Sitia, R. and Molteni, S. N. (2004) Stress, Protein (Mis)folding, and Signaling: The Redox Connection. *Sci. STKE* **2004**, 27

- Stewart, G. R., Robertson, B. D. and Young, D. B. (2003) Tuberculosis: a problem with persistence. *Nat. Rev. Microbiol.* **1**, 97-105

- Stewart, G. R., Snewinm V. A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I. N. and Young, D. B. (2001) Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nat. Med.* **7**, 732-737

- Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Butcher, P. D. and Young, D. B. (2002) The heat shock response of *Mycobacterium tuberculosis*: linking gene expression, immunology and pathogenesis. *Comp. Funct. Genomics* **3**, 348-351

- Storz, G., Opdyke, J. A. and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.* **7**, 140–144

- Straus, D. B., Walter, W. A. and Gross, C. A. (1987) The heat shock response of *E. coli* is regulated by changes in the concentration of sigma 32. *Nature* **329**, 348-351

- Straus, D. B., Walter, W. A. and Gross, C. A. (1990) DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes Dev.* **4**, 2202-2209

- Sunshine, M., Feiss, M., Stuart, J. and Yochem, J. (1977) A new host gene (*groPC*) necessary for lambda DNA replication. *Mol. Gen. Genet.***151**, 27-34

- Swain, J. F., Dinler, G., Sivendran, R., Montgomery, D. L., Stotz, M. and Gierasch, L. M. (2007) Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol. Cell* **26,** 27-39

- Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B., and Hartl, F. U. (1994) The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system-DnaK, DnaJ and GrpE. *Proc. Natl. Acad. Sci. USA* **91**, 10345–10349

- Takano, T. and Kakefuda, T. (1972) Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. *Nat. New. Biol.* **239**, 34-37

- Tanaka, K., Muramatsu, S., Yamada, H. and Mizuno, T. (1991) Systematic characterization of curved DNA segments randomly cloned from *Escherichia coli* and their functional significance. *Mol. Gen. Genet.* **226,** 367–376

- Tang, Y., Chang, H., Hayer-Hartl, M. and Hartl, F. U. (2007) SnapShot: Molecular Chaperones, Part I. *Cell* **128**, 412-412.e1

- Tatsuta, T., Tomoyasu, T., Bukau, B., Kitagawa, M., Mori, H., Karata, K. and Ogura, T. (1998) Heat shock regulation in the *ftsH* null mutant of *Escherichia coli*: dissection of stability and activity control mechanisms of sigma32 in vivo. *Mol. Microbiol.* **30,** 583-593

- Tilly, K., McKittrick, N., Zylicz, M. and Georgopoulos, C. (1983) The dnaK protein modulates the heat-shock response of *Escherichia coli*. *Cell* **34**, 641-646

- Toft, D. O. (1999) Control of hormone receptor function by molecular chaperones and folding catalysts. In: Molecular Chaperones and Folding Catalysts. Regulation, Cellular Functions. *Cell. Mol. Life Sci.* **62**, 313-328

- Tomoyasu, T., Ogura, T., Tatsuta, T. and 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**, 567-581

- Ueno, T., Taguchi, H., Tadakuma, H., Yoshida, M. and Funatsu, T. (2004) GroEL mediates protein folding with a two successive timer mechanism. *Mol. Cell* **14,** 423-434

- Veinger, L., Diamant. S., Buchner, J. and Goloubinoff, P. (1998) The small heat-shock protein IbpB from *E. coli* stabilizes stress-denatured proteins for subsequent refolding by a multi chaperone network. *J. Biol. Chem.* **273**, 11032–11037

- Vierling, E. (1991) The roles of heat shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**, 579–620

- Vogel, M., Mayer, M. P. and Bukau, B. (2006) Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. *J. Biol. Chem.* **281**, 38705–38711

- Waldminghaus, T., Gaubig, L. C., Klinkert, B. and Narberhaus, F. (2009) The *Escherichia coli* ibpA thermometer is comprised of stable and unstable structural elements. *RNA Biol.* **6**

- Wang, Y., Kelly. C. G., Karttunen, J. T., Whittall, T., Lehner, P. J., Duncan, L., MacAry, P., Younson, J. S., Singh, M., Oehlmann, W., Cheng, G., Bergmeier, L. and Lehner, T. (2001)

CD40 is a cellular receptor mediating mycobacterial heat shock protein 70 stimulation of CC-chemokines. *Immunity* **15,** 971-983

- Waters, E. R., Lee, G. J. and Vierling, E. (1995) Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* **47**, 325–338

- Weber-Ban, E. U., Reid, B. G., Miranker, A. D. and Horwich, A. L. (1999) Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90–93

- Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. and Horwich, A. L. (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **83**, 577–587

- Wetzstein, M., Völker, U., Dedio, J., Löbau, S., Zuber, U., Schiesswohl, M., Herget, C., Hecker, M. and Schumann, W. (1992) Cloning, sequencing, and molecular analysis of the *dnaK* locus from *Bacillus subtilis. J. Bacteriol.* **174**, 3300–3310

- Whittall, T., Wang, Y., Younson, J., Kelly, C., Bergmeier, L., Peters, B., Singh, M. and Lehner, T. (2006) Interaction between the CCR5 chemokine receptors and microbial HSP70. *Eur. J. Immunol.* **36**, 2304-2314

- Wickner, S., Maurizi, M. R. and Gottesman, S. (1999) Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins. *Science* **286**, 1888-1893

- Xu, Z., Horwich, A. L. and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)$_7$ chaperonin complex. *Nature* **388**, 741-750

- Young, J. C., Agashe, V. R., Siegers, K. and Hartl, F. U. (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat. Rev. Mol. Cell. Biol.* **5**, 781-791

- Young, J. C., Barral, J. M. and Hartl, F. U. (2003) More than folding: localized functions of cytosolic chaperones. *Trends. Biochem. Sci.* **28**, 541–547

- Young, J. C., Moarefi, I. and Hartl, F. U. (2001) Hsp90: a specialized but essential protein folding tool. *J. Cell Biol.* **154**, 267–273

- Young, L. S., Eliopoulos, A. G., Gallagher, N. J. and Dawson, C. W. (1998) CD40 and epithelial cells: across the great divide. *Immunol. Today* **19,** 502-506

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

- Zhang, Y. and Zuiderweg, E. R. (2004) The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. *Proc. Natl. Acad. Sci. USA* **101**, 10272-10277

- Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. and Hendrickson, W. A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606–1614

- Zuber, U. and 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

In vivo Complementation Studies on Mycobacterial GroEL Homologues

2.1 Introduction

The distinctive feature of the GroES-GroEL chaperonin system in mediating protein folding lies in its ability to exist in a tetradecameric state, form a central cavity and encapsulate the substrate via GroES lid (Bukau and Horwich, 1998). However, recombinant GroELs of *Mycobacterium tuberculosis* (Mtb), GroEL1 and GroEL2 exhibited biochemical features that deviated significantly from the trademark properties of *E. coli* GroEL. Contrary to expectations, when purified from *E. coli*, these recombinant Mtb GroELs existed as lower oligomers (dimers), displayed weak ATPase activities and GroES independence in preventing aggregation of the denatured polypeptides (Qamra and Mande, 2004; Qamra et al., 2004). This atypical behaviour of Mtb GroELs, despite high sequence homology with the well characterized *E. coli* counterpart, could be attributed to two fundamental reasons. Firstly, the intrinsic factor: these chaperones might not be active under the conditions tested or secondly, the extrinsic factor: the chaperones might have lost their activity during the course of purification owing to various grounds including the heterologus expression system. Hence, in order to elucidate the functional role of Mtb GroELs in vivo, activity of Mtb GroELs as chaperones was studied by scoring for the extent these chaperones are able to rescue the Ts phenotype conferred by the *groEL44* allele in *E. coli* SV2.

E. coli SV2 is a strain that depends on mutant GroEL (E191G), the GroEL44, for its growth (Georgopoulos et al., 1972). The mutant strain was isolated initially as an allele blocking the growth bacteriophages λ and T4 (Georgopoulos et al., 1972, Georgopoulos and Hohn, 1978). Additionally, biochemical analysis of the encoded mutant chaperonin, GroEL44, showed that this chaperonin is defective in releasing substrates and is compromised in inter-subunit interactions at elevated temperatures (Richardson et al., 1999, Ang et al., 2001). Hence, the strain *E. coli* SV2 grows well at ambient temperatures but not at elevated temperatures, unless a functional GroEL is supplemented.

2.2 Materials

Molecular biology procedures employed in this study were performed according to the standard protocols (Sambrook et al., 2004). All chemicals, enzymes for biochemical analysis and antibiotics were purchased from Sigma Inc and Amersham Biosciences. Antibodies IT3 and IT56 were procured via TB Vaccine Testing and Research Materials Contract, Colorado State University, USA, and α-Cpn60.1$_{Mtb}$ was a kind gift from Professor Anthony R. M. Coates, St George's Medical Hospital School, UK. Mtb genomic DNA library was a kind gift from Stewart Cole (Cole et al., 1998). Oligonucleotides were procured from Sigma Inc. and MWG Oligos. Different DNA purification kits were procured from Qiagen (Chatsworth CA). Restriction endonucleases, T4 DNA ligase, and protein and DNA markers were purchased from New England Biolabs (Beverly, MA) and Fermentas MBI (Glen Burnie, MD). Reverse transcription PCR kit (SuperScript™ Reverse Transcriptase) and DNaseI were procured from Invitrogen and Amersham Biosciences respectively. Ni-NTA-agarose was purchased from Qiagen Inc. Q-sepharose and pre-packed FPLC columns were from Amersham Biosciences and filtration devises were from Milliopre Inc. Nucleic acid amplifications were performed on GeneAmp® PCR System 2700 (ABI, Foster City, CA).

All reagents were prepared in de-ionised double distilled water. Compositions of the solutions employed for agarose gel electrophoresis, SDS-PAGE and immunoblotting used are as in Tables 2.01, 2.02 and 2.03, respectively.

Reagents	Composition
50X TAE	242 g Tris base +57.1 ml of glacial acetic acid + 100 ml of 0.5 M EDTA per ltr DDW
6X sample loading dye	0.6% Orange-G in 30% glycerol
Ethidium bromide	Stock of 10mg/ml in DDW

Table 2.01: Composition of solutions used for Agarose gel electrophoresis.

Reagents	Composition
30% Acryamide	29.2% acryamide + 0.8% bis-acryamide
4X Separation gel buffer	1.5 M Tris.HCl (pH: 8.8) + 0.4% SDS
4X Stacking gel buffer	1 M Tris.HCl (pH: 6.8) + 0.4% SDS
1X Laemmli sample buffer	10% glycerol +1% β-mercaptoethanol + 2% SDS + 0.1% bromophenol blue in 1X separating buffer
1X Running buffer	3 g Tris.HCl +14.4 g glycine +1 g SDS per liter
Destaining solution	Methanol: Acetic acid: Water :: 5:1:4
Staining solution	0.1 g/ltr of Coomassie brilliant blue R250 in de-staining solution

Table 2.02: Composition of solutions used for SDS PAGE

Reagents	Composition
Electrode Transfer Buffer	12 mM Tris base, 96 mM Glycine, 20% Methanol
Ponceau S solution	0.5 % Ponceau S (w/v) in 1 % Acetic acid (v/v) in water
Tris Buffered Saline (TBS)	100 mM Tris·Cl, (pH 7.5) and 0.9% NaCl
Tris Buffered Saline with Tween (TBST)	0.1% Tween 20 in Tris Buffered saline
Blocking Solution	5 % Fat free milk or 2% BSA in TBST
Developer Solution	Purchased from Kodak
Fixer Solution	Purchased from Kodak

Table 2.03: Composition of solutions used for immunoblotting

2.2.1 Bacterial Strains and Growth Conditions

The strain *E. coli* SV2 is a *groEL44* derivative of the *E. coli* K12 strain B178 (*galE groEL+*) (Georgopoulos et al., 1972). Since strain *E. coli* SV2 has a tendency to revert to the wild type when cultured at 37 °C, it was always maintained at 30 °C, unless otherwise mentioned. *E. coli* LG6 is a derivative of MG1655 wherein chromosomal *groES/L* operon is placed downstream of the lactose/IPTG inducible P_{lac} promoter (Horwich et al., 1993). These strains were kind gifts from Sir Alan Fersht, UK and Dr. Arthur Horwich, USA, respectively. Coliphages λ*c1B2* and T4GT7 starter particles were sourced from laboratory stocks. Plasmids and oligonucleotide primers used in this study are listed in Tables 1 & 2, respectively, in Appendix I. All media and buffer solutions used were prepared in de-ionized double distilled water (DDW) with low conductivity. Compositions of the media and growth supplements used are listed in Tables 2.04 and 2.05 respectively.

Medium	Composition
Luria Bertani (LB)	10 g bactotryptone + 5 g yeast extract + 10 g NaCl per ltr DDW. pH was adjusted to 7.2 with NaOH. The medium was sterilized by autoclaving.
Terrific Broth (TB)	**A.** 12 g bactotryptone + 24 g yeast extract + 4 ml glycerol in 900 ml DDW **B.** 170 mM KH_2PO_4 + 720 mM K_2HPO_4 A and B were autoclaved separately and mixed at the time of innoculation.

Table 2.01: Composition of culture media

Supplement	Stock Concentration	Working Concentration
Ampicillin	100 mg/ml in DDW	100 µg/ml
Kanamycin	30 mg/ml in DDW	30 µg/ml
Chloramphenicol	12.5 mg/ml in ethanol	12.5 µg/ml
Tetracycline	7 mg/ml in ethanol	7 µg/ml
D-glucose	20% in Water	0.2%
L-arabinose	20% in Water	0.2%
Maltose	20% in Water	0.4%
$MgCl_2$	1 M in DDW	5 mM
$CaCl_2$	1 M in DDW	5 mM

All the supplements were sterilized by passing through 0.2 µm filter.

Table 2.02: Stock and working concentrations of growth supplements used in this study

2.2.2 Preparation of Bacteriophage Particles

In this study different coliphages such as T4GT7 and λ*cIb2* were employed for studying functional aspects of GroEL and GroES molecules. The phages were prepared from the wild type *E. coli* strain, MG1655. Fully grown culture of *E. coli* MG1655 was diluted to 1% into 10 ml of LB broth and was incubated at 37 °C with vigorous shaking for 1 to 2 hours. The growth media at this stage was supplemented depending on the phage to be isolated. For the preparation of bacteriophage λ*cIb2,* the media was supplemented with 0.4% maltose since maltose induces the expression of maltose operon and thereby the *lamB* gene, which encodes a cell surface receptor to which bacteriophage phage, λ binds. For the propagation of bacteriophage P1, the media was supplemented with 5 mM $CaCl_2$ since divalent Ca^{++} is required for adsorption of the bacteriophage to host cells. Early-log phage culture was recovered (OD = 0.3 - 0.4) and was infected with different concentrations of phage particles ranging from 10^4 to 10^6 and incubated further for 1 to 2 hours till the culture becomes clear. A few drops of chloroform were added to the lysate, which would disrupt the membranes and complete the cell lysis. Cell debris was removed by centrifugation at 8000 rpm for 10 – 15 min. Taking care not to transfer the debris, cleared supernatant which has the bacteriophage particles was transferred to a sterile Falcon tube. To the lysate, to keep it sterile, a few drops of chloroform (about 300 μl) were added, mixed by agitation and stored at 4 °C. Titer of the prepared bacteriophages was determined by mixing about 1 × 10^3 phage particles and 100 μl of *E. coli* culture in 4 ml soft agar (0.5% agar) and overlaying onto LB plates supplemented with 5 mM $CaCl_2$ and 5 mM $MgSO_4$ and the plates incubated overnight at 37 °C. Number of plaques was counted and total phage particles per ml of phage preparation were calculated.

2.2.3 In vitro Site Directed Mutagenesis

In vitro site directed mutagenesis, a technique which has been employed during the project, is generally carried out to introduce specific mutations such as base changes, deletions, or insertions. This can be carried out into a target gene or region cloned into a plasmid. This technique is effective in studying protein structure-function relationships and gene expression, and for carrying out vector modification. Site directed mutagenesis requires a supercoiled doublestranded DNA (dsDNA) vector with

an insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The desired mutation is designed into the mutagenic primers and should be in the middle of the primer with ~10–15 bases of correct sequence on both sides and therefore these primers should be between 25 and 45 bases in length, with a melting temperature (Tm) of ≥78 °C.

The entire site directed mutagenesis was carried out using the Quickchange™ kit purchased from Promega Inc. This technique involves three steps: a. Mutant Strand Synthesis by PCR; b. DpnI digestion of the parental DNA; c. Transformation into *E. coli*. PCR was set up by mixing 0.5 µM of forward and reverse primers, 25 ng of template DNA, 200 µM of each dNTPs and 2 U of Phusion (Finzymes Inc.) in 1X reaction buffer (10 mM KCl, 10 mM(NH$_4$)$_2$SO$_4$, 20 mM Tris.HCl (pH 8.8), 1.5 mM MgCl$_2$, 1% Triton X-100 and 1 mg/ml nuclease-free BSA). Final volume of PCR reaction was set to 50 µl. The cycles for PCR were set as follows: Initial denaturation of 2 min at 96 °C, followed by 22-25 cycles of denaturation for 10-20 seconds at 96 °C, annealing for 30 seconds at 65-72 °C and extension for 15-30 s/1 kb template at 72 °C. The PCR products thus generated were treated with 10 U DpnI for 2 hours at 37 °C and then 10 µl of the digest was transformed into *E. coli* Top10 cells and were selected on appropriate antibiotic selection. The resulting clones were confirmed for the incorporated mutation by sequencing.

2.3 Cloning the Genes Encoding E. coli and Mtb Chaperonin Homologues

For studying the ability of GroELs from *M. tuberculosis* in complementing for *E. coli* GroEL, expressing the genes encoding these chaperonins from a tightly regulated expression system is required. The L-arabinose inducible P_{BAD} promoter system is known to be tightly regulated and therefore, we have chosen the pBAD plasmid system for the complementation study (Guzman et al., 1995). To this end, the ORFs encoding Mtb GroEL1, GroEL2 and *E. coli* GroEL were cloned in an operonic arrangement with that of their cognate co-chaperonin, GroES under the arabinose inducible P_{BAD} promoter.

2.3.1 Amplifying the ORFs Encoding Mtb and *E. coli* Chaperonins

The ORFs encoding Mtb *groES, groEL1* were amplified from Mtb genomic DNA library (Cole et al., 1998) and *E. coli groEL/S* operon from the genomic DNA of the strain MG1655 using primer pairs SCM1601F/SCM1601R, SCM1602F/SCM1602R and SCM03F/SCM03R, respectively (Figure 2.01). Amplification was carried out by Dynazyme Ext Polymerase (Finzyme Inc.) and the conditions are given in the Table 2.06.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94 °C	300	1
Denaturation	94 °C	60	
Annealing	48 °C	30	5
Extension	72 °C	240	
Denaturation	94 °C	60	
Annealing	50 °C	30	8
Extension	72 °C	240	
Denaturation	94 °C	60	
Annealing	52 °C	30	10
Extension	72 °C	240	
Denaturation	94 °C	60	
Annealing	55 °C	30	17
Extension	72 °C	240	
Final Extension	72 °C	900	1

Table 2.06: PCR cycling conditions for amplifying Mtb *groES*, *groEL1* and *E. coli groESL*.

Amplified DNA fragments were resolved on 1% agarose gel and the fragments corresponding to the desired molecular mass were extracted from the gel (Figure 2.01).
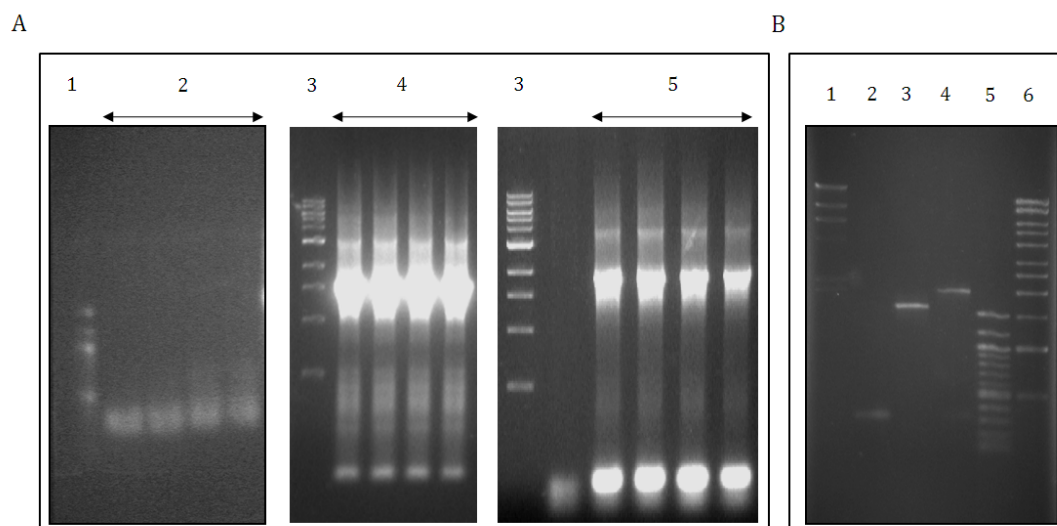
Figure 2.01: PCR Amplification and Extraction of the DNA Fragments. **A.** Mtb *groES*, *groEL1* and *E. coli groES/L* were amplified by PCR were separated by 1% agarose gel and the resulting PCR products were compared with 100 bp marker (NEB). The lanes correspond to: 1, 100 bp marker; 2, PCR for Mtb *groES*; 3, 1 kb marker; 4, PCR for Mtb *groEL1* and 5, PCR for *E. coli groES/L.* **B.** The PCR products were extracted from 1% agarose gel using Qiaquick gel extraction kit (Qiagen Inc.) and were resolved on 1% agarose gel. The lanes correspond to: 1, λ HindIII digest; 2, Mtb *groES*; 3, Mtb *groEL1*; 4, *E. coli groES/L*; 5, 100 bp marker and 6, 1 kb marker.

2.3.2 Cloning the Genes Encoding Chaperonins onto pBAD24

The amplified DNA fragments were cloned under the arabinose inducible P_{BAD} promoter. The plasmids harboring P_{BAD} promoter, pBAD24 and pBAD18 were used for cloning the amplified chaperonin genes (Guzman et al., 1995).

2.3.2.1 Cloning ORF Encoding Mtb GroES

ORF encoding Mtb GroES was cloned into NcoI and SmaI sites on the vector pBAD24 and the resulting plasmid was designated pSCM1600. Two clones of Mtb *groES* in pBAD24 were confirmed by digestion with restriction endonucleases ClaI and HindIII, which have recognition sites upstream and downstream, respectively of the inserted *groES* gene. This would result in a fragment of 1.4 kb for pBAD24 and 1.7 kb if *groES* is inserted (Figure 2.02). Similar digestion with NcoI and SmaI resulted in a fragment of 300 bp.
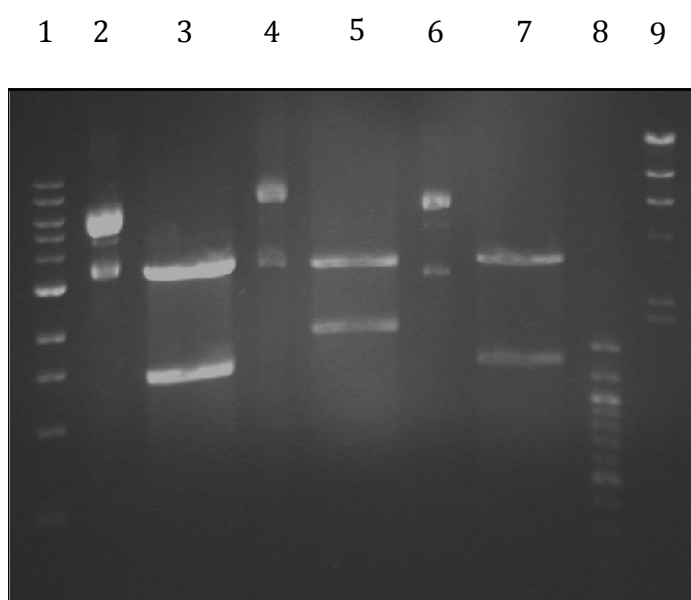
Figure 2.02: Confirmation of Clones Encoding Mtb GroES in pBAD24. Mtb *groES* gene was cloned into NcoI and SmaI sites of pBAD24. Two of the resulting clones were digested with restriction endonucleases ClaI and HindIII and the resulting fragments were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb marker; 2, Clone 1 undigested; 3, Clone 1 digested; 4, Clone 2 undigested; 5, Clone 2 digested 6, pBAD24 undigested; 7, pBAD24 digested; 8, 100 bp marker and 9, λ HindIII digest.

2.3.2.2 Cloning ORFs Encoding Mtb GroEL1 and GroEL2 into pSCM1600

A ribosomal binding site (RBS) was designed into the forward primer of Mtb *groEL1* gene and Mtb *groEL2* gene was excised out from an earlier plasmid pSCM1000 with an optimally placed RBS. Into pSCM1600, Mtb *groEL1* was cloned into SmaI and XbaI sites and *Mtb groEL2* was cloned into XbaI and HindIII sites to generate pSCM1602 and pSCM1603, respectively. Clones of harboring Mtb *groEL1* were confirmed by colony PCR and the clones harboring Mtb *groEL2* were confirmed by digesting with restriction endonucleses XbaI and HindIII (Figure 2.03).
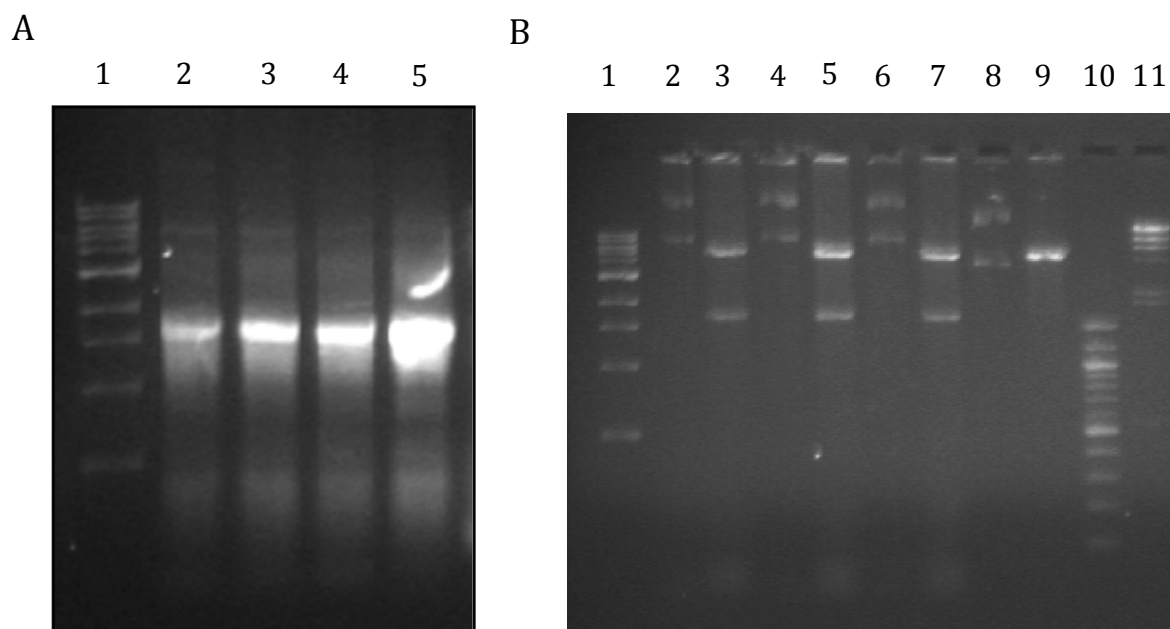
***Figure 2.03: Confirmation of Insertion of Genes Encoding Mtb GroELs.* A.** Mtb *groEL1* gene was cloned into SmaI and XbaI sites of pSCM1600. The resulting clones were confirmed by colony PCR using purified plasmid DNA. The lanes correspond to: 1, 1 kb marker; 2-5 PCR products from the clones encoding Mtb GroEL1. **B.** Mtb *groEL2* gene was cloned into XbaI and HindIII sites on pSCM1600. Three of the resulting clones were digested with restriction endonucleases XbaI and HindIII and the digests were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb marker; 2, Clone 1 undigested; 3, Clone 1 digested; 4, Clone 2 undigested; 5, Clone 2 digested 6, Clone 3 Undigested; 7, Clone 3 digested; 8, pSCM1600 undigested; 9, pSCM1600 digested; 10, 100 bp marker and 11, λ HindIII digest.

2.3.2.3 Cloning ORF Encoding *E. coli* GroES/L into pBAD24

The fragment encoding *E. coli groEL/S* was cloned into NcoI and SmaI sites on pBAD24 to generate pSCM1601. The clones were confirmed by digestions with restriction endonucleases NcoI and HindIII (Figure 2.04)
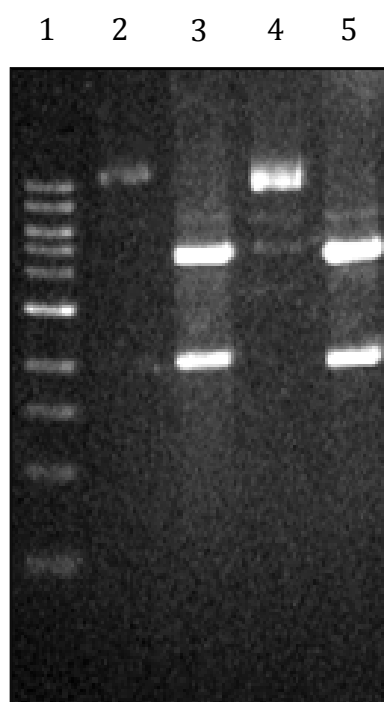
Figure 2.04: Confirmation of Clones Harboring E. coli groES/L in pBAD24. The *E. coli groES/L* operon was cloned into NcoI and HindIII sites on pBAD24. Two of the resulting clones were digested with restriction endonucleases NcoI and HindIII and the resulting fragments were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb marker; 2, Clone 1 undigested; 3, Clone 1 digested; 4, Clone 2 undigested and 5, Clone 2 digested.

2.3.2.4 Cloning ORF Encoding Mtb and *E. coli* GroELs into pBAD24

Additionally, ORFs encoding *E. coli* GroEL, Mtb GroEL1, and GroEL2 were cloned independent of cognate *groES*, under P_{BAD} promoter. For cloning Mtb *groEL1*, pBAD24 was site specifically modified to accommodate NdeI site in place of NcoI site, using primers SCM1604F/SCM1604R and the resulting plasmid was designated as pBAD25 (Figure 2.05).
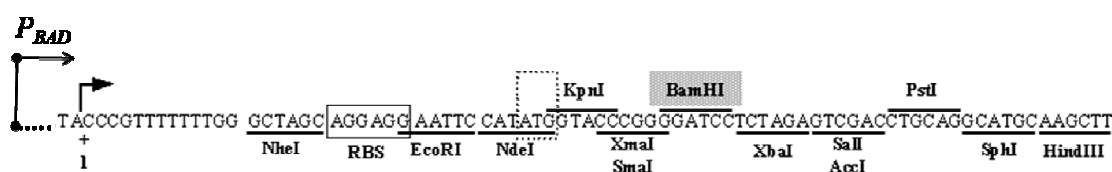
Figure 2.05: Cartoon Showing MCS of pBAD25. Recognition site for the restriction endonuclease NcoI on pBAD24 is altered site specifically to that for NdeI. BamHI site is shaded since it is double.

Mtb *groEL1* was amplified using primers SCM1605F and SCM1605R and cloned into NdeI and SmaI sites of pBAD25 to obtain plasmid pSCM1604. The XbaI and HindIII fragment from pSCM1000 bearing Mtb *groEL2* was cloned into the XbaI and HindIII sites of pBAD18 to generate plasmid pSCM1605 in which *groEL2* ORF is preceded with an optimally placed RBS. *E. coli groEL* was amplified using primers SCM1606F and SCM1603R and cloned into the NcoI and HindIII sites of pBAD24 to generate plasmid pSCM1608. The clones were confirmed by digestions with restriction endonucleases (Figure 2.6).

All clones were confirmed by comparing digestion patterns by restriction endonuclease reactions and by automated sequencing. Primers PBADFOR and PBADREV were used for sequencing (Guzman et al., 1995). Growth media during the course of cloning was supplemented with 0.2% D-glucose to maintain the arabinose inducible P_{BAD} promoter repressed and thereby prevent the untimely expression of cloned genes.
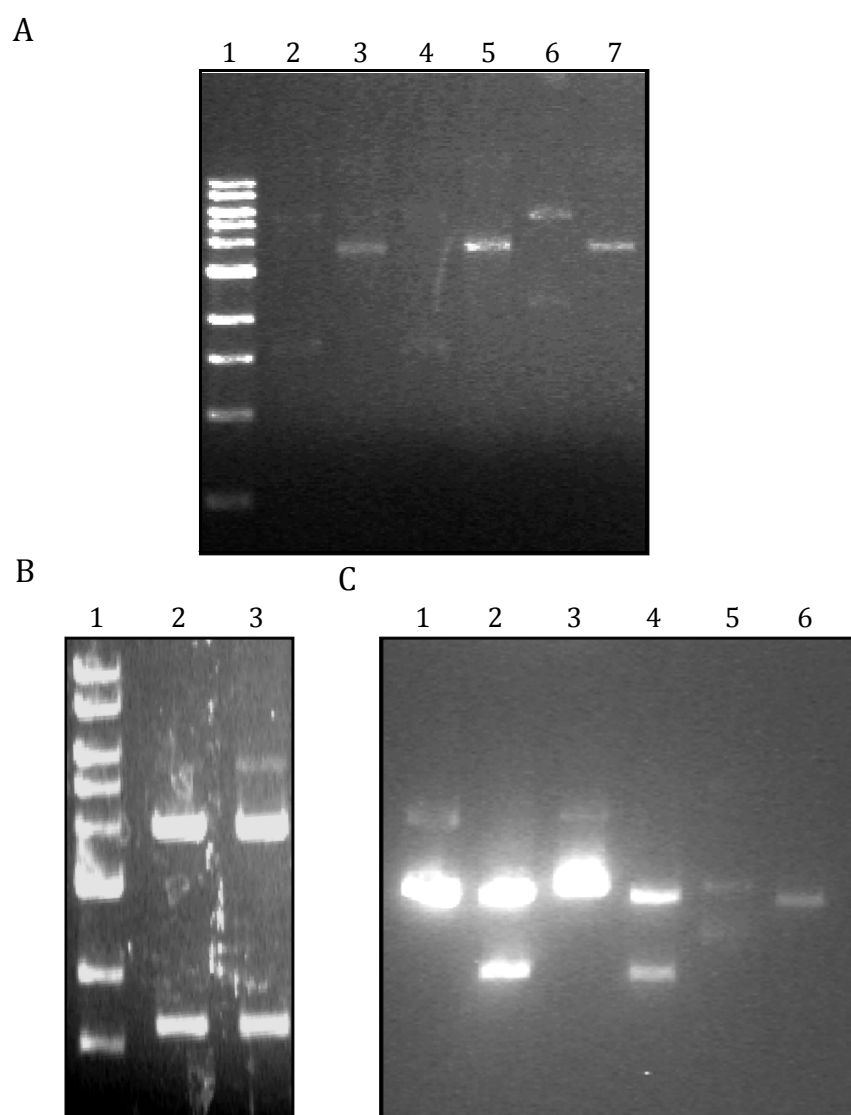
Figure 2.06: Confirmation of Clones Harboring groEL Homologues in pBAD24. The ORFs encoding *E. coli GroEL*, Mtb GroEL1 and GroEL2 were cloned under P_{BAD} control. *E. coli groEL* into NcoI and HindIII sites on pBAD24 **(A)**, Mtb *groEL1* was cloned into SmaI and XbaI on pBAD18 **(B)** and Mtb *groEL2* into NcoI and HindIII sites on pBAD24 **(C)**. Two of the resulting clones were digested with restriction appropriate endonucleases and the resulting fragments were resolved on 1% agarose gel. The lanes correspond to: **A,** 1, 1 kb marker; 2, Clone 1 digested; 3, Clone 1 undigested; 4, Clone 2 digested; 5, Clone 2 undigested; 6, pBAD24 digested and 7, pBAD24 undigested; **B,** 1, 1 kb marker; 2, Clone 1 digested and 3, Clone 2 digested; **C,** 1, Clone 1 undigested; 2, Clone 1 digested; 3, Clone 2 undigested; 4, Clone 2 digested; 5, pBAD24 undigested and 6, pBAD24 digested.

2.4 Complementation Studies on Mtb GroEL Homologues

In vivo function of Mtb GroELs was assayed by three methods:

2.4.1 Assaying the Ability to Rescue Ts Phenotype of *groEL44* Allele

Earlier biochemical and biophysical studies, as discussed in Chapter 1 in detail, had shown that recombinant Mtb GroEL homologues when purified from *E. coli*, exhibited unusual structural and biochemical characteristics (Qamra and Mande, 2004 and Qamra et al., 2004). These observations could be attributed to either inherent property of these proteins which drives them to be deficient in exhibiting typical chaperonin characteristics or to the method employed for the purification of these proteins, which might alter the otherwise active chaperonins into irreversibly inactive chaperonins under the conditions tested. To test the feasibility of these two probabilities, we wished to test if these proteins function in vivo, in a way similar to that exhibited by *E. coli* GroEL. We have employed *E. coli* strain, SV2 that bears a temperature sensitive *groEL44* allele (Georgopoulos and Hohn, 1978). Hence the strain does not grow at elevated temperatures unless a functional GroEL is supplemented. We have chosen this strain to test the in vivo function of the mycobacterial GroELs.

We tested if the two Mtb GroEL homologues could complement the Ts phenotype of the *groEL44* allele. The plasmids that could co-express *groEL* homologues along with their cognate *groES* were employed for this assay. Plasmids encoding the Mtb *groEL1/groES* (pSCM1602), Mtb *groEL2/groES* (pSCM1603) and *E. coli groEL/groES* (pSCM1601) were transformed into *E. coli* SV2 and the transformants were cultured in the presence of 0.2% D-glucose. Stationary phase cultures of *E. coli* strain SV2 containing plasmids pSCM1601, pSCM1602, pSCM1603 and pBAD24 were serially diluted. 5 μl of each serially diluted cultures were spotted onto the surface of LB agar supplemented with either 0.2% L-arabinose to induce the expression of cloned *groEL* genes or 0.2% D-glucose to repress P_{BAD} promoter. The plates were incubated at permissive and restrictive temperatures 30 °C and 42 °C, respectively. Ability to rescue the Ts growth phenotype by the cloned *groEL* genes was estimated by colony forming ability conferred by the *groEL* homologue at the restrictive temperature.

Expression of neither Mtb *groEL1* nor Mtb *groEL2*, from the arabinose inducible P_{BAD} promoter in *E. coli* SV2, could rescue the Ts phenotype associated with the allele

groEL44, even when these Mtb *groELs* were co-expressed with their cognate *groES*. Notably, expression of *E. coli groES/L*, which was employed as a positive control in these experiments, could comfortably rescue the Ts phenotype associated with *groEL44* allele (Figure 2.07). Similar results were obtained when the *groEL* homologues were expressed unaccompanied by the cognate *groES*. Results pertaining to the complementation studies on GroEL homologues alone will be presented at later stages in this book.
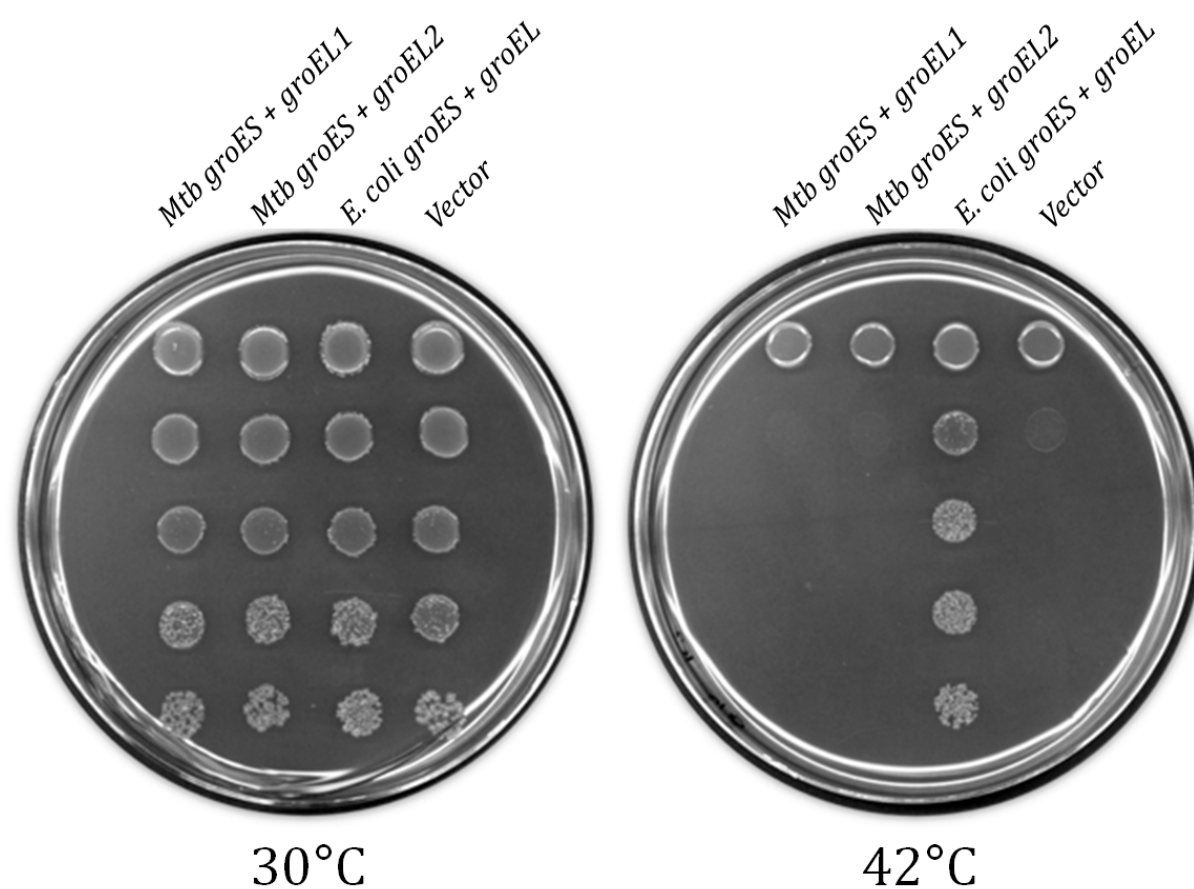
Figure 2.07: Complementation by Mtb GroEL Homologues. Serially diluted cultures of the *E. coli* strain SV2 (*groEL44*) expressing the indicated *groEL/S* genes were spotted onto the surface of LB agar plates supplemented with 0.2% L-arabinose and the plates were incubated at the permissive (30 °C) and restrictive (42 °C) temperatures, as indicated.

2.4.2 Assaying the Ability to Support Bacteriophage Morphogenesis

In *E. coli*, development of bacteriophages like lambda (λ) and T4 requires functional GroEL/S system, specifically for the development and assembly of head portion of bacteriophage particles (Georgopoulos and Hohn, 1978). A functional GroEL could thus support bacteriophage morphogenesis which could be scored as the ability to plaque on the lawn of bacterial culture expressing particular *groEL* homologue. To this end, ability of the Mtb GroEL homologues in supporting bacteriophage morphogenesis was studied as an extension of the complementation of the *groEL44* Ts allele at elevated temperatures.

E. coli SV2 expressing *groEL* homologues as above were grown in LB broth supplemented with 0.4% maltose. The cultures were recovered in stationary phase and were washed with LB broth to remove the residual maltose. These cultures were individually mixed with about 4 ml of LB soft agar (0.5% agar) and were overlayed onto the surface of LB agar plates that were previously supplemented with 0.2% L-arabinose and 5 mM $MgSO_4$. The phage preparations were serially diluted in LB broth supplemented with 5 mM $MgSO_4$ and 5 µl of each dilutions were spotted onto the surface of the overlayed plates. The plates were incubated at 30 °C. Formation of plaques on bacterial lawn was scored as the ability to support bacteriophage morphogenesis.

In agreement with the ability to complement the *groEL44* allele, neither Mtb GroEL1, nor Mtb GroEL2 could promote the development of bacteriophages either λ or T4 in the strain SV2 where as *E. coli* GroEL was able to support morphogenesis of both the phages (Figure 2.08).
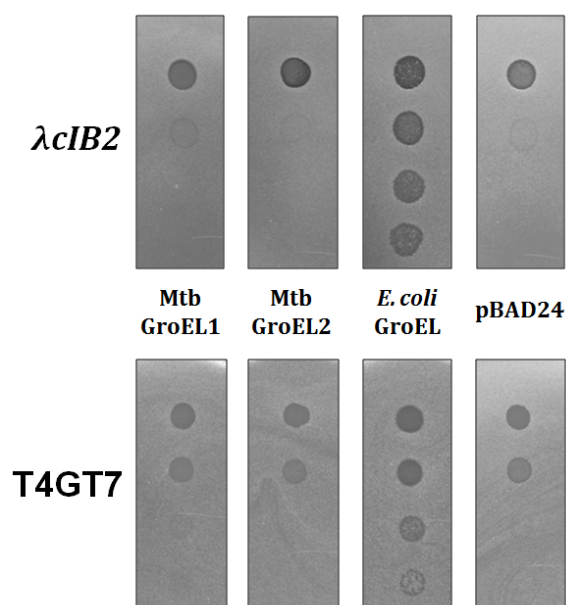
Figure 2.08: Assay for Bacteriophage Morphogenesis. Lawns of the *E. coli* strain SV2 expressing the indicated *groEL* genes were prepared on LB agar plates supplemented with 0.2% L-arabinose. Hundred-fold serial dilutions of bacteriophages *λcIB2* and T4GT7 were spotted on the lawns followed by incubation at 30 °C

2.4.3 Immunoblotting for Detecting the Expression of *groEL/S*

The above two experiments confirmed that the two GroELs of Mtb are not able to function as chaperones in *E. coli* milieu. This observation could either be because these recombinant proteins are not really active or are not getting produced in the cell. To test these hypotheses, expression of the cloned genes was detected by means of immunoblotting employing Mtb GroEL and GroES specific antibodies.

E. coli SV2 expressing *groEL* homologues as above were grown in LB broth supplemented with 0.2% L-arabinose. Cultures were recovered in mid-log phase (A_{600} = 0.5) and the lysates were resolved on 10% SDS-PAGE. To probe for Mtb GroEL1 and GroEL2, the samples were separated on 10% acrylamide gel and for Mtb GroES, on 15% acrylamide gel. Following electrophoresis, the separated protein molecules were transferred by electroelution onto treated Polyvinylidene fluoride (PVDF) membrane. Transfer was set in at 50 V for four hours in 1X electrode transfer buffer at 4 °C. The

membranes were incubated in 5% non-fat milk in 1X TBST for two hours at room temperature. This step is accomplished to block the un-reacted sites on the membrane which would reduce the extent of nonspecific binding of antibodies during subsequent steps and thereby improving the sensitivity of the assay. The transferred protein molecules were probed with Mtb GroEL1, GroEL2 and GroES specific primary antibodies, rabbit polyclonal rabbit α-Cpn60.1$_{Mtb}$, mouse monoclonal IT56, and mouse monoclonal IT3, at 1:10,000, 1:2000 and 1:200 dilutions, respectively for four hours at room temperature. Further, the blots were washed with 1X TBST to remove unbound antibodies and reduce background, thereby increasing the signal-to-noise ratio. The blots were further probed with appropriate secondary antibodies, anti-rabbit or anti-mouse antibodies at 1:10000 dilutions. Following washing with 1X PBST, the blots were developed using ECL (+) Western Blotting Kit (GE Biosciences Inc.).
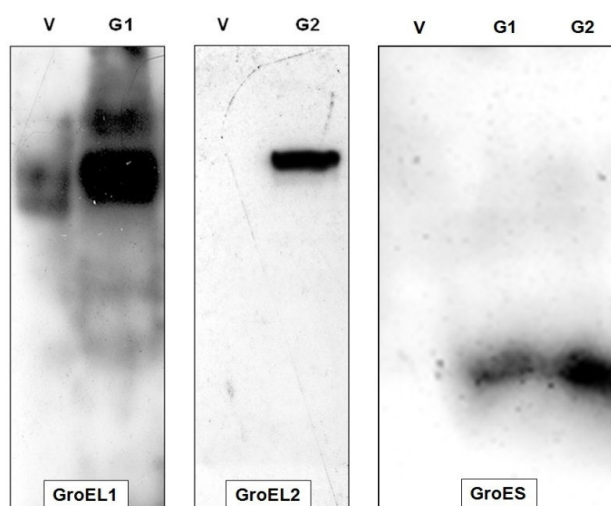
Figure 2.09: Immunodetection of Expression of Chaperonin Homologues. *E. coli* SV2 cultures harboring plasmids, pSCM1602, pSCM1603 and pBAD24 were cultured in the presence of 0.2% L-arabinose. Cells were recovered at mid-log phase and the cell lysates were probed with antibodies specific to Mtb GroES, GroEL1 and GroEL2. V: vector control (pBAD24); G1: GroES + GroEL1 (pSCM1602) and G2: GroES + GroEL2 (pSCM1603).

Immunoblots with GroEL1 and GroEL2 specific antibodies established that the proteins were expressed and hence that the lack of complementation was not due to the lack of expression (Figure 2.09).

2.4.4 Complementation Study in GroEL Depleted Strain

Data in the above experiments clearly show that GroEL homologues from *M. tuberculosis* were inefficient in complementing the loss of *E. coli* GroEL. However, to rule out the possibility that the resident GroES and GroEL44 might be influencing the activity of Mtb GroELs, we have tested the ability of Mtb GroELs in complementing the loss of *E. coli* GroEL in a GroEL depleted strain. Towards this, we have employed *E. coli* strain LG6, wherein the expression of chromosomal *groES/L* is under the P_{lac} promoter (Horwich et al., 1993).

Stationary phase cultures of *E. coli* strain LG6 containing plasmids pSCM1601, pSCM1602, pSCM1603 and pBAD24 were serially diluted. 5 μl of each serially diluted cultures were spotted onto LB agar supplemented with either 0.2% D-glucose (to repress P_{BAD} promoter) or 0.2% L-arabinose (to induce the expression of cloned *groEL* genes) or 1 mM IPTG (to induce the chromosomal copy of *E. coli groES/L* operon). The plates were incubated at 30 °C.
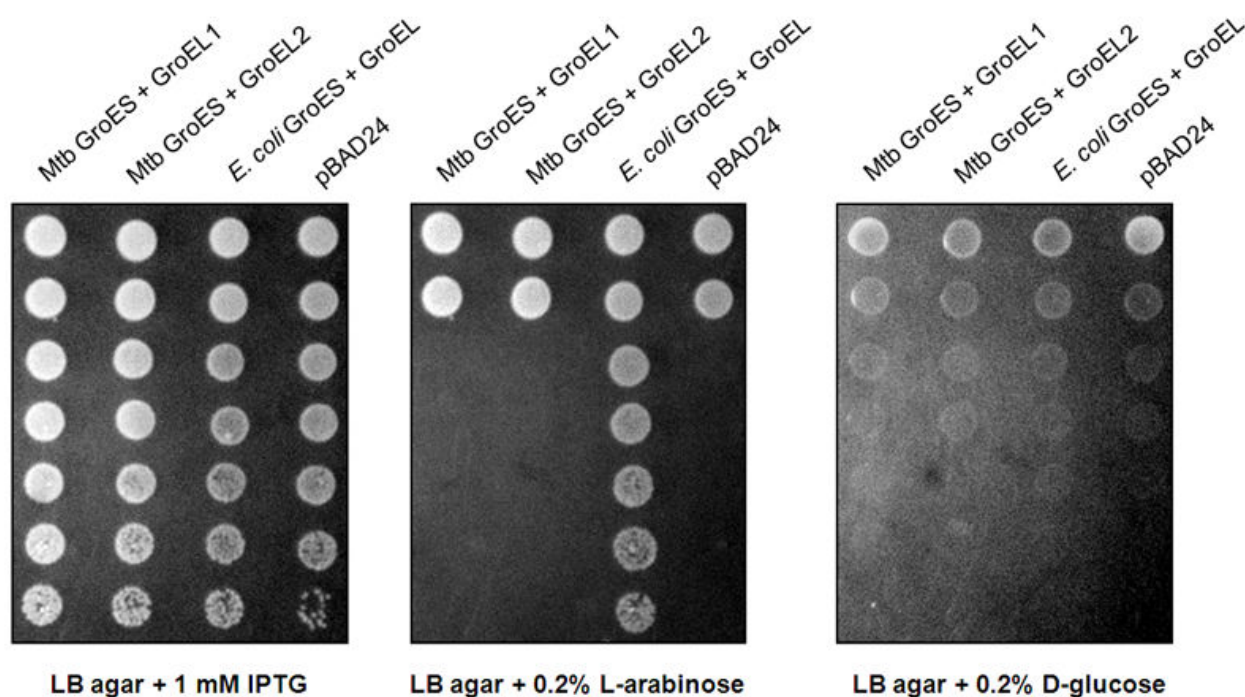
Figure 2.10: Complementation by Mtb GroELs in E. coli LG6. Serially diluted cultures of the *E. coli* strain LG6 expressing the indicated *groEL/S* genes were spotted onto the surface of LB agar plates supplemented as indicated. The plates were incubated at 30 °C.

Data shown in Figure 2.10 clearly shows Mtb GroELs could not support the growth of *E. coli* LG6. This observation is similar to the complementation studies performed in E. coli SV2 as explained in sections 2.4.1 through 2.4.3. Moreover, the inability of Mtb *groEL*s to complement in the loss of GroEL function in *E. coli* is consistent with our earlier in vitro studies on Mtb GroELs that have described their low activity (Qamra et al., 2004).

2.5 Conclusions

ORFs encoding Mtb GroELs, when expressed in *E. coli*, are incapable of complementing a conditional allele of *groEL*, the *groEL44.* In addition, immunoblotting of the cell lysates expressing the cloned Mtb *groEL/S* genes showed the presence of Mtb GroEL/S. These results therefore indicate that this behavior of Mtb GroELs could exclusively be attributed to intrinsic properties of the proteins. Although these observations are in agreement with the reported in vitro analyses (Qamra et al., 2004), it is, however, important to note that the proteins are expressed in heterologous expression system. Moreover, earlier biochemical characterization of these proteins had established that the recombinant Mtb GroELs are incapable of existing in canonical tetradecameric forms, suggesting a connection between the oligomerization and the activity of GroEL. Henceforth, the following three hypotheses were considered as potential causes of lowered chaperone activity of Mtb GroELs:

1. Inspite of high sequence homology with *E. coli* GroEL, variations at numerous critical positions in Mtb GroELs were observed. Studying the potential of these variations in rendering Mtb rGroELs inactive, could be the primary reason for the inefficient chaperone properties of Mtb rGroELs.

2. Either the loss of oligomerization of Mtb rGroELs or the inability to recognize substrate proteins from the heterologus host, *E. coli*, could be attributed to the unusual behavior of Mtb rGroELs, in complementing the Ts phenotype conferred by the *groEL44* allele. Since the said characteristics are conferred upon GroEL by different domains, effect of the individual attribute could be studied.

3. Despite the fact that Mtb rGroELs although show impaired chaperone activities, the genes encoding these proteins in Mtb are not diverted towards becoming pseudo genes, indicating that a function must exist for Mtb GroELs. Therefore, contrary to the condition in *E. coli*, these proteins could be encountering host factor induced activation in the slow growing Mtb. Probing the presence of any such factor and its characterization would unearthen the underlying principles of GroEL metabolism.

These hypotheses were probed individually. Information on the experiments performed and the results obtained are discussed in detail in subsequent chapters.

2.6 References

- Ang, D., Richardson, A., Mayer, M. P., Keppel, F., Krisch, H. and Georgopoulos, C. (2001) Pseudo-T-even bacteriophage RB49 encodes CocO, a cochaperonin for GroEL, which can substitute for *Escherichia coli*'s GroES and Bacteriophage T4's Gp31. *J. Biol. Chem.* **276**, 8720-8726

- Bukau, B. and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366

- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. III., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544

- Georgopoulos, C. P. and Hohn, B. (1978) Identification of a host protein necessary for bacteriophage morphogenesis (the *groE* gene product). *Proc. Natl. Acad. Sci. USA* **75**, 131-135

- Georgopoulos, C. P., Hendrix, R. W., Kaiser, A. D. and Wood, W. B. (1972) Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. *Nat. New. Biol.* **239**, 38-41

- Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose *PBAD* promoter. *J. Bacteriol.* **177,** 4121-4130

- Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N. and Furtak, K. (1993) Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. *Cell* **74,** 909-917

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

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

- Richardson, A., van der Vies, S. M., Keppel, F., Taher, A., Landry, S. J. and Georgopoulos, C. (1999) Compensatory changes in GroEL/Gp31 affinity as a mechanism for allele-specific genetic interaction. *J. Biol. Chem.* **274**, 52-58

- Sambrook, J. and Russell, D. W. (2000) *Molecular Cloning: A Laboratory Manual.* **III Ed:** Cold Spring Harbour Press, Cold Spring Harbour, New York, USA.

CHAPTER III

Gene Shuffling and Domain Swapping of Mycobacterial GroEL

Identification of Critical Sequence Features for Activity

3.1 Introduction

Molecular chaperones assist the folding of naïve and misfolded polypeptides in the cycles of binding and release (Ellis, 1987). GroEL belongs to the family of molecular chaperones that are termed as chaperonins, since they adopt ring like quaternary structure forming a cavity for the sequestration of misfolded polypeptides. (Hemmingsen et al., 1988). Elaborate biochemical and structural studies have shown that GroEL from *E. coli* ologomerizes into a tetradecameric form with two isologous heptameric rings with two cavities for substrate polypeptides to bind (Braig et al., 1994; Hayer-Hartl et al., 1995; Weissman et al., 1995; Xu et al., 1997).

Studies towards biochemical characterization of recombinant GroEL homologues from *Mycobacterium tuberculosis* (Mtb) have revealed weakened chaperone activity exhibited by these molecules (Qamra and Mande, 2004 & Qamra et al, 2004). This deviation from the canonical chaperone behavior, although was speculated due to their impaired oligomerization properties, lacks a direct evidence. Moreover, studying the ability of Mtb GroELs to act as chaperones in vivo was essential. ORFs encoding Mtb GroELs were expressed under the control of arabinose inducible P_{BAD} promoter and the ability of Mtb GroELs to act as chaperones was assessed as described in chapter II. Mtb GroELs were not able to act as effective molecular chaperones, although in the presence of cognate Mtb GroES, when expressed in *E. coli*. This led us to probe the molecular features of Mtb GroELs in detail.

Since oligomerization is compromised in recombinant Mtb GroELs, studying the mutations accumulated in positions responsible for inter-subunit interactions might reveal features important for oligomerization and activity. A directed evolution approach involving gene shuffling was attempted to investigate this aspect, which might further yield useful insights into the functions of Mtb GroELs. Furthermore, Mtb-*E. coli* GroEL chimeras having mutually exchanged oligomerization equatorial domains were generated. Ability of these chimeric molecules to complement *E. coli* GroEL was studied. Details of the experiments and results are discussed in this chapter.

3.2 Materials

All the chemicals, media components and enzymes were purchased from various commercial sources. Strains of *E. coli* were cultured in standard LB supplemented as appropriate. *E. coli* SV2 is a derivative of the *E. coli* K12 strain B178, which bears a temperature sensitive (Ts) allele of *groEL*, namely, *groEL44* (*galE groEL+*) and *E. coli* MGM100 is a derivative of MG1655 wherein the chromosomal *groES/L* operon is placed downstream of L-arabinose inducible P_{BAD} promoter (McLennan and Masters, 1998; Tilly and Georgopoulos, 1982). Coliphages λ*c*IB2 and T4GT7 were sourced from laboratory stocks. Plasmids and oligonucleotide primers used in this study are listed in Appendix II.

3.2.1 Preparation of Electro-competent Cells and Electroporation of *E. coli*

A flask containing 50 ml of LB medium was inoculated with a single colony of *E. coli* DH5α or Top10 from a fresh agar plate and then incubated overnight at 37 °C in a rotary shaker with constant stirring at 250 rpm. 1% of this culture was further inoculated into 500 ml of LB broth and incubated at 37 °C till the mid-log phase (OD_{600} of the cultures reaches 0.4). The culture was transferred to pre-cooled centrifuge tubes and was incubated on ice for 15 – 30 minutes. Cells were harvested by centrifugation at 1000 g for 15 minutes at 4 °C and then washed with 500 ml of ice-cold pure H_2O followed by two washes in 250 ml and 10 ml of ice cold 10% glycerol. The cells were harvested again by centrifugation and the supernatant was carefully decanted. The cell pellet containing the electro-competent cells was resuspended in 1 ml of ice-cold GYT broth (10% Glycerol, 0.125% Yeast extract and 0.25% Tryptone), dispensed as 40-μl aliquots into sterile, ice-cold 0.5-ml microfuge tubes, chilled by transferring into a bath of liquid nitrogen and stored at -70 °C.

10 pg to 25 ng of the DNA was added to 40 μl of the electro-competent cells and incubated on ice for 30-60 seconds. Electroporation apparatus was adjusted to deliver an electrical pulse of 25 μF (field strength of 12.5 kV/cm with a time constant of 4-5 milliseconds) capacitance with a potential of 2.5 kV and a resistance of 200 Ω. DNA-cell mixture was carefully transferred into the bottom of a cold electroporation cuvette and a pulse of electricity was applied to the cells at the settings indicated above.

Immediately 1 ml of SOC medium was added to the electroporation cuvette at room temperature. The cells were transferred to 1.5 ml microfuge tube and were incubated for 1 hour at 37 °C. The cells were spread onto the surface of LB agar supplemented with appropriate antibiotic and the plates were at incubated 37 °C overnight. Transformed colonies appeared in about 12-16 hours.

3.2.2 Preparation of Ultra-competent Cells

A flask containing 50 ml of LB medium was inoculated with a single colony of *E. coli* DH5α or Top10 from a fresh agar plate and then incubated overnight at 37 °C in a rotary shaker with constant stirring at 250 rpm. 1% of this culture was further inoculated into 500 ml of SOB broth and incubated at 37 °C till the mid-log phase (OD$_{600}$ of the cultures reaches 0.55). Cells were transferred to centrifuge tubes and were incubated on ice for ten minutes. Cells were harvested by centrifugation at 2500 g for 10 minutes at 4 °C. The cell pellet was resuspended gently in 80 ml of ice-cold Inoue transformation buffer (55 mM MnCl$_2$, 15 mM CaCl$_2$, 250 mM KCl and 10 mM PIPES pH: 6.7). Cells were harvested by and the cell pellet having the ultra competent cells was resuspended in 20 ml of ice-cold Inoue transformation buffer + 1.5 ml of DMSO. Cell suspension was dispensed into aliquots of 100 μl into pre-chilled, sterile microfuge tubes, snap-frozen by transferring the tubes into the bath of liquid nitrogen. And the tubes were stored at -70 °C. The ultra-competent cells were transformed with plasmid DNA, using the standard bacterial transformation protocol.

3.3 Directed Evolution of Mtb Chaperonins to Investigate Function

Since the recombinant Mtb GroELs do not behave as canonical chaperonins, we investigated functions of these molecules with directed evolution. Towards achieving these objectives, the genes encoding these proteins, *groEL1* and *groEL2*, were subjected to random mutagenesis followed by a functional selection to harvest functional GroEL molecules. Random mutagenesis was carried out either by gene shuffling or by chemical mutagenesis with hydroxyl amine.

3.3.1 Shuffling of Genes Encoding Mtb GroEL Homologues

Gene shuffling involves four major steps:

3.3.1.1 Amplification of DNA Fragments of Interest by PCR

Mtb *groEL1* and *groEL2* ORFs were amplified using primer pairs SCM1607F/SCM1607R and SCM1608F/SCM1608R, respectively (Figure 3.01). Amplification was carried out by Dynazyme Ext Polymerase (Finzyme Inc.) and the cycling conditions were as given in the Table 3.01. Cosmid clones BAC-Rv285 (F7) and BAC-Rv313 (A10) from Mtb genomic DNA library were used as templates for amplifying *groEL1* and *groEL2*, respectively.

Cycle Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94 °C	300	1
Denaturation	94 °C	60	
Annealing	55 °C	60	30
Extension	72 °C	180	
Final Extension	72 °C	600	1

Table 3.01: PCR cycling conditions for amplifying Mtb *groEL1* and *groEL2*

3.3.1.2 Limited DNaseI Digestion of the DNA Fragments

Limited DNaseI digestion was optimized for DNaseI from GE Biosciences Inc. DNaseI obtained from different manufacturers viz Roche Inc., New England Biolabs Inc., were tried and DNaseI from GE Biosciences was selected because of the ease of controlling the reaction. 5-10 μg of the amplified DNA fragments were pooled and were incubated with 0.0015 U/μl DNaseI at 25 °C for 5 min in 1X DNaseI buffer (40 mM Tris.HCl (pH: 7.9), 10 mM NaCl, 6 mM MgCl$_2$ and 1 mM CaCl$_2$). The reaction was stopped by heat denaturation at 75 °C for 20 min. Addition of inhibitors like 1% SDS at or 1mM EDTA improved the efficiency of reaction. However, only 1 mM EDTA was used since presence of SDS adversely affected the recovery of digested DNA fragments from gel. The reaction was optimized to obtain fragments of 50-150 bp by varying concentrations of DNAseI from 0.015 to 2 units per 100μl reaction and the time of incubation from 5 to 15 min. The resulting DNaseI digest was resolved on 1.3% agarose gel and were extracted using QiaexII gel extraction kit (Qiagen Inc.) (Figure 3.01).
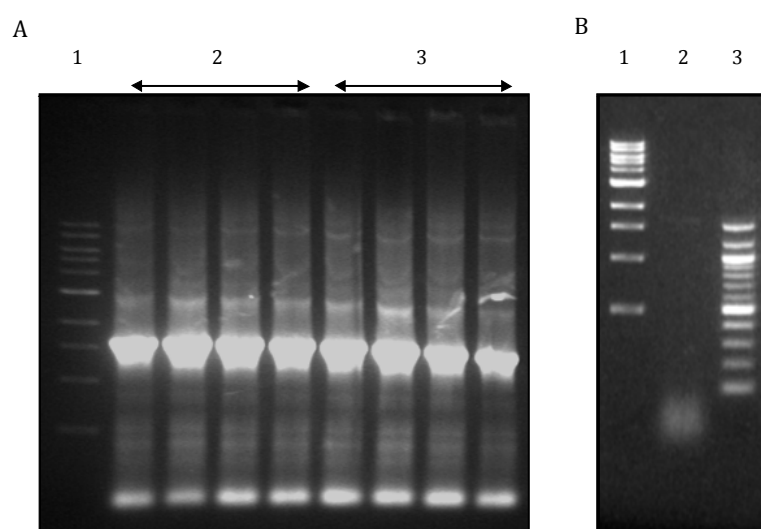
Figure 3.01: Amplification and DNaseI Digestion of Mtb groEL1 and groEL2. **A.** ORFs encoding Mtb GroEL1 *and* GroEL2 were amplified by PCR. PCR products were resolved on 1% agarose gel. 1, 1 kb marker; 2, PCR amplified *groEL1* fragments and 3, PCR amplified *groEL2* fragments. **B.** 5 to 10 μg of the PCR products were subjected to limited DNaseI digestion and the fragments were resolved on 1.3% agarose gel. The lanes correspond to: 1, 1 kb marker; 2, DNaseI digest of GroEL1 and GroEL2 and 3, 100 bp marker.

3.3.1.3 Self-priming Assembly PCR to Randomly Assemble the Fragments after DNaseI Digestion

A self-priming PCR was optimized for randomly assembling the fragments recovered from DNaseI digestion. The conditions in this PCR were modified from the earlier reported protocol (Stemmer, 1994 and Wang et al., 2002). Fragments of 50 to 150 bp from DNaseI digestion were recovered using the Qiaex II kit, diluted ten times and were used as a template for a self-priming step-down PCR using Dynazyme II (Finzyme, Inc). The cycles of reactions for this PCR were optimized empirically by varying the times of extension from 1 to 3 min, number and steps of cycles from 25 to 60 cycles and the annealing temperature ranging from 40 to 65 °C. Amplification at lower temperatures (40 °C) resulted in unusually lengthier products and the amplification at lower times of extension generated the products that were shorter than 1.6 kb, the size of the input gene. The optimized PCR was carried out with 7.5 µl/100 µl template and the reaction cycles were as given in the Table 3.02. Migration pattern of the PCR product was analyzed on agarose gel and was used as a template for subsequent rounds of PCR.

Cycle Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94 °C	300	1
Denaturation	94 °C	30	
Annealing	55 °C	30	15
Extension	72 °C	90	
Denaturation	94 °C	60	
Annealing	50 °C	30	45
Extension	72 °C	240	
Final Extension	72 °C	900	1

Table 3.02: PCR cycling conditions for self-priming PCR.

As anticipated, the resulting PCR product was composed of fragments of varied sizes indicating extremely random assembly. These products were resolved on 1% agarose gels, which showed a smear, indicating the randomness of fragment assembly (Figure 3.02).
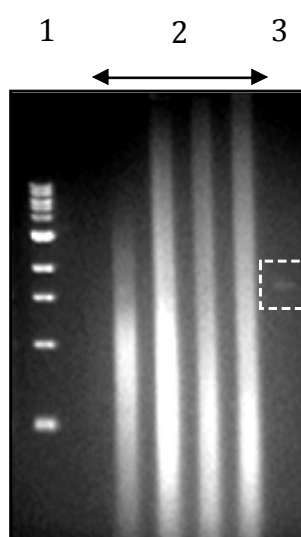
Figure 3.02: Self Priming Assembly PCR. DNaseI digest was subjected to primer-less PCR and the resulting PCR products were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb marker; 2, Self-priming PCR products and 3, PCR amplified GroEL fragment (shown boxed) loaded for comparison.

3.3.1.4 PCR to Amplify *groEL*-like Molecules Using Specific Primers

Self-assembled PCR product was used as a template in the error prone step-up PCR with 30 μM of specific primers, SCM06F and SCM03R, using a 2.5 U of 1:1 mix of Pfu turbo and Taq (MBI). The method described by Zhao and Arnold, 1997, was employed for this set of PCRs. Employing Dynazyme II for this final PCR did not result in a significant amplification whereas the mixture of Pfu turbo/Taq gave a decent amplification. The parameters used in the cycles for this final round of PCR are given in Table 3.03. In this round of PCR using specific primers for the amplification of 1.6 kb DNA, *E. coli groEL* specific primers SCM4F and SCM3R were used taking into consideration the fairly hydrophobic GGM repeats at the carboxy-terminus of the *E. coli* GroEL. Many of the proteobacterial GroELs contain a 13 residue (GGM)$_4$M carboxyl terminus, alteration in size and chemical nature of which is thought to affect GroEL function in folding a few of the substrate polypeptides (Tang et al., 2006, Farr et al., 2007). However, neither of the mycobacterial GroELs possess typical GGM repeat at the C-terminus, though GroEL2 bears an imperfect repeat. A fragment of around 1.6 kb, approximately the size of *groEL* gene was recovered from this round of PCR. Similar

reactions with Mtb *groEL1* and *groEL2* were also set which also resulted in amplification, but with a lesser efficiency than that with *E. coli groEL* specific PCR.

Cycle Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	96 °C	120	1
Denaturation	94 °C	30	
Annealing	55 °C	30	10
Extension	72 °C	60	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	100	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	140	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	180	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	220	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	260	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	300	
Denaturation	94 °C	60	
Annealing	50 °C	30	2
Extension	72 °C	340	
Final Extension	72 °C	900	1

Table 3.03: Cycling conditions for specific PCR according to Zhao and Arnold, 1997

Other PCR reactions were similarly set following the protocol reported by Stemmer, 1994. Self-assembled PCR product was used as a template in multiplex step-up PCR with the primers specific to both *groEL1* and *groEL2*, SCM1607F/SCM1607R and SCM1608F/SCM1608R, respectively. Amplification was carried out by Dynazyme II (Finzyme Inc.) and the cycling conditions are as given in Table 3.04 This reaction was optimized for the combination of primers to be used since it was noted that the reactions, where primers used were in mixed combination or as a multiplex, resulted in fragments longer than 1 kb, but not in reactions where the primers were specific to only one gene, i. e., either to *groEL1* or to *groEL2*.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94 °C	300	1
Denaturation	94 °C	60	
Annealing	40 °C	30	8
Extension	72 °C	180	
Denaturation	94 °C	60	
Annealing	42 °C	30	10
Extension	72 °C	180	
Denaturation	94 °C	60	
Annealing	45 °C	30	12
Extension	72 °C	180	
Denaturation	94 °C	60	
Annealing	55 °C	30	15
Extension	72 °C	180	
Final Extension	72 °C	600	1

Table 3.04: Cycling conditions for Specific PCR according to Stemmer, 1994

To select the desired fragment from the library of random fragments, we opted initially to recover the fragments of *groEL* size and amplify them by PCR. These trials were not successful as the PCRs ended up in primer dimers. So the complete library was used as a template for the subsequent specific PCRs. Further, to increase the chances of PCR we incorporated primers specific for both *groEL1* and *groEL2.* Surprisingly we observed a bright fragment of the size of *groEL* being amplified (Figure 3.03). Specificity

and yield of similar amplified product was increased remarkably by employing primers specific to *E. coli groEL*.
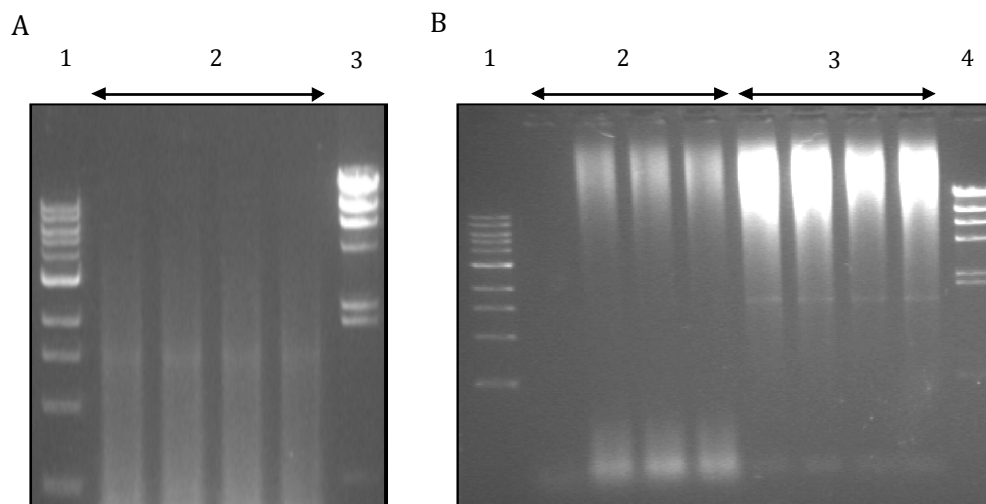
Figure 3.03 Final Round of PCR to Amplify groEL-like Molecules. **A.** PCR employing the method described by Stemmer, 1994. The lanes correspond to: 1, 1 kb marker; 2, PCR amplified *groEL1* fragments and 3, λ HindIII marker. **B.** PCR employing the method described by Zhao and Arnold, 1997 using primers specific to Mtb *groEL*s and *E. coli groEL*. The lanes correspond to: 1, 1 kb marker; 2, Amplification using primers specific to Mtb *groEL*s; 3, Amplification using primers specific to *E. coli groEL and 4,* λ HindIII marker.

3.3.2 Chemical Mutagenesis of Genes Encoding Mtb GroEL Homologues

Chemical mutagenesis of DNA was followed as described in Sikorski and Boeke, 1991. Plasmids harboring Mtb *groEL1* and *groEL2*, pSCM1604 and pSCM1605, respectively were subjected to chemical mutagenesis using Hydroxylamine (NH$_2$OH) as the mutagen, which converts a GC base pair to an AT base pair by tautaumerization. About 10 µg plasmid DNA was incubated with 500 µl Hydroxylamine solution containing 1 M Hydroxylamine-HCl, 50 mM Sodium pyrophosphate, 100 mM NaCl and 2 mM EDTA for 20 h at 37 °C. Mutagenized DNA was purified from Hydroxylamine using QIAquick PCR Purification Kit (Qiagen Inc.). Purified mutagenized DNA was employed for selection of active clones as described below.

3.3.3 Cloning of Random Mutagenesis Products

DNA library obtained by gene-shuffling was digested with restriction endonucleases NcoI and HindIII and cloned onto pBAD24 (Guzman et al., 1995) by low proportion ligation. The ligation mix was transformed into DH5α by electroporation and the resulting transformants were selected in the presence of 100 μg/ml ampicillin. Transformants were scrapped and grown to stationary phase in 100 ml of LB broth supplemented with appropriate antibiotic and 0.2% D-glucose. A proportion of the culture was preserved as cloned library. Plasmid DNA was purified from this culture and was then transformed into *E. coli* SV2 strain (Georgopoulos et. al., 1972). The transformants were directly selected at 42 °C in the presence of 0.2% L-arabinose (to induce the P_{BAD} promoter), which would enable selecting the clones capable of encoding active versions of GroEL, that would in turn be capable of complementing the Ts phenotype associated with the *groEL44* allele. Plasmid DNA was prepared from each of the resulting transformants. Variants so recovered were designated as *groELSp01* to *groELSp40*. Of these, 9 plasmids (pSCM1622 to pSCM1637) were chosen for further study.

To rule out the possibility of contaminating DNA encoding *E. coli* GroEL, restriction digestion profiles of the said clones were compared to those of *E. coli* GroEL in two different reactions. In one reaction the clones were subjected to digestion by a combination of restriction endonucleases, BamHI and ClaI and in the other by KpnI. The profiles shown in Figure 3.04, clearly demonstrate that the selected clones are not contaminated by DNA encoding *E. coli* GroEL.
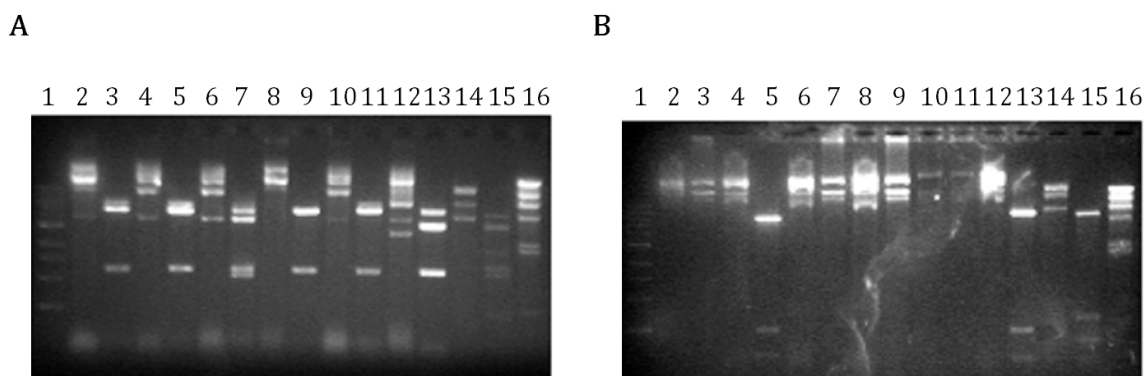
Figure 3.04: Restriction Digestion Profiling of Gene-shuffled Variants. Plasmids harboring gene-shuffled *groEL* variants were digested with restriction endonucleases BamHI/ClaI (A) and KpnI. The resulting fragments were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb Marker; 2, pSCM1622 Undigested; 3, Clone pSCM1622 digested; 4, Clone pSCM1624 Undigested; 5, Clone pSCM1624 digested; 6, Clone pSCM1625 Undigested; 7, Clone pSCM1625 digested; 8, Clone pSCM1626 Undigested; 9, Clone pSCM1626 digested; 10, pSCM1627 Undigested; 11, pSCM1627 digested; 12, pSCM1632 Undigested; 13, pSCM1632 digested; 14, pSCM1608 Undigested; 15, pSCM1608 digested; 16, λ/HindIII Digest.

3.3.4 Complementation Studies on Random Mutagenesis Products

Having established that the clones obtained by gene-shuffling are not false positives, we assessed the activity of these clones by quantitative analysis of the extent of their ability to complement the Ts phenotype harbored by the *groEL44* allele. The gene shuffled clones were transformed into *E. coli* SV2 and the resulting transformants were grown in the presence of 0.2% D-glucose. Serially diluted stationary phase cultures of *E. coli* SV2 harboring the said clones were spotted onto LB agar plates supplemented with 0.2% L-arabinose. The plates were incubated at permissive (30 °C) and at restrictive temperatures (42 °C). A quantitative phenotypic analysis of nine of the gene shuffled *groEL* variants is shown in Figure 3.05. The clones showed different extent of complementation which in turn would reflect the differences in the activity of the GroEL variants produced. Hence to investigate the specific variations that resulted in the gain of function, sequences encoding the gene-shuffled GroEL variants were analyzed.
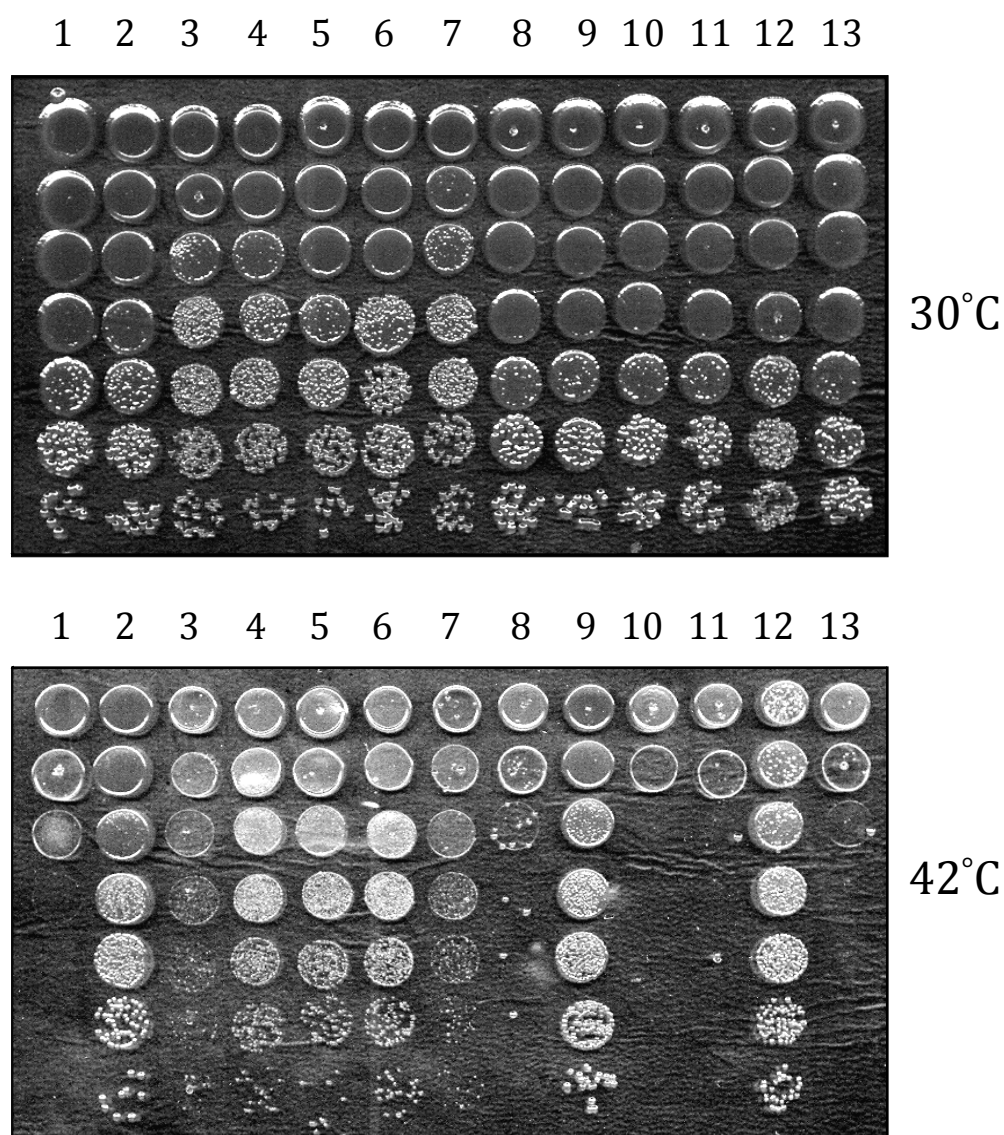
Figure 3.05: Phenotypic Analysis of the GroEL Mutants Generated by Gene Shuffling. Cultures of SV2 expressing the indicated GroEL variants were ten-fold serially diluted and spotted onto the surface of LB plates supplemented with 0.2% L-arabinose and incubated at the permissive (30 °C) and restrictive (42 °C) temperatures, as indicated. 1-9, GroEL variants obtained by gene shuffling, GroELSp22, 24, 25, 26, 27, 32, 35, 36 and 37; 10, *Mtb* GroEL1; 11, *Mtb* GroEL2; 12, *E. coli* GroEL; 13, Vector (pBAD24).

3.3.5 Sequence Analysis of Random Mutagenesis Products

ORFs encoding gene-shuffled *groEL* variants were sequenced using ABI Prism sequencer. Analysis of the sequences showed that the mutants generated by gene shuffling are chimeras of Mtb *groEL1*, *groEL2* and *E. coli groEL*. Sequence analysis of the clones and the comparison of G + C contents show that the gene shuffled *groEL* variants are in fact derived from Mtb *groEL*s (Table. 3.05).

groEL variant	G + C content (%)
groELSp22	68.1
groELSp24	69.4
groELSp25	67.8
groELSp26	68.1
groELSp27	67.5
groELSp32	68
groELSp35	69.3
groELSp36	67.9
groELSp37	67.9
E. coli groEL	52.9
Mtb *groEL1*	65.1
Mtb *groEL2*	65.4

Table 3.05. GC Contents in the Gene Shuffled *groEL* Variants. Nucleotide sequences of the Gene shuffled *groEL* variants were analyzed for their G + C contents. These are further compared with *E. coli* and Mtb GroELs for the G + C contents.

Translated polypeptide sequences were compared with the Mtb and *E. coli* GroEL sequences using Clustal X (1.81) (Thompson et al., 1997). Multiple sequence alignment of the polypeptides encoded by the said *groEL* alleles with *E. coli* and *Mtb* GroELs' sequences interestingly showed that their putative apical domains are subject to considerable variation in amino acid sequence with some variants bearing fairly large deletions and insertions whereas their putative equatorial domains are conserved among the variants and are analogous to *E. coli* GroEL (Figure 3.06).

A
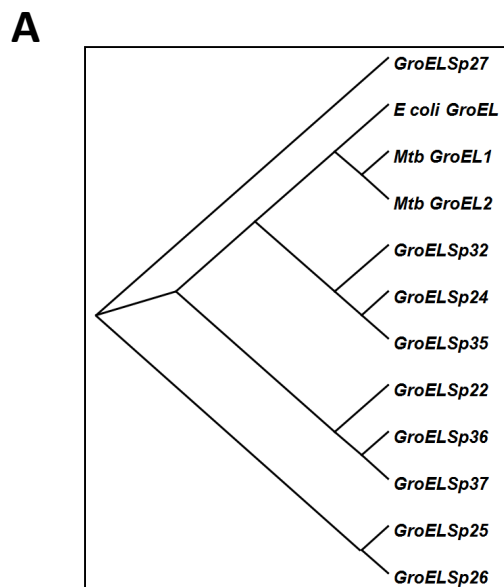
Figure 3.06: Sequence Analysis of the GroEL Mutants Generated by Gene Shuffling. **A.** Cladogram generated by distance calculation from the multiple sequence alignment shown below. **B.** Multiple sequence alignment of the polypeptide sequences of gene-shuffled GroEL variants with sequences of *E. coli* and Mtb GroEL homologues. The scale indicates amino acid positions. Deletions (interrupted lines) extending from the polypeptide region 190 to 210 (GroELSp22, 25, 32) and an insertion of an amino acid segment in one of the variants (GroELSp27) are apparent. The variants GroELSp24 and GroELSp35 bear multiple segmental deletions.

B
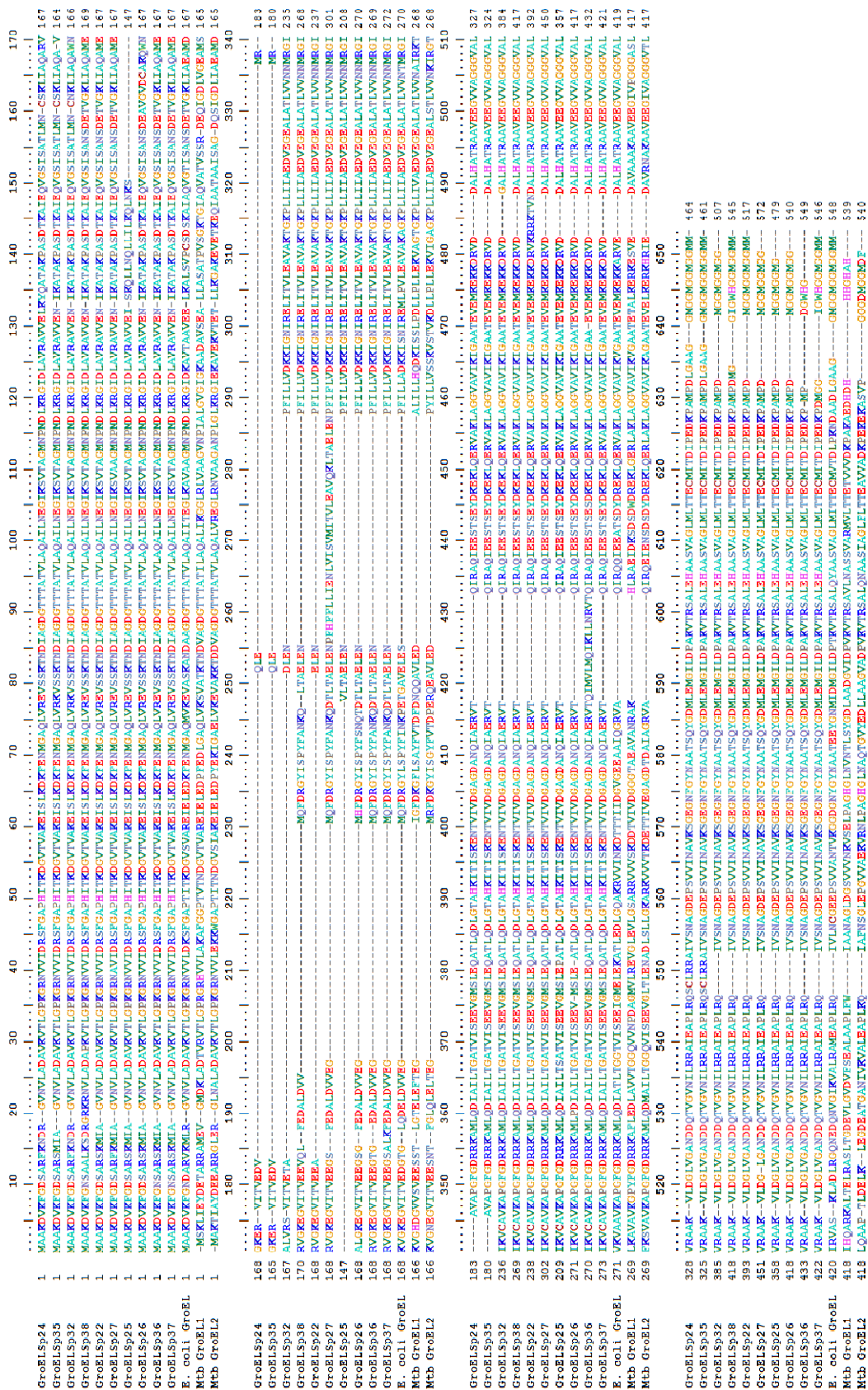
3.3.6 Mapping Mutations of Gene-shuffled GroEL Variants onto Structure

Surprised by the diversity observed in the gene-shuffled GroEL mutants, as revealed by the multiple sequence alignment, and to understand functional significance for the occurrence of these deviations, the observed insertions and deletions were mapped on to the monomeric GroEL structure. Co-ordinates for monomeric GroEL were adopted from Xu et al., 1994 (PDB ID: 1AON). The deletions in GroELSp22, GroELSp25, and GroELSp32 spanned the intermediate and apical domains, though the substrate and GroES interacting regions were retained. On the other hand the deletion observed in GroELSp24 spanned the substrate interacting domain. Moreover, insertions observed in GroELSp27 and GroELSp36 were also located in the apical domain (Figure 3.07). Despite stark variations observed in the sequences of apical domains, all the mutants bore an equatorial domain similar to *E. coli* GroEL. Since all the mutants tested are capable of complementing the Ts phenotype of the *groEL44* allele, this observation lends additional support to the supposition that ability to oligomerize, due to the presence of an "*E. coli* GroEL-like" equatorial domain, correlates with biologically relevant GroEL activity and suggests that the apical domain can tolerate substantial variations in its amino acid sequence.

GroELSp22 **GroELSp24** **GroELSp25**

GroELSp27 **GroELSp32** **GroELSp36**
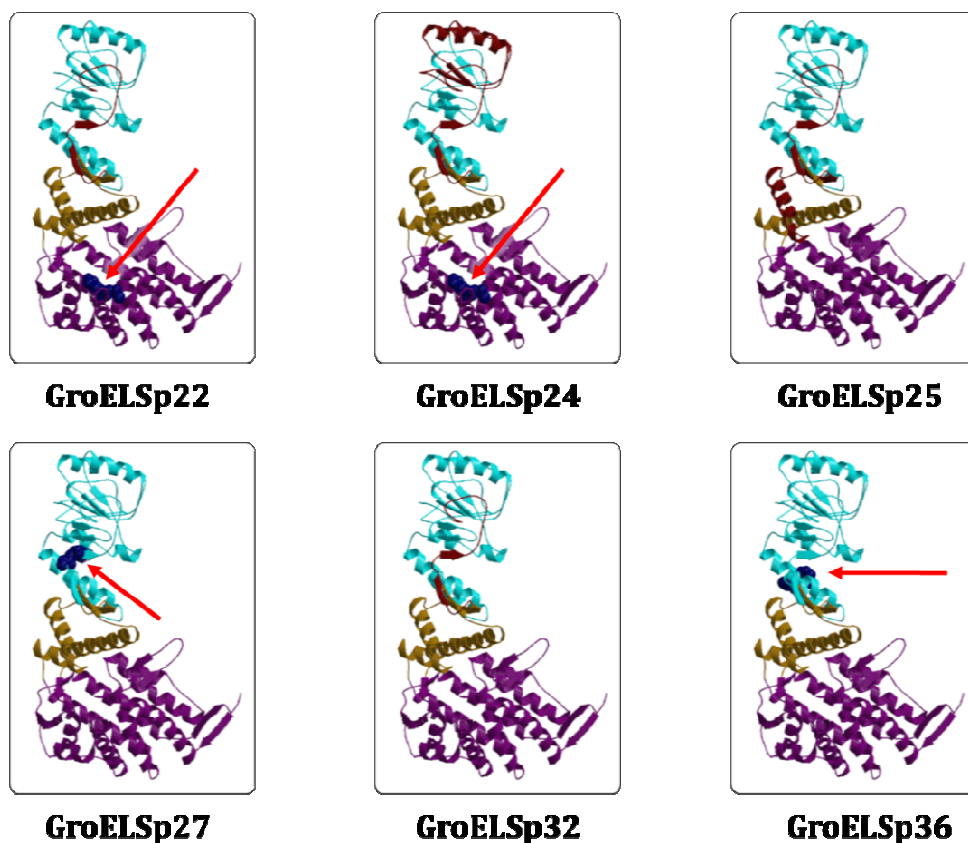
Figure 3.07: Location of Mutations in Gene-shuffled GroEL Variants. *E. coli* GroEL monomer ribbon diagrams, illustrating the sites of insertions and deletions observed in the indicated gene shuffled GroEL variants. Variations observed in the GroEL variants are color-coded. Regions marked in blue represent insertions and those in maroon represent deletion. Insertions are also indicated with an arrow for easy distinction. GroEL domains are also shown in three different colors to highlight the occurrence of the mutations in different domains. Cyan and magenta respectively represent the apical and equatorial domain, while brown represents the intermediate domain. The molecular representations were made using Molscript and the co-ordinates for monomeric GroEL are as reported previously (Xu et al., 1994 PDB ID: 1AON A chain)

Analysis of the sequences indicates that the selection pressure on the GroEL equatorial domain is relatively stringent than that on the apical domain (Figure 3.06). Consequently the apical domain is capable of buffering several mutations. In addition, it is noteworthy that the equatorial domains of Mtb GroEL homologs have diverged to a large extent from that of the *E. coli* GroEL. This deviation in the amino acid sequences of equatorial domains is speculated to be the potential cause of the impaired oligomerization properties displayed by Mtb GroELs (Qamra et al., 2004).

These observations led us to study the effect of exchanging the equatorial domains between Mtb GroEL1 and *E. coli* GroEL with the following two questions:

1. Would Mtb GroEL1 harboring equatorial domain from *E. coli* GroEL be able to complement *E. coli* GroEL and vice versa?
2. Would *E. coli* GroEL harboring equatorial domain from Mtb GroEL1 be able to display characteristics similar to Mtb GroEL?

These questions are addressed in the following part of the chapter.

3.4 Mtb GroEL1 Regains Chaperonin Function by Equatorial Domain Substitution

Experiments with the gene shuffled GroEL mutants showed that the substrate interacting apical domain can tolerate variations whereas the oligomerization equatorial domain needs to be similar to *E. coli* GroEL's equatorial domain. Since the selected variants are functional entities, we addressed the question: if GroELs of *M. tuberculosis* can be made functional by alterations in the equatorial domains. To achieve this, exchanging the equatorial domain of *M. tuberculosis* GroEL1 with that of *E. coli* GroEL, was attempted. Likewise, we wished to check if *E. coli* GroEL can be made non-functional upon exchanging its equatorial domain with that from *M. tuberculosis* GroEL1. Details of the experiments in generation of the chimeras and analysis of their function are described below.

3.4.1 Generation of GroEL Chimeras

For the generation of Mtb-*E. coli* GroEL chimeras, we have undertaken the overlap extension PCR. This was accomplished in three steps.

3.4.1.1 Amplification of Domain Regions of GroELs

The designed chimeric molecules with regions used for replacement between Mtb GroEL1 and *E. coli* GroEL are illustrated in Figure 3.09. The equatorial domains of Mtb GroEL1 and *E. coli* GroEL were mutually exchanged employing overlap extension PCR (Warrens et al., 1997 and Kondoh et al., 2002). The N-terminal and C-terminal equatorial sub-domains of *E. coli groEL* and Mtb *groEL1* were amplified individually by PCR using primer pairs SCM1606F/SCMJ1 and SCMJ2/SCM1609R for *E. coli groEL* and SCM1605F/SCMJ3 and SCMJ4/SCM1605R for Mtb *groEL1*, respectively. Amplification in these reactions was carried out by Phusion polymerase (Finzymes Inc.) and the cycling conditions are as given in table 3.06.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	96 °C	120	1
Denaturation	96 °C	10	
Annealing	55 °C	60	35
Extension	72 °C	45	
Final Extension	72 °C	600	1

Table 3.06: PCR cycling conditions for amplifying equatorial domain fragments of *E. coli groEL* and Mtb *groEL1*

The PCR products were resolved on 1 % agarose gel and fragments of ~400 kb were extracted using standard procedures (Figure 3.08). In addition the fragments of Mtb and *E. coli groELs* encoding apical-intermediate domain were amplified in a similar manner. Fragments of ~700 kb were recovered and were used in the next round of overlap extension PCR.
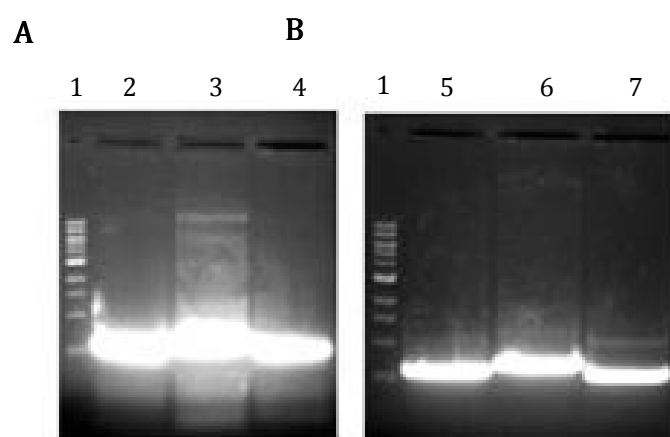
Figure 3.08: Amplification of DNA Fragments Encoding GroEL Domains. DNA fragments encoding equatorial and apical-intermediate domains of *E. coli* GroEL **(A)** and Mtb GroEL1 **(B)** were amplified. The PCR products were resolved on 1 % agarose gel. Different lanes are as follows: 1, 1 kb ladder (New Englands Biolabs Inc.); 2, Fragment encoding N-terminal equatorial sub-domain of *E. coli* GroEL; 3, Fragment encoding apical-intermediate domains of *E. coli* GroEL; 4, Fragment encoding C-terminal equatorial sub domain of *E. coli* GroEL; 5, Fragment encoding N-terminal equatorial sub-domain of Mtb GroEL; 6, Fragment encoding apical-intermediate domains of Mtb GroEL and 7, Fragment encoding C-terminal equatorial sub domain of Mtb GroEL.

3.4.1.2 Overlap Extension PCR to Generate Chimeric Molecules

The ORFs encoding the equatorial sub-domains of *E. coli* GroEL and Mtb GroEL1 were recovered and were used as primers to generate the domain fusion constructs in the overlap extension PCR, conditions of which are listed in Table 3.07. ORFs encoding the apical-intermediate domains of Mtb GroEL1 and *E. coli* GroEL, respectively were used as templates in these reactions to generate the Mtb *groEL1* bearing *E. coli* GroEL's equatorial domain, *groELMEF* and the *E. coli groEL* bearing Mtb GroEL1's equatorial domain, *groELMER*, respectively. Thus, *groELMEF* bears DNA sequence encoding apical and intermediate domains of Mtb GroEL1 (residues 134-408) and equatorial domain of *E. coli* GroEL (residues 1-133 & 409-548) and likewise *groELMER* bears the DNA sequence encoding apical and intermediate domains of *E. coli* GroEL (residues 134-408) and equatorial domain of Mtb GroEL1 (residues 1- 133 & 409-539) (Figure 3.09 and 3.10).

Amplification of *groELMEF*			Amplification of *groELMER*		
Temperature in °C	Time in Seconds	No. of Cycles	Temperature in °C	Time in Seconds	No. of Cycles
96 °C	120	1	96 °C	120	1
96 °C	30		96 °C	30	
55 °C	30		55 °C	30	
72 °C	180	22	72 °C	240	25
96 °C	30		96 °C	30	
72 °C	210		72 °C	270	
72 °C	600	1	72 °C	600	1

Table 3.07: PCR cycling conditions for amplifying *groELMEF* and *groELMER*
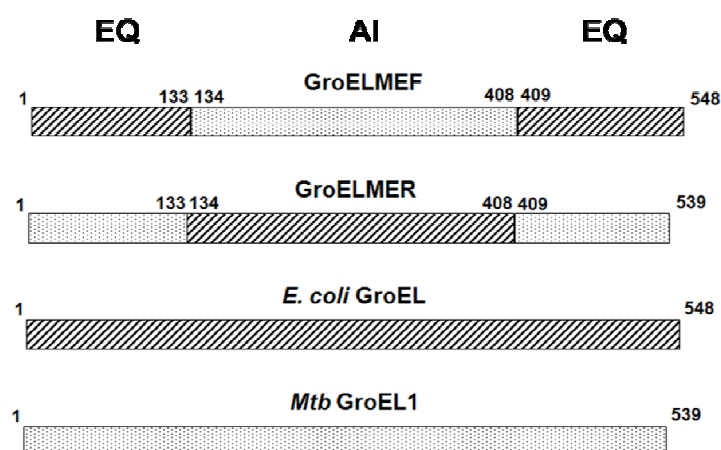
Figure 3.09: Schematic Representation of Domain Allocation in GroEL Variants, GroELMEF and GroELMER. Numbering denotes amino acid residue positions of the parental Mtb GroEL1 and *E. coli* GroEL polypeptides. Equatorial domain regions are indicated as EQ and those spanning apical and intermediate domains as AI.
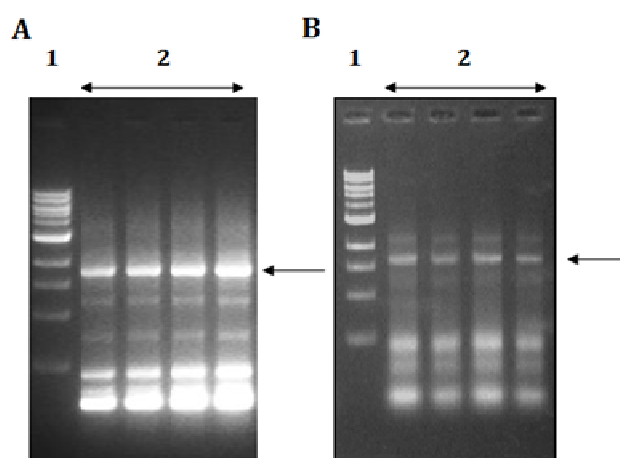
Figure 3.10: Generation of groELMEF and groELMER. ORFs encoding GroELMEF **(A)** and GroELMER **(B)** were generated by employing overlap extension PCR. PCR products were resolved on 1% agarose gel. DNA fragments matching the theoretical molecular mass of GroELMEF and GroELMER are indicated with arrow. The lanes correspond to: 1, 1 kb ladder (New England Biolabs Inc.) and 2, PCR Products.

3.4.1.3 Cloning of the Chimeric Molecules *groELMEF* and *groELMER*

The ORFs encoding the two chimeric GroEL molecules, GroELMEF and GroELMER were recovered from agarose gel. GroELMEF was cloned into NcoI and SalI sites on pBAD24 (Guzman et al., 1994). GroELMER was cloned into NdeI and SmaI sites on pBAD25. Resulting plasmids were designated pSCM1609 and pSCM1611, respectively. The clones were confirmed by digestions with restriction endonucleases and automated sequencing (Figure 3.11).
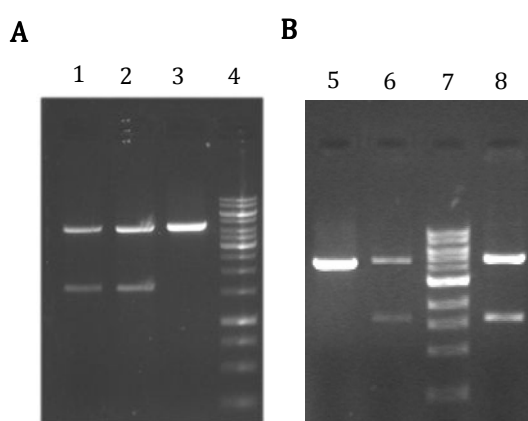
Figure 3.11: Confirmation of Cloning of groELMEF and groELMER. Clones of pSCM1609 was digested with restriction endonucleases NcoI and SalI and pSCM1610 with NdeI and SmaI. The digests were resolved on 1% agarose gel. The lanes are: 1 & 2, clones of pSCM1609 digested; 3, pBAD24 digested; 4, 1 kb ladder (New England Biolabs Inc.); 5, pBAD25 digested; 6 and 8, pSCM1610 clenes digested and 7, 1 kb ladder (MBI Fermentas Inc.).

3.4.2 Complementation Studies on GroELMEF and GroELMER in *E. coli* SV2

In vivo function of GroELMEF and GroELMER were assessed by estimating the ability of these variants in:

a. complementing the Ts phenotype associated with the *groEL44* allele in *E. coli* SV2,

b. restoring the Ts phenotype of SV2 upon plasmid curing and

c. supporting morphogenesis of bacteriophages λ and T4

3.4.2.1   Assaying the Ability to Rescue Ts Phenotype of *groEL44* Allele

Briefly, SV2 cultures expressing GroELMEF and GroELMER independently were ten-fold serially diluted in LB. Each dilution was spotted on LB agar plates supplemented with 0.2% L-arabinose. The plates were incubated at permissive (30 °C) and restrictive (42 °C and 45 °C) temperatures. Complementation of the Ts growth phenotype associated with *groEL44* allele by the GroEL variants, GroELMEF and GroELMER, was estimated by colony forming ability conferred by a given plasmid at the restrictive temperature (Figure 3.12). SV2 cultures expressing *E. coli groEL*, Mtb *groEL1*, Mtb *groEL2* and harboring vector pBAD24 were included as controls.
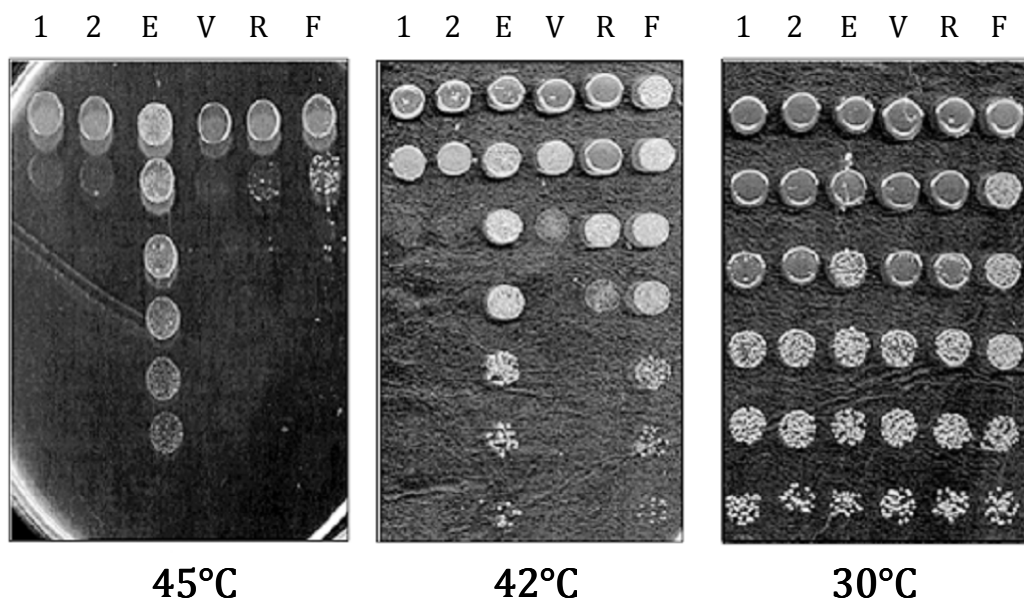
Figure 3.12: Complementation of the groEL44 Allele by GroEL Variants. Ten fold serially diluted cultures of the *E. coli* strain SV2 (*groEL44*) expressing the genes encoding Mtb GroEL1 (1), Mtb GroEL2 (2), *E. coli* GroEL (E), GroELMER (R) and GroELMEF (F) were spotted on LB agar plates supplemented with 0.2% L-arabinose and the plates were incubated at the indicated temperatures.

Data in Figure 3.12 showing the complementation studies with GroELMEF and GroELMER confirmed that the exchange of the equatorial domain in Mtb GroEL1 with that of *E. coli* GroEL, specified by the GroELMEF, turns it into a functional GroEL, in vivo. GroELMEF was able to complement the *groEL44* allele at the restrictive temperature of 42 °C to an extent similar to that exhibited by *E. coli* GroEL, whereas at 45 °C GroELMEF displayed weak complementation. In contrast, GroELMER, in which the DNA encoding

equatorial domain of *E. coli* GroEL is replaced with that encoding the corresponding domain from Mtb GroEL1, was not able to exhibit any complementation at 42 °C.

3.4.2.2   Plasmid Curing Experiments

Furthermore, to ascertain the exhibited phenotype to the plasmid borne GroEL chimeric molecules, curing of the plasmids from the SV2 cultures was attempted. SV2 expressing *groELMEF* and *groELMER* were cultured overnight in the absence of ampicillin, such that cell would not require to maintain the plasmid. These cultures were streaked onto the surface of LB agar and incubated at 30 °C. Colonies that were cured off the plasmid were obtained by screening for the loss of antibiotic resistance. The plasmid cured cultures were streaked on LB agar plates with and without the supplement of 0.2% L-arabinose. The plates were incubated at 30 °C and 42 °C. Ability of the plasmid cured cultures of SV2 in rescuing the Ts phenotype associated with *groEL44* allele of SV2 was compared with that of the uncured cultures (Figure 3.13).

42 °C   **30 °C**
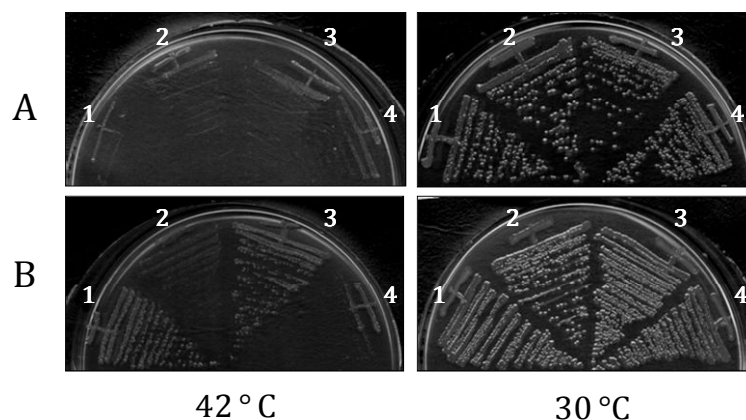
Figure 3.13: Plasmid Curing from SV2 Cultures. SV2 cultures that were expressing *E. coli groEL* (1), *groELMEF* (3) and *groELMER* (4) were cured of the plasmids. Plasmid cured cultures (A) were streaked on LB agar plates supplememted with 0.2% L-arabinose and plasmid retained cultures (B) were streaked onto LB agar plates supplemented with 100 µg ampicillin and 0.2% L-arabinose. The plates were incubated at the indicated temparatures. SV2 harboring vector pBAD24 (2) was included as a control.

Curing the plasmids from the SV2 cultures expressing plasmid borne *groELMEF* and *groELMER* resulted in loss of complementation. This confirms that the phenotype exhibited by GroELMEF in rescuing the Ts phenotype exhibited by SV2 and supporting morphogenesis of bacteriophages, is indeed vector borne. Thus, GroELMEF, and not GroELMER, is able to function in vivo in a manner similar to that of *E. coli* GroEL.

3.4.2.3 Assaying the Ability to Support Bacteriophage Morphogenesis

In *E. coli*, development of bacteriophages like lambda (λ) and T4 requires functional GroEL/S system (Zeilstra-Ryalls et al., 1993). To gauge the extent of restoration of GroEL function present in GroELMEF, its ability to support phage morphogenesis was studied. Cultures of *E. coli* SV2 bearing plasmids pSCM1604 (Mtb GroEL1), pSCM1605 (Mtb GroEL2), pSCM1608 (*E. coli* GroEL), pSCM1609 (GroELMEF), pSCM1610 (GroELMER) and pBAD24 obtained after culturing in 0.4% D-maltose with appropriate antibiotic were overlaid in soft agar on LB plates containing either 0.2% D-glucose or 0.2% L-arabinose. The plates were incubated at 30 °C. Plaque forming ability of bacteriophages λ*c1B2* and T4GT7 was assessed by spotting 5 µl each of hundred-fold diluted suspension of bacteriophage stocks.
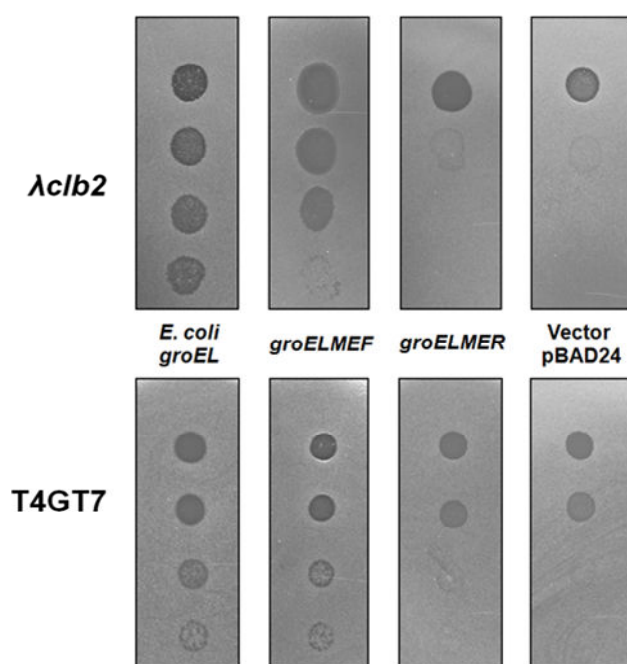
Figure 3.14: Assay for Bacteriophage Morphogenesis. Lawns of the *E. coli* strain SV2 expressing indicated *groEL* variants were prepared on LB agar plates supplemented with 0.2% L-arabinose. Hundred-fold serial dilutions of bacteriophages λcIB2 and T4GT7 were spotted on the lawns. The plates were incubated at 30 °C.

Expression of *groELMEF* like that of *E. coli groEL* was able to support growth of both the phages λ and T4 in the *groEL44* mutant strain SV2, whereas expression of Mtb *groEL1*, *groEL2* and *groELMER* did not promote phage development (Figure 3.14). These studies thus support the notion that the reason for the observed functional difference between the *E. coli* and Mtb GroELs must be a consequence of differences between the equatorial domains of the two molecules.

Having established the ability of GroELMEF in complementing the Ts phenotype conferred by *groEL44* allele of *E. coli* SV2, and further since GroELMEF harbors the substrate and GroES interacting domain derived from Mtb GroEL1, we were led to the following fundamental queries.

1. Whether GroELMEF can supplement the *E. coli* GroEL?
2. Whether the ability of GroELMEF in rescuing the Ts phenotype is independent of the resident GroEL44 or does GroELMEF forms a hetero tetradecamer with the GroEL44?
3. Whether GroELMEF requires GroES for its activity?
4. Whether GroELMEF can function in concert with Mtb GroES?

These queries were answered by undertaking experiments in GroES/L depletion strain. These experiments are described in the following paragraphs.

3.4.3 Complementation Studies on GroELMEF in GroES/L Depletion Strain

For this set of experiments a GroES/L depleted strain of *E. coli*, MGM100 was chosen. *E. coli* MGM100 is derivative of MG1655 in which expression of chromosomal *groES/L* operon is under the control of L-arabinose inducible P_{BAD} promoter (McLennan and Masters, 1998). Therefore, the strain fails to form colonies in the absence of L-arabinose unless a functional copy of GroES/L is expressed. For experiments in *E. coli* MGM100, the GroEL and GroES variants were cloned into pTrc99A in which the expression of cloned genes is under the influence of lactose/IPTG inducible P_{tac} promoter (Amann et al., 1988).

3.4.3.1 Cloning of GroEL Variants for Complementation in GroES/L Depleted Strain

ORFs encoding *E. coli* GroES/L operon, GroELMEF, were cloned into NcoI and SalI sites on pTrc99A to generate pSCM1611 and pSCM1612, respectively. ORF encoding *E. coli* GroEL was cloned into NcoI and HindIII sites on pTrc99A to generate pSCM1613. The clones were confirmed by digestions with restriction endonucleases and sequencing (Figure 3.15).
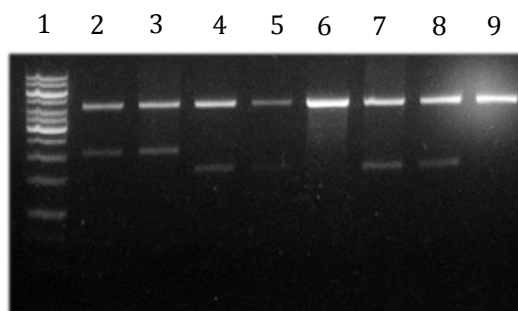
Figure 3.15: Confirming Cloning of E. coli GroES/L, GroEL and GroELMEF Genes. Clones of pSCM1611 (2 and 3), pSCM1612 (4 and 5) and pTrc99A (6) were digested with restriction endonucleases, NcoI and SalI. Clones of pSCM1613 (7 and 8) and pTrc99A (9) were digested with restriction endonucleases NcoI and HindIII. The digests were resolved on 1% agarose gel and molecular weights of the released fragments were compared with the 1 kb ladder (1, MBI Fermentas Inc.).

Clones of pSCM1611 were subjected to site directed mutagenesis for incorporating sites for the recognition of XbaI and HindIII restriction endonucleases, individually, thereby generating pSCM1614 and pSCM1615, respectively. The said sites were engineered at the intergenic region between *groES* and *groEL* employing site directed mutagenesis using the primer pairs, SCM16012F/SCM16014R and SCM16013F/SCM16013R, respectively. Digestion of pSCM1614 with XbaI and HindIII releases the fragment encoding *E. coli* GroEL retaining that of *E. coli* GroES. Clones were confirmed by digestions with restriction endonucleases and sequencing (Figure 3.16).
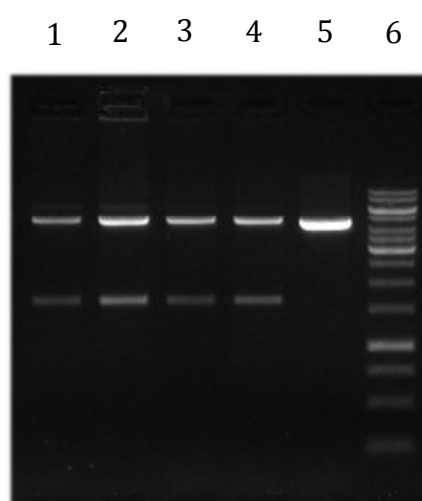
Figure 3.16: Confirming the Site Directed Mutagenesis in pSCM1614 and pSCM1615. Plasmid pSCM1614 was digested with restriction endonucleases XbaI and HindIII and pSCM1615 with HindIII. The digests were resolved on 1% agarose gel. The lanes correspond to: 1 & 2, clones of pSCM1614 digested; 3 & 4, pSCM1615 digested; pTrc99A digested with restriction endonucleases XbaI and HindIII and 6, 1 kb ladder (MBI Fermentas Inc.).

Following this, further steps of cloning involving ORFs encoding *E. coli* GroEL and GroELMEF in operonic arrangement with *E. coli* and Mtb *groES* were followed, results of which are illustrated in Figure 3.17. The clones thus generated were confirmed by digestions with restriction endonucleases (Figure 3.18).
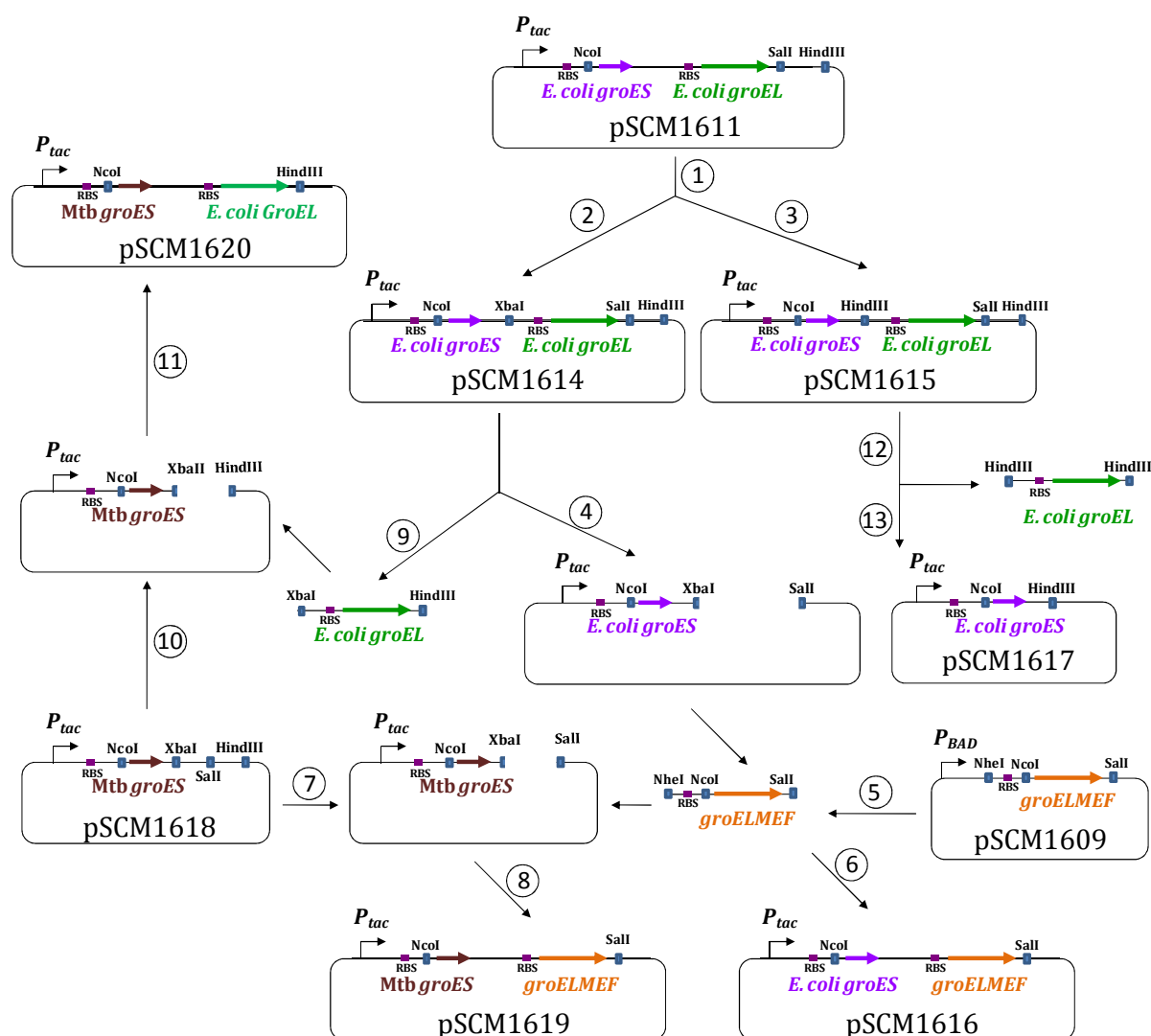
Figure 3.17: Schematic Representation of Cloning of groEL Variants into pTrc99A.
Plasmid pSCM1611 was subjected site directed mutagenesis (1). Recognition sites for restriction endonucleases XbaI (2) and HindIII (3) were introduced individually at the *groEL/groES* intergenic space into the plasmid pSCM1611 resulting in plasmids pSCM1614 and pSCM1615, respectively. Plasmid pSCM1614 was digested with XbaI/SalI to excise out *E. coli groEL* (4). *groELMEF*, obtained by digesting pSCM1609 with NheI and SalI (5), was cloned into XbaI and SalI digested pSCM1614 (6) to obtain pSCM1616. Mtb *groES* was cloned into NcoI and SmaI sites on pTrc99A to generate pSCM1618. Plasmid pSCM1618 was digested with XbaI and SalI (7) and into this, *groELMEF* that was obtained in reaction 5 was inserted (8) to generate pSCM1619. Plasmid pSCM1614 was digested with XbaI and HindIII (9) to excise the *E. coli groEL,* which was cloned into XbaI and HindIII digested (10) pSCM1618 to generate pSCM1620 (11). Plasmid pSCM1615 was digested with HindIII to excise out *E. coli groEL* (12) and the resulting backbone harboring *E. coli groES* was self-ligated to obtain pSCM1617 (13).
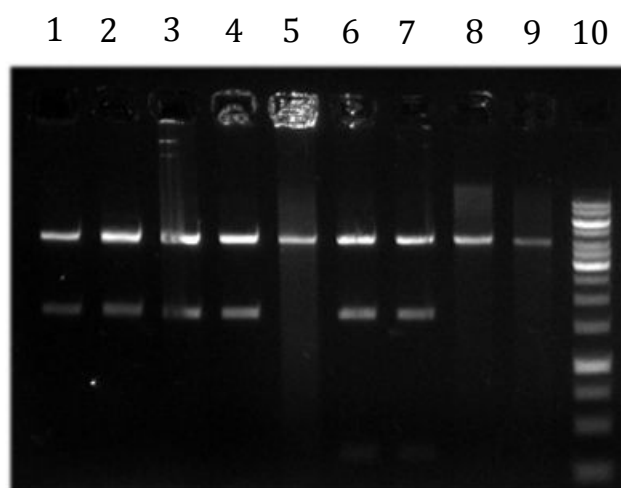
Figure 3.18: Confirming the Cloning of groEL Variants into pTrc99A. Plasmid pSCM1620 (1 & 2), pSCM1619 (3 & 4), pSCM1618 (5), pSCM1616 (6 & 7), pSCM1617 (8) and pTrc99A were digested with corresponding restriction endonucleases and the digests were resolved on 1% agarose gel and molecular weights of the released fragments were compared with the 1 kb ladder (1, MBI Fermentas Inc.).

3.4.3.2   Assessing if GroELMEF is Independent of GroEL44

To avoid potential interference of the indigenous GroEL polypeptide encoded by the resident *groEL44* allele on the complementation exhibited by *groELMEF,* the effect of *groELMEF* expression under conditions of endogenous GroEL depletion was studied. Towards this end plasmids bearing *E. coli groEL* and *groELMEF*, individually and in operonic arrangement with *E. coli groES*, under the expression control of *lac* promoter (P_{tac}) were generated as described before (Figures 3.17 and 3.18). The strain MGM100 wherein expression of the *groES/L* operon is under P_{BAD} promoter, fails to form colonies in absence of arabinose unless a functional copy of GroES/L is expressed. Hence, the ability of GroELMEF to support growth of MGM100 in absence of exogenous L-arabinose supplementation was assessed.

E. coli MGM100 was transformed with plasmids pSCM1611, pSCM1612, pSCM1613, pSCM1616, pSCM1619 and pTrc99A. Resulting transformants were grown under permissive conditions in the presence of 0.2% L-arabinose, washed with LB broth to remove traces of L-arabinose, diluted serially in LB broth and 6 μl of dilutions were spotted onto plates supplemented with 0.2% L-arabinose and 0.2% D-glucose + 0.1 mM

IPTG representing respectively permissive and restrictive conditions and the plates were incubated at 30 °C.

LB + 0.2% L-arabinose　　　　**LB + 0.2% D-glucose + 1 mM IPTG**
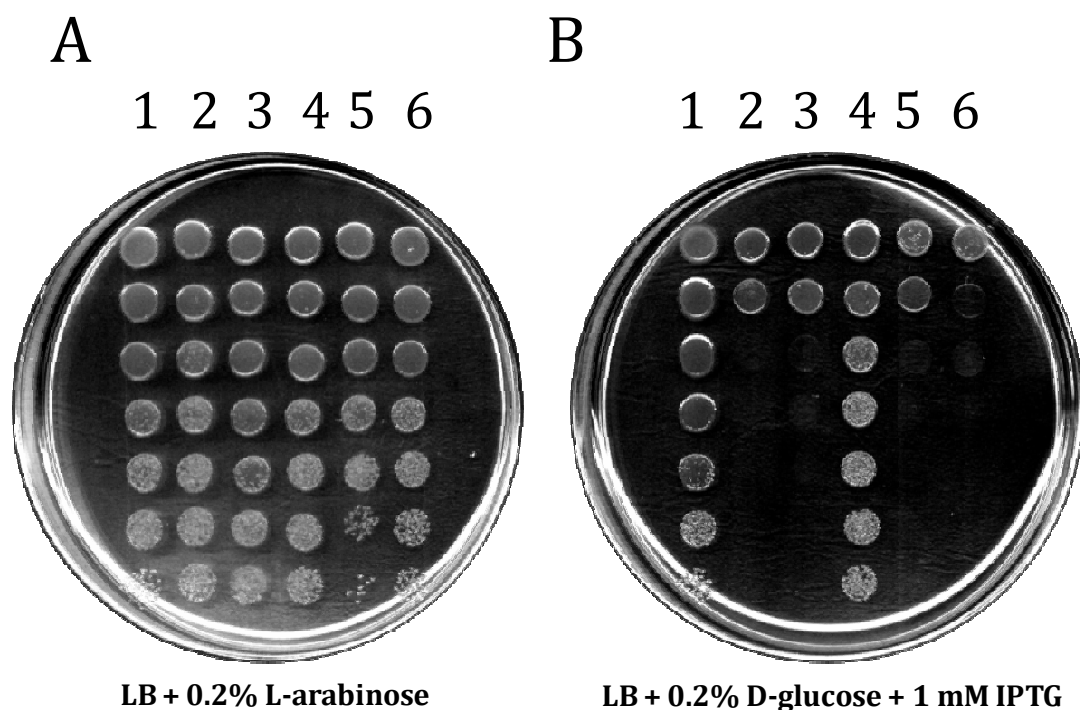
Figure 3.19: Assesment of GroELMEF Independence. Cultures of MGM100 expressing *E. coli groES*, *groEL*, *groELMEF* and combinations were serially diluted, spotted onto the surface of LB agar plates and incubated at 30 °C under permissive (A) and restrictive (B) conditions. 1, MGM100 expressing *E. coli groES + E. coli groEL*; 2, *E. coli groEL*; 3, *E. coli groES*; 4, *E. coli groES + groELMEF*; 5, *groELMEF* and 6, pTrc99A.

Of the various plasmids tested only the plasmids co-expressing *E. coli groES* with either *groELMEF* or *E. coli groEL* allowed colony formation in absence of L-arabinose (Figure 3.19).

These results clearly show that:

1. GroELMEF can substitute for *E. coli* GroEL in vivo,
2. GroELMEF can act as the chaperonin and its activity is not influenced by resident GroEL44 of the strain SV2 and
3. GroELMEF requires *E. coli* GroES for its activity.

3.4.3.3   Assessing if GroELMEF Could Function with Mtb GroES

Having established that GroELMEF requires the presence of *E. coli* GroES for its activity in vivo, and driven by the fact that GroELMEF harbors Mtb GroEL1 borne substrate and GroES interacting domain, effect of Mtb GroES on the in vivo activity of GroELMEF was assessed. Towards this end plasmids bearing *E. coli groEL* and *groELMEF*, individually and in operonic arrangement with *E. coli* and Mtb *groES*, under the expression control of *lac* promoter (P_{tac}) were generated as described before (Figures 3.17 and 3.18). The ability of GroELMEF to support growth of MGM100 in absence of exogenous L-arabinose supplementation and in the presence of *E. coli* or Mtb *groES* was assessed.

E. coli MGM100 was transformed with plasmids pSCM1611, pSCM1612, pSCM1613, pSCM1616, pSCM1619, pSCM1620 and pTrc99A. Resulting transformants were grown under permissive conditions in the presence of 0.2% L-arabinose, washed with LB broth to remove traces of L-arabinose, diluted serially in LB broth and 6 μl of dilutions were spotted onto plates supplemented with 0.2% L-arabinose and 0.2% D-glucose + 0.1 mM IPTG representing respectively permissive and restrictive conditions and the plates were incubated at 30 °C (Figure 3.20).

A

1 2 3 4 5 6 7

B

1 2 3 4 5 6 7

LB + 0.2% L-arabinose

LB + 0.2% D-glucose + 1 mM IPTG
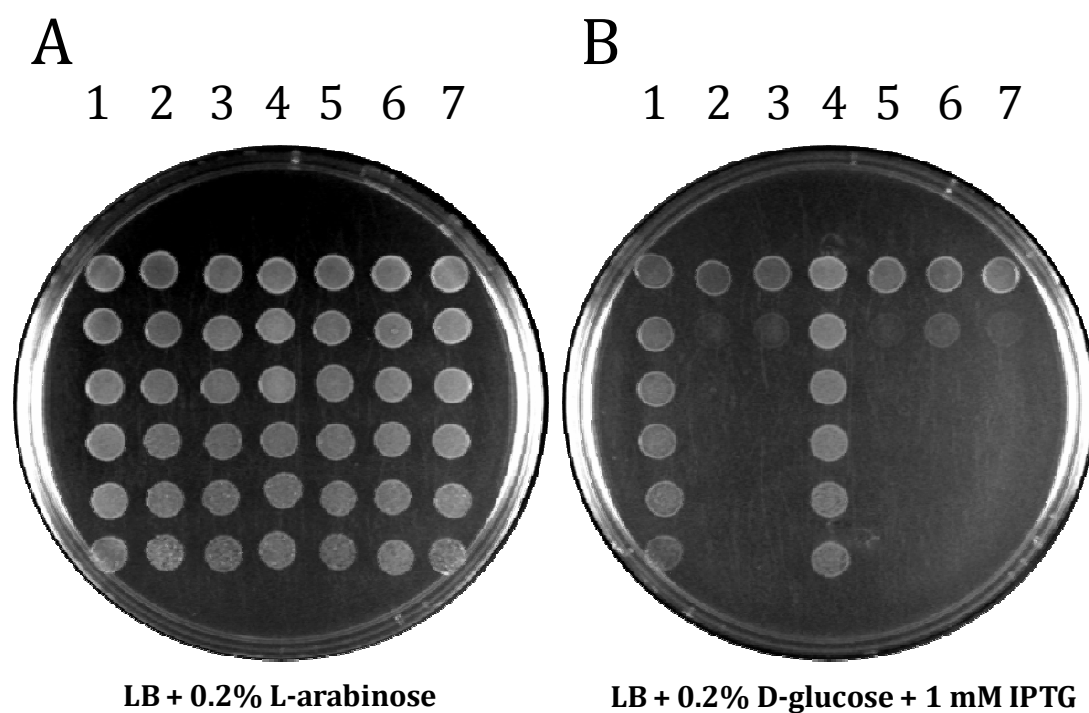
Figure 3.20: Effect of Mtb GroES on GroELMEF. Cultures of MGM100 expressing Mtb and *E. coli groES*, *groEL*, *groELMEF* and combinations were serially diluted, spotted onto the surface of LB agar plates and incubated at 30 °C under permissive (A) and restrictive (B) conditions. 1, MGM100 expressing *E. coli groES + E. coli groEL*; 2, Mtb *groES + E. coli groEL*; 3, *E. coli groEL*; 4, *E. coli groES + groELMEF*; 5, Mtb *groES + groELMEF*; 6, *groELMEF* and 7, pTrc99A.

Of the various plasmids tested only the plasmids co-expressing *E. coli groES* with either *groELMEF* or *E. coli groEL* allowed colony formation in absence of L-arabinose. These results clearly show that GroELMEF requires the presence of *E. coli* GroES for its activity. Surprisingly, GroELMEF, despite possessing the apical domain of Mtb GroEL1, fails to complement along with its cognate Mtb GroES. The reasons for this observation are not clear currently.

3.4.4   Isolation of GroEL Variant by Ligation Based Domain Swapping

Another variant of GroELMEF, GroELDS was generated by ligating DNA fragments encoding and N-terminal and C-terminal equatorial sub-domains of *E. coli* GroEL (EQ1 and EQ2, respectively) and apical-intermediate domains of *M. tuberculosis* GroEL1 (AI).

3.4.4.1   Amplification of DNA Fragments Encoding EQ1, EQ2 and AI Domains

For this, DNA fragments encoding *E. coli* GroEL's equatorial domains (1-399 (EQ1) and 1225-1647 (EQ2) nucleotides encoding 1-133 and 409-549 amino acid residues, respectively) were fused with the fragment encoding Mtb GroEL1's apical and intermediate domains (400-1224 (AI) nucleotides encoding 134-408 amino acid residues). The fragments EQ1, AI and EQ2 were amplified by PCR, using primer pairs SCM1614F/SCM1614R, SCM1615F/SCM1615R and SCM1616F/SCM1616R, respectively (Figure 3.21). PCR Cycling conditions are given in Table 3.08.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	96 °C	120	1
Denaturation	96 °C	10	
Annealing	55 °C	30	30
Extension	72 °C	60	
Final Extension	72 °C	600	1

Table 3.08: PCR cycling conditions for amplifying DNA fragments encoding *E. coli* GroEL's equatorial domains and *Mtb* GroEL1's apical and intermediate domains.
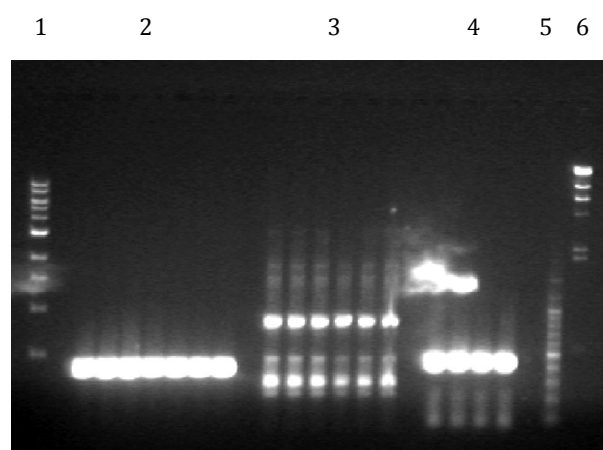
Figure 3.21: PCR Amplification of GroEL Domains for Generation of GroELDS.
DNA fragments corresponding to the equatorial domain of *E. coli* GroEL and apical-intermediate regions of Mtb GroEL1 were amplified by PCR. The PCR products were resolved on 1 % agarose gel. 1, 1 kb ladder (New England Biolabs Inc.); 2, Fragment corresponding to 1-399 bases of *E. coli* GroEL; 3, fragment corresponding to 400-1224 bases of Mtb GroEL1; 4, Fragment corresponding to 1225-1647 bases of *E. coli* GroEL; 5, 100 bp ladder (New England Biolabs Inc.) and 6, λ/HindIII digest.

3.4.4.2   Construction of the Gene Encoding GroELDS

The PCR products described above were digested by restriction endonucleases as given in Table 3.09 and were ligated into pETDuet-1 (Novagen) in a four-component ligation reaction, to generate pSCM1621, which was further digested with restriction endonuclease NcoI and EcoRV to excise the fragment containing the cloned gene, which was sub-cloned into pBAD24 that was digested with NcoI and SmaI to generate pSCM1623. The resulting clones were confirmed by restriction digestion with the four enzymes and by sequencing (Figure 3.22).

Fragment	Restriction Endonuleases Used
EQ1	NcoI/HindIII
AI	HindIII/NdeI
EQ2	NdeI/BglII
pETDuet-1	NcoI/BglII

Table 3.09: Restriction endonucleases employed for generating GroELMEF variant.
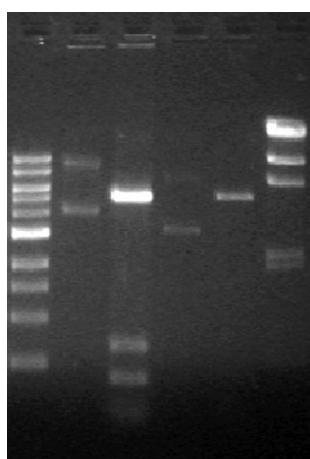
Figure 3.22: Clone Confirmation of pSCM1623. Plasmid pSCM1622 was digested with restection endonucleases NcoI, HindIII, NdeI, BglII. The digests were resolved on 1% agarose gel. The lanes correspond to: 1, 1 kb ladder (New England Biolabs Inc.); 2, pSCM1623 undigested; 3, pSCM1623 digested; 4, pBAD24 undigested; 5, pBAD24 digested and 6, λ/HindIII digest.

3.4.4.3 Complementation Studies on GroELDS

Ability of this clone to complement the Ts phenotype associated with *groEL44* allele was performed in the similar fashion as explained before above. GroELDS was shown to complement the Ts phenotype associated with *groEL44* allele of *E. coli* SV2 (Figure 3.23).

30 °C   42 °C
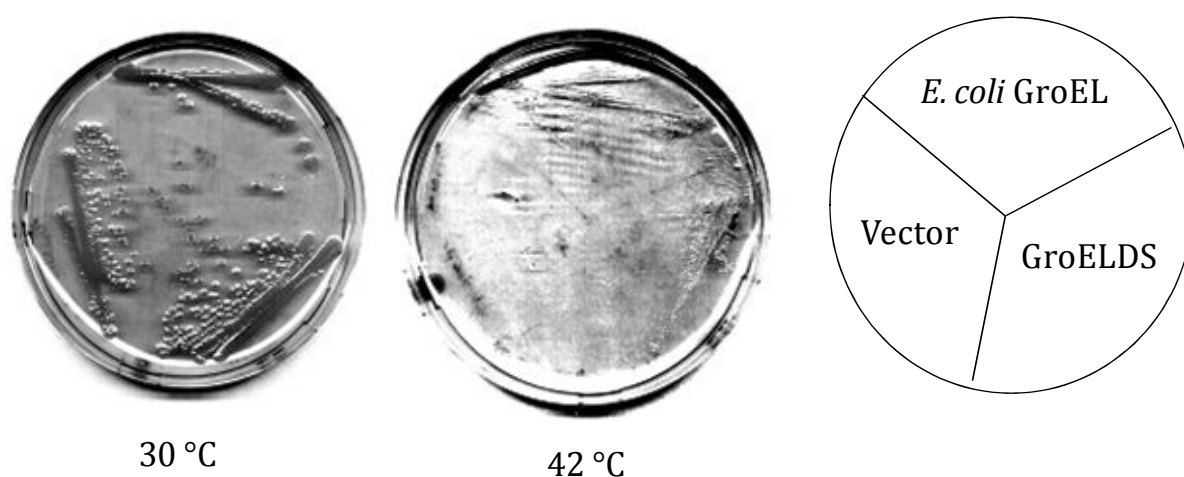
Figure 3.23: Complementation Studies on GroELDS. SV2 cultures expressing the genes encoding indicated GroEL variants were streaked onto LB agar plates supplemented with 0.2% L-arabinose. The plates were incubated at the indicated temperatures.

The data presented in the Figure 3.23 clearly indicates that GroELDS is able to rescue the Ts phenotype associated with the *groEL44* allele. This implies that GroELDS is a functional chaperone although GroELDS exhibits weak complementation. However, the gene encoding GroELDS harbors 12 extra nucleotides due to the cloning strategy, which leads to the formation of polypeptide with four extra residues at the domain boundaries. Since this construct would not generate GroEL variant with sharp domain boundaries, we wished to switch to the overlap extension PCR for the generation of chimeric GroEL variants, as explained in sections 3.4.1 through 3.4.3.

3.5 Conclusions

Directed evolution of Mtb GroELs involving gene-shuffling approach resulted in GroEL variants that are capable of complementing the Ts phenotype conferred by the *groEL44* allele in *E. coli* SV2. Furthermore, sequence analysis of these variants showed that the putative equatorial domains of the variants are conserved and show high homology to that of *E. coli* GroEL. However, the apical domain is capable of absorbing considerable variations in its amino acid sequence. The large deletions and insertions observed in the shuffled gene products occur mostly in the apical domain (Figure 3.6). Tolerance of the apical domain to variations can be reconciled with the observation that GroEL can recognize a wide repertoire of substrates with almost no sequence or structural specificity. Thus the said domain may bear a latent plasticity.

Domain swapping studies show that an inactive chaperonin, recombinant Mtb GroEL1, can be rendered active by replacement of its native equatorial domain with the corresponding one from an active chaperonin, *E. coli* GroEL. Furthermore, *E. coli* GroEL can be rendered non-functional and indistinguishable from Mtb GroEL1 by replacing its equatorial domain with the one present in GroEL1 (Figure 3.12). Since GroELMEF can substitute for the lack of GroEL function in *E. coli* (Figure 3.19), it suggests that despite the heterologous apical domain of Mtb GroEL1 borne on GroELMEF, the chimeric protein (and therefore Mtb GroEL1) may recognize the same cellular substrates in vivo, as does its *E. coli* counterpart. However, the observation that Mtb GroELMEF could not function in the presence of Mtb GroES (Figure 3.20) needs to be probed further in future.

The genetic studies reported in this chapter were followed by in vitro biochemical characterizations of the mutants to understand the molecular basis of the observed phenotypes. Results of these biochemical characterizations are discussed in chapter VI.

3.6 References

- Amann, E., Ochs, B. and Abel, K. J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301-315

- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. and Sigler, P. B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371,** 578-586

- Ellis, R. J. (1987) Proteins as molecular chaperones. *Nature* **328**, 378–379

- Farr, G. W., Fenton, W. A. and Horwich, A. L. (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. USA* **104**, 5342-5347

- Georgopoulos, C. P., Hendrix, R. W., Kaiser, A. D. and Wood, W. B. (1972) Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. *Nat. New. Biol.* **239**, 38-41

- Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**, 4121-4130

- Hayer-Hartl, M. K., Martin, J. and Hartl, F. U. (1995) Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding. *Science* **269,** 836-841

- Hemmingsen, S. M.,Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, G. C., Hendrix, R. W. and Ellis, R. J. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**, 330–334

- Kondoh, O., Takasuka, T., Arisawa, M., Aoki, Y. and Watanabe, T. (2002) Differential Sensitivity between Fks1p and Fks2p against a Novelβ-1,3-Glucan Synthase Inhibitor, Aerothricin1. *J. Biol. Chem.* **277**, 41744–41749

- McLennan, N. and Masters, M. (1998) GroE is vital for cell-wall synthesis. *Nature* **392**, 159.

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

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

- Sikorski, R. S. and Boeke, J. D. (1991) In vitro mutagenesis and plasmid shuffling: from cloned gene to mutant yeast. *Methods Enzymol.* **194**, 302-318

- Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747-10751

- Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U. and Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903-914

- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882

- Tilly, K. and Georgopoulos, C. (1982) Evidence that the two *Escherichia coli groE* morphogenetic gene products interact in vivo. *J. Bacteriol.* **149**, 1082-1088

- Wang, J. D., Herman, C., Tipton, K. A., Gross, C. A. and Weissman, J. S. (2002) Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell* **111**, 1027-1039
- Warrens, A. N., Jones, M. D. and Lechler, R. I. (1997) Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186**, 29-35
- Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. and Horwich, A. L. (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **83**, 577–587
- Xu, Z., Horwich, A. L. and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)$_7$ chaperonin complex. *Nature* **388,** 741-750
- Zeilstra-Ryalls, J., Fayet, O., Baird, L. and Georgopoulos, C. (1993) Sequence analysis and phenotypic characterization of *groEL* mutations that block lambda and T4 bacteriophage growth. *J. Bacteriol.* **175**, 1134-1143
- Zhao, H. and Arnold, F. H. (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307-1308

Chapter IV

Biochemical Characterization of GroEL variants

Evidence for Phosphorylation Mediated Oligomerization in Mycobacterial GroEL1

4.1 Introduction

Molecular chaperones belong to a distinct class of proteins that are ubiquitous and conserved throughout different life forms. These proteins facilitate the correct folding, assembly, transport and degradation of other proteins *in vivo* (Saibil and Ranson, 2002)*.* Chaperonins are a group of molecular chaperones which form large oligomeric structures and mediate the folding of substrate proteins via encapsulation within a cavity formed due to their oligomeric structures. Critical features of chaperonin function are characterized by ATP driven cycles of binding, encapsulation and controlled release of substrate polypeptides leading to productive folding (Roseman et al., 1996; Rye et al., 1999; Sigler et al., 1998). Detailed studies have revealed mechanistic and physiological characteristics of isologus ring form of chaperonins, GroEL/S (Tilly and Georgopolous 1982; Richardson et al., 1998; Chen and Sigler, 1999; Tirumalai and Lorimer, 2001). According to the current understanding, the following properties of GroEL are significant in controlling its activity: i) oligomerization mediated by the equatorial domain, resulting in the formation of folding chamber and encapsulation of substrate polypeptides (Sigler et al., 1998; Wang et al., 1998), ii) recognition of substrate polypeptides mediated by the apical domain (Buckle et al., 1997; Stan et al., 2006), iii) conformational changes between the said two domains driven by ATP and GroES binding/release (Xu et al., 1997; Yokokawa et al., 2006; Kipnis et al., 2007), and iv) ATP hydrolysis (Todd et al., 1994; Galan et al., 2001). Impairment of any of these properties significantly alters the functioning of GroEL (Mendoza et al., 2000; Qamra and Mande, 2004).

Recent genome annotation studies on various bacteria have revealed that a few bacterial genomes bear potential to encode multiple copies of *groEL* genes (Fisher et al., 1993; Karunakaran et al., 2003; Barreiro et al., 2005). The *Mycobacterium tuberculosis* (Mtb) genome bears two copies of *groEL* genes (*groEL*s). One of these, *groEL1,* is arranged in an operon, with the cognate co-chaperonin *groES*, being the first gene, while the second copy, *groEL2* exists separately on the genome (Goyal et al., 2006; Qamra et al, 2004). Recombinant mycobacterial GroELs were shown to possess biochemical features that deviated significantly from the trademark properties of *E. coli* GroEL. The most striking feature of Mtb GroELs, however, was their oligomeric state, where contrary to expectations, *in vitro* they did not form the canonical tetradecameric assembly when

purified from *E. coli*. The proteins rather existed as lower oligomers (dimers) irrespective of the presence or absence of cofactors such as the cognate GroES or ATP (Qamra and Mande, 2004; Qamra et al., 2004). Furthermore, they displayed weak ATPase activities and GroES independence in preventing aggregation of the denatured polypeptides.

In the in vivo functional studies discussed in chapter 3, GroELMEF, which bears the substrate interacting apical domain from *M. tuberculosis* GroEL1, appeared to complement *E. coli* GroEL. Whereas, GroELMER despite having the apical domain from *E. coli* GroEL failed to complement the loss of *E. coli* GroEL. In addition, the gene shuffled GroEL variants, GroELSp24 and GroELSp32, were able to complement *E. coli* GroEL despite atypical variations in their apical domains. In order to understand the basis for the behavior exhibited by the GroEL variants, it was therefore important to understand the mechanism of action of these proteins.

Biochemical and biophysical characterization of proteins attempts to explore the mechanism of action of these macromolecules and thereby offers complete understanding of biological functions in vivo. Availability of proteins in their pure form and in sufficiently large quantities becomes a prime requirement for accomplishment of the different in vitro studies. Recombinant DNA technology has made it possible to obtain large amounts of any desired protein. Cloning of the desired gene in a suitable expression vector and its overexpression in a bacterial or eukaryotic host allows the production of the protein of interest. Moreover, the current protein expression methods have been highly effective in production of recombinant proteins, thus making possible their availability in large quantities for the various studies intended.

The genes encoding *E. coli* GroEL, GroELMEF, GroELMER, GroELSp24, GroELSp32 and *M. tuberculosis* GroEL1 were cloned into expression vectors. These were over expressed and the resulting over produced proteins were purified from *E. coli.* Detailed biochemical characterizations performed on these molecules is presented in this chapter.

4.2 Materials

The chaperonin variants were over-produced in *E. coli* BL21 (DE3) using Auto Induction System I (EMD Biosciences Inc.). All the chemicals, reagents and enzymes employed during the project were purchased from commercial sources. Several resins employed for purification of proteins were purchased from GE Life Sciences and the purification was carried out on Duoflow FPLC (Bio-Rad Laboratories, USA) and Pharmacia FPLC.

M. tuberculosis H37Rv was cultured in Middlebrook 7H9 broth (Difco) supplemented as appropriate with ADC (BD Biosciences). Reagents and films for developing and detection of signal in immunoblotting experiments were purchased from GE Life Sciences. Antibody IT56 was procured via TB Vaccine Testing and Research Materials Contract, Colorado State University, USA. Anti-Cpn60.1$_{Mtb}$ was a kind gift from ARM Coates. Phosphoaminoacyl specific antibodies, Anti-phosphoserine polyclonal antibody and anti-phosphothreonine monoclonal antibody were purchased from Assay Designs, USA and Rockland Immunochemicals, USA respectively.

4.2.1 Estimation of Concentration of Proteins by Bradford's Assay

Concentration of purified protein samples and the total protein content of *M. tuberculosis* cell lysate were estimated by Bradford's Protein assay (Bradford, 1976). In a reaction of 1 ml, 800 μl of protein sample was mixed with 200 μl of 5X Bradford's reagent (0.05% Coomassie Brilliant Blue G250, 25% ethanol and 50% phosphoric acid) and incubated at room temperature for five minutes. The concentrations of individual proteins were determined by reading optical density at 595 nm employing a Perkin Elmer Lambda 35 UV/Vis Spectrophotometer. A standard curve was set concurrently with each experiment with Bovine serum albumin (BSA) ranging from 20 to 100 μg.

4.3 Expression and Purification of Chaperonin Variants

For the purification of GroEL variants, ORF encoding the variants were cloned into *E. coli* expression vectors. ORF encoding Mtb GroEL1 was cloned into NdeI and BamHI sites and that encoding Mtb GroEL2 into NcoI and HindIII sites on pRSET B (Invitrogen) and the resulting plasmids were designated as pSCM1628 and pSCM1629. Similarly, the ORF encoding GroELMER was cloned into NdeI and EcoRV sites on pET-20b(+) and that encoding GroELSp24 into NcoI and HindIII sites of plasmid pTrc99A and the resulting plasmids were designated as pSCM1630 and pSCM1631, respectively. The clones were confirmed by digestions with restriction endonucleases (Figure 4.01). ORFs encoding *E. coli* GroES and GroEL were expressed from the pSCM1611 and that encoding GroELMEF from pSCM1615. ORF encoding GroELSp32 was expressed in the presence of 0.2 % L-arabinose from pSCM1632.
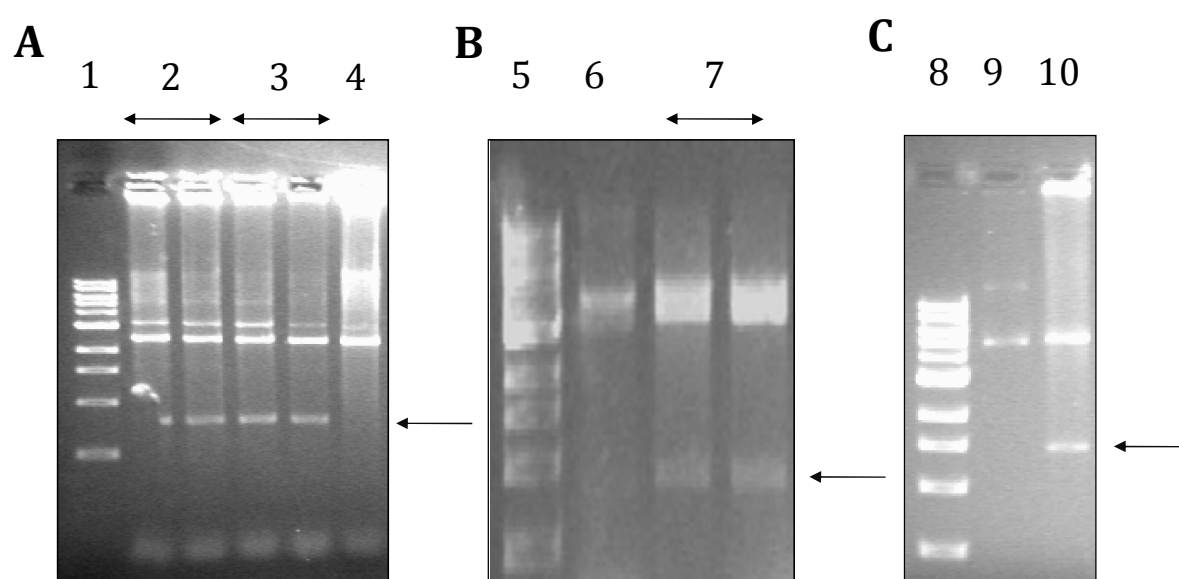
Figure 4.01: Clone Confirmation of GroEL Variants. Restriction digestion patterns of pSCM1628 and pSCM1629 (A), pSCM1630 (B) and pSCM1631 (C) showing the excised DNA fragment, as shown by the arrow, corresponding to *M. tuberculosis groEL1* (2), *groEL2* (3), *groELMER* (7) and *groELSp24* (10), respectively.

The said plasmids were transformed into BL21 (DE3) and the cloned genes were expressed using Auto Induction System I (Novagen). Purification of GroEL variants mentioned above was carried out in three steps:

4.3.1 Protein Extraction and Clarification by Ammonium Sulphate

Cells expressing the GroEL variants mentioned above were recovered at late log-phase, re-suspended in Lysis Buffer I, containing 50 mM Tris.HCl (pH: 8.0), 150 mM NaCl, 1 mM EDTA and 1 mM PMSF and the cells were lysed by sonication. The resulting soluble fraction bearing the GroEL variants were subjected to ammonium sulfate extraction. The protocol for purification of *E. coli* GroEL and GroES was followed as described earlier (Clark et al., 1998). *E. coli* GroEL and GroES were salted-in at 30% and salted-out at 65% saturated ammonium sulfate. GroELMEF and GroELSp32 were salted-out at 40% and 50% respectively, where as GroELSp24 was found to be soluble till 75% saturated ammonium sulfate hence a 40% salted-in sample was used for further purification. GroELMER was salted-in at 35% and the supernatant was used for further processing.

4.3.2 Enrichment Using Ion Exchange or Hydrophobic Interaction Chromatography

All proteins were desalted using PD10 columns (GE Life Sciences). *E. coli* GroEL and GroES were purified further by employing ion exchange chromatography, using HiLoad 16/60 Q Sepharose HP column (GE Life Sciences). The protein sample pre-equilibrated with the loading Buffer I, containing 50 mM Tris.HCl (pH: 8.0) and 1 mM EDTA, was loaded onto the column and eluted by applying step gradient of 50 mM NaCl increments. *E. coli* GroES and GroEL were eluted at 250 mM NaCl and 400 mM NaCl, respectively.   GroELSp24, GroELMEF and GroELMER were purified employing hydrophobic interaction chromatography using Hiprep 16/60 Phenyl FF column (GE Life Sciences) and GroELSp32 using Hiprep 16/60 Octyl FF column (GE Life Sciences). The said proteins pre-equilibrated with the loading Buffer II containing 50 mM Tris.HCl (pH: 8.0), 1 M NaCl and 1 mM EDTA, were loaded onto the appropriate column and were eluted by subjecting to a linear negative gradient of NaCl concentration. GroELSp24 and GroELSp32 were eluted at 50 mM Tris·HCl (pH: 8.0), 1 mM EDTA. GroELMEF and GroELMER did not elute in NaCl-less conditions either and hence were eluted in Milli-Q grade water.

4.3.3   Determination of Quaternary Structure

All the GroEL variants were pre-equilibrated with the loading Buffer III, containing 50 mM Tris.HCl (pH: 8.0), 150 mM NaCl and 1 mM EDTA. The said protein samples were further loaded onto HiPrep 16/60 Sephacryl S-300 HR column (GE Life Sciences). The elution profiles of these GroEL variants were further used for estimating molecular weight and oligomeric status.
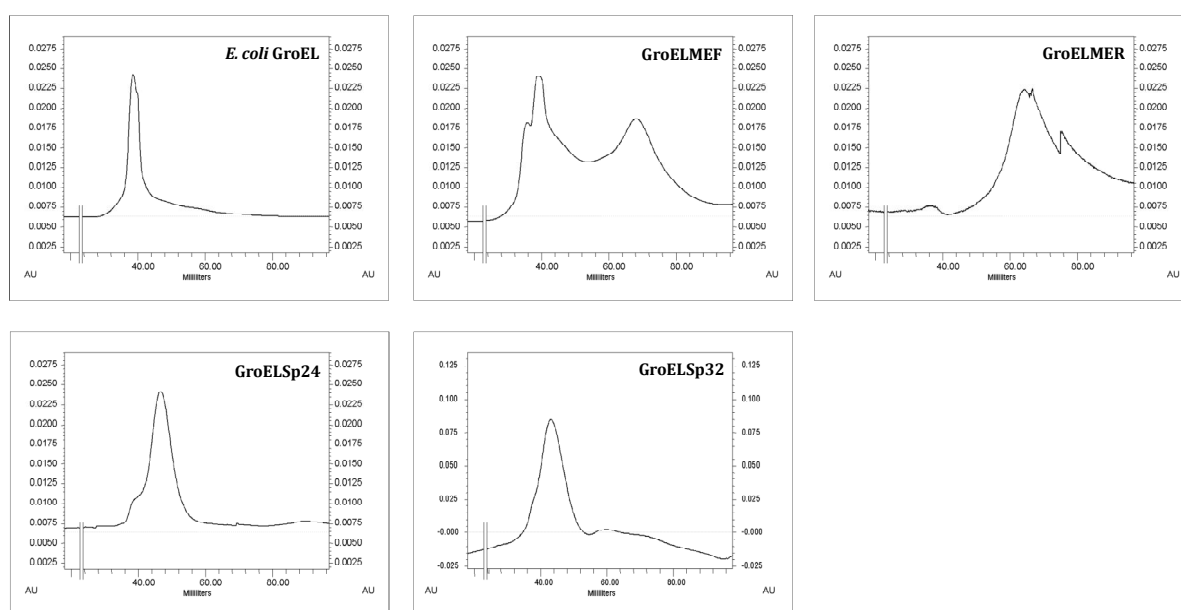
Figure 4.02: Oligomeric States of GroEL Variants. Gel filtration chromatograms of *E. coli* GroEL, GroELMEF, GroELMER, GroELSp24 and GroELSp32. The indicated proteins were separated on HiPrep 16/60 Sephacryl S-300 HR column (GE Life Sciences) in Biologic DuoFlow FPLC System (Bio-Rad Laboratories).

4.3.4 The Equatorial Domain of GroELMEF is Responsible for Attaining a Higher Oligomeric State

Having shown that GroELMEF is functional in vivo, we wished to test whether the ability of GroELMEF in complementing a defect in *groEL44* allele, is a consequence of existing in higher-order oligomeric form and thus being able to form cavity for encapsulation, features that are lacking in the parental Mtb GroEL1 (Qamra et al., 2004). Since the polypeptides encoded by the gene shuffled variants of *groEL, groELSp24* and *groELSp32* bear considerable variation (Figure 3.07), similar attributes of the encoded

polypeptides were studied. Gel filtration studies with HiPrep 16/60 Sephacryl S-300 HR column (GE Life Sciences) showed that GroELMEF was capable of existing in a higher oligomeric state, similar to that seen for *E. coli* GroEL, whereas GroELMER displayed a lower oligomeric character (Figure 4.02), a property reminiscent to that seen for *M. tuberculosis* GroEL1 (Qamra et al., 2004). One noticeable aspect on the oligomeric properties of GroELMEF is that it existed in equilibrium between higher and lower oligomeric states, with the higher oligomeric state being the predominant species, which is presumably related to its weakened ability to substitute for *E. coli* GroEL in vivo at 45 °C (Figure 3.12). Since the higher oligomeric state of GroELMEF displays gel permeation characteristics similar to that of *E. coli* GroEL, it is reasonable to presume that the said state corresponds to a teteradecameric assembly. We have also examined the oligomeric states of GroELSp24 and GroELSp32, variants that bear deletions in the putative intermediate-apical domain boundary. Both the variants displayed tendencies to exist in a higher oligomeric state and their gel permeation profiles were consistent with predicted molecular weights.

4.4 Estimation of Secondary Structural Composition of GroEL Variants

Quaternary structure information on the GroEL variants established their correlation to in vivo activity. GroELMER, which existed in a dimeric form, was not able to exhibit activity. On the other hand, the gene shuffled variants, GroELSp24 and GroELSp32, despite losing significant portion of the apical domain were capable of existing as higher order oligomers and hence were capable of exhibiting activity. The possibility that the dimeric nature of GroELMER and the instability of GroELMEF in exhibiting oligomer-monomeric equilibrium might have arisen due to the loss of secondary and tertiary structure was tested by measuring the CD spectrum of the proteins. In addition, owing to the loss of helices H and I in the apical domains of GroELSp24 and GroELSp32, CD spectrum for these proteins was measured to check the extent of secondary structure information.

Secondary structural compositions of the purified GroEL variants were studies by Circular Dichroism spectroscopy using the standard procedures (Motojima and Yoshida, 2003; Boris et al., 2005). CD spectra of the variants were recorded using a Jasco J-715 spectropolarimeter at 25 °C. The GroEL variant proteins at 100- 150 nM each were adjusted to 10 mM Tris.Cl buffer (pH 8.0). Far UV-CD spectrum from 240 nm through 190 nm was recorded with a path length of 0.1 mm. Secondary structure composition of each mutant was estimated according to Chen and Yang, 1971 and Reed and Reed, 1997.

The far UV-CD spectrum of the GroEL variants (Figure 4.03) is characteristic of highly helical proteins with signature bands for helical structure at 208 and 222 nm, although GroELSp24 and GroELSp32 showed lower helical nature (Johnson, 1990). The CD data suggest that the all the GroEL variants possess significant secondary structure conformation. Hence, differences in the oligomeric nature exhibited by the variants GroELMEF and GroELMER do not appear to be due to absence of secondary structure of the proteins. Moreover, despite bearing lower helical content, the gene shuffled GroEL variants, GroELSp24 and GroELSp32 retained significant secondary structure information.
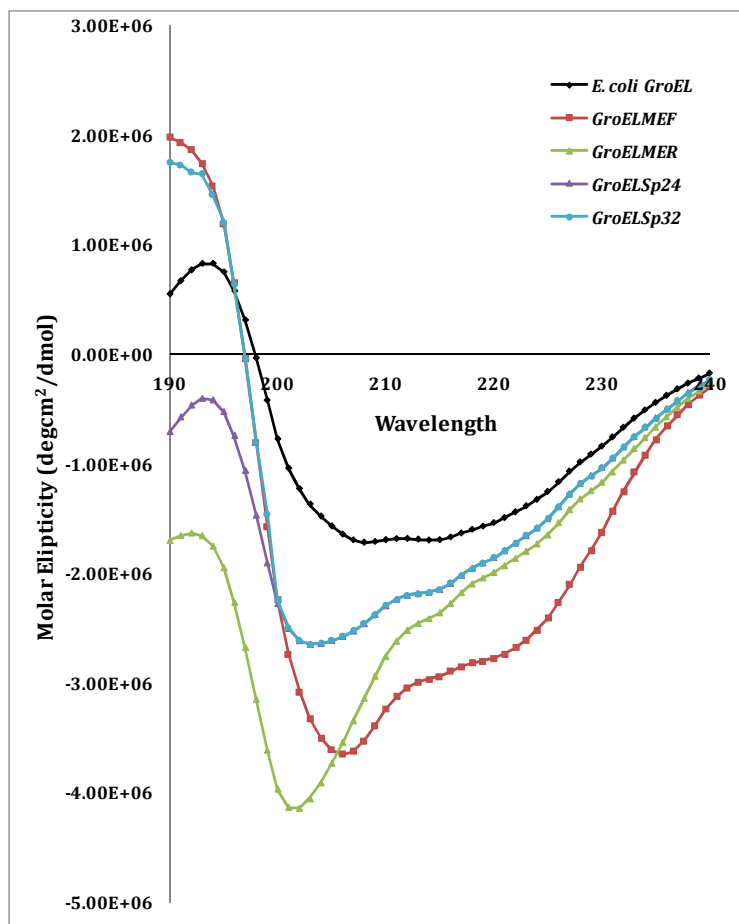
Figure 4.03: Far UV-CD Spectrum of GroEL Variants. GroEL variants at 100 -150 nM were equilibrated with 10 mM Tris (pH: 8.0) and molar ellipticity for each variant was measured versus wavelength in far UV range (190 nm - 240 nm) at 25 °C. Mean residue ellipticity was calculated and plotted against wavelength.

4.5 Estimation of Exposed Hydrophobicity by bis-ANS Fluorescence Assay

Apical domains in a GroEL oligomer are known to possess exposed hydrophobic area where substrate polypeptides and GroES have been shown to bind (Xu et al., 1997). Due to the distinct oligomeric features exhibited by the GroEL variants, presence and extent of such hydrophobic patches on the surface of the GroEL variants was probed by the ability to bind to 1,1-bis (4-anilino) naphthalene-5, 5'-disulfonic acid (bis-ANS).

Exposed hydrophobicity on each of the chaperonins was estimated as binding of equimolar ratios of 4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid (bis-ANS, $C_{32}H_{22}K_2N_2O_6S_2$) and the chaperonins according to standard procedures (Shi et al., 1994; Lee et al., 1997; Sharma et al., 1998; Panda et al., 2001; Smoot et al., 2001). Briefly, binding of bis-ANS with GroEL variants and *E. coli* GroES at 20 μM each was monitored by exciting the probe at 395 nm and recording the emission spectra in the range of 400-600 nm. The fluorescence intensity measurements were carried out in 100mM Tris-Cl (pH: 8.0) at 25 °C on an FP-6500 High Performance Research Fluorescence Spectrophotometer. Buffer alone and bis-ANS alone reactions were set as controls.

As shown in Figure 4.04, the chaperonin variants showed comparable exposed hydrophobic nature, although GroELMER exhibited slightly higher exposed hydrophobic area, which can be correlated with its dimeric nature. The fluorescence enhancement exhibited by the chaperonin variants is comparable to that observed for *E. coli* GroEL. The results therefore clearly indicate the presence of hydrophobic surfaces on the GroEL variants.
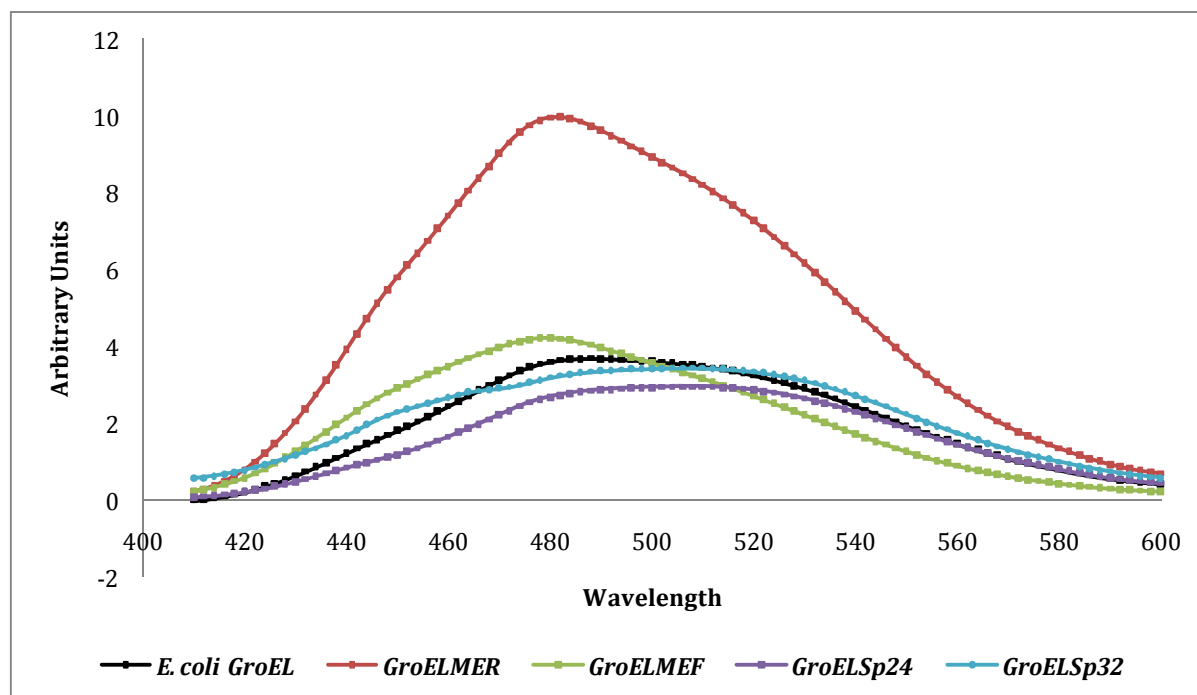
Figure 4.04: Assessment of Exposed Hydrophobicity of GroEL Variants.
Fluorescence intensity increase upon interaction of individual GroEL variants
with bis-ANS was measured from 400 nm - 600 nm. The spectral values were
subtracted for buffer and bis-ANS alone.

4.6 ATPase Activity Assay

Binding to ATP and hydrolysing it, is an essential feature of the GroEL mediated protein folding cycle. *E. coli* GroEL binds to ATP via the pseudo-Walker motif located in its equatorial domains and possesses a weak potassium ion (K^+) stimulated ATPase activity (Viitanen et al., 1990). Owing to the distinct quaternary structures displayed by the GroEL variants and moreover, since the pseudo-Walker motif is similar in all the variants, measuring the ATPase activity of these proteins was an important test for their functional characterization.

ATPase activity of the purified GroEL variants was quantified by a colorimetric assay performed as described (Henkel et al., 1988; Viitanen et al., 1990). Briefly, 50 µl of the reaction buffer containing 50 mM Tris.HCl (pH 8.0), 10 mM KCl, 10 mM $MgCl_2$ and 2.5 µM of GroEL variant individually was incubated with 1 mM ATP at 37 °C for 20 minutes. Enzymatic reactions were terminated by addition of 200 µl of the acidic solution of Malachite green containing 741.5 µM Malachite green oxalate ($C_{52}H_{54}N_4O_{12}$), 291.8 mM Ammonium molybdate ($(NH_4)_6Mo_7O_{24}.4H_2O$), 1 M HCl and 0.387% Polyvinyl alcohol ($(C_2H_4O)_n$). Amount of inorganic phosphate liberated was measured at 655 nm using Nanodrop ND-1000. Control reactions without ATP and GroEL were performed. A standard curve with 200 - 1000 µM monobasic potassium phosphate was generated concurrently with each experiment.

As shown in the Figure 4.05, GroELMEF and the gene shuffled variants, GroELSp24 and GroELSp32, displayed significantly higher intrinsic ATPase activities. On the other hand, GroELMER displayed weak ATPase activity. Since all the variants have similar ATP binding pocket, variation in the hydrolyzing ability might be correlated to the requirement of oligomerization for ATP hydrolysis.
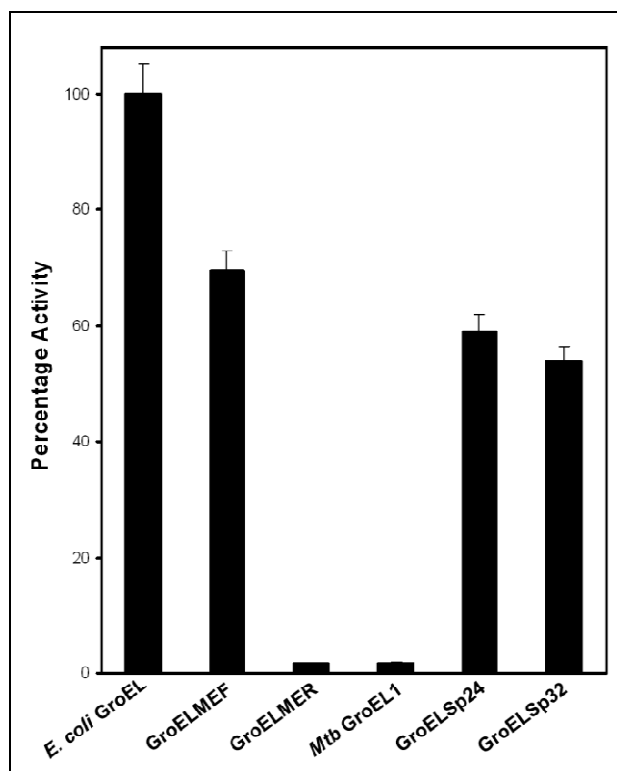
Figure 4.05: ATPase Activities of Purified GroEL Variants. The indicated GroEL variants were assayed by malachite green calorimetric assay. Amount of inorganic phosphate released was quantified at 655 nm. Mean for individual data sets was calculated and plotted along with the standard deviation considering activity of *E. coli* GroEL as 100%.

4.7 Prevention of Aggregation of Citrate Synthase by Chaperonins

GroEL assisted folding of substrate proteins typically involves i) binding of the polypeptides by virtue of exposed hydrophobic interactions, ii) sequestration of polypeptides into the cavity during stress, thereby preventing irreversible aggregation. The former is typically studied by measuring prevention of aggregation of substrate polypeptides, while the latter can be studied by monitoring recovery of folded substrate proteins. *E. coli* GroEL has been shown to facilitate refolding of denatured citrate synthase by suppressing its aggregation (Buchner et al., 1998). Exposed hydrophobic patches on the GroEL variants suggested that these might exert similar effect on substrate proteins. Therefore, the effect of various chaperonins on the aggregation of Citrate synthase was monitored.

Prevention of aggregation of pig heart Citrate synthase at elevated temperature, by various chaperonin homologues was studied as described earlier (Buchner et al., 1998). The GroEL variants were suspended in 40 mM HEPES (pH: 7.5) for this reaction. 0.15 μg/ml citrate synthase was incubated at 43 °C in the presence or absence of equimolar oligomer ratios of different GroEL variants, in 40 mM HEPES–KOH buffer (pH: 7.5). Ability of the chaperonin variants in preventing of aggregation of citrate synthase was monitored on-line for 20 minutes on Hitachi F-4000 spectrofluorimeter with emission and excitation wavelengths set at 465 nm and corresponding band passes set at 3.0 nm. Temperature of the sample was maintained with Julabo circulating water-bath and was monitored using Physitemp type T microcouple.

Remarkably, GroELMEF was poor in preventing aggregation of the substrate protein, citrate synthase (CS), at elevated temperatures for more than five minutes. On the other hand GroELMER prevented CS aggregation for at least 20 minutes (Figure 4.06). Similarly, GroELSp32 was able to prevent aggregation of citrate synthase whereas GroELSp24 was not (Figure 4.06).
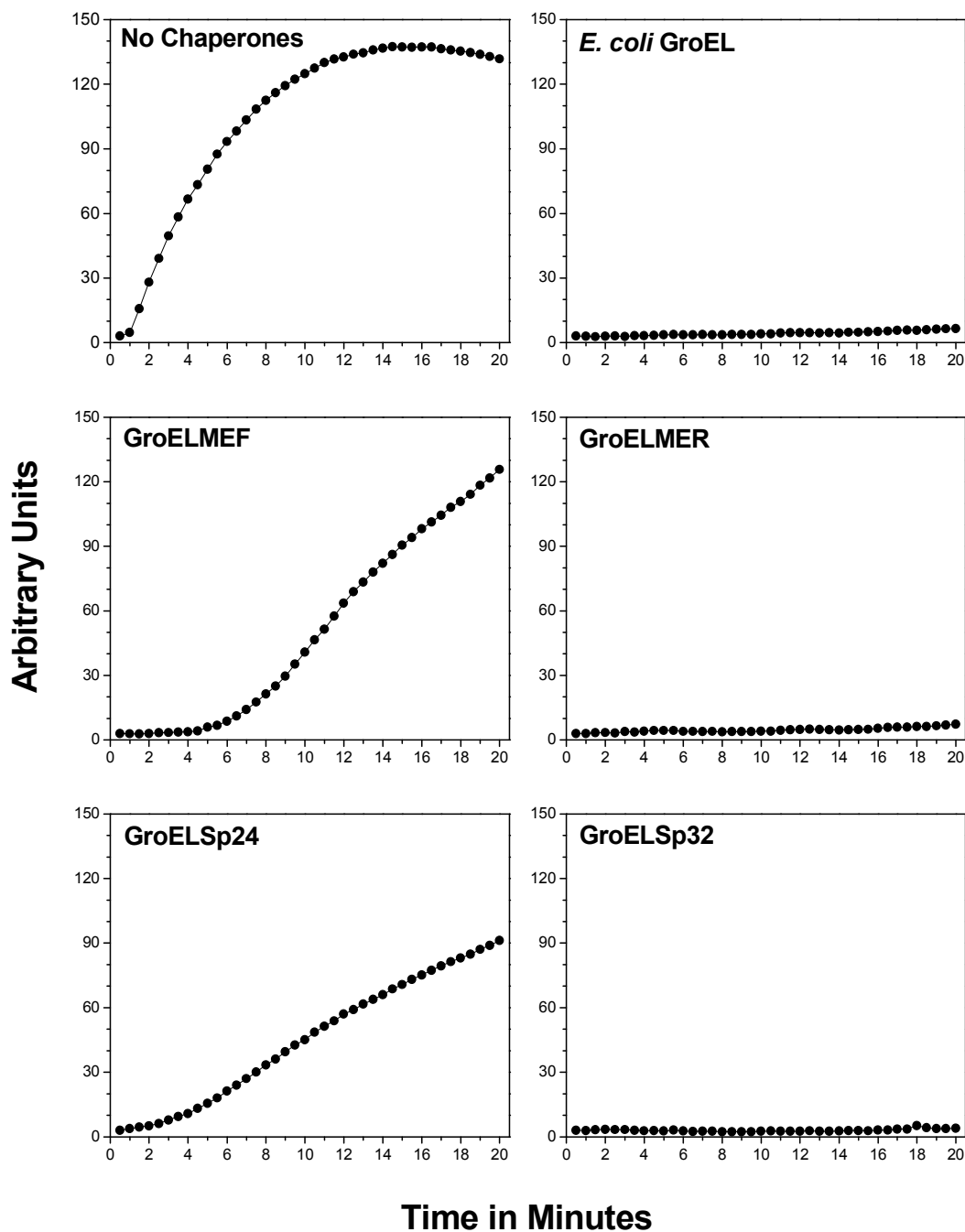
Figure 4.06: Prevention of Aggregation of Citrate Synthase by the Chaperonin Variants as a Function of Time. Aggregation of citrate synthase at 43 °C in the absence or the presence of equimolar ratios of the indicated GroEL variants was measured as a function of light scattered at 465 nm for twenty minutes.

4.8 Chaperonin Assisted Refolding of Chemically Denatured Citrate Synthase

The protocol for denaturation, refolding and assaying the activity of pig heart mitochondrial Citrate synthase (CS, EC 4.1.3.7) was modified from Buchner et al., 1998, which involves four steps:

4.8.1 Preparation of CS

400 µL of the ammonium sulfate slurry of CS was centrifuged in 1.5 mL tubes at 12000 rpm for 10 min, resulting pellet was resuspended into 400 µL of 20 mM potassium phosphate buffer (pH: 7.4) and transferred to YM-10 Centricon concentrator. The protein solution was brought to a volume of 1.5 ml with 20 mM potassium phosphate buffer (pH: 7.4), and centrifuged at 2800 rpm for 30 min at 4°C till the volume comes to ~100 µl. This step was repeated three times to remove residual ammonium ions. Concentrated CS was collected using the manufacture's protocol and concentration was determined spectroscopically at 280 nm, using molar extinction co-efficient of CS as 8.9×10^4/M.cm.

4.8.2 Dentaturation of CS

CS at 15 µM was denatured for 30 minutes at 25 °C in a 50 µl reaction mixture containing 100 mM Tris.HCl (pH: 8.0), 6 M Guanidine.HCl and 1 mM DTT. An additional mock denaturation mixture was prepared by substituting 20 µL potassium phosphate buffer (pH: 7.4) and CS.

4.8.3 Renaturation of CS

The denatured CS was diluted 100-fold into a 100 µl refolding buffer containing 20 mM Potassium phosphate buffer (pH 7.5), 10 mM MgCl$_2$, 1 mM Oxaloacetic acid (C$_4$H$_4$O$_5$) and 1 µM of different GroEL variants in the presence or absence of 2 µM *E. coli* GroES. 2 mM ATP was added to initiate the chaperone mediated refolding reaction. The reaction was further incubated at 37 °C for 1 h.

4.8.4 Activity Assay for CS

CS catalyzes the condensation of Acetyl Coenzyme A ($C_{23}H_{38}N_7O_{17}P_3S$) and Oxalacetic Acid to liberate Citric acid as shown in the following reaction;

$$Oxaloacetate^{2-} + Acetyl\ Coenzyme\ A + H_2O \longleftrightarrow Citrate^{3-} + CoA + H^+$$
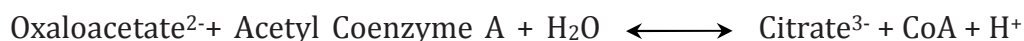
Hence, CS activity was measured as the decrease in absorption at 233 nm owing to the disappearance of Acetyl Coenzyme A, for 90 seconds (Srere et al., 1966; Steede et al., 2000). 100 µl of the reaction mix contained 10 µL of 4 mM Oxalacetate, 10 µL of 1.6 mM Acetyl Coenzyme A, 75 µL of 20 mM potassium phosphate buffer (pH: 7.4), all pre-equilibrated to room temperature. The reaction was initiated by adding 5 µL of refolding mixture, and the absorbance was monitored at 233 nm for 0–90 seconds using a SpectraMax Plux[368] UV visible Spectrophotometer (Molecular Devises). The percentage of CS recovered after renaturation was calculated from slope of the curve and considering the activity exhibited by mock denatured Citrate synthase as 100%.
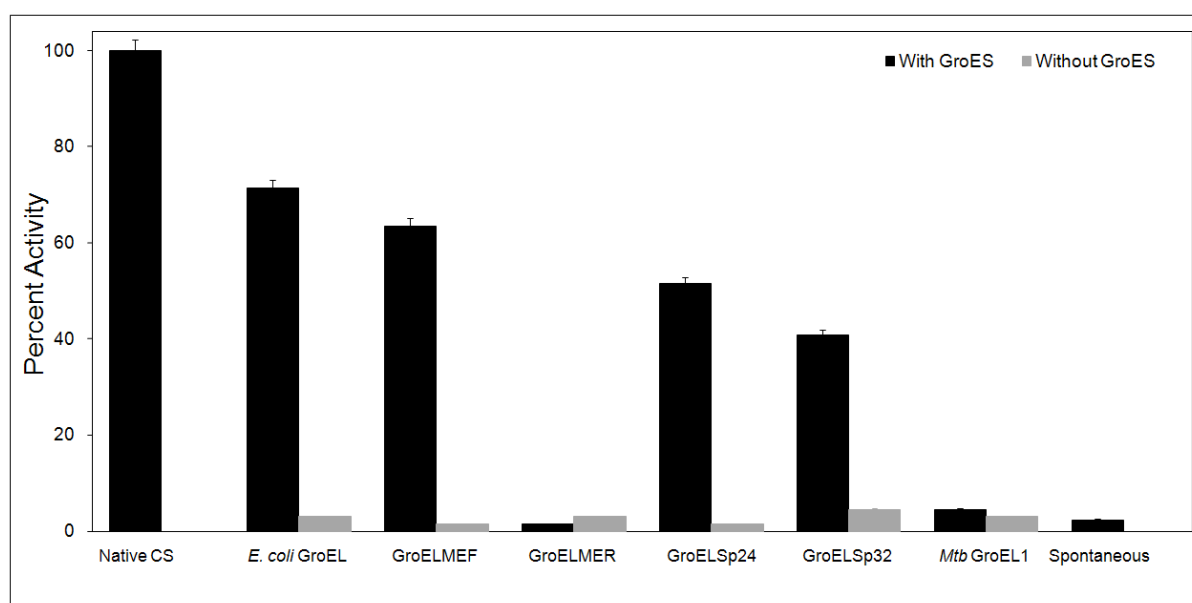
Figure 4.07: Effect of Chaperonins in Refolding Chemically Denatured Citrate Synthase. Citrate synthase was chemically denatured, refolded with the indicated GroEL variants in the presence or absence of *E. coli* GroES. Activity of the refolded enzyme was measured after 60 min of refolding at 25 °C as decrease in absorbance at 233 nm due to the consumption of acetyl CoA. Yield of refolded enzyme is expressed as percentage of the activity determined for an equal quantity of non-denatured native CS.

As shown in the Figure 4.07, GroELMEF and the gene shuffled variants, GroELSp24 and GroELSp32, which form higher order oligomers, were capable of refolding the model substrate protein, citrate synthase (CS). On the other hand, GroELMER was weak in refolding CS.

Moreover, GroELMEF despite not being able to efficiently prevent substrate aggregation was able to fold substrates due to its competence in encapsulation. On the other hand GroELMER, despite being able to prevent aggregation, was not able to promote refolding of substrates due to its inability to encapsulate. Similarly, GroELSp32 was able to prevent aggregation of citrate synthase whereas GroELSp24 was not (Figure 4.06). The results on GroELMER suggest that lower oligomeric GroELs possess substrate binding activity, but yet are in inefficient in promoting refolding. These observations therefore reinforce a correlation between oligomerization ability and chaperonin function.

4.9 Analysis of GroEL in M. tuberculosis Lysates

The phenotypic and biochemical analyses of GroEL variants obtained by domain swapping studies suggest that the impaired chaperoning ability of recombinant *M. tuberculosis* GroELs is a consequence of their inability to form higher order oligomers in *E. coli* and that oligomerization is the prelude to the formation of an active GroEL chaperonin. Moreover, evolutionary studies on *M. tuberculosis groEL* sequences have suggested rapid evolution of the *groEL1* gene, yet not turning these into pseudogenes (Goyal et al., 2006). The other hypothesis suggests that *M. tuberculosis* being a slow growing organism might require GroEL function that does not utilize ATP rapidly, but rather with a slow turnover rate. Alternately, additional mechanisms might exist in *M. tuberculosis*, which could mediate regulated oligomerization of *M. tuberculosis* chaperonins. Such regulation might help in the controlled utilization of ATP in nutrient deprived *M. tuberculosis*, as observed for other chaperones such as small heat shock proteins (Haslbeck et al., 2005). Hence, we tested the nature of oligomerization of native *M. tuberculosis* GroELs and explored the possibility of existence of regulated oligomerization for these molecules in their natural setting.

For these studies we considered three aspects concerning the biology of *M. tuberculosis* and GroEL. First, multiple sequence alignment of bacterial GroEL homologues showed that several positions in Mtb GroELs' equatorial domains were altered from different residues to serine\threonine indicating potential phosphorylation sites (Qamra and Mande, 2004; Qamra et al., 2004). Second, *M. tuberculosis* genome encodes 11 eukaryotic like serine threonine kinases (ELK) and emerging evidence that ELKs play a role in different cellular processes including stress adaptation across bacterial genera. This tempted us to speculate a role for ELKs in GroEL oligomerization (Zhang et al., 1998; Narayan et al., 2007; Wehenkel et al., 2008). Third, some heat shock proteins including mitochondrial and chloroplast Hsp60 homologues exhibit heat induced or concentration dependent oligomerization. The mitochondrial Hsp60 homologue although predominantly found as a single ring entity is known to display a concentration dependent tetradecamer formation (Dickson et al., 2000; Levy-Rimler et al., 2001). Hence we attempted to explore if *M. tuberculosis* GroEL1 exhibits regulated oligomerization and if so, to determine the source of its regulation.

4.10 Culturing M. tuberculosis for Cell Lysate Proteins

Culturing and extracting cell lysate proteins from *Mycobacterium tuberculosis* H37Rv (ATCC 27294) was followed under Bio-safety Level 3 (BSL-3) facility at the University of Delhi South Campus, New Delhi.

4.10.1 Starter Culture of *M. tuberculosis*

M. tuberculosis was streaked on a 7H11 agar plates and incubated at 37 °C for 24 days. Flasks containing 50 mL of Middlebrook 7H9 broth which was supplemented with Albumin Dextrose Catalase Complex (ADC) were inoculated with about 10^8 CFU/mL and incubated in an orbital shaker at 37 °C with constant shaking at 200 rpm for 10-15 days or until an OD_{600} of 1.6 - 2.0 is reached. A few drops of the culture from each flask were inoculated from the sample onto 7H11 agar plate to test for contamination by other bacteria, and the plate was incubated for 24 h at 37 °C. The following day, after inspecting the 7H11 agar plates, the flask corresponding to no contamination was used directly for starter culture.

4.10.2 Culturing *M. tuberculosis* for Cell Lysate Preparation

2 x 300 mL of 7H9 broth which was supplemented with ADC complex and Tween 80 was inoculated with the starter culture corresponding to 10^8 CFU/mL and incubated in an orbital shaker at 37 °C with constant shaking at 200 rpm. The culture was grown till mid- to late-log phase (OD_{600}: 0.9 - 1.2), which takes about 5-6 days. One flask containing the mid-log phase culture of *M. tuberculosis* was transferred to orbital shaker at 42 °C with constant shaking at 200 rpm for 2-6 hrs to induce heat shock response. The other flask was maintained at 37 °C.

4.10.3 Preparation of Cell Lysate

Cell lysate proteins of *M. tuberculosis* cells were recovered following standard protocols (Hirschfield et al., 1990; Lee et al., 1992). Briefly, *M. tuberculosis* cells from the cultures above were recovered by centrifugation at 5000 rpm for 15 min and were washed twice in 50 ml 1X PBS (pH: 7.2). The cells were resuspended in 1 ml of lysis buffer containing 50 mM Tris (pH: 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM PMSF. The cell suspension was transferred to 2 ml screw-cap microcentrifuge tube containing 1 ml of

acid washed 0.1 mm silica/ceramic beads. Cells were lysed using 6-10 pulses of 30 sec at high speed in the Mini Bead-Beater with intermittent cooling on ice. Cell lysate was separated from the cellular debris by centrifugation at 13000 rpm for 20 min. supernatant containing the cell lysate was filtered through 0.2 μm filter, to remove particulate substances and residual live cells, if any. Furthermore the lysate was subjected to ultra sonication to remove lipid content and the sonicate was centrifuged at 13000 rpm for 20 min. Protein fraction of the cell lysate was separated from the top lipid layer by collecting from the bottom of the tube. The lysate thus generated was estimated for the total protein content using Bradford's Assay for protein estimation.

4.11 Mycobacterial GroEL1 is Phosphorylated on Serine but not on Threonine

Earlier studies on the comparative analysis of *M. tuberculosis* GroELs with *E. coli* counterpart had shown the presence of systematic mutations of Glutamates to Serines or Threonines at different positions in their sequences (Qamra et al., 2004). This suggested a possibility that the Serine or Threonine residues are potential sites of phosphorylation. Therefore, investigating the presence and nature of phosphorylation was attempted by immunoprecipitation of GroEL1 and GroEL2 from Mtb cell-lysates followed by immunoblotting with anti-phosphoseryl and anti-phosphothreonyl antibodies.

4.11.1 Immuno-precipitation of *M. tuberculosis* GroEL1 and GroEL2

Using antibodies specific to Mtb GroEL1 and GroEL2, these proteins were immunoprecipitated from Mtb cell-lysates using standard protocols (Bonifacino et al., 2001). Briefly, 50-100 µg total protein of cell lysates from the normally grown (37 °C) and heat shocked (42 °C) cultures were mixed separately with anti-Cpn60.1$_{Mtb}$ and IT56 antibodies at equal titer and were incubated at 4 °C overnight with constant mixing. 40 µl of Protein A-coupled Sepharose beads (GE Life Sciences) that were pre-equilibrated in Buffer P (50 mM Tris (pH: 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM PMSF and 0.5% NP40) was added to each tube and was incubated for three hours at 4 °C with constant mixing. The Protein A-coupled Sepharose beads, which were bound by the antibody-GroEL complex, were collected by centrifugation and the pellet was washed thrice in Buffer P to remove the unbound cell proteins. Equal proportions of the protein-bound beads were resolved on 10 % SDS-PAGE to quantitative the level of GroELs in Mtb.
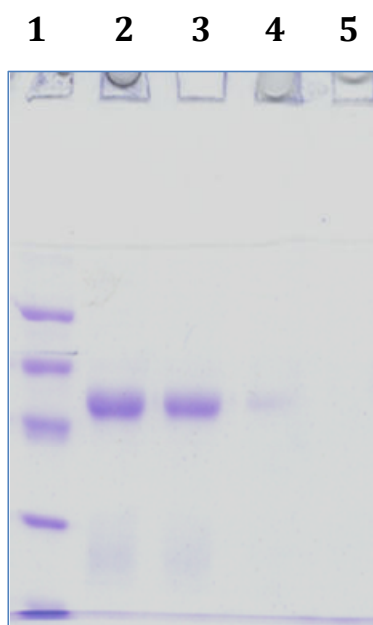
Figure 4.08: Immunoprecipitation of M. tuberculosis GroEL1 and GroEL2.
GroEL1 and GroEL2 were immuno-precipitated from *M. tuberculosis* cell-lysates using GroEL1 and GroEL2 specific antibodies. Equal fractions of the precipitated proteins were resolved on 10 % SDS PAGE and were stained in Coomassie brilliant blue R-250. The lanes are 1, Molecular weight marker; 2, Immunoprecipitated GroEL1 from cells grown at 37 °C; 3, Immunoprecipitated GroEL1 from cells grown at 42 °C; 4, Immunoprecipitated GroEL2 from cells grown at 37 °C; 5, Immunoprecipitated GroEL2 from cells grown at 42 °C.

4.12.2 Determination of Phosphorylation by Immunoblotting

Immunoprecipitated samples containing *M. tuberculosis* GroEL1 and GroEL2 were resolved on 10% SDS PAGE and were further probed by immunoblotting for the presence of phsphorylation using antibodies specific to phosphoseryl and phosphothreonyl residues. The said antibodies were used at 1:1000 dilutions.
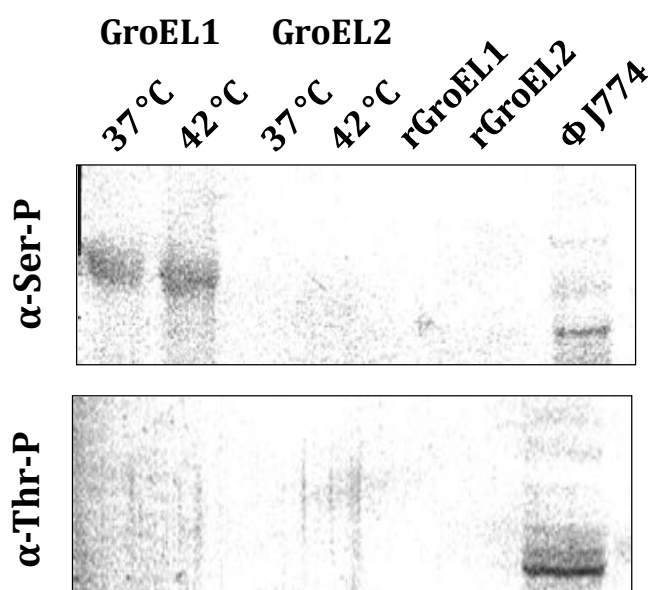
Figure 4.09. Detection of Phosphorylation of M. tuberculosis GroEL1 and GroEL2. GroEL1 and GroEL2 were immunoprecipitated from the *M. tuberculosis* cell-lysates. Equal fractions of the precipitated proteins were resolved on 10% SDS PAGE, followed by probing for phosphorylation using antibodies specific to phospho-seryl and phosphor-threonyl residues. Cell lysate from human macrophage cell-line J774 and recombinant *M. tuberculosis* GroELs, purified from *E. coli* were employed as positive and negative controls, respectively.

The data in Figures 4.08 and 4.09 clearly show that GroEL1 from *M. tuberculosis* is phosphorylated. The phosphrylation is on a Serine residue (s) but not on Threonine residue (s). Phosphorylation on GroEL2, however, was not detectable.

4.12 Mycobacterial GroEL1 Exhibits Phosphorylation Mediated Oligomerization

Having established that *M. tuberculosis* GroEL1 exhibits phosphorylation on Serine residue, we attempted to explore if *M. tuberculosis* GroEL1 exhibits regulated oligomerization and if so, to determine the role of phosphorylation in regulation of oligomerization.

4.12.1 Separation of Oligomeric Forms of GroEL1 by Gel Filtration

The cell lysates were prepared from normally grown (at 37 °C) and heat shocked (at 42 °C) cultures of *M. tuberculosis* that were normalized for OD_{600}. 2-3 mg total protein of the cell-lysates from both the cultures were resolved by gel filtration chromatography using Superdex S200 16/60 on AKTA FPLC chromatographic system (GE Life Sciences).
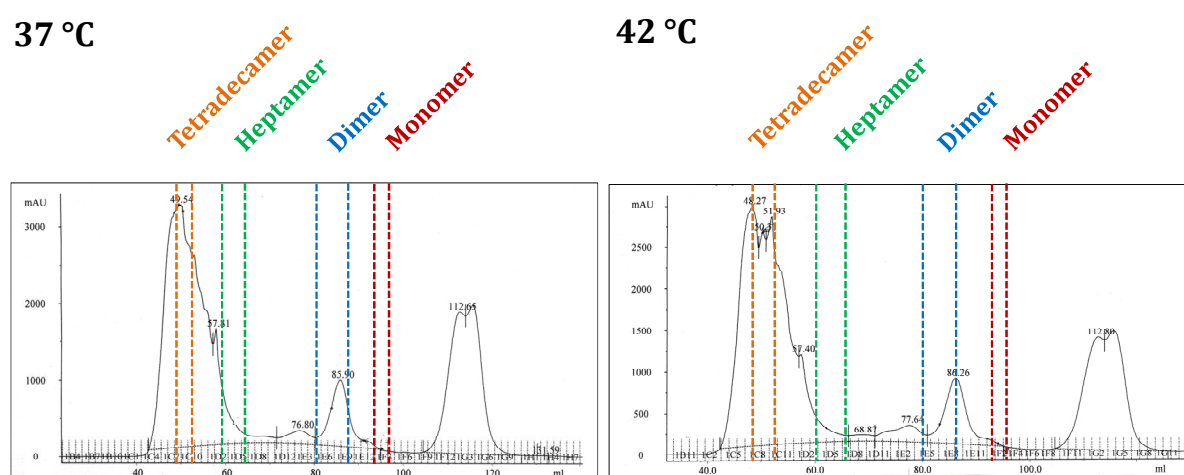
Figure 4.10: Separation of M. tuberculosis Cell Lysates by Gel-filtration. 2-3 mg of total cell protein from the cells grown at 37 °C and 42 °C were resolved on Superdex S200 16/60. This facilitated the segregation of different oligomeric forms of *M. tuberculosis* GroEL1. Fractions collected were further probed by immunoblotting for the presence of GroEL1. Peak regions of different oligomeric forms are indicated in different color.

4.12.2 Phosphorylation of Different Oligomeric Forms

Further, to determine the peak position of different oligomeric forms of *M. tuberculosis* GroEL1, each fraction from chromatography was resolved on 10 % SDS-PAGE and probed by immunoblotting, using anti-Cpn60.1$_{Mtb}$ (rabbit polyclonal) antibody at 1:10,000 dilution for the presence of *M. tuberculosis* GroEL1. Fractions bearing individual oligomeric forms of Mtb GroEL1, as compared to the molecular weight standards that were also resolved on gel filtration, were pooled and equal proportion of each pool was probed by immunoblotting using anti-Cpn60.1$_{Mtb}$, for quantitative determination of the level each oligomeric forms.
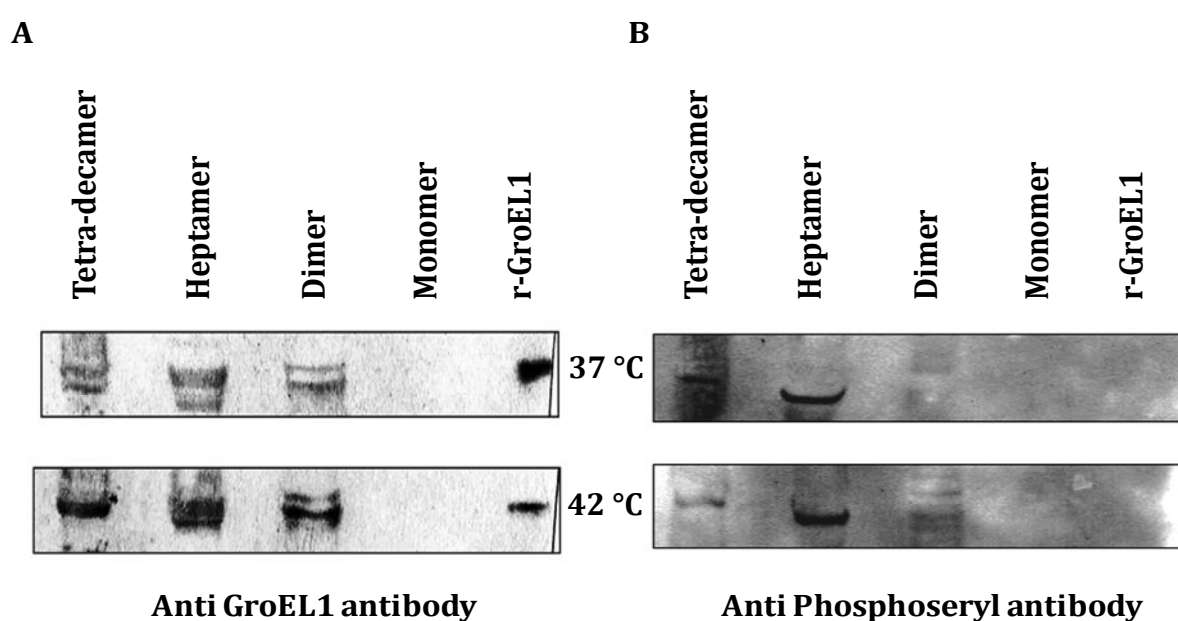
Figure 4.11: Detection of Phosphorylation of Multiple Oligomeric Forms of M. tuberculosis GroEL1. **(A)** Mtb cell lysates from cells grown at 37 °C and 42 °C were resolved on gel filtration. Equal proportions of the indicated oligomeric forms of GroEL1 from the cultures grown at indicated temperatures were probed with anti-GroEL1 specific antibody. **(B)** Different oligomeric forms of GroEL1, as indicated in part A, were resolved by 10% SDS PAGE and were probed for the presence of phosphorylation using antibody specific to Phospho-seryl residue. Mass spectrometric analysis of the protein corresponding to the band in the heptametrical fraction, which is migrating faster than GroEL1, revealed it to be Glutamine synthatase A1 (GlnA1, Rv2220) of *M. tuberculosis.*

The data in Figure 4.11 A clearly show that GroEL1 exhibited multiple oligomeric forms in its natural biological settings, which, to the best of our knowledge, is the first ever observation for bacterial chaperones. GroEL homologues from mammalian mitochondria and from choloroplasts have been shown to exhibit multiple oligomeric forms where the conversion from single ring form to double ring form is concentration and GroES dependent (Levy-Rimler et al., 2001; Dickson et al., 2000). Attempts to make single ring version of *E. coli* GroEL by site directed mutagenesis of four critical residues at the inter-ring surface had yielded mutants, SR1, that are compromised in GroES release (Weissman et al., 1995). Furthermore conversion of a temperature sensitive GroEL mutant, GroEL44 into the single ring form, SR44, however resulted in an active chaperonin (Chatellier et al., 2000). Studies on the assembly of bacterial chaperonin were attempted with purified *E. coli* GroEL involving co-factors and denaturing agents such as Mg^{++}, GroES, urea and guanidium chloride, which showed the existence of bacterial chaperonin in multiple oligomeric forms in various denaturing conditions tested (Kusmierczyk and Martin, 2001).

The tetradecameric form existed at both 37 °C and 42 °C (Fig 4.11). Presence of tetradecameric form in cultures grown at 37 °C indicates that the switch between heptameric to tetradecameric forms is not solely temperature-mediated. The dimeric form also was detected in both the conditions and monomeric form, although present, is extremely low when compared to the other higher order forms. These results indicate that equilibrium exists between the monomeric, dimeric and heptameric forms.

Strikingly, we observed that only the tetradecameric form was phosphorylated but the heptameric and dimeric forms were not phosphorylated (Figure 4.11 B), indicating that a phosphorylation switch might be mediating the conversion from heptameric single ring form to the tetradecameric double ring form.

4.12.3 Determination of the Site of Phosphorylation on *M. tuberculosis* GroEL1 by Mass-spectroscopy

Having established that the conversion of single ring (heptameric) GroEL1 to double ring (tetradecameric) form is mediated by phosphorylation, we proceeded to identify the phosphorylated Serine (s). To this end, the tetradecameric fraction of GroEL1 was subjected to mass-spectrometry. In order to further confirm the absence of phosphorylation in the heptameric, dimeric and monomeric forms of GroEL1, these fractions were also subjected to mass-spectrometry.

Different oligomeric forms of recombinant *M. tuberculosis* GroEL1 were resolved on 10% SDS-PAGE and the coomassie stained bands corresponding to GroEL1 were sliced out as polyacrylamide gel plugs. These were subjected to in-gel digestion using Trypsin Gold (Promega) followed by MALDI-TOF study for Peptide Mass Fingerprinting (PMF) and Protein Sequencing using Ultraflex MALDI-TOF/TOF (Bruker Daltonics). The peptides which hosted a Seryl residue were subjected to MS/MS using ion trap mass spectrometer (Agilent), to determine the presence of phosphorylation. MALDI/TOF and MS/MS data was analyzed using MASCOT search engine or Spectrum Mill software. In these experiments purified recombinant GroEL1 was included as a control for peptide identification. Mass spectrometric studies were carried out in collaboration with The Centre for Genomic Applications (TCGA), New Delhi, India.

Mass spectrometric analysis of Mtb GroEL1 by MALDI-TOF followed by MS/MS fragmentation of the resulting peptides confirmed that Serine-393 is phosphorylated in the tetradecameric form (Figure 4.12). However, the possibility that a few more Serines could be phosphorylated, and if so, these might be located in the critical positions governing the inter-ring contacts cannot be ruled out. These results thus reveal that there exist multiple oligomeric forms of GroEL1 molecules in Mtb, and that the conversion from the heptameric to tetradecameric forms is mediated by phosphorylation of Serine-residues.
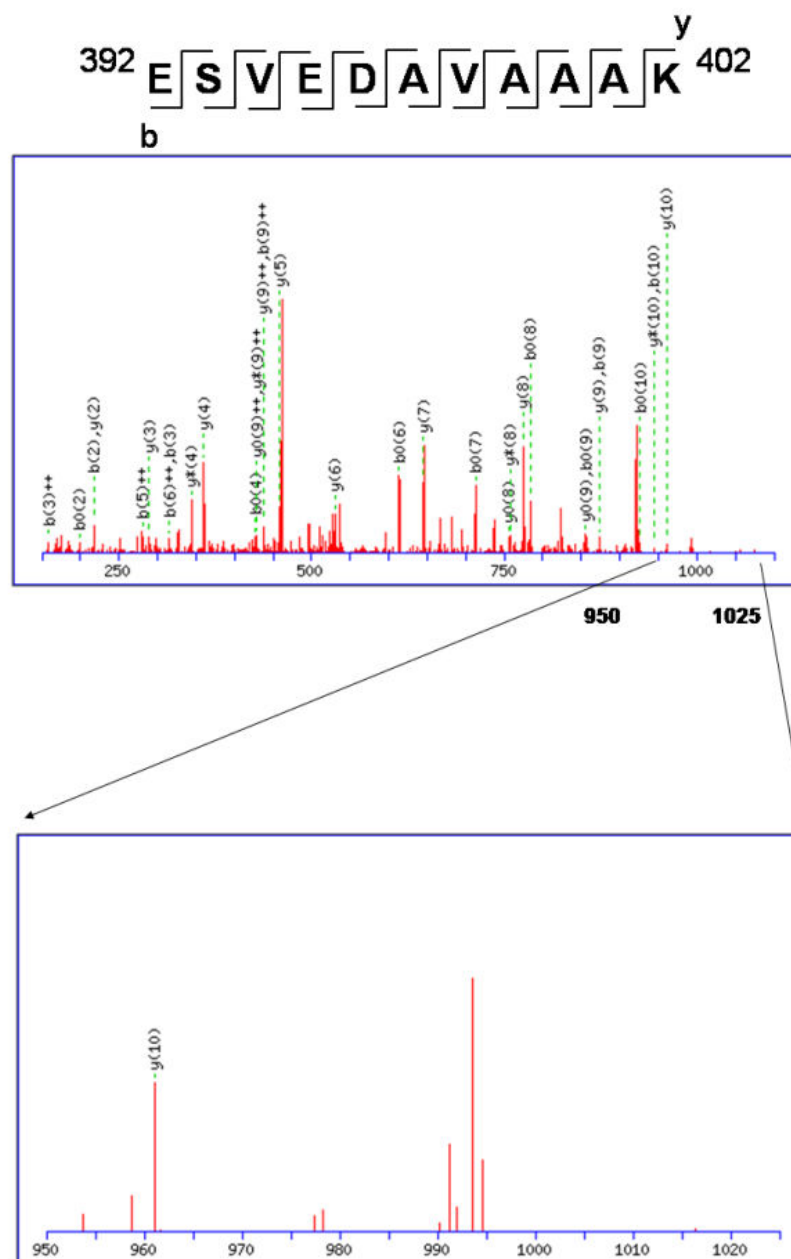
Figure 4.12: Detection of Phosphorylation on M. tuberculosis GroEL1.
MALDI TOF/TOF tandem MS fragmentation analysis of the tryptic peptide
ESVEDAVAAAK ([MH]+ m/z 1088.912724) corresponding to amino acid
residues 392-402 of *M. tuberculosis* GroEL1. Phosphorylation on the peptide
is confirmed by a mass shift of 80-Da owing to the loss of HPO_3^-.

4.13 Conclusions

The work presented in this chapter shows the essential requirement of oligomerization for GroEL function. Although, GroEL1 from *M. tuberculosis* is capable of oligomerization in its native environment, the recombinant protein, when purified from *E. coli* is not capable of oligomerization. However, exchanging the equatorial domains renders the *M. tuberculosis* GroEL active, suggesting the specific oligomerization capability of *M. tuberculosis* GroEL's equatorial domain only in the *M. tuberculosis* cellular milieu. Furthermore, *E. coli* GroEL can be rendered non functional and indistinguishable from *M. tuberculosis* GroEL1 by replacing its equatorial domain with the one present in GroEL1. Since GroELMEF can substitute for the lack of GroEL function in *E. coli*, it suggests that despite the heterologous apical domain of Mtb GroEL1 borne on GroELMEF, the chimerical protein (and therefore *M. tuberculosis* GroEL1) may recognize the same cellular substrates in vivo, as does its *E. coli* counterpart. Despite the large functional difference between *M. tuberculosis* GroEL1 and *E. coli* GroEL, a noteworthy aspect is that GroEL1 retains the highly conserved ATP binding pocket that bears amino acid residues identical to those in *E. coli* GroEL. This tempts us to speculate that oligomerization may be a prerequisite for GroEL ATPase activity and therefore the ability to assist refolding of substrate polypeptides. These studies therefore reveal that the basis of reduced activity of recombinant *M. tuberculosis* GroEL1 in *E. coli* was as a result of its impaired oligomerization.

For the first time it is also shown that mycobacterial GroEL1 exists in different oligomeric forms and the switch between the single ring (heptamer) and double ring (tetradecameric) GroEL forms is mediated by phosphorylation. It is therefore tempting to propose that the naturally synthesized GroEL1 from *M. tuberculosis* exists in equilibrium between a dimer and a heptamer, and that heptamer to tetradecamer conversion is mediated by phosphorylation (Figure 4.12). Since a similar strategy of chaperonin oligomerization operates in mitochondrial GroEL, this study, in addition to probing Mtb GroEL biology, might also shed light on the evolution of mitochondria and chloroplasts, which are supposed to be the bacterial symbionts of the eukaryotes (Allen, 2003).

4.15 References

- Allen, J. F. (2003) Why chloroplasts and mitochondria contain genomes. *Comp. Funct. Genom.* **4,** 31–36

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

- Bonifacino, J. S., Dell'Angelica, E. C. and Springer, T. A. (2001) Immunoprecipitation In *Current Protocols in Immunology* **8.3.1-8.3.28,** New York: John Wiley.

- Boris, M. G., Jeffrey, W. S. and Paul, M. H. (2005) Residual Structure in Urea-Denatured Chaperonin GroEL. *Biochemistry* **34**, 13928–13933

- Bradford, M. M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal. Biochem.* **72**, 248-254

- Buchner, J., Grallert, H. and Jakob, U. (1998) Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods Enzymol*. **290**, 323-338

- Buckle, A. M., Zahn, R. and Fersht, A. R. (1997) A structural model for GroEL polypeptide recognition. *Proc. Natl. Acad. Sci. USA* **94**, 3571-3575

- Chatellier, J., Hill, F., Foster, N. W., Goloubinoff, P. and Fersht A. R. (2000) From Minichaperone to GroEL 3: Properties of an Active Single-ring Mutant of GroEL. *J. Mol. Biol.* **304**, 897-910

- Chen, Y. H. and Yang, J. T. (1971) A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochem. Biophys. Res. Commun.* **44**, 1285–1291

- Chen, L. and Sigler, P. B. (1999) The Crystal Structure of a GroEL/Peptide Complex: Plasticity as a Basis for Substrate Diversity. *Cell* **99**, 757-768

- Clark, A. C., Ramanathan, R. and Frieden, C. (1998) Purification of GroEL with low fluorescence background. *Methods Enzymol.* **290**, 100–118

- Dickson, R., Weiss, C., Howard, R. J., Alldrich, S. P., Ellis, R. J., Lorimer, G. H., Azem, A. and Viitanen, P. V. (2000) Reconstitution of Higher Plant Chloroplast Chaperonin 60 Tetradecamers Active in Protein Folding *J. Biol. Chem.* **275**, 11829-11835

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

- Galan, A., Sot, B., Llorca, O., Carrascosa, J. L., Valpuesta, J. M. and Muga, A. (2001) Excluded volume effects on the refolding and assembly of an oligomeric protein. GroEL, a case study. *J. Biol. Chem.* **276**, 957-964

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

- Haslbeck, M., Franzmann, T., Weinfurtner, D. and Buchner, J. (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12**, 842-846

- Henkel, R. D., Vander-Berg, J. L. and Walsh, R. A. (1988) A microassay for ATPase. *Anal. Biochem.* **169**, 312-318

- Hirschfield, G., McNeil, M. and Brennan, P. (1990) Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis. J. Bacteriol.* **172,** 1005-1013

- Johnson, W. C. Jr. (1990) Protein secondary structure and circular dichroism: a practical guide. *Proteins: Struct. Func. Genet.* **7**, 205-214

- Karunakaran, K. P., Noguchi, Y., Read, T. D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C. C. and Brunham, R. C. (2003) Molecular analysis of the multiple GroEL proteins of Chlamydiae. *J. Bacteriol.* **185,** 1958-1966

- Kipnis, Y., Papo, N., Haran, G. and Horovitz, A. (2007) Concerted ATP-induced allosteric transitions in GroEL facilitate release of protein substrate domains in an all-or-none manner. *Proc. Natl. Acad. Sci. USA* **104**, 3119-3124

- Kusmierczyk, A. R. and Martin, J. (2001) Assembly of chaperonin complexes. *Mol. Biotechnol.* **19**, 141-153

- Lee, B., Hefta, S. and Brennan, P. (1992) Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect. Immun.* **60,** 2066-2074

- Lee, G. J., Roseman, A. M., Saibil, H. R. and Vierling, E. (1997) A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**, 659-671

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

- Mendoza, J. A., Dulin, P. and Warren, T. (2000) The lower hydrolysis of ATP by the stress protein GroEL is a major factor responsible for the diminished chaperonin activity at low temperature. *Cryobiology* **41,** 319-323

- Motojima, F. and Yoshida, M. (2003) Discrimination of ATP, ADP, and AMPPNP by chaperonin GroEL: hexokinase treatment revealed the exclusive role of ATP. *J. Biol. Chem.* **278**, 26648-26654

- Narayan, A., Sachdeva, P., Sharma, K., Saini, A. K., Tyagi, A. K. and Singh, Y. (2007) Serine threonine protein kinases of mycobacterial genus: phylogeny to function. *Physiol. Genomics* **29**, 66-75

- Panda, M., Smoot, A. L. and Horowitz, P. M. (2001) The 4,4'-dipyridyl disulfide-induced formation of GroEL monomers is cooperative and leads to increased hydrophobic exposure. *Biochemistry* **40**, 10402-10410

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

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

- Reed, J. and Reed, T. A. (1997) A set of constructed type spectra for the practical estimation of peptide secondary structure from CD. *Anal. Biochem.* **254**, 36-40

- Richardson, A., Landry, S. J. and Georgopoulos, C. (1998) The ins and outs of a molecular chaperone machine. *Trends Biochem. Sci.* **23**, 138-143

- Roseman, A. M., Chen, S., White, H., Braig, K. and Saibil, H. R. (1996) The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* **87**, 241-251

- Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. and Horwich, A. L. (1999) GroEL-GroES cycling: ATP and non-native polypeptide direct alternation of folding-active rings. *Cell* **97**, 325-338

- Saibil, H. R. and Ranson, N. A. (2002) The chaperonin folding machine. *Trends Biochem. Sci.* **27**, 627-632

- Sharma, K. K., Kumar, G. S., Murphy, A. S. and Kester, K. (1998) Identification of 1,1'-bi(4-anilino)naphthalene-5,5' -disulfonic acid binding sequences in α-crystallin. *J. Biol. Chem.* **273**, 15474-15478

- Shi, L., Palleros, D. R. and Fink, A. L. (1994) Protein conformational changes induced by 1,1'-bis(4-anilino-5-naphthalenesulfonic acid): Preferential binding to the molten globule of DnaK. *Biochemistry* **33**, 7536-7546

- Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A. and Horwich, A. L. (1998) Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67**, 581-608

- Smoot, A. L., Panda, M., Brazil, B. T., Buckle, A. M., Fersht, A. R. and Horowitz, P. M. (2001) The binding of bis-ANS to the isolated GroEL apical domain fragment induces the formation of a folding intermediate with increased hydrophobic surface not observed in tetradecameric GroEL. *Biochemistry* **40**, 4484-4492

- Srere, P. A. (1966) Citrate-condensing enzyme oxalacetate binary complex, studies on its physical and chemical properties. *J. Biol. Chem.* **241**, 2157-2165

- Stan, G., Brooks, B. R., Lorimer, G. H. and 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. USA* **103**, 4433-4438

- Steede, N. K., Temkin, S. L. and Landry, S. J. (2000) in *Chaperonin Protocols* (*Methods in Molecular Biology*) **140** (Schneider, C., ed), 133-138, Humana Press, Totowa, NJ

- Thirumalai, D. and Lorimer, G. H. (2001) Chaperonin-mediated protein folding. *Annu. Rev. Biophy. Biomol. Struc.* **30**, 245-269

- Tilly, K. and Georgopoulos, C. (1982) Evidence that the two *Escherichia coli groE* morphogenetic gene products interact in vivo. *J. Bacteriol.* **149**, 1082-1088

- Todd, M. J., Viitanen, P. V. and Lorimer, G. H. (1994) Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science* **265**, 659-666

- Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P. and Lorimer, G. H. (1990) Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (*groEL*) are K⁺ dependent. *Biochemistry* **29**, 5665-5671

- Wang, J. D., Michelitsch, M. D. and Weissman, J. S. (1998) GroEL-GroES-mediated protein folding requires an intact central cavity. *Proc. Natl. Acad. Sci. USA* **95**, 12163-12168

- Wehenkel, A., Bellinzoni, M., Graña, M., Duran, R., Villarino, A., Fernandez, P., Andre-Leroux, G., England, P., Takiff, H., Cerveñansky, C., Cole, S. T. and Alzari, P. M. (2008) Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. *Biochem. Biophys. Acta* **1784**, 193-202

- Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. and Horwich, A. L. (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **83**, 577–587

- Xu, Z., Horwich, A. L. and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)$_7$ chaperonin complex. *Nature* **388**, 741-750

- Yokokawa, M., Wada, C., Ando, T., Sakai, N., Yagi, A., Yoshimura, S. H. and Takeyasu, K. (2006) Fast-scanning atomic force microscopy reveals the ATP/ADP-dependent conformational changes of GroEL. *EMBO J.* **25**, 4567-4576

- Zhang, C. C., Gonzalez, L. and Phalip, V. (1998) Survey, analysis and genetic organization of genes encoding eukaryotic- like signaling proteins on a cyanobacterial genome. *Nucleic Acids Res.* **26**, 3619-3625

CHAPTER V

Isolation and Characterization of GroES Independent GroEL Mutants

5.1 Introduction

GroEL belongs to the class I family of chaperonins, which require the assistance from co-chaperonin GroES. Function of GroEL in facilitating protein folding requires the formation of asymmetric complex of GroEL and GroES. *E. coli* GroEL exists in double-ring form enclosing cavities for substrate polypeptides to bind. GroES acts as a lid in usually encapsulating the protein bound cavity (Bukau and Horwich, 1998). Class II chaperonins comprise the Hsp60 homologues located in archea and cytosol of eukaryotes (Horwich et al., 2007). These chaperonins bear a built-in lid in the substrate interacting apical domain and hence are independent of an external lid like structure.

Elegant genetic studies by Georgopuolos and colleagues on the genes encoding GroES and GroEL have established that both the genes are essential in *E. coli* under all growth conditions (Fayet et al., 1989). Proteomic studies aimed at identifying the cellular substrates of *E. coli* GroEL and GroES system have attempted to suggest the reasons for this essentiality. These studies have revealed that eleven substrates, which follow an obligatory GroEL-mediated folding path, are known to be essential in *E. coli* (Kerner et al., 2005; Houry et al., 1999). Although this reason for the essentiality of GroEL is justified, the same for GroES is not evident from the proteomic studies. It may be hypothesized that capping of a substrate-encapsulated cavity of GroEL is a strict requirement for the productive folding of many substrates.

The existence of co-chaperonin independent group II chaperonins, provide an interesting path to explore if *E. coli* GroEL could be made GroES independent. Towards this goal, we have generated a pool of GroEL variants as described in chapter III. This pool was searched to select the GroEL variants which could rescue GroEL + GroES depletion. For this we have employed *E. coli* LG6 strain, in which the expression of *groEL* and *groES* is under the control of lactose inducible P_{lac} promoter (Horwich et al., 1993). This strain therefore fails to form colonies in the absence of IPTG or lactose.

Details of the experiments performed and the results obtained are described in the following sections of this chapter.

5.2 Materials

All the chemicals, media components and enzymes were purchased from various commercial sources. Strains of *E. coli* were cultured in standard LB supplemented as appropriate. *E. coli* LG6 is a derivative of MG1655 wherein the chromosomal *groES/L* operon is placed downstream of IPTG and lactose inducible P_{lac} promoter (Horwich et al., 1993). Coliphages P1 and λ*c*IB2 were sourced from laboratory stocks. Plasmids and oligonucleotide primers used in this study are listed in Appendix II.

5.3 Isolation of GroES independent GroES variants

GroES and GroEL are the only chaperones that are essential in *E. coli* under all the physiological conditions (Fayet et al., 1989; Horwich et al., 1993). GroEL functions by encapsulating unfolded proteins in its cavities and GroES acts as a lid for the cavity (Mande et al., 1996; Hunt et al., 1996). To select for GroES independent GroEL variants, a library of random mutants of GroEL was generated followed by selecting for GroES independent GroEL variants in a GroES depleted strain.

5.3.1 Gene Shuffling for Generation of GroEL Mutant Library

The library of GroEL mutants was generated by employing gene shuffling technique, as described in chapter III. Briefly, 5-10 μg of the 1.6 kb PCR amplified Mtb *groEL1* and *groEL2* ORFs, using primer pairs SCM7F/SCM7R and SCM8F/SCM8R respectively, were used for the multi-gene DNA shuffling following the reported method (Stemmer, 1994; Wang et al., 2002). The PCR products were subjected to limited DNaseI digestion, to obtain the 50-150 bp fragments. Fragments of 50-150 bp were recovered from agarose gels, diluted ten times and were used as a template for a primer-less assembly PCR. The assembled PCR product was diluted 40 times and was used for the next round of PCR with *E. coli groEL* specific primers. The shuffling protocol at this step was modified according to Zhao & Arnold, 1997. During the shuffling protocol, in the final round of PCR using specific primers for the amplification of 1.6 kb DNA, *E. coli groEL* specific primers SCM4F and SCM3R were used taking into the consideration of the fairly hydrophobic GGM repeats at the carboxy-terminus of the *E. coli groEL* (Tang et al., 2006; Farr et al., 2007). The product obtained after the final PCR was digested and

cloned into NcoI and HindIII sites of pBAD24, thus placing the shuffled ORFs under the control of the P_{BAD} promoter and the library of plasmids was recovered in *E. coli* DH5α.

5.3.1 Selection and Isolation of GroES Independent GroEL Variants

The library of GroEL variants generated in *E. coli* DH5α was pooled and plasmids were prepared. The pooled plasmid preparation was used to transform the *E. coli* strain LG6 and the transformants were selected on LB agar plates supplemented with 0.5 mM IPTG. About seventy transformants were screened in the presence of 0.2% L-arabinose for the clones that could complement the loss of GroES/L in LG6. Two clones of the seventy colonies screened, exhibited complementation and these were labeled as pSCM1632 and pSCM1633 and resulting GroEL variants were labeled GLLG18 and GLLG25, respectively.

The ability of the GroEL variants, GLLG18 and GLLG25 in complementing GroES/L depletion was confirmed by repeating the complementation experiments with fresh preparation of plasmids pSCM1632 and pSCM1633.

5.4 Phenotypic Analysis of GroES Independent GroEL Mutants

In vivo activity of the isolated GroEL variants was quantified by scoring for their ability to rescue the GroES + GroEL depletion and supporting the morphogenesis of bacteriophages.

5.4.1 Complementation Studies in GroES/L Depletion Strain

Having established that the GroEL variants, GLLG18 and GLLG25 are not false positives, we set to assess the activity of these clones by quantitative analysis of the extent of their ability to rescue the GroES/L depletion in LG6. Plasmids encoding the said variants were transformed into *E. coli* LG6 and the resulting transformants were grown in the presence of 1 mM IPTG. Cells were recovered at the stationary phase, to remove the traces of IPTG, washed three times with and serially diluted in LB. 5 μl of each serially diluted culture was spotted on LB agar supplemented with either 0.2% L-arabinose (to induce the expression of plasmid cloned genes) or 1 mM IPTG (to induce the chromosomal copy of *E. coli groES/L* operon). The plates were incubated at 30 °C. LG6 expressing *E. coli groES + groEL* and *E. coli groEL* were included as controls. A quantitative phenotypic analysis of nine of the gene shuffled *groEL* variants is shown in Figure 5.01.

Data shown in Figure 5.01 clearly shows that the GroEL variants, GLLG18 and GLLG25 are capable of supporting *E. coli* LG6 despite the depletion of chromosomally encoded GroEL and GroES. Nonetheless, *E. coli* GroEL could not support the growth of LG6 unless supplemented with the *E. coli* GroES. On the other hand neither of the GroEL variants, which are able to support the growth of LG6, was supplemented with GroES. Hence, these results suggest that the said GroEL variants might be able to substitute for both GroEL and GroES. These observations therefore suggest that the tested GroEL variants might act independent of a GroES supplementation.

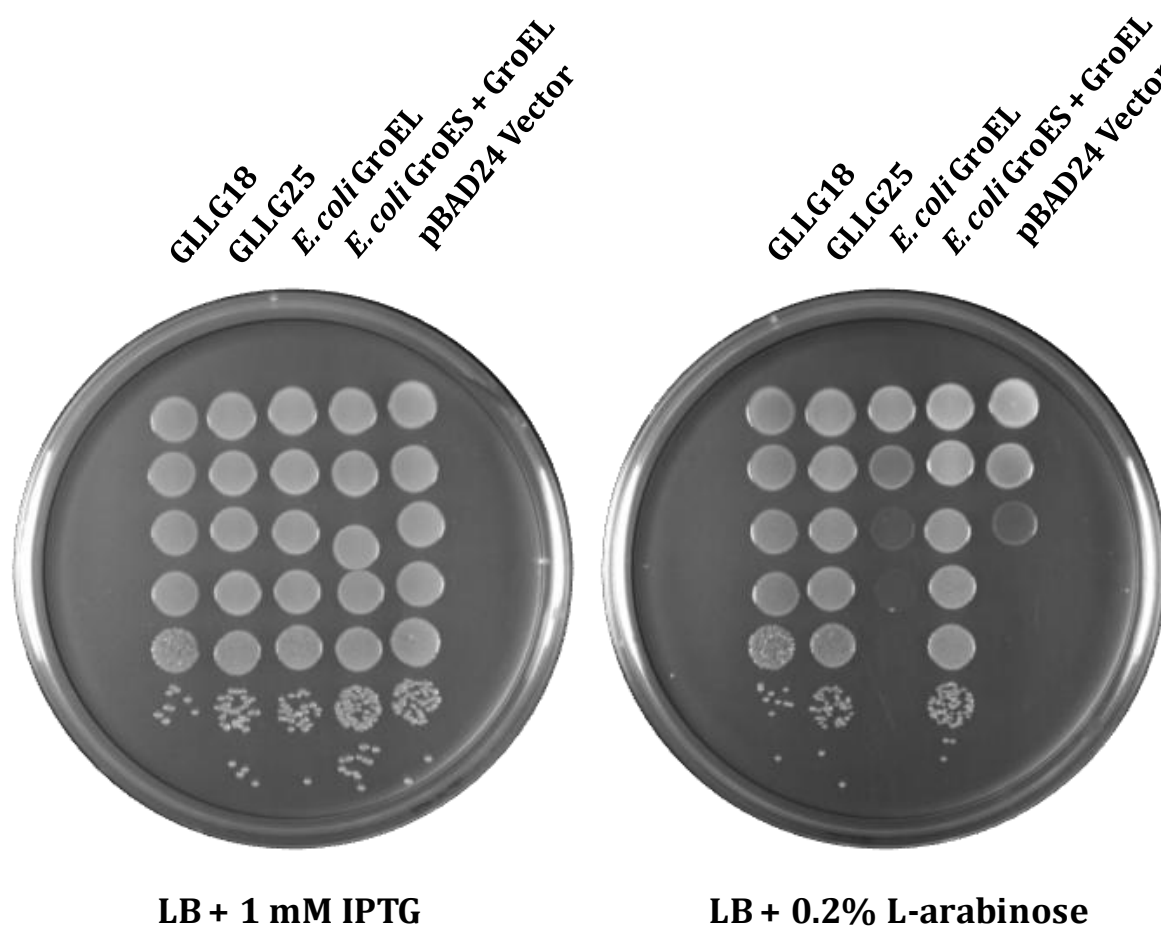**LB + 1 mM IPTG**   **LB + 0.2% L-arabinose**

Figure 5.01: Complementation of GroES/L Depletion by GLLG18 and GLLG25. Serially diluted cultures of the *E. coli* LG6 expressing the indicated *groEL/S* genes were spotted on LB agar plates supplemented as indicated. The plates were incubated at 30 °C.

5.4.2 Assay for Bacteriophage Morphogenesis

Development of several coliphages requires functional GroEL/S system, specifically for folding and assembly of phage capsid proteins (Georgopoulos and Hohn, 1978; Zeilstra-Ryalls et al., 1993). Ability of chaperone system could therefore be scored as the ability to support bacteriophage plaquing the lawn of bacterial culture expressing particular *groEL* homologue. To this end, ability of the GroEL variants in supporting bacteriophage morphogenesis was studied as an extension to the complementation studies in LG6.

LG6 expressing *groEL* mutants as above were grown in LB broth supplemented with 1 mM IPTG. The cultures were further supplemented individually with 0.4% maltose for studying bacteriophage λ morphogenesis and with 5 mM CaCl$_2$ + 5 mM MgSO$_4$ for bacteriophage P1. The cultures were recovered in stationary phase and were washed with LB broth to remove the residual IPTG. These cultures were individually mixed with about 4 ml of LB soft agar (0.5% agar) and were overlaid onto LB agar plates that were previously supplemented with 0.2% L-arabinose, 5 mM CaCl$_2$ and 5 mM MgSO$_4$. The phage preparations were serially diluted in LB broth supplemented with 5 mM MgSO$_4$ and 5 µl of each dilutions were spotted on the plates. The plates were incubated at 30 °C. Formation of plaques on bacterial lawn was scored as the ability to support bacteriophage morphogenesis. Representative results of experiments repeated three times are presented in Figure 5.02

In contrast to the complementation studies in LG6, neither GLLG18, nor GLLG25 could promote the development of bacteriophages either λ or P1 in the strain LG6. However, LG6 culture expressing *E. coli groES + E. coli groEL* was able to support morphogenesis of both the phages (Figure 5.02). These results suggest that the GroEL variants have the ability to assist the folding of proteins encoded by *E. coli* genome but might have lost the ability to support the folding of the phage encoded proteins.
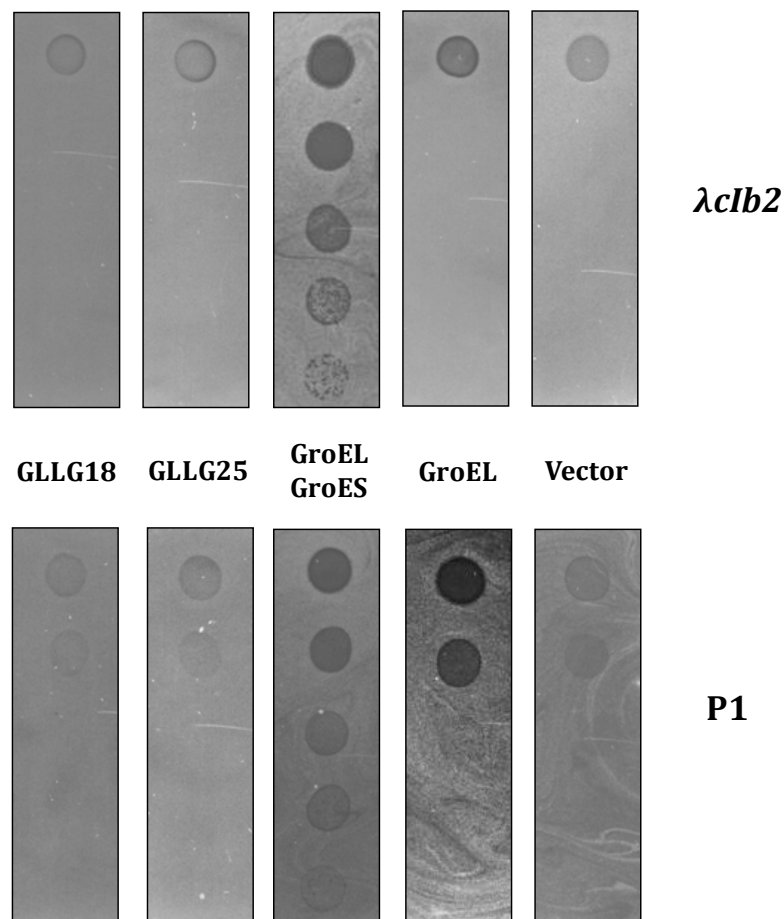
Figure 5.02: Assay for Bacteriophage Morphogenesis. Lawns of the *E. coli* strain LG6 expressing indicated *groEL* variants were prepared on LB agar plates supplemented with 0.2% L-arabinose. Hundred-fold serial dilutions of bacteriophages λcIB2 and P1 were spotted on the lawns. The plates were incubated at 30 °C.

5.5 Biochemical Characterization of GroEL Variants

In vivo activity of the GroEL variants, GLLG18 and GLLG25 showed that these proteins are capable of supporting the growth of *E. coli* LG6 under GroEL/S depletion. However, these variants were incapable of supporting the morphogenesis of bacteriophages. Understanding the basis of this disparity in the activities of GroEL variants requires a comprehensive biochemical characterization. To this end, we set to purify the GroEL variants there by assessing their chaperone properties.

5.5.1 Purification of GroEL variants

To evade the potential contamination from *E. coli* GroEL and GroES, GLLG18 and GLLG25 were purified from LG6. Towards this, the plasmids encoding the said variants were transformed into LG6 and throughout the expression procedure these cultures were maintained in 0.2% L-arabinose, to avoid the contamination by resident GroEL and GroES. Cells expressing the GroEL variants were recovered at late log-phase, re-suspended in Lysis Buffer I, containing 50 mM Tris.HCl (pH: 8.0), 1 mM EDTA and 1 mM PMSF and the cells were lysed by sonication. The resulting soluble fraction bearing the GroEL variants were subjected to ammonium sulfate $((NH_4)_2SO_4)$ extraction. The protocol for purification of *E. coli* GroEL and GroES was followed as described earlier (Clark et al., 1998). *E. coli* GroEL and GroES were salted-in at 30% and salted-out at 65% saturated ammonium sulfate. GLLG18 and GLLG25 were salted-out at 40% and 50% respectively. The proteins were desalted using PD-10 Desalting columns (GE Life Sciences). *E. coli* GroEL and GroES were purified further by employing ion exchange chromatography, using HiLoad 16/10 Q Sepharose HP column (GE Life Sciences). The protein sample pre-equilibrated with the loading Buffer I, containing 50 mM Tris.HCl (pH: 8.0) and 1 mM EDTA, was loaded onto the column and eluted by applying step gradient of 50 mM NaCl increments. *E. coli* GroES and GroEL were eluted at 250 mM NaCl and 400 mM NaCl, respectively. However, GLLG18 and GLLG25 were further enriched by heat precipitation at 70 °C (Figure 5.03).
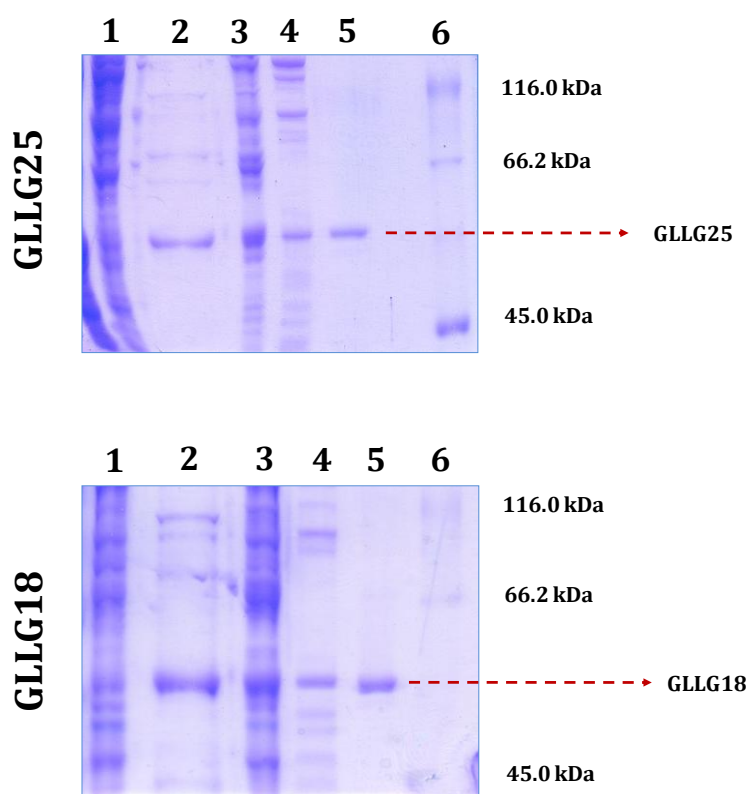
Figure 5.03: Purification of GLLG18 and GLLG25. Lanes showing the steps followed for the purification of GroEL variants. Lane 1 and 2 are showing the supernatant and pellet, respectively after ammonium sulphate extraction; lanes 3 and 4 are showing supernatant and pellet, respectively after heat precipitation; Lane 5 showing the fraction eluted in gel filtration; and lane 6 shows the Unstained Protein Molecular Weight Marker (MBI Fermentas, Canada), wherein the molecular masses of the corresponding protein bands are denoted. Bands corresponding to GLLG18 and GLLG25 are indicated by a maroon coloured arrow.

5.5.2 Determination of Quaternary Structures of GroEL Variants

Further, quaternary structure of the GroEL variants was determined using gel filtration chromatography. The said protein samples were resolved on HiPrep 26/60 Sephacryl S-300 HR column (GE Life Sciences). The elution profiles of these GroEL variants were further used for estimating molecular weight and oligomeric status (Figure 5.04).
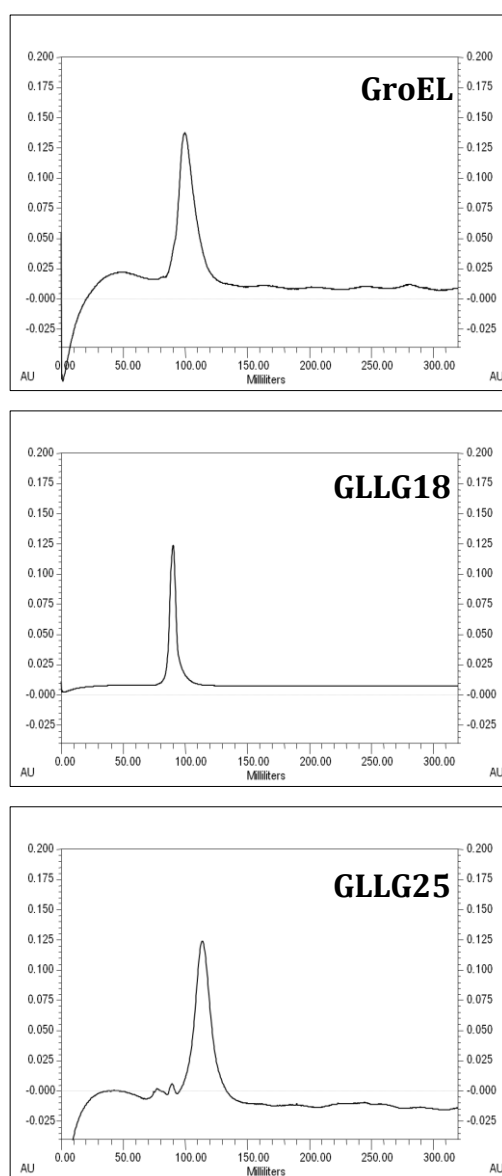
Figure 5.04: Oligomeric States of GroEL Variants. Gel filtration chromatograms of *E. coli* GroEL, GLLG18 and GLLG25. The indicated proteins were separated on 16/60 Sephacryl S-300 HR column (GE Biosciences) in Biologic DuoFlow FPLC System (Bio-Rad Laboratories).

Gel filtration profiles of the GroEL variants suggest that they display gel filtration characteristics similar *E. coli* GroEL. Since both the variants displayed tendencies to exist in a higher oligomeric state and their gel permeation profiles were consistent with predicted molecular weights, it is therefore reasonable to presume that the said state corresponds to a teteradecameric assembly (Figure 5.04).

5.5.3 ATPase Activity Assay

ATPase activity of the purified GroEL variants was quantified by a colorimetric assay performed as described (Henkel et al., 1988; Viitanen et al., 1990). Briefly, 50 µl of the reaction buffer containing 50 mM Tris.HCl (pH 8.0), 10 mM KCl, 10 mM $MgCl_2$ and 2.5 µM of GroEL variant individually was incubated with 1 mM ATP at 37 °C for 20 minutes. Enzymatic reactions were terminated by addition of 200 µl of the acidic solution of Malachite green. Amount of inorganic phosphate liberated was measured at 655 nm using Nanodrop ND-1000 (Figure 5.05). Control reactions without ATP and GroEL were performed. A standard curve with 200 - 1000 µM monobasic potassium phosphate was generated concurrently with each experiment.

Figure 5.05: ATPase Activities of Purified GroEL Variants. The indicated GroEL variants were assayed by malachite green calorimetric assay. Amount of inorganic phosphate released was quantified at 655 nm. Mean for individual data sets was calculated and plotted along with the standard deviation considering activity of *E. coli* GroEL as 100%.

As shown in the figure 5.05, the gene shuffled variants, GLLG18 and GLLG25, displayed similar ATPase activities as the wild type Gro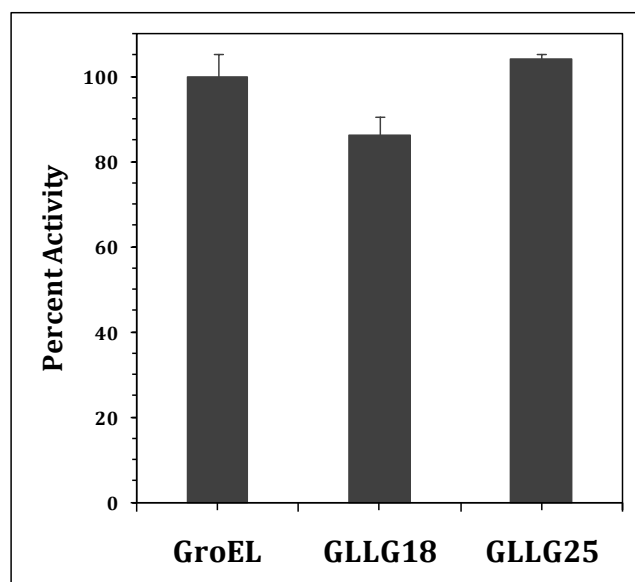EL. Although GLLG18 displayed weaker ATPase activity, the activity exhibited by GLLG25 is comparable to *E. coli* GroEL.

5.5.4 Estimation of Exposed Hydrophobicity on the GroEL Variants by bis-ANS Fluorescence Assay

Hydrophobicity is the principal virtue of a chaperone, owing to which it interacts with the substrates that are into a non-native state, thereby preventing the substrates from taking the non-productive folding route. Apical domains in *E. coli* GroEL are known to possess an exposed hydrophobic area where substrate polypeptides and GroES have been shown to bind (Xu et al., 1997). Due to the variation exhibited by the GroEL variants in the in vivo activities, presence and extent of such hydrophobic patches on the surface of the GroEL variants was probed by the ability to bind to 1,1-bis (4-anilino) naphthalene-5, 5'-disulfonic acid (bis-ANS, $C_{32}H_{22}K_2N_2O_6S_2$).

Exposed hydrophobicity on each of the chaperonins was estimated as binding of equimolar ratios of bis-ANS and the chaperonins according to standard procedures (Shi et al., 1994; Lee et al., 1997; Sharma et al., 1998; Panda et al., 2001; Smoot et al., 2001). Briefly, binding of bis-ANS with GroEL variants and *E. coli* GroES at 20 μM each was monitored by exciting the probe at 390 nm and recording the emission spectra in the range of 400-600 nm. The fluorescence intensity measurements were carried out in 100 mM Tris.HCl (pH: 8.0) at 25 °C in SpectraMax M5 Spectrofluorimeter (Molecular Devices Inc., USA). Buffer alone and bis-ANS alone reactions were set as controls.

As shown in Figure 5.06, the chaperonin variants showed comparable exposed hydrophobic nature, although GLLG18 exhibited slightly lower exposed hydrophobic area. The fluorescence enhancement exhibited by the chaperonin variants is comparable to that observed for *E. coli* GroEL. The results therefore clearly indicate the presence of hydrophobic surfaces on the GroEL variants.
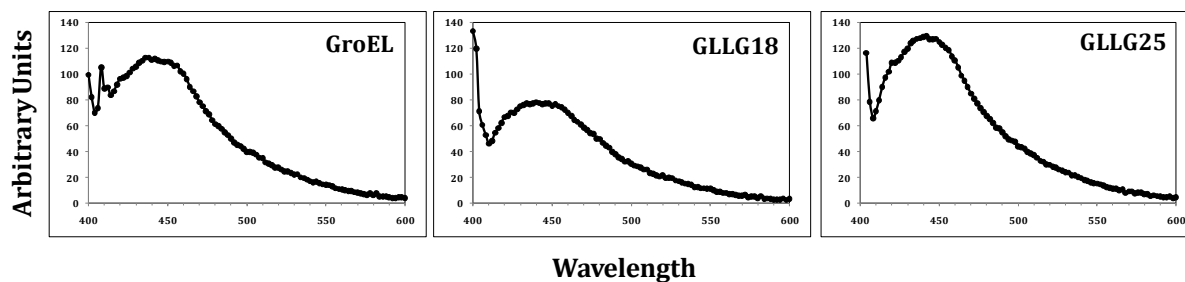
Figure 5.06: Assessment of Surface Hydrophobicity of GroEL Variants.
Fluorescence intensity increase upon interaction of individual GroEL variants
with bis-ANS was measured from 400 nm - 600 nm. The spectral values were
subtracted for buffer and bis-ANS alone.

5.5.5 Prevention of Aggregation of Citrate Synthase

As discussed in Chapter VI, GroEL assisted folding of substrate proteins typically
involves binding of the polypeptides by virtue of exposed hydrophobic interactions and
sequestration of polypeptides into the cavity during stress, thereby preventing
irreversible aggregation. Presence of surface hydrophobicity on the GroEL variants
suggested if these variants could bind the substrate proteins. Hence, the effect of various
chaperonins on the aggregation of Citrate synthase (CS) was monitored following the
standard methods (Buchner et al., 1998). Briefly, 15 ng/ml CS was incubated at 43 °C in
the presence or absence of equimolar ratios of GroEL variants. Ability of the chaperonin
variants in preventing of aggregation of citrate synthase was monitored on-line for 20
minutes on SpectraMax M5 spectrofluorimeter (Molecular Devices Inc., USA) with
emission and excitation wavelengths set at 465 nm and corresponding band passes set
at 3.0 nm. Temperature of the sample was maintained with the built-in temperature
regulator in the spectrofluorimeter.

Time in Minutes
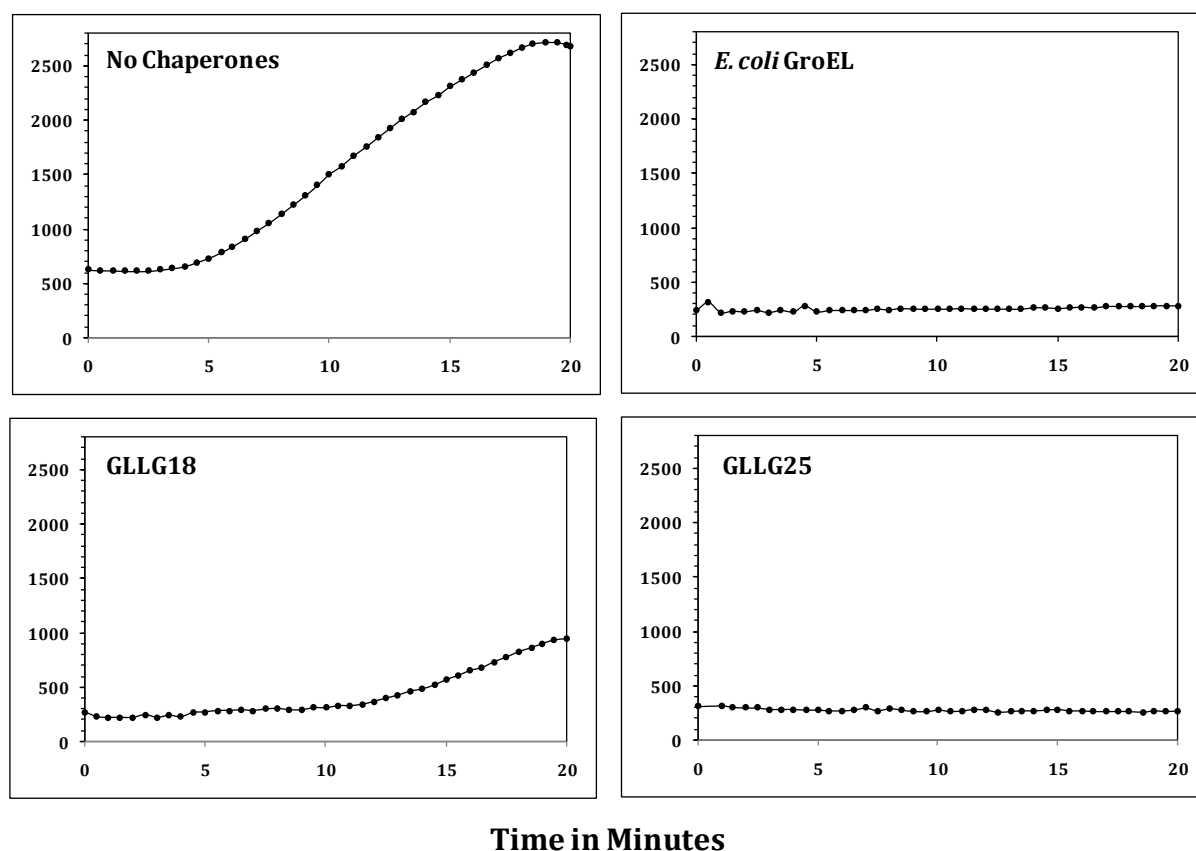
Figure 5.07: Prevention of Aggregation of Citrate Synthase by the Chaperonin Variants as a Function of Time. Aggregation of citrate synthase at 43 °C in the absence or the presence of equimolar ratios of the indicated GroEL variants was measured as a function of light scattered at 465 nm for twenty minutes.

Remarkably, the ability of GLLG25 in preventing aggregation of the substrate protein, Citrate synthase, at elevated temperatures, was comparable with *E. coli* GroEL (Figure 5.07). On the other hand, GLLG18 exhibited a weaker ability, which is in correlation with the reduced hydrophobicity on its surface (Figure 5.06).

5.5.6 Refolding the Chemically Denatured Substrate Proteins

The protocol for denaturation, refolding and assaying the activity of pig heart mitochondrial Citrate synthase (CS) was modified from Buchner et al., 1998. Briefly, CS at 15 µM was denatured at 25 °C in the presence of 6 M Guanidine.HCl and 1 mM DTT. The denatured CS was diluted into the refolding buffer containing 1 µM of different

GroEL variants in the presence or absence of 2 μM *E. coli* GroES. 2 mM ATP was added to initiate the chaperone mediated refolding reaction and incubated for 1 h at 37 °C.

Ability of GroEL variants in refolding the denatured CS was measured as the activity of the recovered CS in the presence of different chaperonins. CS activity was measured as the decrease in absorption at 233 nm owing to the disappearance of Acetyl Coenzyme A, for 90 seconds (Srere et al., 1966; Steede et al., 2000). The reaction was initiated by adding 5 μL of refolding mixture to 100 μL of the assay mixture containing 400 μM Oxalacetate and 160 μM Acetyl Coenzyme A, and the absorbance was monitored at 233 nm for 0–90 seconds using a SpectraMax M5 Spectrophotometer. The percentage of CS recovered after renaturation was calculated from slope of the curve and considering the activity exhibited by mock denatured Citrate synthase as 100% (Figure 5.08).
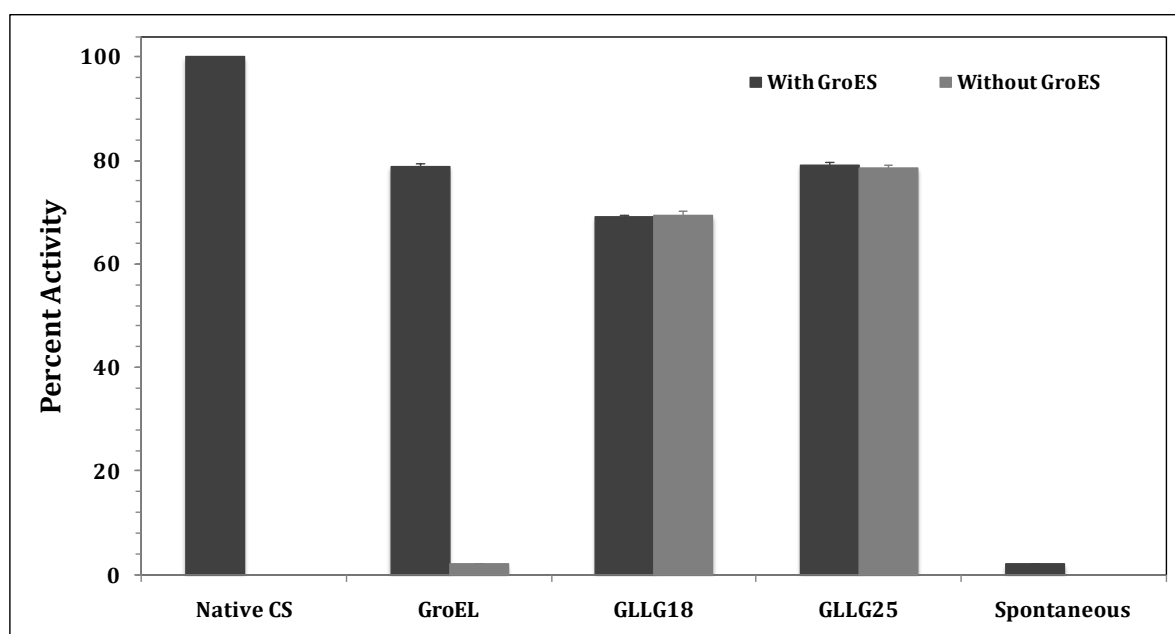
Figure 5.08: Effect of chaperonins in Refolding Chemically Denatured Citrate Synthase. Citrate synthase was chemically denatured, refolded with the indicated GroEL variants in the presence or absence of *E. coli* GroES. Activity of the refolded enzyme was measured as a decrease in absorbance at 233 nm due to the consumption of acetyl CoA. Yield of refolded enzyme is expressed as percentage of the activity determined for an equal quantity of non-denatured native CS.

As shown in the Figure 5.07, GLLG18 and GLLG25, which form higher order oligomers, were capable of refolding the model substrate protein, citrate synthase (CS), although GLLG18 displayed a lower activity. Moreover, in contrast to the *E. coli* GroEL, the activity exhibited by GLLG18 and GLLG25 is not influenced by the addition of *E. coli* GroES, suggesting that these chaperones are independent of GroES for their activity, at least under the conditions tested.

5.6 Conclusions

GroEL variants GLLG18 and GLLG25 could rescue the GroEL and GroES depletion in LG6. However, bacteriophage morphogeneis is not supported by these proteins. This disparity might be due to specific requirements of the phage proteins for their folding. These chaperones were active in vitro and exhibited chaperone characteristics. They display surface hydrophobicity and hence could prevent aggregation of substrate proteins. Moreover, owing to their oligomeric nature, these proteins are efficient in hydrolysing ATP and assisting the folding of chemically denatured substrate proteins, in GroES independent manner.

These observations therefore suggest that GLLG18 and GLLG25 are efficient chaperonins and could act independent of the assistance from GroES. However, precise understanding of the basis of such behaviour could be discerned by observing their sequence and structure. Future studies on the sequencing of these proteins, including MALDI TOF and further structural studies would disclose the basis for this behaviour.

5.7 References

- Buchner, J., Grallert, H. and Jakob, U. (1998) Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods Enzymol.* **290**, 323-338

- Bukau, B. and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351–366

- Clark, A. C., Ramanathan, R. and Frieden, C. (1998) Purification of GroEL with low fluorescence background. *Methods Enzymol.* **290,** 100–118

- Farr, G. W., Fenton, W. A. and Horwich, A. L. (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. USA* **104**, 5342-5347

- Fayet, O., Ziegelhoffer, T. and 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

- Georgopoulos, C. P. and Hohn, B. (1978) Identification of a host protein necessary for bacteriophage morphogenesis (the *groE* gene product). *Proc. Natl. Acad. Sci. USA* **75**, 131-135

- Henkel, R. D., Vander-Berg, J. L. and Walsh, R. A. (1988) A microassay for ATPase. *Anal. Biochem.* **169**, 312-318

- Horwich, A. L., Fenton, W. A., Chapman, E. and Farr, G. W. (2007) Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell. Dev. Biol.* **23**, 115-145

- Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N. and Furtak, K. (1993) Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. *Cell* **74,** 909-917

- Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F. U. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**, 147-454

- Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. and Deisenhofer, J. (1996) The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature* **379**, 37-45

- Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. and Hartl, F. U. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli. Cell* **122**, 209-220

- Lee, G. J., Roseman, A. M., Saibil, H. R. and Vierling, E. (1997) A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**, 659-671

- Mande, S. C., Mehra, V., Bloom, B. R. and Hol, W. G. (1996) Structure of the heat shock protein chaperonin-10 of *Mycobacterium leprae. Science* **271**, 203-207

- Panda, M., Smoot, A. L., and Horowitz, P. M. (2001) The 4,4'-dipyridyl disulfide-induced formation of GroEL monomers is cooperative and leads to increased hydrophobic exposure. *Biochemistry* **40**, 10402-10410

- Sharma, K. K., Kumar, G. S., Murphy, A. S. and Kester, K. (1998) Identification of 1,1'-bi(4-anilino)naphthalene-5,5' -disulfonic acid binding sequences in α-crystallin. *J. Biol. Chem.* **273**, 15474-15478

- Shi, L., Palleros, D. R. and Fink, A. L. (1994) Protein conformational changes induced by 1,1'-bis(4-anilino-5-naphthalenesulfonic acid): Preferential binding to the molten globule of DnaK. *Biochemistry* **33**, 7536-7546

- Smoot, A. L., Panda, M., Brazil, B. T., Buckle, A. M., Fersht, A. R. and Horowitz, P. M. (2001) The binding of bis-ANS to the isolated GroEL apical domain fragment induces the formation of a folding intermediate with increased hydrophobic surface not observed in tetradecameric GroEL. *Biochemistry* **40**, 4484-4492

- Srere, P. A. (1966) Citrate-condensing enzyme oxalacetate binary complex, studies on its physical and chemical properties. *J. Biol. Chem.* **241**, 2157-2165

- Steede, N. K., Temkin, S. L. and Landry, S. J. (2000) in *Chaperonin Protocols* (*Methods in Molecular Biology*) **140** (Schneider, C., ed), 133-138, Humana Press, Totowa, NJ

- Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747-10751

- Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U. and Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903-914

- Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P. and Lorimer, G. H. (1990) Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (*groEL*) are K+ dependent. *Biochemistry* **29**, 5665-5671

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

- Xu, Z., Horwich, A. L. and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. *Nature* **388,** 741-750

- Zeilstra-Ryalls, J., Fayet, O., Baird, L. and Georgopoulos, C. (1993) Sequence analysis and phenotypic characterization of *groEL* mutations that block lambda and T4 bacteriophage growth. *J. Bacteriol.* **175**, 1134-1143

CHAPTER VI

Conclusions

6.1 Conclusions

Understanding of the biology of chaperonin function is dominated by the information on the GroEL-GroES system of *E. coli*. Since the discovery of chaperonin function, genetic, structural and functional studies by several groups on *E. coli* GroEL-GroES system have led to the knowledge of various aspects of its function (Georgopoulos et al., 1972; Takano and Kakefuda, 1972; Thirumalai and Lorimer, 2001; Horwich et al., 2006; Horwich et al., 2007). Since *E. coli* GroEL was the first prokaryotic chaperonin to be identified, the understanding from the said studies was generalized to the prokaryotic chaperonins, which have been classified as Group I chaperonins.

According to the present understanding, GroEL exists as a homo-teteradecamer with two heptameric rings, each of which encloses a central cavity, where the unfolded polypeptide substrates are encapsulated. GroEL interacts with a wide range of unfolded or partially folded proteins, and with the assistance of its co-chaperonin, GroES, helps them to reach their fully folded active state. Co-immunoprecipitation experiments combined with mass spectroscopy have attempted to classify GroEL substrates into particular structural types (Houry et al., 1999; Kerner et al., 2005). However, as GroEL has been shown to interact with about half of the *E. coli* proteome (Viitanen et al., 1992), it is believed to display no substrate specificity. In its reaction cycle, GroEL displays binding to ATP, substrate polypeptide and GroES at various stages (Weissman et al., 1995, 1996; Mayhew et al., 1996; Xu et al., 1997). Structural changes at each stage of binding thereby result in encapsulation of the substrate into the cavity, wherein the bound protein is helped to fold in a sequestered hydrophilic environment of the cavity (Rye et al., 1997; 1999).

Although studies on the *E. coli* GroEL have illuminated an intricate mechanism of chaperonin-assisted protein folding, homologues of GroEL from other species exhibit altered properties. For example, those from mammalian mitochondria and from choloroplasts have been shown to exhibit multiple oligomeric forms such as single ring heptameric and double ring teteradecameric forms, where the conversion from the single ring to the double ring form is concentration and GroES dependent (Dickson et al., 2000; Levy-Rimler et al., 2001). Attempts to make a single ring version of *E. coli* GroEL by site directed mutagenesis of four critical residues at the inter-ring surface had

yielded a mutant, SR1 that is compromised in GroES release (Weissman et al., 1995). Furthermore conversion of a temperature sensitive GroEL mutant, GroEL44, into a single ring form, SR44, resulted in an active chaperonin (Chatellier et al., 2000). These studies therefore have proved that a double ring form of the GroEL in its protein folding function is an absolute requirement.

Recent genome annotation studies on various bacteria have revealed that several bacterial genomes possess multiple copies of *groEL* genes such as bacteria belonging to the genus Actinobacteria, where the multiple chaperonins were reported first (Rinke de Wit et al., 1992; Kong et al., 1993; Barriero et al., 2005). Multiple copies have also been observed in alpha proteobacteria (Rhizobium) (Fischer et al., 1993; 1999; Rusanganwa and Gupta, 1993; Wallington and Lund, 1994; Ogawa and Long, 1995; Kaneko et al., 2000 a & b; Galibert et al., 2001; González et al., 2006; Young et al., 2006) and Chlamydiae (Karunakaran et al., 2003). In these genomes one (or more) of the multiple genes is arranged in an operon, with the cognate co-chaperonin *groES*, being the first gene. This *groEL* gene is termed as *groEL1*, while the others that exist separately on the genome are named as *groEL2*, *groEL3* and so on. The high sequence conservation among GroELs from different species is an indication that the mechanism of GroEL is universally conserved. Consequently, several chaperonin homologues from other bacteria have been shown to function in *E. coli*, showing that the spectra of substrate proteins must overlap considerably with those in *E. coli*.

The *E. coli* GroEL/GroES system, despite serving as a paradigm in chaperone-mediated protein folding, is unable to address the multiple occurrence of chaperonin genes in many bacteria. These multiple genes are likely to have arisen through either gene duplication or horizontal gene transfer, either of which is believed to allow subsequent evolution of new roles for different copies of chaperonins, wherein one of them is thought to maintain a housekeeping function in the cell (Ojha et al., 2005; Goyal et al., 2006). For example, in Rhizobia different GroELs are supposed to play different roles in several aspects nitrogen fixation, probably by encountering different substrates, although evidence for significant specificity of function is lacking (Fischer et al., 1993; 1999; Ogawa and Long, 1995; Bittner and Oke, 2006). The analysis of Goyal et al., 2006, has shown that the multiple occurrences of *groEL* genes has arisen from independent

gene duplications in different lineages, and probably these multiple copies have discrete histories of adaptation and selection (Goyal et al., 2006). Moreover, evolution of novel function in chaperonins was reported earlier where the *E. coli groEL* and *groES* were subjected to random mutagenesis and variants with an enhanced ability to fold GFP *in vivo* were selected, with cellular function being sheltered by the wild type GroEL (Wang et al., 2002). Variants that showed improved folding of GFP were shown to have a reduced ability to function as general chaperones in the cell. This study elegantly demonstrated a conflict between the ability of GroEL in recognizing specific substrates on one hand and a wide range of substrates on the other. Such a conflict would be alleviated following gene duplication event. Moreover, although the number of studies that have been carried out on GroEL homologues from different bacteria are limiting, nonetheless they suggest that understanding of chaperonin function may not be generalized and therefore suggest the need for specific and comprehensive studies on these proteins. We have taken up to study the case of GroELs from *Mycobacterium tuberculosis*, causative agent of the deadly disease, tuberculosis.

Mycobacterium tuberculosis H37Rv genome encodes two chaperonin homologues *groEL1* and *groEL2*, both of which are shown to be induced by not only heat shock but also other stresses including nutritional, osmotic and oxidative stresses (Stewart et al., 2002; Hu et al., 2008). Consequently, these are believed to be regulated by the heat shock repressor HrcA owing to the presence of sequences with good matches to the consensus CIRCE elements in the presumptive promoters of the *groES-groEL1* and the *groEL2* genes (Narberhaus, 1999). However, neither of the genes was induced when *M. tuberculosis* was grown in phagosomes, despite the presumably stressful environment (Schnappinger et al., 2003). It is interesting to note that *M. leprae,* Mtb's closest sequence relative, reported to lack a heat shock response, which was supposed to be due to the lack of SigE, and thus the duplicated *groEL* genes in this organism are not induced by heat shock (Williams et al., 2007). Moreover, GroEL2 with an N-terminal truncation, probably either via the envelope-associated protease, Rv2224c, or the proteolytic activity ascribed to GroEL2 (Portaro et al., 2002), was detected in the culture supernatants. GroEL2 deletion mutant was shown to reduce the pathogenicity of *M. tuberculosis* (Rengarajan et al., 2008). Moreover, a role for both the GroELs in enhancing the activity of the *M. tuberculosis* heat shock repressor HspR, suggesting a chaperone-

like activity for these proteins and a role in heat shock regulation (Das Gupta et al., 2008).

Biochemical and biophysical characterization of the recombinant Mtb GroELs showed that despite possessing a high sequence homology with *E. coli* GroEL Mtb GroELs possess biochemical features that deviated significantly from the trademark properties of *E. coli* GroEL. The most striking feature of Mtb GroELs, however, was their oligomeric state, where contrary to expectations, in vitro they did not form the canonical tetradecameric assembly, when purified from *E. coli*. The proteins rather existed as lower oligomers (dimers) irrespective of the presence or absence of cofactors such as the cognate GroES or ATP (Qamra and Mande, 2004; Qamra et al., 2004). Furthermore, they displayed weak ATPase activities and GroES independence in preventing aggregation of the denatured polypeptides. These studies therefore suggested a stringent requirement of host environment for oligomerization. Moreover, evolutionary studies on Mtb *groEL* sequences have suggested that the two *groEL* genes are evolving at different rates with the *groEL1* evolving rapidly, yet not turning these into pseudogenes (Goyal et al., 2006), in contrast to the Streptomyces *groEL*s (Hughes, 1993). Recent studies have shown that in *Mycobacterium smegmatis*, GroEL1, a paralog of Mtb GroEL1 is involved in biofilm formation, whereas GroEL2 is thought to provide housekeeping chaperonin function (Ojha et al., 2005). The observation that GroEL1 can physically associate with KasA, a component of the myclolic acid synthesis pathway, is thought to support the notion that GroEL1 may play chaperonin role in the process of biofilm formation. Furthermore, Mtb GroEL1 is also shown to induce host inflammatory responses (Hu et al., 2008).

Since Mtb GroELs did not turn into pseudogenes during evolution, the possibility of a chaperonin requirement, in the Mtb cytoplasm might be existing (Goyal et al., 2005). Mtb being a slow growing organism, it might, however, require a GroEL function that does not utilize ATP rapidly, but rather with a slow turnover rate. Alternately, additional mechanisms might exist in Mtb, which could mediate regulated oligomerization of Mtb chaperonins. Such regulation might help in the controlled utilization of ATP in nutrient deprived Mtb, as observed for other chaperones such as Hsp90 and small heat shock proteins, which were shown to exhibit temperature or

phosphorylation mediated regulation in oligomerization (Halsbeck et al., 2005). Exploring the oligomeric status and the possibility of regulated oligomerization of Mtb GroEL1 and the source of regulation, if any, was attempted employing immunochemical assays following mass spectrometric analysis and reported in this thesis. It was shown that mycobacterial GroEL1 exists in different oligomeric forms and the switch between the single ring (heptamer) and double ring (tetradecameric) GroEL forms is mediated by phosphorylation. Thus, it is attractive to propose a hypothesis that the naturally synthesized GroEL exists in equilibrium between a dimer and a heptamer, and that heptamer to tetradecamer conversion is mediated by phosphorylation (Figure 6.01). Since a similar strategy of chaperonin oligomerization operates in mitochondrial GroEL, this study, in addition to probing Mtb GroEL biology, might also shed light on the evolution of mitochondria and chloroplasts, which are supposed to be the bacterial symbionts of the eukaryotes (Allen, 2003).
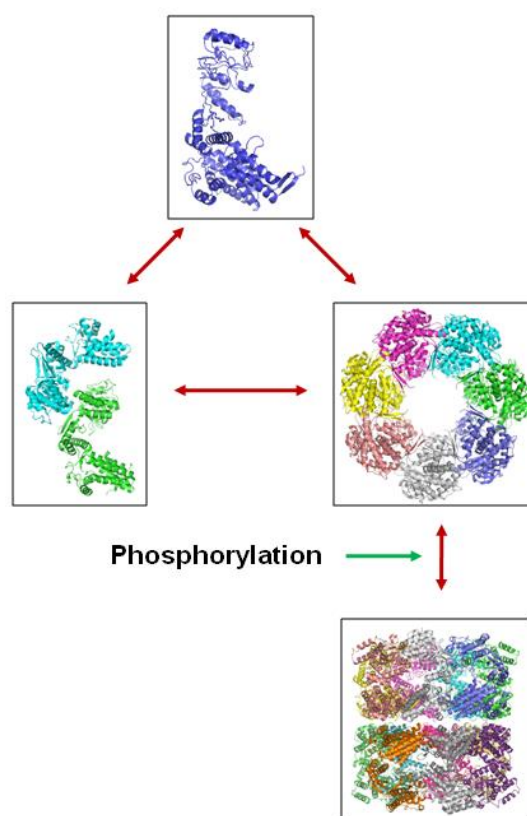
Figure 6.01: Model for the regulation of oligomerization in Mtb GroEL1 mediated by phosphorylation.

Several questions remain unanswered from these observations such as:

- What is the role of these chaperonins in vivo?
- What is the structural basis for this the potential functional specialization of the chaperones from different organism?
- How these genes are regulated and are the regulatory controls for multiple chaperonins are same or different?
- Do the multiple chaperonins form homo-oligomers or hetero-oligomers?
- What is the evolutionary connection between the homologues from different organisms?

Answering these questions requires comprehensive studies and the answers certainly have the potential to enhance our understanding on the biology of chaperonins in particular and the nature of cellular protein folding, in general.

6.2 References

- Allen, J. F. (2003) Why chloroplasts and mitochondria contain genomes. *Comp. Funct. Genom.* **4:**31–36.

- Barreiro, C., González-Lavado, E., Pátek, M. and Martín, J. F. (2004) Transcriptional analysis of the *groES-groEL1*, *groEL2*, and *dnaK* genes in *Corynebacterium glutamicum*: characterization of heat shock-induced promoters. *J. Bacteriol.* **186**, 4813-4817

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

- Chatellier, J., Hill, F., Foster, N. W., Goloubinoff, P. and Fersht, A. R. (2000) From Minichaperone to GroEL 3: Properties of an Active Single-ring Mutant of GroEL. *J. Mol. Biol.* **304,** 897-910

- Das Gupta, T., Bandyopadhyay, B. and Das Gupta, S. K. (2008) Modulation of DNA-binding activity of *Mycobacterium tuberculosis* HspR by chaperones. *Microbiology* **154**, 484-490

- Dickson, R., Weiss, C., Howard, R. J., Alldrich, S. P., Ellis, R. J., Lorimer, G. H., Azem, A. and Viitanen, P. V. (2000) Reconstitution of Higher Plant Chloroplast Chaperonin 60 Tetradecamers Active in Protein Folding *J. Biol. Chem.* **275,** 11829-11835

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

- Fischer, H. M., Schneider, K., Babst, M. and Hennecke, H. (1999) GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum*. *Archives Micro.* **171,** 279-289.

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

- Georgopoulos, C. P., Hendrix, R. W., Kaiser, A. D. and Wood, W. B. (1972) Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. *Nature New. Biol.* **239,** 38-41

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

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

- Haslbeck, M., Franzmann, T., Weinfurtner, D. and Buchner, J. (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12,** 842-846

- Horwich, A. L., Farr, G. W. and Fenton, W. A. (2006) GroEL-GroES-mediated protein folding. *Chem. Rev.* **106,** 1917-1930

- Horwich, A. L., Fenton, W. A., Chapman, E. and Farr, G. W. (2007) Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell. Dev. Biol.* **23**, 115-145

- Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F. U. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**, 147-454

- Hu, Y., Henderson, B., Lund, P. A., Tormay, P., Ahmed, M. T., Gurcha, S. S., Besra, G. S. and Coates, A. R. (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,** 1535-1546

- Hughes, A. L. (1993) Contrasting evolutionary rates in the duplicate chaperonin genes of *Mycobacterium tuberculosis* and *M. leprae. Mol. Biol. Evol.* **10,** 243-255

- 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., Takeguchi, C., Yamada, M. and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti. DNA Res* **7,** 331-338

- 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. and Tabata, S. (2000) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110 *DNA Res* **9,** 189-197

- Karunakaran, K. P., Noguchi, Y., Read, T. D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C. C. and Brunham, R. C. (2003) Molecular Analysis of the Multiple GroEL Proteins of Chlamydiae. *J. Bacteriol.* **185**, 1958-1966

- Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. and Hartl, F. U. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli. Cell* **122**, 209-220

- Kong, H., Coates, A. R., Butcher, P. D., Hickman, C. J. and Shinnick, T. M. (1993) *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc. Natl. Acad. Sci. USA* **90**, 2608–2612

- Kusmierczyk, A. R. and Martin, J. (2001) Assembly of chaperonin complexes. *Mol. Biotechnol.* **19,** 141-153

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

- Maiwald, M., Lepp, P. W. and Relman, D. A. (2003) Analysis of conserved non-rRNA genes of *Tropheryma whipplei. Syst. Appl. Microbiol. 26, 3-12*

- Mayhew, M., da Silva, A. C., Martin, J., Erdjument-Bromage, H., Tempst, P. and Hartl, F. U. (1996) Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* **379**, 420-426

- Narberhaus, F. (1999) Negative regulation of bacterial heat shock genes. *Mol. Micro.* **31,** 1-8

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

- Ojha, A., Anand, M., Bhatt, A., Kremer, L., Jacobs, W. R. Jr. and Hatfull, G. F. (2005) GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* **123,** 861-873

- Portaro, F. C. V., Hayashi, M. A. F., de Arauz, L. J., Palma, M. S., Assakura, M. T., Silva, C. L. and de Camargo, A. C. M. (2002) The *Mycobacterium leprae hsp65* displays proteolytic activity. Mutagenesis studies indicate that the *M-leprae hsp65* proteolytic activity is catalytically related to the Hs1VU protease. *Biochemistry* **41,** 7400-7406

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

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

- Rengarajan, J., Murphy, E., Park, A., Krone, C. L., Hett, E. C., Bloom, B. R., Glimcher, L. H. and Rubin, E. J. (2008) *Mycobacterium tuberculosis* Rv2224c modulates innate immune responses. *Proc. Natl. Acad. Sci. USA* **105,** 264-269

- Rinke de Wit, T. F., Bekelie, S., Osland, A., Miko, T. L., Hermans, P. W., van Soolingen, D., Drijfhout, J. W., Schoningh, R., Janson, A. A. and Thole, J. E. (1992) Mycobacteria contain two *groEL* genes: the second *Mycobacterium leprae groEL* gene is arranged in an operon with *groES. Mol. Microbiol.* **6,** 1995-2007

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

- Rye, H. S., Burston, S. G., Fenton, W. A., Beechem, J. M., Xu, Z., Sigler, P. B. and Horwich, A. L. (1997) Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. *Nature* **388,** 792-798

- Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. and Horwich, A. L. (1999) GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97,** 325-338

- Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., Dolganov, G., Efron, B., Butcher, P. D., Nathan, C. and Schoolnik, G. K. (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. *J. Exp. Med.* **198,** 693-704

- Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Butcher, P. D. and Young, D. B. (2002) The heat shock response of *Mycobacterium tuberculosis*: linking gene expression, immunology and pathogenesis. *Comp. Funct. Genomics* **3**, 348-351

- Takano, T. and Kakefuda, T. (1972) Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. *Nat. New. Biol.* **239**, 34-37

- Thirumalai, D. and Lorimer, G. H. (2001) Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30,** 245-269

- Viitanen, P. V., Gatenby, A. A. and Lorimer, G. H. (1992) Purified chaperonin 60 (*groEL*) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1,** 363-369

- Wallington, E. J. and Lund, P. A. (1994) *Rhizobium leguminosarum* contains multiple chaperonin (*cpn60*) genes. *Microbiology* **140,** 113-122

- Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H. R., Fenton, W. A. and Horwich, A. L. (1995) Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **83,** 577-587

- Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. and Horwich, A. L. (1996) Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* **84,** 481-490

- Williams, D. L., Pittman, T. L., Deshotel, M., Oby-Robinson, S., Smith, I. and Husson, R. (2007) Molecular basis of the defective heat stress response in *Mycobacterium leprae. J. Bacteriol.* **189,** 8818-8827

- Xu, Z., Horwich, A. L. and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)$_7$ chaperonin complex. *Nature* **388,** 741-750

- Young, J. P. W., Crossman, L. C., Johnston, A. W. B., Thomson, N. R., Ghazoui, Z. F., Hull, K. H., Wexler, M., Curson, A. R. J., Todd, J. D., Poole, P. S., Mauchline, T. H., East, A. K., Quail, M. A., 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. and Parkhill, J. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7,** R34

Functional Analysis of

Mycobacterium tuberculosis Hsp70

Protein-Protein Interaction Studies

A1.1 Introduction

In this part of the thesis, I would like to elaborate on the work done on Hsp70 of *Mycobacterium tuberculosis.* As detailed in chapter I, Hsp70 belongs to the family of 70 kDa heat shock proteins, which are well conserved throughout the evolution and are involved in a variety of cellular processes such as the conventional protein folding, export and translocation (Horwich and Bukau, 1998). For this study on Mtb Hsp70 we have considered two observations.

1. Mtb Hsp70 is reported to bind, via its substrate binding domain, to the extracellular domain of human CD40 and thereby blocks various host cell processes that could combat the pathogen (Wang et al., 2001; Lazarevic et al., 2003; MacAry et al., 2004).

2. Mtb genome encodes a repressor for the *hsp70* operon, the HspR. The gene encoding HspR is in operonic arrangement with *hsp70*, *grpE* and *hsp40* (Cole et al., 1998; Stewart et al., 2001). HspR homologues from Streptomycetes were shown to bind Hsp70 and the complex formed was demonstrated to repress *hsp70* operon (Bucca et al., 1995; Grandvalet et al., 1997; Das Gupta et al., 2008).

Having this in mind, we wished to study the molecular interactions between Hsp70 and CD40, and Hsp70 and HspR. Towards this end, the proteins were purified and attempts of crystallizations were carried out. Hsp70 and HspR were individually purified from *E. coli*. Since CD40 is a human membrane bound glycoprotein with several disulphide linkages, we have also attempted to purify this protein employing different expression systems such as yeast and insect cell expression systems. Moreover, the complex of Hsp70 with CD40 was purified from *E. coli.* Detailed methodology followed and the results obtained are discussed in this appendix.

A1.2 Materials

All the chemicals and enzymes were purchased from various commercial sources. Plasmids and oligonucleotide primers used in this study are listed in Appendix II. *E. coli* strains for expression studies were purchased from several commercial sources. Media recipes for expression studies in *E. coli* are given in table A1.01. Components for the *Pichia pastoris* expression system were purchased from Invitrogen, CA, USA. Transformation and competent cells preparation in *P. pastoris* was followed as per manufactures guidelines. Strains of *P. pastoris,* media recipes and stock solutions are listed in tables A1.02, A1.03 and A1.04, respectively.

Media	Composition
SOB	20 g Bactotryptone, + 5 g Yeast extract + 0.5 g NaCl + 10 ml of a 250 mM solution of KCl in 900 ml DDW. Media sterilized by autoclaving. 5 ml of a sterile solution of 2 M MgCl$_2$ was added after autoclaving.
2x YT	16 g Bactotryptone + 10 g Yeast extract + 5 g NaCl in I ltr DDW. Media sterilized by autoclaving.

Table A1.01: Media recipes for *E. coli*.

Strains	Genotype	Phenotype
X-33	wild-type	Mut$^+$
GS115	*his4*	His$^+$, Mut$^+$
KM71H	*arg4 aox1::ARG4*	Mut^S, Arg$^+$
GS115/Albumin	*HIS4*	Mut^S
GS115/pPICZ/*lac*Z	*his4*	His$^-$, Mut$^+$

Table A1.02: Strains of *Pichia pastoris* employed in this study.

Media	Composition
Yeast Extract Peptone Dextrose Medium (YPD)	1% Yeast extract + 2% Peptone Media was sterilized by autoclaving. After autoclaving, dextrose was added at 2%.
Minimal Dextrose Medium (MD)	1.34% Yeast Nitrogen base (YNB) + 4 x 10^{-5}% biotin + 2% dextrose
Minimal Methanol Medium (MM)	1.34% YNB + 4 x 10^{-5}% biotin + 0.5% methanol

Table A1.03: Media recipes for yeast expression system.

Stock solutions	Composition
13.4% Yeast Nitrogen Base (10x)	134 g of yeast nitrogen base (YNB) with ammonium sulfate and without amino acids dissolved in 1000 ml of water and sterilized by filtration. Stored at +4°C.
0.02% Biotin (500x)	20 mg biotin in 100 ml of water and sterilized by filtration. Store at +4°C.
0.4% Histidine (100x)	Dissolve 400 mg of L-histidine in 100 ml of water and sterilize by filtration. Stored at +4°C.
20% Dextrose (10x)	200 g of D-glucose in 1000 ml of water and sterilize by filtration. Stored at +4°C.
5% Methanol (10x)	Mix 5 ml of methanol with 95 ml of water and sterilize by filtration. Stored at +4°C.

Table A1.04: Media components for yeast expression system.

Reagents, bacterial strains, plasmids and Sf9 cell lines for baculovirus based expression were purchased from Invitrogen Inc., CA, USA. The *E. coli* host strain, DH10Bac (genotype: F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 ΔlacX74 *rec*A1 *end*A1 *ara*D139 Δ(*ara,leu*)7697 *gal*U *gal*K λ⁻ *rps*L *nup*G/bMON14272/pMON7124) hosts the transposition event from the donor plasmid onto the baculovirus shuttle vector (bacmid). Recombinant bacmid was purified using Spin columns (Qiagen Inc.).

A1.3 Expression, Purification and Crystallization of Mtb Hsp70

A1.3.1 Cloning the ORF encoding Mtb Hsp70

ORF encoding *Mycobacterium tuberculosis* H37Rv Hsp70 was amplified from Mtb genomic library clone BAC-Rv285 (A8), using primers SCM1640F, SCM1640R. The polymerase employed for amplification was Dynazyme Ext (Finzymes Inc.) and the cycling conditions were as given in table A1.05.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94°C	300	1
Denaturation	94°C	60	
Annealing	48°C	30	5
Extension	72°C	240	
Denaturation	94°C	60	
Annealing	50°C	30	8
Extension	72°C	240	
Denaturation	94°C	60	
Annealing	52°C	30	10
Extension	72°C	240	
Denaturation	94°C	60	
Annealing	55°C	30	17
Extension	72°C	240	
Final Extension	72°C	900	1

Table A1.05: PCR cycling conditions for amplifying ORF encoding Mtb Hsp70

The resulting PCR product was extracted and was cloned into NdeI and BamHI sites on pET-23a(+) (Novagen) to generate pSCM1638. Clones were confirmed by digestion with restriction endonucleases and DNA sequencing (Figure A1.01).
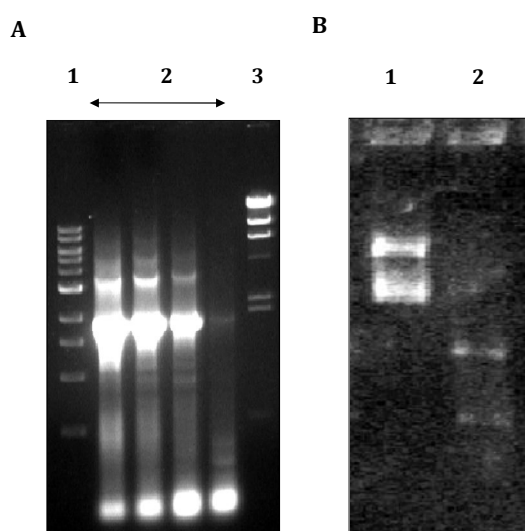
Figure A1.01: PCR Amplification and Cloning of the ORF Encoding Mtb Hsp70.
A. Mtb *hsp70* amplified by PCR were separated by 1% agarose gel and the resulting PCR products were compared with 1kb marker (NEB). The lanes correspond to: 1, 1 kb marker; 2, PCR for Mtb *hsp70*; 3, λ HindIII digest. **B.** Mtb *hsp70* gene was cloned into NdeI and BamHI sites on pET-23a (+). Resulting clone was digested with restriction endonucleases NdeI and BamHI and the digest was resolved on 1% agarose gel. The lanes correspond to: 1, pSCM1638 undigested and 2, pSCM1638 digested.

A1.3.2 Expression of Mtb *hsp70*

Expression of the cloned Mtb *hsp70* gene from pSCM1638 was studied in two strains of *E. coli*, BL21 (DE3) and BL21 (DE3) pLys S. Plasmid encoding Mtb Hsp70, pSCM1641 was transformed into the said strains and the resulting transformants were cultured at 37°C in standard LB broth and terrific broth (TB) in the presence of varying concentrations of IPTG, 0.1 mM through 1mM IPTG (Figure A1.02).

As shown in Figure A1.02, expression of Mtb *hsp70* was detected in *E. coli* strain BL21 (DE3) pLys S, cultured in Luria broth and terrific broth in the presence of various concentrations of IPTG ranging from 0.1 mM through 1mM IPTG. However, leaky expression was detected in cells cultured in TB. Therefore, we have proceeded with culturing in LB.
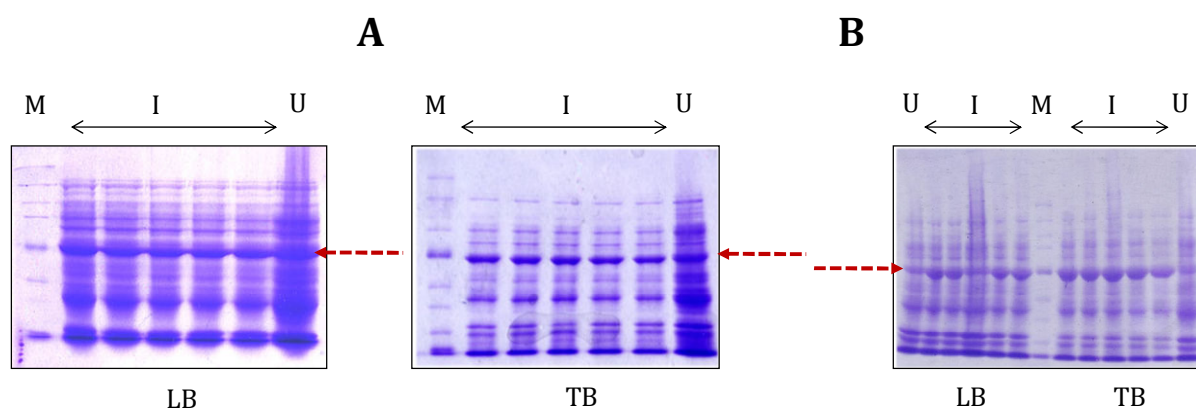
Figure A1.02: Expression analysis of Mtb hsp70. The plasmid pSCM1638 was transformed into BL21 (DE3) **(A)** and BL21 (DE3) pLys S **(B)**. Resulting transformants were cultured in the indicated media and were induced in the presence of 0.1 to 1 mM IPTG (I) for six hours. Resulting cell lysates were resolved on 10% SDS-PAGE and compared with the culture grown in the absence of IPTG (U). Bands corresponding to the Hsp70 are indicated by arrows and M indicates the Broad range protein marker (NEB).

A1.4.3 Analysis of Solubility of the Produced Hsp70

Having shown that Mtb Hsp70 is produced in *E. coli,* we have analyzed if the protein is soluble when produced in the heterologus host environment. The BL21 (DE3) pLys S cultures expressing Mtb *hsp70* were resuspended in Lysis buffer containing 25 mM Tris.HCl (pH: 8.5), 1 mM EDTA and 1 mM PMSF. The suspension was subjected to ultra-sonication. Resulting cell lysates were cleared by centrifugation. Solubility of the protein was determined by resolving the supernatant and precipitated fractions of the lysates on 10% SDS-PAGE (Figure A1.03).

As shown in Figure A1.03, Mtb Hsp70 was found to be soluble when produced at 37 °C. Therefore, this condition was chosen as starting point for further steps of purification.
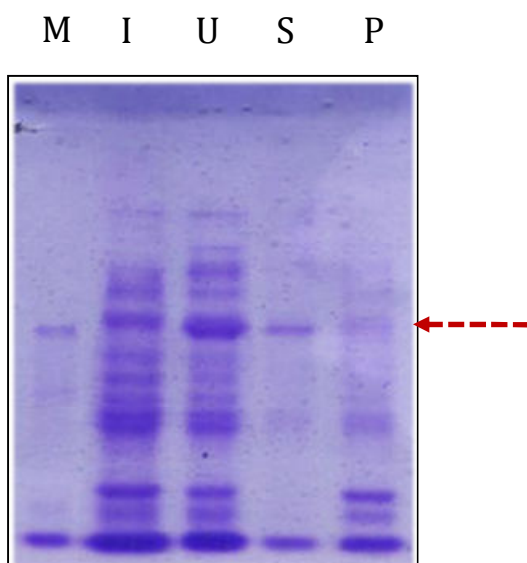
Figure A1.03: Solubility of Mtb Hsp70. *E. coli* BL21 (DE3) pLys S cultures expressing Mtb *hsp70* were lysed by ultra-sonication. Insoluble precipitate (P) and soluble supernatant (S) were seperated by centrifugation and were resolved on 10% SDS-PAGE. The lanes corresponding to Broad range protein marker (M), induced whole cell pellet (I), uninduced whole cell pellet (U) and an arrow indicating the molecular mass of Mtb Hsp70 are presented.

A1.3.4 Purification of Mtb Hsp70

Purification of Mtb Hsp70 was accomplished in five steps.

A1.3.4.1 Enrichment by Heat Precipitation

Heat precipitation is one of the classical methods employed in protein purification, wherein heat resistant proteins are enriched and the heat susceptible proteins get aggregated upon exposure to elevated temperatures. The aggregated proteins are precipitated out by centrifugation, to obtain supernatant having the enriched protein. Since Mtb Hsp70 is a homologue of heat shock proteins, we wanted to test the heat stability of this protein, which would assist further in the process of purification.

Having shown that Mtb Hsp70 is soluble at 37 °C, owing to its homology to heat stable *E. coli* DnaK, stability of Mtb Hsp70 to heat was assessed to aid in further steps of purification. Cell lysates expressing Mtb Hsp70 were incubated at different temperatures ranging from 40 °C to 70 °C for one hour. Stability of Mtb Hsp70 was assayed by resolving the soluble and precipitated protein fractions on 10% SDS-PAGE.
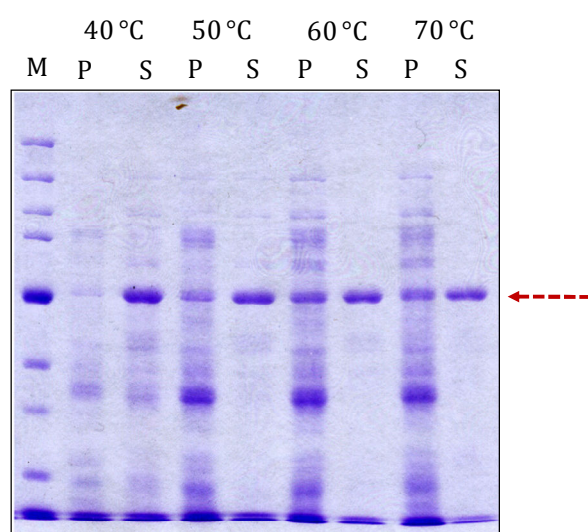

Figure A1.04: Stability of Mtb Hsp70 to Heat. BL21 (DE3) pLys S lysates expressing Mtb Hsp70 were incubated at the indicated temperatures for one hour. Soluble supernatant (S) and insoluble precipitated (P) fractions were resolved on 10% SDS-PAGE. Lane corresponding to Broad range protein marker (NEB) (M) and an arrow indicating the molecular mass of Mtb Hsp70 are presented.

As shown in the Figure A1.04, the protein was found to be extremely tolerant to heat as it could withstand temperatures till 70 °C. Moreover, several heat labile proteins are increasingly precipitated with the increase in temperature, thereby enriching Hsp70 in solution. However, since significant amount of Hsp70 was precipitated, when incubated at temperatures 60 °C and 70 °C, we have proceeded with the incubation at 50 °C.

A1.3.4.2 Capture by Ion Exchange Chromatography

Theoretical pI of Mtb Hsp70 at 4.7 led us to employ Ion exchange chromatography for further purification of the protein. Cell lysate following heat precipitation was loaded onto Mono Q 10/100 GL column (GE Life sciences). The column was subjected to linear gradient from start buffer comprising 25 mM Tris.HCl (pH: 8.5), 1 mM EDTA through Elution buffer comprising 25 mM Tris.HCl (pH: 8.5), 1 mM EDTA and 1 M NaCl (Figure A1.05).
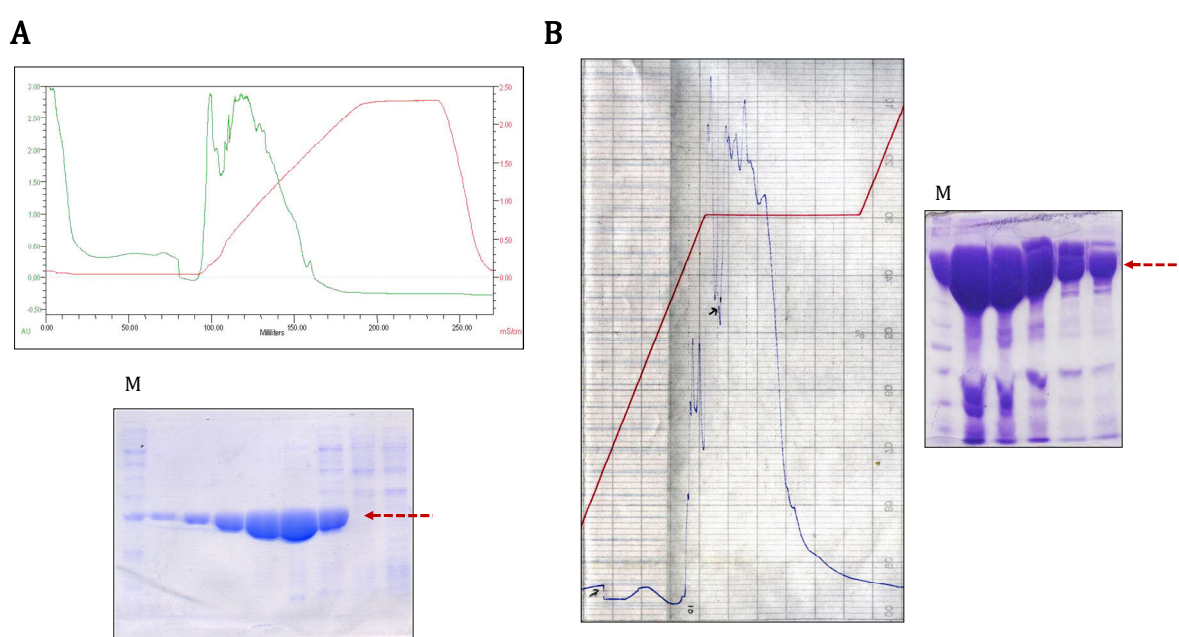
A

B

Figure A1.05: Ion Exchange Chromatography of Hsp70. Chromatograms showing the elution profiles of Mtb Hsp70 in linear gradient from 0 to 1 M NaCl **(A)** and a step gradient at 600 mM NaCl **(B)**. Corresponding eluates were resolved on 10% SDS-PAGE and the molecular mass was compared with the broad range protein marker (NEB) (M), as indicated by the arrows.

As shown in figure A1.05A, Hsp70 was eluted at 600 mM NaCl. Therefore, step gradients at 600 mM NaCl were set in subsequent rounds of purification to increase the yield of Hsp70 (Figure A1.05B).

A1.3.4.3 Further Purification by Hydrophobic Interaction Chromatography

Hsp70 as a chaperone molecule recognizes its substrates by the virtue of hydrophobic patches in its substrate binding domain (Swain et al., 2007). This property of Hsp70 has led us to employ hydrophobic interaction chromatography for further purification of the protein. In hydrophobic interaction chromatography the proteins are bound in high salt solvents and eluted by lowering the salt concentration.

Samples from ion exchange chromatography were suspended in the start buffer comprising 25 mM Tris.HCl (pH: 8.5), 1 mM EDTA, 1 M NaCl and loaded onto HiPrep Phenyl FF 16/10 column (GE Life sciences). The protein was eluted in water (Figure A1.06).
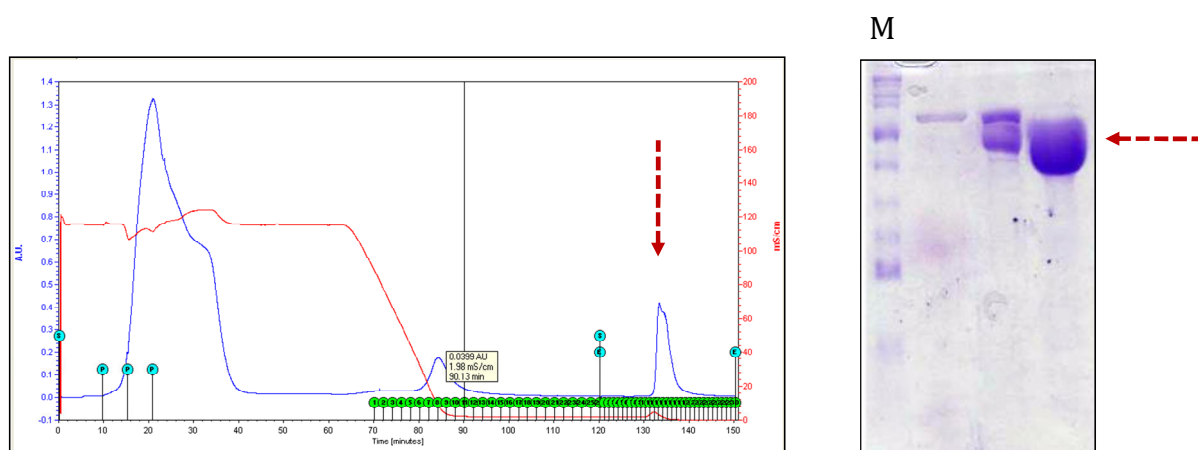

Figure A.06: Hydrophobic Interaction Chromatography of Hsp70. Mtb Hsp70 was loaded onto HiPrep Phenyl FF 16/10 column (GE Life sciences). The protein was eluted in water. Peak and bands corresponding to Hsp70 are indicated by arrows. The fractions from the peak were resolved on 10% SDS-PAGE.

A1.3.4.4 Determination of Oligomeric Status by Gel Filtration Chromatography

Eluates from Ion exchange chromatography were loaded on to Superdex S200 10/100 GL (GE Life Sciences). The molecular mass and oligomeric status of Hsp70 was determined by comparing with the molecular weight standards. Hsp70 was found to exist as monomer and the molecular weight was determined to be 66 kDa (Figure A1.07).
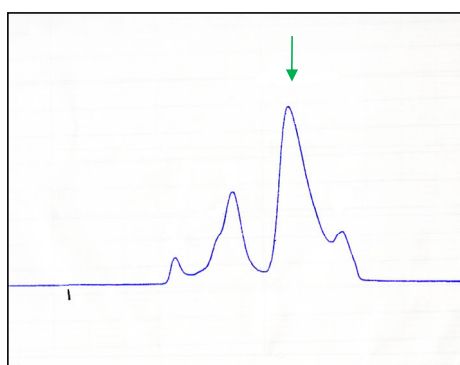

Figure A1.07: Molecular Exclusion Chromatography of Hsp70. Mtb Hsp70 was resolved on Superdex S 200 10/100. Peak corresponding to Hsp70 is indicated by an arrow.

A1.3.4.5 Determination of Purity of Hsp70 by Immunoblotting

Purified Hsp70 was resolved on 10% SDS-PAGE and was transferred onto charged PVDF membrane. Presence of Mtb Hsp70 was probed using mouse monoclonal antibody specific to Mtb Hsp70, IT40 (TBVTRM, Colorado State Univeristy), at 1:50 dilutions using the standard protocol as explained in chapter II (Figure A1.08).
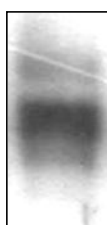

Figure A1.08: Immunoblotting of Mtb Hsp70. Mtb Hsp70 was resolved 10% SDS PAGE and the presence of proteins was probed by Hsp70 specific antibody, IT41. The blot was developed by ECL + Westernblotting kit (GE Life sciences).

A1.3.5 Crystallizations of Mtb Hsp70

In order to elucidate the structural features of Mtb Hsp70, crystallization trials of the protein were attempted.Hsp70 at different concentrations ranging from 10 through 18 mg/ml was subjected to crystallizations. Crystallizations were performed in Microbatch plates (Hampton Inc.) employing the crystallization matrices Crystal Index, Crystal screens I and II (Hampton Inc.). The plates were incubated at 4 °C and 25 °C.

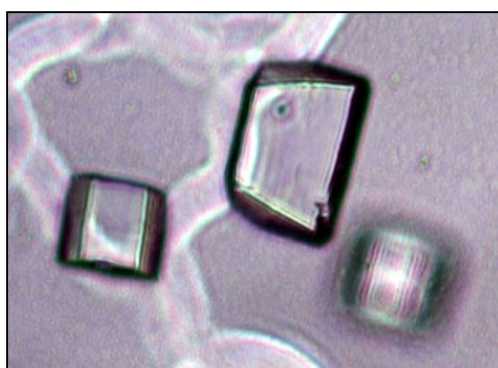*Figure A.08:* **Crystallizations of Mtb Hsp70.** Crystallizations were set with Crystal index screen in the Microbatch plate. The plate was incubated at 25 °C. Presence of crystals was recorded after 14 days.

Figure A1.08 shows crystals of Mtb Hsp70. Crystals appeared after 14 days, in the H10 condition (0.1 M Bis-Tris pH 6.5, 45% v/v Polypropylene glycol P 400) of Crystal index (Hampton Inc.) at 25 °C.

A1.4 Cloning, Expression and Purification of Human sCD40

Human CD40 is a member of TNF receptor family and plays important role in humoral and adaptive immunity. Recent immunological studies showed that Mtb Hsp70 interacts with the extracellular domain of human CD40 and elicits cytokine and CC chemokine response in PBMC, THP-1 and HEK 293 cell lines (Wang et al., 2001; Lazarevic et al., 2003; MacAry et al., 2004). Domain organization in CD40 constitutes three domains: 22-kDa extracellular domain, 22-residue long trans-membrane helix and a 42 residue long cytoplasmic tail (Banchereau et al., 1994; Foy et al., 1996; Young et al., 1998).

Towards understanding the structural basis of the interaction of Hsp70 with CD40, we have attempted to co-express the extracellular domain of CD40 (sCD40) with Mtb Hsp70 and crystallize the complex. Likewise, since CD40 is a human glycoprotein, expressing and purifying this protein in eukaryotic expression systems, such as yeast and Insect cell expression systems were attempted. The details of the experiments are presented here.

A1.4.1 Generation of cDNA for Human sCD40

The cDNA for Human sCD40 was generated as described below.

A1.4.1.1 Isolation of Human Total RNA from Human Macrophages

Human macrophages HEK 299 cells were cultured and recovered by centrifugation. To the cell pellet was added 500 µl of TRIzol reagent (GIBCO BRL) and the cells were homogenized by vigorous vortexing. To this was added add 100 µl of chloroform and the tube was mixed by gentle vortexing and was allowed to stand for 5 minutes at room temperature. The mixture was centrifuged at 12,000 g for 15 minutes. Pink colored aqueous phase was transferred into a new tube containing 500 µl of isopropanol, mixed gently and centrifuged further at maximum speed for 15 minutes at 4 °C. The pellet containing RNA was washed with 500 µl of 70% ethanol and air dried. The pellet was dissolved in 50 µl of DEPC treated water. Concentration of the RNA was determined by measuring the absorbance at 260 nm, which was 144 ng/µl.

A1.4.1.2 Amplification of cDNA Encoding Human CD40 by RT PCR

To 500 ng of RNA, 1 μl of RNase Stop solution was added and was incubated at 70 °C for 10 minutes for denaturation. The contents were chilled on ice to lock the conformation. To this was added oligo dT primer, mixed and incubated at 25 °C for ten minutes for annealing. This mix was again chilled and to this were added 10 mM dNTPs, DTT and 1x MMLV Reverse transcriptase. The contents of the tube were mixed and incubated at 37 °C for 50 minutes. This is accomplished for the extension. After the extension, the reaction mix was incubated at 70 °C for 15 min to stop the reaction. The cDNA so obtained was stored at -20 °C.

A1.4.1.3 Amplification of sCD40 by PCR

Human cDNA was employed for the amplification of the ORF encoding human sCD40. *sCD40* at this stage was amplified using different primer pairs SCM1641F/SCM1641R, SCM1642F/1642R, SCM1643F/SCM1643R and SCM1644F/SCM1644R to incorporate (His)$_6$ tag at N-terminus and at C-terminus, respectively with the presence or absence of the signal peptide. The conditions for the PCR are given in table A1.06.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94°C	300	1
Denaturation	94°C	30	
Annealing	55°C	30	30
Extension	72°C	120	
Final Extension	72°C	600	1

Table A1.06: PCR cycling conditions for amplifying ORF encoding Human CD40.

Data in Figure A1.09 shows the PCR product corresponding to sCD40. This was further used for cloning into different vectors as detailed in the subsequent paragraphs.
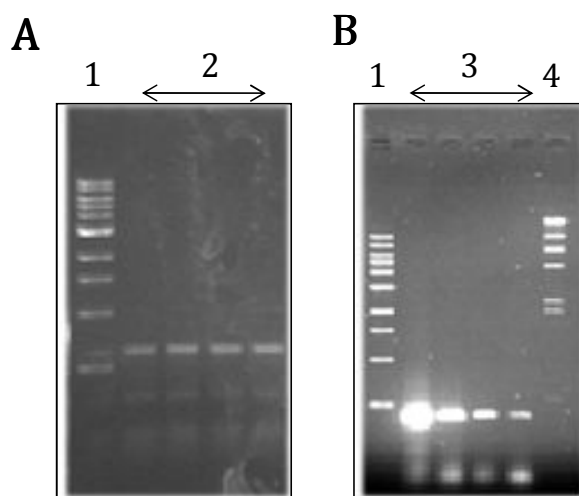
Figure A1.09: Amplification of sCD40. **A.** Human total cDNA was amplified using human total RNA employing reverse transcriptase reaction. From this pool cDNA for Human sCD40 was amplified using CD40 specific primers. **B.** Human sCD40 cDNA was further amplified using different primer sets. The samples were resolved on 1 % agarose gel. The lanes correspond to: 1, 1 kb ladder (NEB); 2, human sCD40 cDNA; 3, human sCD40 PCR amplified product; 4 λ/HindIII digest.

A1.4.2 Cloning, Co-expression, Purification of sCD40 with Mtb Hsp70

Human sCD40 was cloned and co-expressed with Mtb Hsp70. For this, the said ORFs were cloned into pETDuet-1 (Novagen).

A1.4.2.1 Cloning of sCD40 into pETDuet-1

Human sCD40 was cloned into NcoI and XhoI of MCS2 in pETDuet-1 with the (His)$_6$ tag at the N-terminus and the C-terminus of resulting protein. Further, two more constructs were generated without the signal peptides to generate clones pSCM1639-pSCM1642 (Appendix II). The clones were confirmed by digestions by restriction endonucleases and sequencing (A1.10).
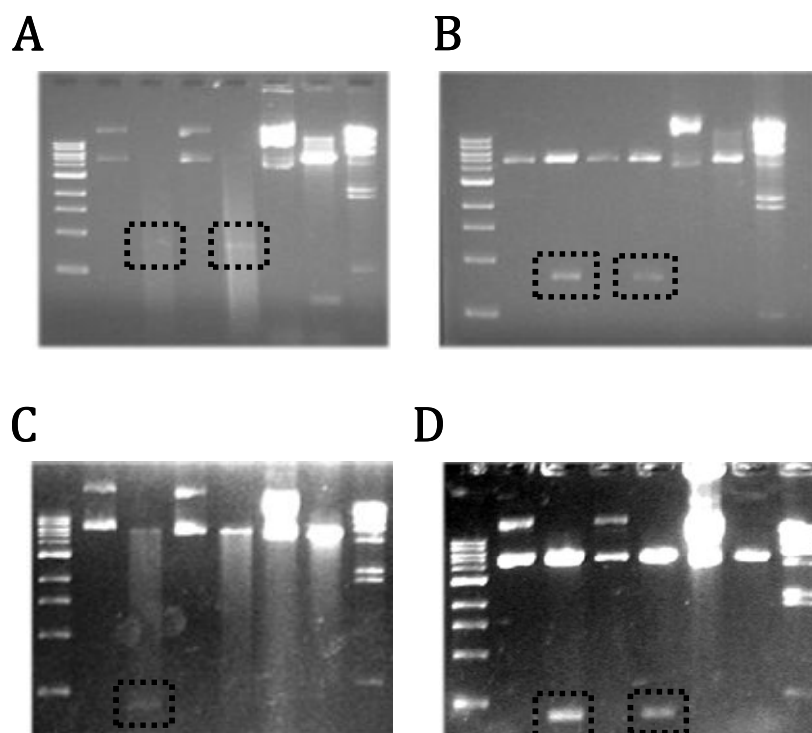
Figure A.10: Cloning of CD40 into pETDuet-1. Human sCD40 was cloned onto NdeI and XhoI sites on MCS2 of pETDuet-1**.** Resulting clones, pSCM1639 **(A)**, pSCM1640 **(B)**, pSCM1641 **(C)** and pSCM1642 **(D)**, were digested with the restriction endonucleases NdeI and XhoI and digests were resolved on 1% agarose gel. Bands corresponding to CD40 are boxed.

A1.4.2.2 Cloning of *hsp70* into pETDuet-1

Mtb *hsp70* was cloned into NcoI and BamHI sites on MCS1 of pSCM1639, pSCM1640, pSCM1641 and pSCM1642 to generate pSCM1643, pSCM1644, pSCM1645 and pSCM1646, respectively. Likewise the peptide binding domain (PBD) of Hsp70 was amplified and cloned into NcoI and BamHI sites on pSCM1641 and pSCM1642, to generate pSCM1647 and pSCM1648, respectively. The clones were confirmed by digestion with restriction endonucleases and sequencing (Figure A1.11)
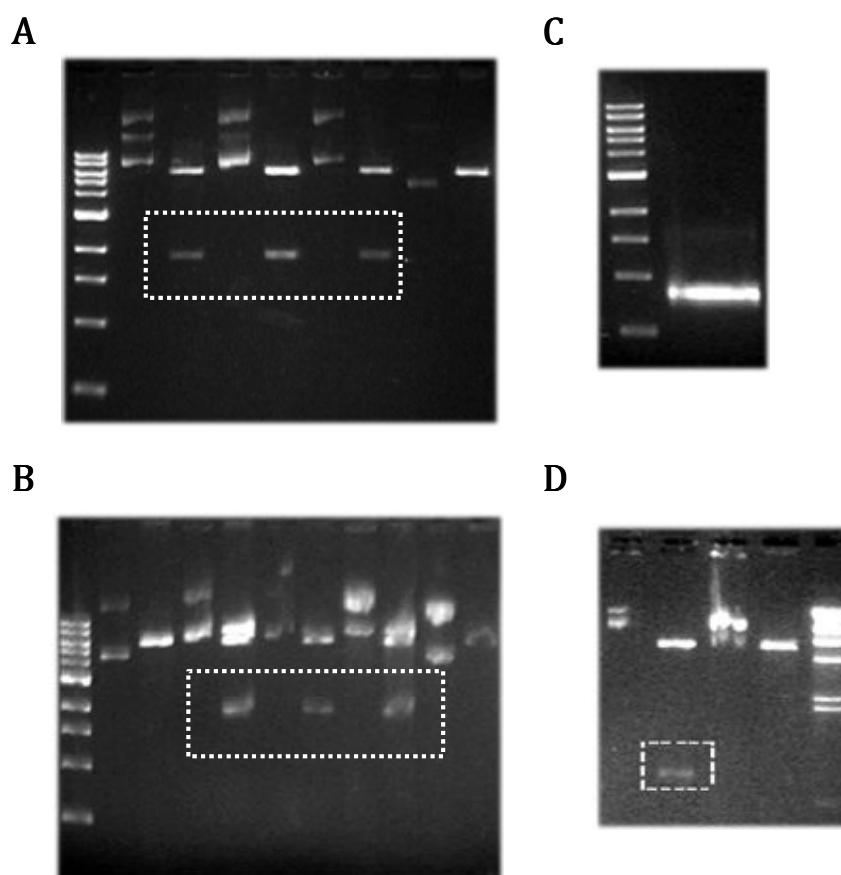
A

C

B

D

Figure A1.11: Cloning of Mtb hsp70. Plasmids pSCM1645 **(A)** and pSCM1646 **(B)** were digested with NcoI and BamHI and the digests were resolved on 1% agarose gel. **C.** Amplification of peptide binding domain of Mtb Hsp70. **D.** pSCM1647 was digested with NcoI and BamHI and the digests were resolved on 1% agarose gel. Fragments corresponding to Mtb *hsp70* and *pbd* are boxed.

A1.4.2.3 Cloning of sCD40 into pET-23a(+)

For the comparison of structural studies we attempted to purify CD40 individually, using different expression systems. Initially, the human sCD40 was cloned individually into NdeI and XhoI sites of pET-23a(+) to generate pSCM1649, which was used for expression and purification of sCD40 from *E. coli*. The clone pSCM1649 was confirmed by digestion by restriction endonucleases and sequencing (Figure A1.12).
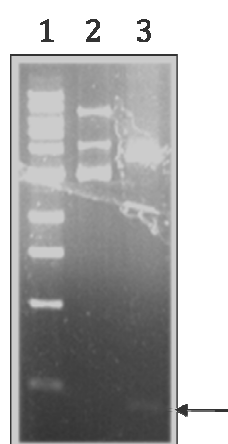
Figure A1.12: Cloning into pET23a(+). Plasmid pSCM1649 was digested with the restriction endonucleases NdeI and XhoI and the digests were resolved on 1% agarose gel. The lanes correspond to: 1, 1 Kb ladder (NEB); 2, pSCM1649 undigested; 3, pSCM1649 digested. Fallout of CD40 is indicated by an arrow.

A1.4.3 Co-expression, Purification and Crystallization of sCD40 - Hsp70 Complex

Plasmid pSCM1643 was transformed into different expression strains of *E. coli*. Expression profiles, co-purification of the complex and crystallizations were attempted.

A1.4.3.1 Co-expression Profiles of Human sCD40 and Mtb Hsp70

Co-expression of the ORFs encoding sCD40 and Hsp70 was checked in different strains of *E. coli* such as BL21 (DE3), BL21 (DE3) pLys S, BL21 (DE3) Rosetta and BL21 (DE3) Codon Plus RIL and in different growth media such as LB, TB, SOB and 2XYT (Figure A1.13).

LB   TB   SOB   LB   TB   SOB

I   U   I   U   I   U   M   U   I   U   I   U   I

BL21 (DE3) Codon Plus   BL21 (DE3) pLys S
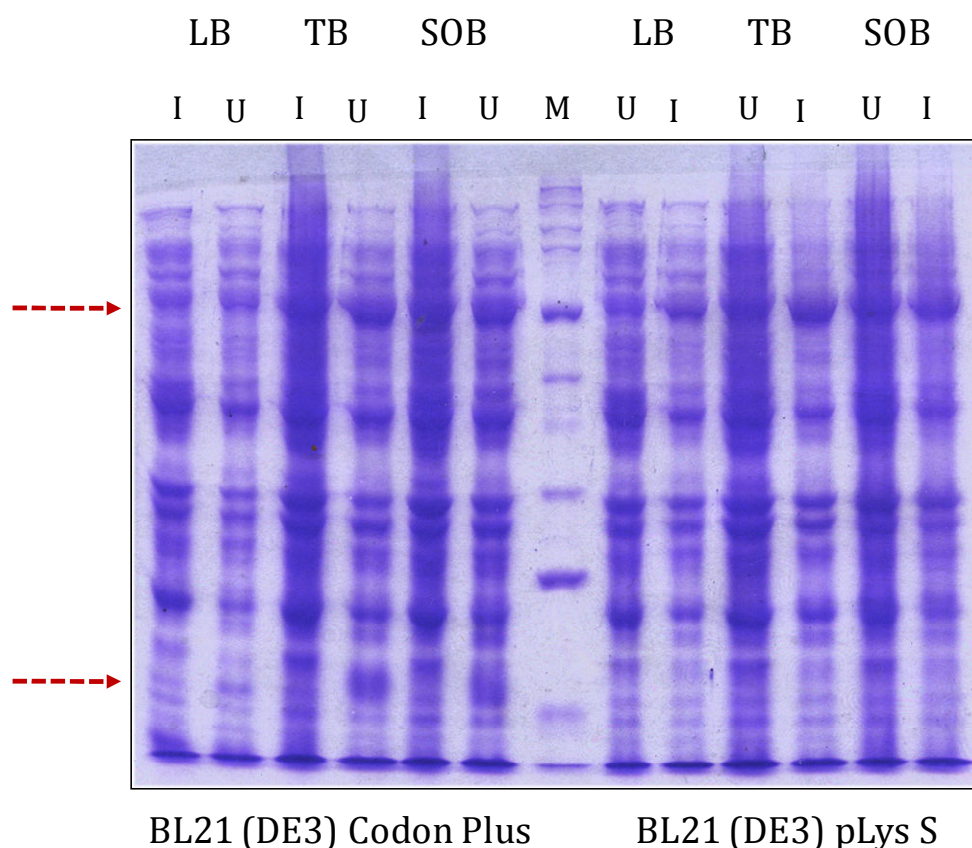
Figure A1.13: Expression Profiles of of Hsp70 and sCD40. BL21 (DE3) Codon Plus RIL and BL21 (DE3) pLys S transformants were cultured in the indicated media and were induced in late log phase with 1 mM IPTG for 6 hrs. The cells were lysed and the samples were resolved on 10% SDS-PAGE. The lanes correspond to: M, Broad range protein marker (NEB); I, Induced and U, uninduced whole cell pellets.

As shown in the Figure A1.13, expression of ORF encoding Hsp70 was detected in all the conditions tested. However, ORF encoding sCD40 was expressed in *E. coli* BL21 (DE3) Codon Plus RIL when cultured in TB and SOB media. Therefore, we have chosen to *E. coli* BL21 (DE3) Codon Plus RIL to be cultured in TB media, for the further studies.

A1.4.3.2 Solubility and Co-purification of Hsp70 and sCD40

Plasmid pSCM1649 was transformed into *E. coli* BL21 (DE3) Codonplus RIL and the cultures co-expressing the hCD40 and Hsp70 were lysed in 50 mM Tris.HCl (pH: 8.0), 150 mM NaCl, 20 mM Imidazole and 1 mM PMSF. The insoluble and soluble fractions were resolved on 10% SDS-PAGE to check the solubility of the proteins. The complex of sCD40 and Hsp70 was eluted in 50 mM Tris.HCl (pH: 8.0), 150 mM NaCl and 150 mM Imidazole. The protein was dialyzed to remove imidazole and stored at 4 °C (Figure A1.14).
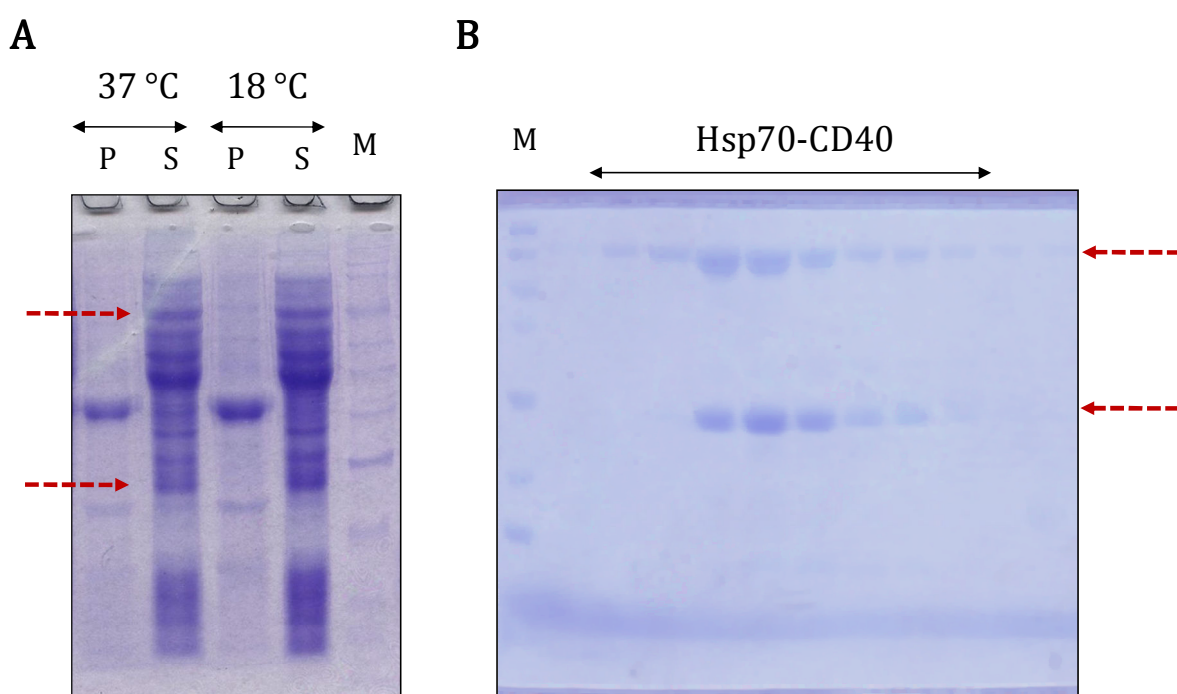
Figure A1.14: Solubility and Co-purification of Hsp70 and sCD40: **A.** BL21 (DE3) Codon Plus RIL expressing Hsp70-CD40 was cultured at the indicated temperatures. The cells were lysed and the insoluble (P) and supernatant (S) fractions were resolved on 12% SDS-PAGE. The lanes correspond to: M, Broad range protein marker (NEB); P, pellet and S, supernatant. **B.** Co-Purification of Hsp70-CD40. Cell lysates harboring the complex were loaded onto Ni-NTA resin in 20 mM imidazole and was eluted in 150 mM imidazole. The molecular weights of the proteins were compared with the broad range protein marker (M) by resolving on 12% SDS-PAGE.

A1.4.3.3 Crystallizations of Hsp70-CD40 complex

16.4 mg/ml of the Hsp70-CD40 complex was subjected to crystallizations in Microbatch Crystallization plates. The crystals for Hsp70-CD40 complex appeared in the E3 (25% Ethylene Glycol) condition of the Crystal screen II at 4 °C (Figure A1.15).
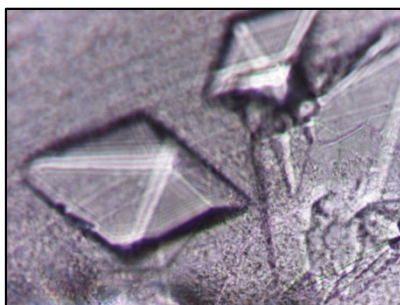

Figure A1.15: Co-crystallization of Hsp70-CD40. The complex was subjected to crystallizations and the plates were incubated at 4 °C. Crystals appeared after 11 days.

A1.4.4 Co-expression of ORFs encoding sCD40 and PBD and the Solubility of the Complex

Since Hsp70 interacts with sCD40 via the peptide binding domain, studying the interaction between the PBD and sCD40 was attempted. To this end, plasmid pSCM1647 was transformed into BL21 (DE3) Codon Plus RIL and the transformants were cultured in TB and SOB separately. Co-expression profiles of the cloned genes were assessed in the presence of various concentrations of IPTG ranging from 0.1 mM to 1mM (Figure A1.16A). Further, to analyze the solubility of the complex, these cells were lysed by ultra-sonication and the fractions corresponding to supernatant and precipitate were resolved on 10% SDS-PAGE (Figure A1.16B).

The data in figure A1.16 shows that the ORFs encoding sCD40 and PBD got expressed, when cultured in TB and SOB. However, both the proteins were found to be insoluble. Moreover, the presence of Sodium Lauroyl Sarcosine at 10 mM did not improve the solubility.
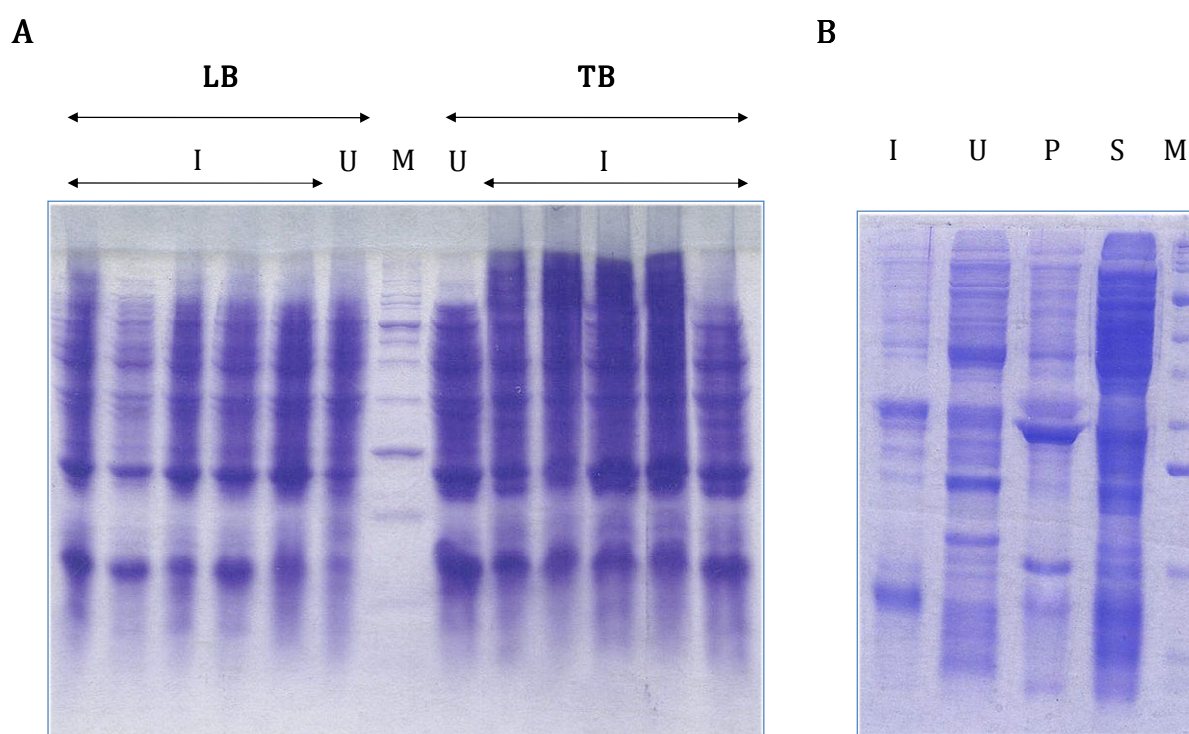
Figure A1.16: Co-expression of Hsp70 PBD and sCD40 ORFs. A. Co-expression of *hsp70 PBD* and *sCD40* was attempted in the indicated media. Whole cell lysates were resolved on 10% SDS-PAGE. **B.** Solubility of Hsp70 PBD and sCD40 was assessed by sonicating the cell suspension in the presence of 10 mM Sodium Lauroyl Sarcosine. The lanes indicate: M, Broad range protein marker (NEB); I, Induced whole cell lysate; U, uninduced whole cell lysate; P, insoluble precipitate and S, soluble supernatant.

A1.4.5 Expression, Purification and Refolding of sCD40 From *E. coli*

Since CD40 is speculated to form eight disulphide bridges in the extracellular domain, expression of the gene encoding sCD40 was attempted in *E. coli* BL21 (DE3) Rosetta-gami and *E. coli* BL21 (DE3) Rosetta-gami2 (Novagen) strains in LB. These strains are *trx* and *gor* mutants, where the two principal redox systems of *E. coli* are silenced. Therefore, the cytoplasm in these strains was shown to be oxidizing rather than reducing, thereby favoring disulphide formation. Human sCD40 was later purified under denaturing conditions and was refolded by gradually removing the denaturant.

A1.4.5.1 Expression of ORF Encoding of sCD40

Plasmid pSCM1649 was transformed into the said strains and the resulting transformants were cultured in LB broth at 37 °C. Cultures were induced at late log phase with 0.7 mM IPTG for 3 hrs. Whole cell pellets were resolved on 15% SDS-PAGE (Figure A1.17A). Further, for checking the solubility of the proteins, the cells were lysed in the presence of 10 mM Sodium Lauroyl Sarcosine and the fractions were resolved on 15% SDS-PAGE (Figure A1.17B).

A1.4.5.2 Purification of sCD40 under Denaturing Conditions

Human sCD40 was found to be insoluble when expressed in *E. coli*. Henceforth we have purified the protein under denaturing conditions using Ni-NTA chromatography. For this, the cells expressing *sCD40* were resuspended in lysis buffer comprised of 50 mM Tris.HCl (pH: 8.0), 150 mM NaCl, I mM PMSF and were lysed by ultra-sonication. Insoluble precipitate bearing sCD40 was resuspended in denaturing buffer (50 mM Tris.HCl (pH: 8.0), 150 mM NaCl, 20 mM Imidazole, 8 M Urea and 20 mM β-Mercaptoethanol) and was loaded onto Ni-NTA. Unbound proteins were washed with 20 mM imidazole and sCD40 was eluted in the presence of 150 mM Imidazole (Figure A1.17C).

A1.4.5.3 Refolding of sCD40

Refolding of the protein was carried out by subjecting 10 μg/ml of sCD40 to step-dialysis with reducing the concentrations of the denaturants. The buffers for the serial dialysis were in the order listed below:

1. 50 mM Tris (pH: 8.0), 150 mM NaCl, 6 M Urea and 10 mM β-Mercaptoethanol,

2. 50 mM Tris (pH: 8.0), 150 mM NaCl, 4 M Urea and 5 mM β-Mercaptoethanol,

3. 50 mM Tris (pH: 8.0), 150 mM NaCl and 2 M Urea,

4. 50 mM Tris (pH: 8.0) and 150 mM NaCl.

First three steps of dialysis were set at room temperature for 2 hours each. Fourth step was set at 4 °C. Efficiency of the refolding reaction was assessed by

estimating the secondary structural composition of the refolded protein using Circular Dichroism (Figure A1.18).
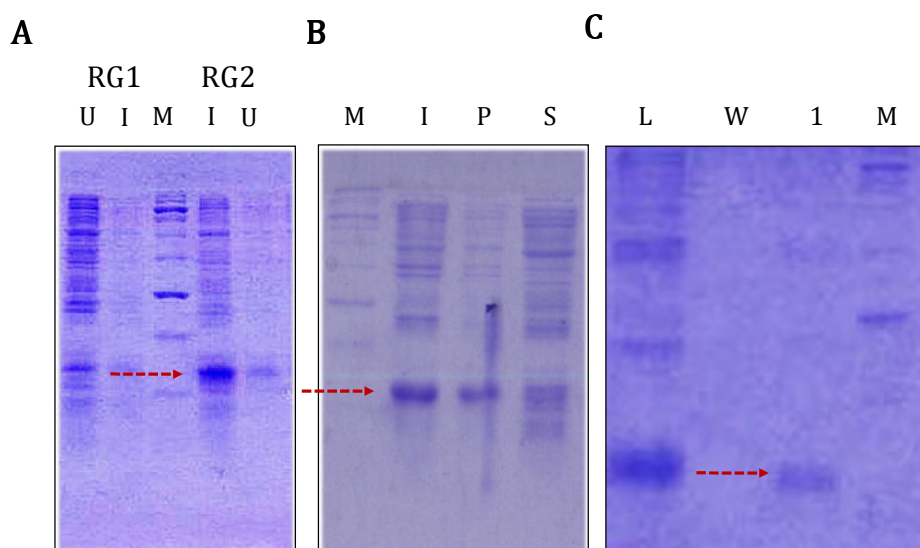
Figure A1.17: Purification of sCD40 from E. coli. **A.** Expression of *sCD40* was attempted in Rosetta-gami (RG1) and Rosetta-gami 2 (RG2) cells. **B.** solubility of the protein comparing the soluble and precipitated fractions, resolved on 12% SDS-PAGE. **C.** Purification of sCD40 was performed in denaturing conditions and the fractions were resolved on 12% SDS-PAGE. The lanes correspond to: U, Uninduced and I, induced cell lysate; P, insoluble fraction; S, soluble fraction; L, load; W, wash; 1, purified protein and M, Broad range protein marker (NEB).
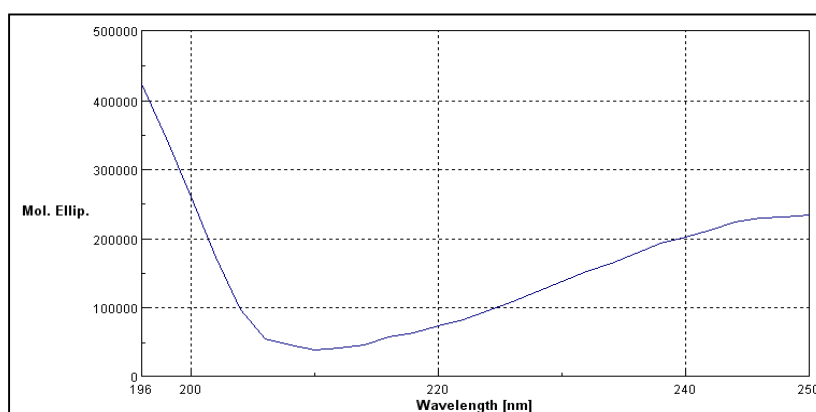
Figure A1.18: Circular Dichroism of Refolded sCD40. Human sCD40 was dialyzed against 10 mM Tris.Hcl (pH: 8.0). Molecular elipticity was measured from 196 nm to 250 nm.

A1.4.6 Cloning and Recombination of hCD40 for Yeast Expression System

Yeast expression systems based on *P. pastoris* are gaining importance for purifying eukaryotic proteins with post translational modifications. Post translational modifications profiles in *P. pastoris* were demonstrated to be equivalent to the human profiles (Cregg et al., 1993). Since human CD40 is a glycoprotein, we have attempted to produce the protein in *P. pastoris* expression system.

A1.4.6.1 Cloning of Human CD40 into pPICZα A

ORF encoding sCD40 was cloned into the KpnI and XhoI sites of pPICZα A (Invitrogen) to generate the clones, pSCM1650. The clones were selected in *E. coli* Top10F' on LB agar plates supplemented with 30 μg/ml of zeocin. The clones were confirmed by colony PCR and sequencing (Figure A1.19).
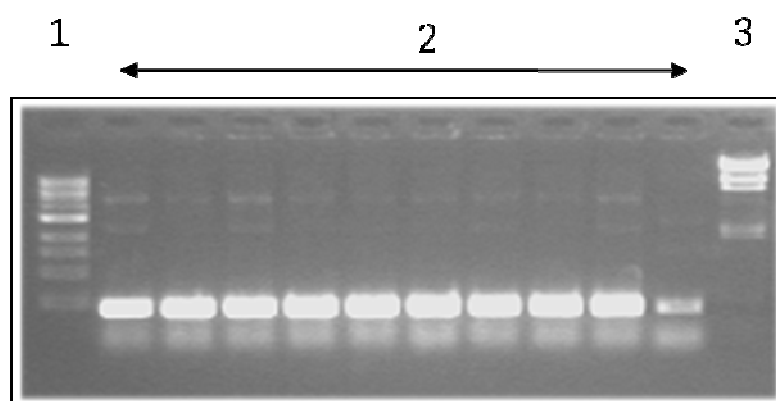
Figure A1.19: Cloning of CD40 into pPICZα. Colony PCR was performed with the clones harboring sCD40. The PCR products were resolved on 1% agarose gel. The lanes correspond to: 1, 1 Kb ladder (NEB); 2, PCR products from clones pSCM1650-pSCM1659, showing sCD40 fragment and 3, λ/HindIII digest (NEB).

A1.4.6.2 Recombination of Human CD40 into *P. pastoris*

One of the confirmed clones of pSCM1650, was linearised by digesting with the restriction endonucleases, SacI. 3 μg of the linearised product was transformed into the *P. pastoris* strains X-33, GS115 and KM71H. The transformants were selected on YPD

plates supplemented with 100 µg/ml of zeocin. The resulting recombinants were checked for the Mut (Methanol utilization) phenotype by spotting the cultures on to Minimal Dextrose and Minimal Methanol plates supplemented with/without histidine. Among the zeocin resistant transformants eight recombinants from X-33 and ten from GS115 were screened for the Mut phenotype (Figure A1.20). Two strains, GS115 Albumin is Mut^S and GS115/pPICZ/*lac*Z is Mut$^+$ were included as controls.

Minimal + Dextrose   Minimal + Methanol

Recombinants of GS115

Recombinants of X33

Histidine +

Recombinants of GS115

Recombinants of X33

Histidine -
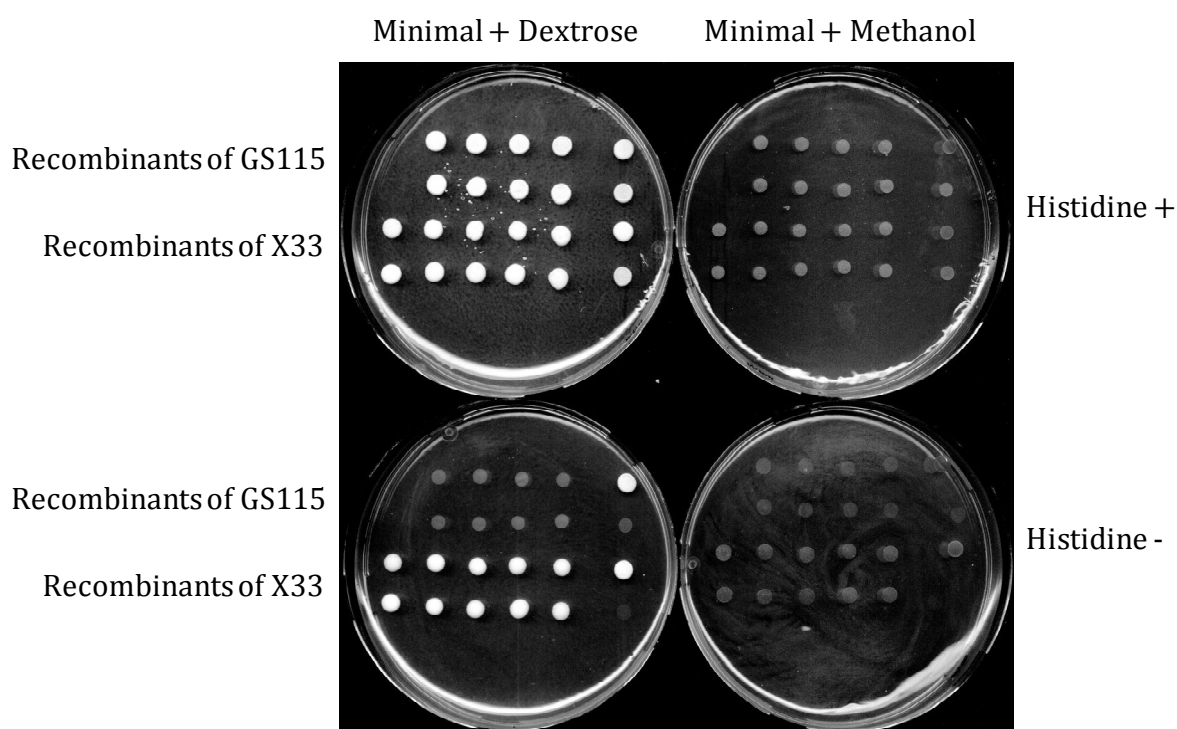
Figure A1.20: Mut phenotype of the P. pastoris recombinants.
Recombinants of strains GS115 and X33 were cultured in minimal dextrose media. The cultures were spotted onto the minimal agar plates supplemented with the indicated growth components. The plates were incubated for 5 days at 30 °C. Control strains for recombination events, GS115 Albumin (Mut^S) and GS115/pPICZ/*lac*Z (Mut$^+$) were spotted on the extremity on each plate.

The clones showed the recombination. Recombinants from KM71H were not tested, as they are constitutively Mut^S. Expression studies with the recombinants are in progress.

A1.4.7 Cloning, Expression and Purification of sCD40 from Insect Cell Expression System

Baculovirus based expression is a widely employed method for producing glyco-proteins including several candidate vaccines, from mammalian sources (Altman et al., 1999; Kost et al., 1999; Betting et al., 2009). This method involves generating recombinant baculovirus with the cloned genes of interest. The recombinant virus is employed to infect the insect cell lines of fall armyworm, *Spodoptera frugiperda* Sf9.

Since CD40 is also a glycoprotein, we have employed the baculovirus expression system for purification of sCD40 using Bac-to-Bac Baculovirus Expression System (Invitrogen Inc.). This involves cloning of the gene of interest into the donor plasmid. The resulting clone is used to transform *E. coli* DH10Bac, where the cloned gene is transposed from the donor plasmid to the baculovirus shuttle vector (bacmid), which later is transfected into the Sf9 cells to generate the recombinant baculovirus. The recombinant baculovirus is infected to Sf9 cells to produce the protein of interest. The steps followed in the expression and purification of sCD40, are explained below.

A1.4.7.1 Cloning of sCD40 into Donor Plasmid

Human *sCD40* fragment, with and without signal peptide, was cloned into NcoI and BamHI sites of the donor plasmid pFastBacHT A, generating pSCM1651 and pSCM1652, respectively. The clones were confirmed by digesting with restriction endonucleases (Figure A1.21).

A   **B**
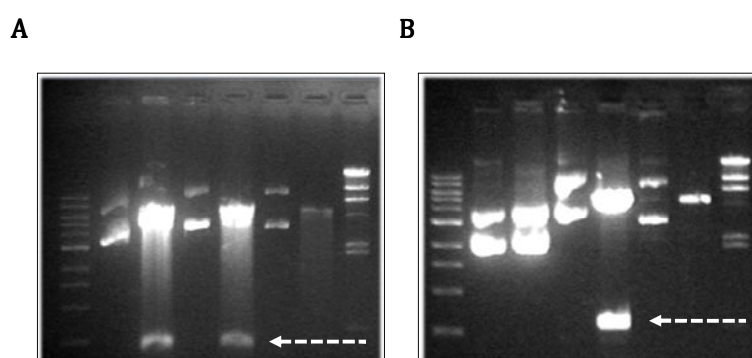
Figure A1.21: Cloning of sCD40 into Donor Plasmid. Plasmids pSCM1651 **(A)** and pSCM1652 **(B)** were digested with NcoI and BamHI. The digests were resolved on 1% agarose gel. Arrows indicate fragments corresponding to CD40.

A1.4.7.2 Generation of Recombinant Bacmid

Plasmids pSCM1651 and pSCM1652 were transformed into DH10Bac and the transformants were selected on LB plates supplemented with 50 μg/ml kanamycin, 7 μg/ml gentamicin, 10 μg/ml tetracycline, 100 μg/ml X-gal, and 40 μg/ml IPTG for blue white screening. Recombinant bacmids were purified from the white colonies and the event of transpositions was confirmed by PCR (Table A1.06), using the flanking primers pUC/M13 forward and pUC/M13 reverse (Invitrogen, Inc.) (Figure A1.22).

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94°C	300	1
Denaturation	94°C	45	
Annealing	55°C	45	30
Extension	72°C	180	
Final Extension	72°C	600	1

Table A1.06: PCR cycling conditions for determining recombinant Bacmid.

Figure A1.22: Confirmation of Transposition Event. PCR was performed on bacmids purified from white colonies harboring the recombinant bacmid hosting sCD40 with **(A)** or without **(B)*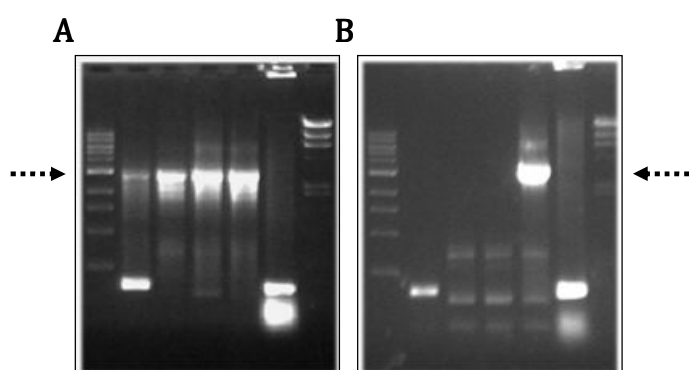* signal peptide. PCR products were resolved on 1% agarose gel. Arrows indicate the bands corresponding to the transposition event.

The recombinant bacmids generated from pSCM1651 and pSCM1652 were named pSCM1653 and pSCM1654, respectively. These bacmids were used for transfecting the Sf9 cells.

A1.4.7.3 Transfection of Recombinant Baculovirus into Sf9 Cells

Transfection of the Sf9 cells was followed according to the manufactures guidelines. About 9×10^5 cells from mid-log phase culture of Sf9 were seeded in 2 ml of Sf-900 II SFM medium containing 0.5x streptomycin and incubated at 27 °C allowing the cells adherence. Cells were washed with 2 ml of Sf-900 II ICM. 5 μl of bacmid DNA preparation was diluted into 100 μl of Sf-900 II ICM containing 6 μl of Lipofectin Reagent and incubated at room temperature for 30 min. The mixture containing bacmid DNA was overlayed onto the cell layer and incubated for 5 h at 27 °C. Following this, the transfection mixture was removed and the cells were supplemented with Sf-900 II ICM media containing appropriate antibiotics. The cells were further incubated at 27 °C for 72h. Virus was harvested and its titre was estimated. Virus titre was increased by repeated rounds of infection.

A1.4.7.4 Expression and Purification of hCD40 from Sf9 Cells

Sf9 cells after six rounds of infection with recombinant baculovirus were washed with PBS. Expression of sCD40 was detected by resolving fractions of cells from III round infection and VI round infection on 15% Tricine-PAGE (Figure A1.23). Increased levels of was observed for sCD40 after VI round of infection, when compared to the III round.

Having established that sCD40 is produced, attempts to purify the protein were carried out. Cells expressing the sCD40 were suspended in PBS and lysed by freez-thaw method. The soluble supernatant was loaded onto Ni-NTA spin column (Qiagen Inc.). The protein was eluted in the presence of 200 mM imidazole (Figure A1.23).
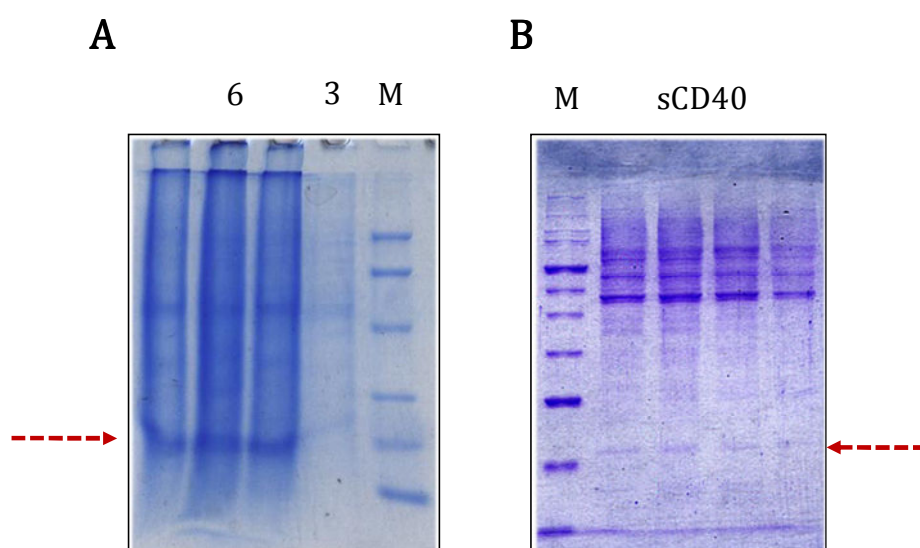
Figure A1.23: Expression and Purification of sCD40 from Sf9 Cells. **A.** ORF encoding *sCD40* was expressed in the Sf9 cells infected with recombinant baculo virus of different passages. Cells were recovered in mid-log phase and the cell lysates were resolved on 15% tricine gel. The lanes correspond to: 6, VI round of viral passage; 3, III round of viral passage; M, PAGE-Ruler protein marker (MBI Fermentas Inc.). **B.** Human sCD40 was purified employing Ni-NTA chromatography. Protein was eluted at 200 mM Imidazole. M denotes the broad range protein marker (NEB) and the arrows indicate the band corresponding to sCD40.

A1.5 Cloning, Expression and Purifcation of Mtb HspR

HspR gene of Mtb is in operonic arrangement with Hsp70 and is demonstrated to interact with Hsp70 thereby involved in repressing the Hsp70 operon (Cole et al., 1998; Stewart et al., 2001). We wished to study the molecular features of the interaction between Mtb Hsp70 and HspR. To this end, we have cloned and expressed the ORF encoding Mtb HspR and purified the protein.

A1.5.1 Cloning the ORF encoding Mtb HspR

ORF encoding Mtb HspR was amplified using primers SCM1648F/SCM1648R and SCM1649F/SCM1649R, respectively, to incorporate (His)$_6$ tag at the amino and carboxy termini in the resulting protein. Mtb genomic library clone BAC-Rv285 (A8) was used as template. The PCR cycling conditions are given in table A1.07.

Cylce Step	Temperature in °C	Time in Seconds	No. of Cycles
Initial Denaturation	94°C	300	1
Denaturation	94°C	60	
Annealing	54°C	120	30
Extension	72°C	90	
Final Extension	72°C	600	1

Table A1.07: PCR cycling conditions for amplifying ORF encoding Mtb *hspR*.

The PCR products were cloned into NcoI and XhoI sites on the MCS2 of pETDuet-1 to generate pSCM1654 and pSCM1655, respectively. The clone was confirmed by restriction digestion and sequencing (Figure A1.24).

A

B

N terminal
Histidine tag

C terminal
Histidine tag
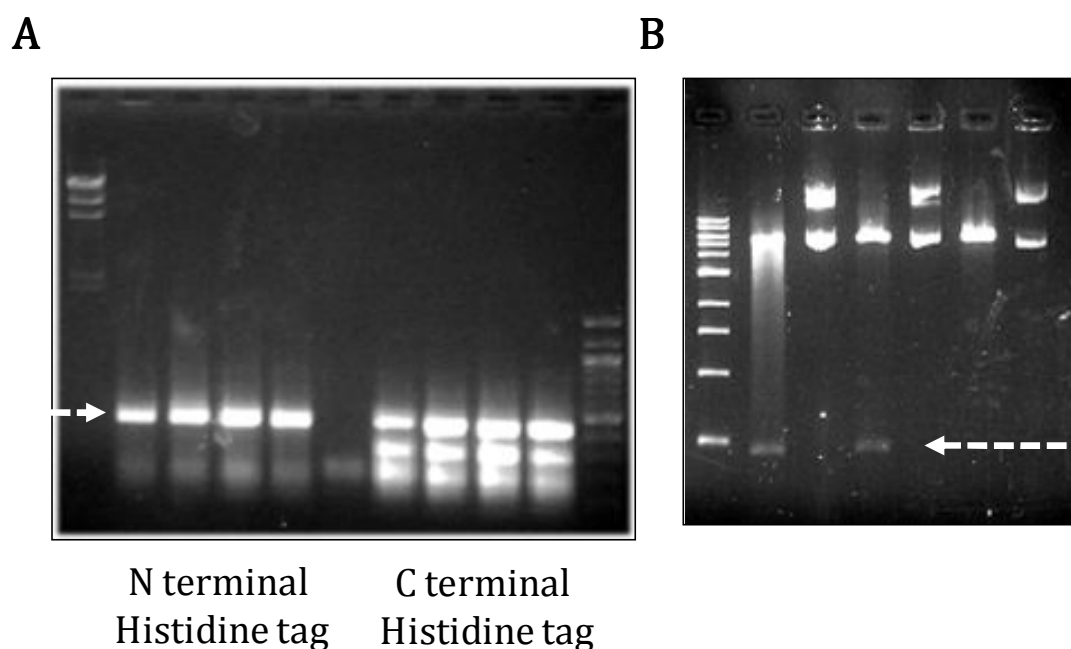
Figure A1.24: Cloning of Mtb hspR. **A.** Mtb *hspR* was amplified using different primer sets to incorporate the (His)$_6$ tags at amino and carboxy terminus of the protein. The PCR products were resolved on 1% agarose gels. **B.** Plasmids pSCM1654 and pSCM1655 were digested with restriction endonucleases NcoI and XhoI. Arrows indicate the bands corresponding to *hspR.*

A1.5.2 Expression and Purification of HspR

Expression of the ORF encoding HspR was analyzed in BL21 (DE3) cultured in three media LB, TB, and 2XYT. The protein was found to be insoluble in the presence of 10 mM Sodium Lauroyl Sarcosine. HspR, therefore, was purified by Ni-NTA chromatography under denaturing conditions with 6M urea in the buffers (Figure A1.25).
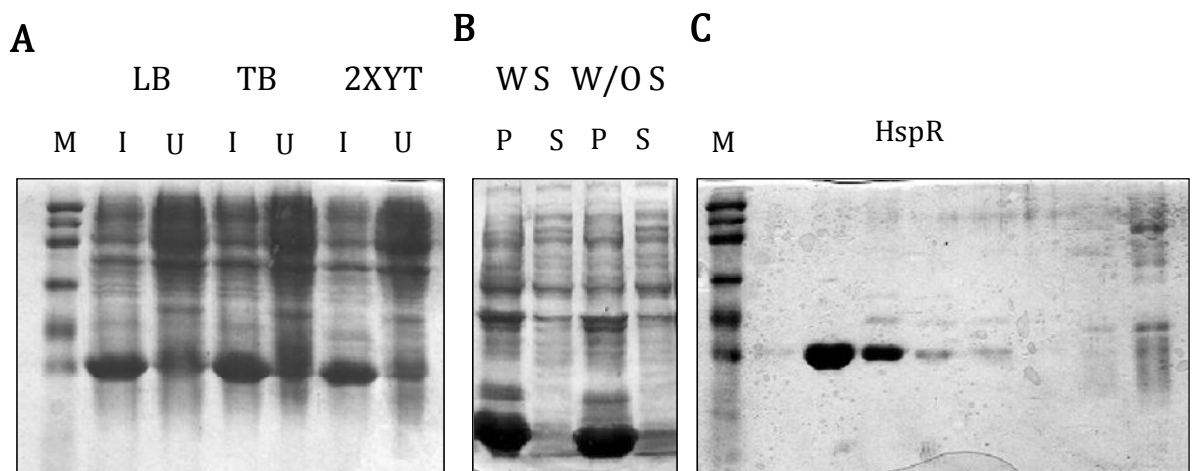
Figure A1.25: Production of Mtb Hsp70. **A.** BL21 (DE3) harboring plasmid pSCM1654 were cultures in the indicated media. The cultures were induced in mid-log phase with 1 mM IPTG. Induced (I) and uninduced (U) cell lysates were resolved on 15% SDS-PAGE. **B.** BL21 (DE3) expressing *hspR* was lysed by ultra-sonication in the presence (W S) and absence (W/O S) of 10 mM Sodium Lauryl Sarcosene. The soluble (S) and insoluble (P) fractions in each case were resolved on 15% SDS-PAGE. **C.** HspR was purified using Ni-NTA chromatography. Protein was eluted at 150 mM Imidazole. M denotes broad range protein marker.

A1.6 Conclusions

Several attempts were made towards the purification of CD40 and these are ready to be taken forward. The complex made between Hsp70 and CD40 shows promising crystals and likewise the individual Hsp70 protein. Elucidating the structure would illuminate differences that govern the interactions of Hsp70 and the native ligand CD40L.

A1.7 References

- Altmann, F., Staudacher, E., Wilson, I. B. and Marz, L. (1999) Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconjugate J.* **16**, 109–123

- Banchereau, J., Bazan, F., Blanchard, D., Brière, F., Galizzi, J. P., van Kooten, C., Liu, Y. J., Rousset, F. and Saeland, S. (1994) The CD40 and its antigen. *Annu. Rev. Immunol.* **12**, 881-922

- Betting, D. J., Mu, X. Y., Kafi, K., McDonnel, D., Rosas, F., Gold, D. P. and Timmerman, J. M. (2009) Enhanced immune stimulation by a therapeutic lymphoma tumor antigen vaccine produced in insect cells involves mannose receptor targeting to antigen presenting cells. *Vaccine* **27**, 250-259

- Bucca, G., Ferina, G., Puglia, A. M. and Smith, C. P. (1995) The *dnaK* operon of *Streptomyces coelicolor* encodes a novel heat-shock protein which binds to the promoter region of the operon. *Mol. Microbiol.* **17,** 663–674

- Bukau, B. and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351–366

- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. III., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544

- Cregg, J. M., Vedvick, T. S. and Raschke, W. C. (1993) Recent Advances in the Expression of Foreign Genes in *Pichia pastoris. Bio/Technology* **11**, 905-910

- Das Gupta, T., Bandyopadhyay, B. and Das Gupta, S. K. (2008) Modulation of DNA-binding activity of *Mycobacterium tuberculosis* HspR by chaperones. *Microbiology* **154**, 484-490

- Foy, T. M., Aruffom A., Bajorath, J., Buhlmann, J. E. and Noelle, R. J. (1996) Immune regulation by CD40 and its ligand GP39. *Annu. Rev. Immunol.* **14**, 591-617

- Grandvalet, C., Servant, P., Mazodier, P. (1997) Disruption of *hspR*, the repressor gene of the *dnaK* operon in *Streptomyces albus* G. *Mol. Microbiol.* **23**, 77–84

- Kost, T. A., Condreay, J. P. (1999) Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr. Opin. Biotechnol.* **10**, 428–433

- Lazarevic, V., Myers, A. J., Scanga, C. A. and Flynn, J. L. (2003) CD40, but not CD40L, is required for the optimal priming of T cells and control of aerosol *M. tuberculosis* infection. *Immunity* **19**, 823 - 835

- MacAry, P. A., Javid, B., Floto, R. A., Smith, K. G., Oehlmann, W., Singh, M. and Lehner, P. J. (2004) HSP70 peptide binding mutants separate antigen delivery from dendritic cell stimulation. *Immunity* **20**, 95-106

- Stewart, G. R., Snewinm V. A., Walzl, G., Hussell, T., Tormay, P., O'Gaora, P., Goyal, M., Betts, J., Brown, I. N. and Young, D. B. (2001) Overexpression of heat-shock proteins reduces survival of *Mycobacterium tuberculosis* in the chronic phase of infection. *Nat. Med.* **7**, 732-737

- Wang, Y., Kelly. C. G., Karttunen, J. T., Whittall, T., Lehner, P. J., Duncan, L., MacAry, P., Younson, J. S., Singh, M., Oehlmann, W., Cheng, G., Bergmeier, L. and Lehner, T. (2001) CD40 is a cellular receptor mediating mycobacterial heat shock protein 70 stimulation of CC-chemokines. *Immunity* **15**, 971-983

- Young, L. S., Eliopoulos, A. G., Gallagher, N. J. and Dawson, C. W. (1998) CD40 and epithelial cells: across the great divide. *Immunol. Today* **19**, 502-506.

APPENDIX II

Oligonucleotide Primers and Plasmid Vectors

A2.1 Introduction

Genetic and biochemical studies on Hsp60 and Hsp70 homologues from *Mycobacterium tuberculosis* H37Rv described in the previous chapters were assisted by the molecular biology techniques in generating vectors for the complementation studies and for the purification of proteins. Generation and confirmation of the said clones involved several oligonucleotide primers and appropriate vectors. The sequences of oligonucleotide primers used for cloning and sequencing are listed in the following sections, A2.2 and A2.3, respectively. Likewise, Plasmid vectors used for cloning and the plasmids vectors generated in this study are listed in the following sections, A2.4 and A2.5, respectively.

A2.2 Oligonucleotide Primers Used for Cloning in this Study

Name	Sequence (in 5' - 3' direction)
SCM1601F	5' GACAATCCATGGCGAAGGTGAACATCAAGCC 3'
SCM1601R	5' GCATCACCCGGGCTACTTGGAAACGACGGCCAGC 3'
SCM1602F	5' GGCAACCCCGGGAGGAAATAAACATGAGCAATACGACG 3'
SCM1602R	5' GAGAAGTCTAGATCAGTGCGCGTCCCGTGG 3'
SCM1603F	5' GCCTAGCCCGGGATGAATATTCGTCCATTGCATGATCGCG 3'
SCM1603R	5' GAGTTCAAGCTTTACATCATGCCGCCCATGCCACC 3'
SCM1604F	5' GCTAGCAGGAGGAATTCCATATGGTACCCGGGGATCC 3'
SCM1604R	5' GGATCCCCGGGTACCATATGGAATTCCTCCTGCTAGC 3'
SCM1605F	5' CGTCGTCATAGTAGCAAGCTGATCGAATACGACG 3'
SCM1605R	5' ACGTCTCCCGGGTCAGTGCGCGTGTCCGTGG 3'
SCM1606F	5' AAGGATCCATGGCAGCTAAAGACGTAAAATTCGGTAACG 3'
SCM1607F	5'ATGAGCAAGCTGATCGAATACGACGA 3'
SCM1607R	5'ACGTCTCCCGGGTCAGTGCGCGTGTCCGTGG 3'
SCM1608F	5'ATGGCCAAGACAATTGCGTACGACG 3'
SCM1608R	5'ACTTCCTCTAGATCAGAAATCCATGCCACCCATGTCGC 3'
SCM1609R	5' ACTGCTGTCGACTTACATCATGCCGCCCATGCCACC 3'
SCM16J1	5'CGGACACCGGCGTGGCCGCTTTCAGTTCTTCAACTGCAGC 3'
SCM16J2	5'CGCGGTCGAGGAGGGCGAAGGCGTGGTTACTGGTGG 3'
SCM16J3	5'TAGAGCTGAGAGCATGGTACGGACAGCGATGCCAGCAGTGCCT 3'
SCM16J4	5'GCGACCCGTGCTGCTGTAGAAATCGTCCCTGGTGGGGGAG 3'
SCM1612F	5' CGCACGACACTGAACTCTAGAATTTAAGGAATAAA 3'
SCM1612R	5' TTTATTCCTTAAATTCTAGAGTTCAGTGTCGTGCG 3'
SCM1613F	5' CGACACTGAACATACGAAGCTTTAAGGAATAAAGATA 3'
SCM1613R	5' TATCTTTATTCCTTAAAGCTTCGTATGTTCAGTGTCG 3'
SCM1614F	5' CTTGAACCATGGCAGCTAAAGACGTAAAATTCGGTAACG 3'
SCM1614R	5' CGGACTAAGCTTTCAGTTCTTCAACTGCAGCGGTAAC 3'
SCM1615F	5' CCATGAAAAGCTTCCACTCCCGTGTCCGGCAAGACC 3'
SCM1615R	5' ATCGTCCATATGCCCTCCTCGACCGCGGCC 3'
SCM1616F	5' GATCGTCATATGGAAGGCGTGGTTGCTGGTGGTGGT 3'
SCM1616R	5' TTAAGATCTTTAGTGGTGGTGGTGATGGTGCATCATGCCGCCCATTCCACCC 3'
SCM1640F	5' GCTATACATATGGCTCGTGCGGTC 3'
SCM1640R	5' AATAAGGATCCTCACTTGGCCTCC 3'
SCM1641F	5'AACCAGATATCATGCACCACCACCACCACCACGTTCGTCTGCCTCTGCAGTGCGTCTCT 3'
SCM1641R	5' CGCACACTCGAGTCATCTCAGCCGATCCTGGGGACCACAG 3'

Name	Sequence (in 5' – 3' sequence)
SCM1642F	5' AGCAAGATATCATGGTTCGTCTGCCTCTGCAGTGCGT 3'
SCM1642R	5'CGCACGCTCGAGTCAGTGGTGGTGGTGGTGGTGTCTCAGCCGATCCTGGGGACCACAG 3'
SCM1643F	5' GCGCGACCATATGCATCATCATCATCATCATCCTTGCGGTGAAAGCGAATTCCTAGACAC 3'
SCM1644F	5' GAGCGTCCATATGCCTTGCGGTGAAAGCGAATTCCTAGACAC 3'
SCM1645F	5' AGGTAGGCCATGGCTCGTGCGGTCGGGATCGAC 3'
SCM1645R	5' ATACATGGATCCTCACTTCGCCTCCCGGCCGTCG 3'
SCM1646F	5'CGACTCCATGGGTCATCATCATCATCATCATCCTTGCGGTGAAAGCGAATTCCTAGACAC 3'
SCM1647F	5' CGACTCCATGGGTCCTTGCGGTGAAAGCGAATTCCTAGACAC 3'
SCM1648F	5' AACCAGATATCATGCACCACCACCACCACCACGGCTCCGCGGCGAAGAACCCAAAGGACGGCG 3'
SCM1648R	5' ATCATCCTCGAGTCACCGGCGCGGTTTCCAGAC 3'
SCM1649F	5' ATCATCCTCGAGTCACCGGCGCGGTTTCCAGAC 3'
SCM1649R	5'CAGAACCATATGCACCACCACCACCACCACGGCTCCGCGGCGAAGAACCCAAAGGACGGCG 3'

A2.3 Oligonucleotide Primers Used for Sequencing

Name	Sequence (in 5' - 3' direction)	Source/Reference
PBADFOR	5' CTGTTTCTCCATACCCGTT 3'	Guzman et al., 1995
PBADREV	5' CTCATCCGCCAAAACAG 3'	Guzman et al., 1995
T7 Promoter Primer	5' TAATACGACTCACTATA 3'	Novagen Inc., USA
T7 Terminator Primer	5' GCTAGTTATTGCTCAGCGG 3'	Novagen Inc., USA
pET Upstream Primer	5' ATGCGTCCGGCGTAGA 3'	Novagen Inc., USA
DuetDOWN-1 Primer	5' GATTATGCGGCCGTGTACAA 3'	Novagen Inc., USA
DuetUP2 Primer	5' TAATACGACTCACTATAGGG 3'	Novagen Inc., USA
pUC/M13 Forward Primer	5' CCCAGTCACGACGTTGTAAAACG 3'	Invitrogen Inc, USA
pUC/M13 Reverse Primer	5' AGCGGATAACAATTTCACACAGG 3'	Invitrogen Inc, USA
5' *AOX1* sequencing primer (5' *Pichia* primer)	5' GACTGGTTCCAATTGACAAGC 3'	Invitrogen Inc, USA
3' *AOX1* sequencing primer (3' *Pichia* primer)	5' GCAAATGGCATTCTGACATCC 3'	Invitrogen Inc, USA
α-Factor sequencing primer	5' TACTATTGCCAGCATTGCTGC 3'	Invitrogen Inc, USA

A2.4 Plasmid Vectors Used in this Study

Name	Description	Reference/Source
pBAD18	Arabinose inducible expression vector for expression in *E. coli*	Guzman et al., 1995
pBAD24	Arabinose inducible expression vector for expression in *E. coli*	Guzman et al., 1995
pET-20b(+)	P_{T7} based expression vector for expression in *E. coli*	Novagen Inc., USA
pET-23a(+)	P_{T7} based expression vector for expression in *E. coli*	Novagen Inc., USA
pET-23d(+)	P_{T7} based expression vector for expression in *E. coli*	Novagen Inc., USA
pET-28a(+)	P_{T7} based expression vector for expression in *E. coli*	Novagen Inc., USA
pETDuet-1	P_{T7} based expression vector, dual MCS for expressing two genes in *E. coli*	Novagen Inc., USA
pTrc99A	P_{tac} based expression vector for expression in *E. coli*	Amann et al., 1988
pRSET B	P_{T7} based expression vector for expression in *E. coli*	Invitrogen Inc., USA
pFastBacHT A	Donor plasmid for baculovirus expression	Invitrogen Inc., USA
pPICZ A	AOX1 promoter based expression vector for intracellular protein production in *Pichia pastoris*	Invitrogen Inc., USA
pPICZα A	AOX1 promoter based expression vector for secreted protein production in *Pichia pastoris*	Invitrogen Inc., USA
bMON14272	Bacmid for the production of recombinant baculovirus	Invitrogen Inc., USA

A2.5 Plasmid Vectors Generated in this Study

Name	Description
pBAD25	NcoI site of pBAD24 converted to NdeI site by site directed mutagenesis
pSCM1000	Mtb *groEL2* cloned in NcoI and HindIII sites of pET-28a(+)
pSCM1600	Mtb *groES* cloned in NcoI and SmaI sites of pBAD24
pSCM1601	*E. coli groESL* operon cloned in NcoI and HindIII sites of pBAD24
pSCM1602	*Mtb groEL1* cloned in SmaI and XbaI sites of pSCM1600
pSCM1603	*Mtb groEL2* cloned in XbaI and HindIII sites of pSCM1600
pSCM1604	*Mtb groEL1* cloned in NdeI and XbaI sites of pBAD25
pSCM1605	*Mtb groEL2* cloned in XbaI and HindIII sites of pBAD18
pSCM1608	*E. coli groEL* cloned in NcoI and HindIII sites of pBAD24
pSCM1609	*groELMEF* cloned in NcoI and SalI sites of pBAD24
pSCM1610	*groELMER* cloned in NdeI and HindIII sites of pBAD25
pSCM1611	*E. coli groESL* operon cloned in NcoI and HindIII sites of pTrc99A
pSCM1612	*groELMEF* cloned in NcoI and SalI sites of pTrc99A
pSCM1613	*E. coli groEL* cloned in NcoI and HindIII sites of pTrc99A
pSCM1614	XbaI site incorporated between *groES* and *groEL* genes in pSCM1611

Name	Description
pSCM1615	HindIII site incorporated between *groES* and *groEL* genes in pSCM1611
pSCM1616	*groELMEF* cloned in XbaI and SalI of pSCM1614
pSCM1617	*E. coli groES* clone in pTrc99A, generated from pSCM1615
pSCM1618	Mtb *groES* cloned in NcoI and XbaI sites of pTrc99A
pSCM1619	*groELMEF* cloned into XbaI and SalI sites of pSCM1618
pSCM1620	*E. coli groEL* cloned in XbaI and HindIII sites on pSCM1618
pSCM1621	*groELDS* cloned in NcoI and BglII sites of pETDuet-1
pSCM1622	*groELSp22* cloned in NcoI and HindIII sites of pBAD24
pSCM1623	*groELDS* from pETDuet-1 cloned into NcoI and SmaI sites of pBAD24
pSCM1624	*groELSp24* cloned in NcoI and HindIII sites of pBAD24
pSCM1625	*groELSp25* cloned in NcoI and HindIII sites of pBAD24
pSCM1626	*groELSp26* cloned in NcoI and HindIII sites of pBAD24
pSCM1627	*groELSp27* cloned in NcoI and HindIII sites of pBAD24
pSCM1628	Mtb *groEL1* cloned in NdeI and BamHI sites of pRSET B
pSCM1629	Mtb *groEL2* from pSCM1000 cloned in XbaI and HindIII sites of pRSET B
pSCM1630	*groELMER* cloned in NdeI and EcoRV sites of pET-20b(+)
pSCM1631	*groELSp24* cloned in NcoI and HindIII sites of pTrc99A
pSCM1632	*groELSp32* cloned in NcoI and HindIII sites of pBAD24
pSCM1632	ORF encoding GLLG18 cloned in NcoI and Hind III sites of pBAD24
pSCM1633	ORF encoding GLLG25 cloned in NcoI and Hind III sites of pBAD24
pSCM1635	*groELSp35* cloned in NcoI and HindIII sites of pBAD24
pSCM1636	*groELSp36* cloned in NcoI and HindIII sites of pBAD24
pSCM1637	*groELSp37* cloned in NcoI and HindIII sites of pBAD24
pSCM1638	Mtb *dnaK* cloned in NdeI and BamHI sites of pET-23a(+)
pSCM1639	Human sCD40 cDNA with signal peptide and N-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1
pSCM1640	ORF of Human sCD40 cDNA with signal peptide and N-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1
pSCM1641	ORF of Human sCD40 cDNA Δ signal peptide and N-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1
pSCM1642	ORF of Human sCD40 cDNA Δ signal peptide and N-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1
pSCM1643	Mtb *dnaK* cloned in NcoI and BamHI sites of pSCM1639
pSCM1644	Mtb *dnaK* cloned in NcoI and BamHI sites of pSCM1640
pSCM1645	Mtb *dnaK* cloned in NcoI and BamHI sites of pSCM1641
pSCM1646	Mtb *dnaK* cloned in NcoI and BamHI sites of pSCM1642

Name	Description
pSCM1647	Mtb *dnaK PBD* cloned into NcoI and BamHI sites of pSCM1641
pSCM1648	Mtb *dnaK PBD* cloned into NcoI and BamHI sites of pSCM1642
pSCM1649	ORF of Human sCD40 cloned in NcoI and XhoI sites on pET-23d(+)
pSCM1650	Human sCD40 cDNA was cloned in pPICZα A
pSCM1651	Human sCD40 cDNA was cloned into pFastBac 1
pSCM1652	Human fCD40 cDNA was cloned into pFastBac 1
pSCM1653	Recombinant bacmid derived from pSCM1651
pSCM1654	Recombinant bacmid derived from pSCM1652
pSCM1655	Mtb *hspR* with N-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1
pSCM1656	Mtb *hspR* with C-terminal six-Histidine tag cloned in NdeI and XhoI sites of pETDuet-1

A2.5 References

- Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**, 4121-4130
- Amann, E., Ochs, B. and Abel, K. J. (1988) Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301-315