ANALYSIS OF THE MULTIPLE CHAPERONINS OF *MYCOBACTERIUM SMEGMATIS*

By

TARA RAO

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
The University of Birmingham
November 2009

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 3: Analysis of chaperonin function in *M. smegmatis*

Chapter 4: Biofilms of *M. smegmatis*

Chapter 5: Characterization of *M. smegmatis* chaperonins in *E. coli*

Chapter 6: Purification and analysis of oligomeric states of Cpn60.1 and Cpn60.2

List of tables

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 4: Biofilms of *M. smegmatis*

Chapter 5: Characterization of *M. smegmatis* chaperonins in *E. coli*

Chapter 6: Purification and analysis of oligomeric states of Cpn60.1 and Cpn60.2

ABSTRACT

Proteins form the building blocks of every biological system. In order to be biologically active and functional, proteins need to be in their native or stable states. However, during the folding process, some proteins tend to misfold and aggregate and are non-functional. Molecular chaperones are a group of specialised proteins which play a significant role in preventing or reversing aggregation as part of this protein folding process. Members of the 60kDa class of molecular chaperones are more commonly known as the chaperonins. Although most bacteria contain a single gene for the essential oligomeric chaperonin Cpn60, many contain two or more *cpn60* genes. The non-pathogenic *Mycobacterium smegmatis* has three *cpn60* homologues, while the pathogenic *Mycobacterium tuberculosis* has two. This study is a functional characterisation of the chaperonins of *M. smegmatis*.

In the first part of this study, the *M. smegmatis* chaperonins were characterised based on their sequences and their expression under stress conditions. Expression of *cpn60.1*, *cpn60.2* and *cpn10,* but not *cpn60.3,* was found to be induced under stress conditions, particularly heat shock. A combination of RACE and promoter probe studies were used to define the transcription of the *cpn10-cpn60.1* operon, and it was concluded that transcription is from a single promoter, with a subsequent post-transcriptional cleavage of the mRNA between the two genes. Cpn60.1 in *M. smegmatis* has been reported to be required for biofilm maturation. Using this assay, we intended to test various Cpn60 homologues for their ability to function in *M. smegmatis*. However, the assay could not be reproduced due to unknown difficulties. Extensive analysis was not able to determine the reasons for the differences between the two labs (Prof. Hatfull's lab, Department of

Biological Sciences, University of Pittsburgh, USA and Lund lab, University of Birmingham, UK). Preliminary results obtained from experiments conducted in the Pittsburgh lab revealed that only the *M. tuberculosis* Cpn60.1 can fully complement for loss of the *M. smegmatis* Cpn60.1. While *E. coli* GroEL and Cpn60.3 appear to only partially complement, Cpn60.2 shows no complementing ability.

The next part of the study involved the cloning and expression of the three *cpn60* and the co-chaperonin *cpn10* to determine if they could functionally replace the *E. coli groES* and *groEL* genes. The three *cpn60* genes were successfully cloned and expressed in *E. coli* together with the *E. coli* co-chaperonin *groES* or the *M. smegmatis cpn10*. Cpn60.2 expresses well and complements for loss of the *E. coli* Cpn60 homologue GroEL even at higher temperatures (42°C). Neither Cpn60.1 nor Cpn60.3 complemented for loss of GroEL. As it has been claimed that the Cpn60 proteins from *M. tuberculosis* do not have the typical oligomeric structure of chaperonins, the *M. smegmatis* Cpn60.2 protein was purified to homogeneity and its oligomeric status analysed using native gels and analytical ultracentrifugation. The results showed that Cpn60.2 did not oligomerise under normal conditions, however in the presence of nucleotide and high concentrations of salt, the formation of large oligomers was observed.

This study has confirmed the role of Cpn60.2 as the housekeeping chaperonin in *M. smegmatis*, and also highlighted the possible substrate conservation between proteins folded by the chaperonins of *M. smegmatis* and *E. coli*. The formation of large oligomers of *M. smegmatis* Cpn60.2 has been reported for the first time in this study.

Although the role of Cpn60.1 in biofilm formation could not be fully investigated, this study provides a platform for further experiments. No role of Cpn60.3 was conclusively established from this study.

Acknowledgements

First and foremost, I would like to thank the Darwin Trust of Edinburgh for funding my PhD and giving me the opportunity to fulfil my dream.

I could not have completed this PhD without the constant support, encouragement and guidance from my supervisor, Dr. Pete Lund. He has been an inspiration to me and I am very grateful for the opportunity to have worked in his lab.

A big thank you to Andy, from whom I have learned so much, not only about science but various other subjects, far removed from science, as well - I did enjoy our "debates"! Thanks to MingQi for his help with the many issues encountered during my PhD. I thank Dr. Apoorva Bhatt and members of his lab for their help and useful discussions on Smegmatis related issues. I would also like to thank Dr. Anil Ojha for his help with the rather stubborn biofilms. Thanks to the members of the May lab for allowing use of their microscopes.

For loads of fun times and for sharing the highs and lows of research, I would like to thank past and present members of the Lund lab, especially Preethy, Neil, Tabish, Matt, Elsa, Bevan, Misbah and Mala. Thanks also to the members of the Henderson lab for making the past few years in lab T101 very memorable. And, most importantly, I would like to thank my family for their support throughout these three wonderful years.

List of abbreviations

ADC	Albumin dextrose catalase
ADP	Adenosine diphosphate
Amp	Ampicillin
ATP	Adenosine triphosphate
APS	Ammonium persulphate
AUC	Analytical ultracentrifugation
CCT	Chaperonin-containing TCP-1
CIRCE	Controlling inverted repeats of chaperone expression
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediamine tetra-acetic acid
GFP	Green fluorescent protein
HAIR	HspR associated inverted repeat
HrcA	Heat regulation at CIRCE
Hsc70	Heat shock cognate protein
Hsp	Heat shock protein
HspR	Heat shock protein regulator
Hyg	Hygromycin
IPTG	Isopropyl-□-D-thiogalactoside
Kan	Kanamycin
kDa	Kilodalton
LB	Luria-Bertani

M	Molar
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
TAE	Tris-acetate EDTA
TCP-1	T-complex polypeptide-1
TEMED	N.N.N', N'-tetramethylethylenediamine
Tris	2-amino-1-(hydroxymethyl)-1, 3-propanediol
X-GAL	5-bromo-4-chloro-3-indolyl β-D-galactoside

1.1 Introduction

The main focus of this study is the multiple chaperonins of *Mycobacterium smegmatis*. The introduction will begin with a brief background on the Actinobacteria, with emphasis on the Mycobacteria in particular. This will be followed by a discussion of protein folding and chaperonin function. Finally, both these aspects will be discussed when introducing the chaperonins of *M. smegmatis*.

1.1.1 Actinobacteria

The phylum Actinobacteria, comprising 39 families and 130 genera represent one of the largest taxonomic groups within the bacteria (Boone *et al.*, 2001). Since the Actinobacteria appear to have diverged from other bacteria a long while ago, phylogenetically it is not possible to identify the closest related bacterial group (Embley and Stackebrandt, 1994). The presence of certain conserved insertions and deletions (indels) has recently been reported in three widely distributed proteins and is said to be unique to the Actinobacteria (Gao and Gupta, 2005). This distinctive characteristic (indels) which distinguishes all members of the Actinobacteria from other groups of bacteria was probably introduced into this group by a common ancestor and further characterisation of these unique sequences could provide insights into the novel biochemical and physiological characteristics of the Actinobacteria.

This group is characterised as Gram-positive bacteria with a high GC content ranging from 50% in some *Corynebacteria* to more than 70% in *Streptomyces* and *Frankia*. The pathogen *Tropheryma whipplei* is an exception as its genome has a GC content of less

than 50% (Ventura *et al*., 2007). Morphologically, the Actinobacteria are varied with some species being coccoid (*Micrococcus*), rod-coccoid (*Arthrobacter*), fragmenting hyphal forms (*Nocardia*) or permanently and highly differentiated branched mycelium (*Streptomyces*) (Atlas, 1997). Some members of this group possess fungal characteristics, such as hyphal branching and formation of spores for dissemination, while some members possess bacterial type flagella, exhibit sensitivity to phages and antibacterial antibiotics, and have cell walls containing muropeptides (Lechevalier and Lechevalier, 1967). The Actinobacteria produce a number of extracellular enzymes and secondary metabolites. The Streptomycetes in particular are known for their antibiotic production which is the main reason they are exploited by the pharmaceutical industry (Lechevalier and Lechevalier, 1967). The members of the Actinobacteria have very different lifestyles ranging from human pathogens (*Mycobacteria* spp, *Nocardia*, *Tropheryma* and *Corynebacteria*), to soil inhabitants (*Streptomyces*), to plant commensals (*Leifsonia*), to nitrogen fixing symbionts (*Frankia*) to even gastrointestinal tract inhabitants (*Bifidobacterium* spp) (Ventura *et al*., 2007).

The genera Corynebacteria, Mycobacteria and Nocardia form a well defined subgroup (CMN) of *Actinobacteria* (Embley and Stackebrandt, 1994). These three subgroups share an unusual cell envelope which is characterised by the presence of a waxy coating rich in mycolic acids and responsible for the acid-fast nature of members of these groups (see table 1.1). The *Nocardia* spp. are filamentous soil saprophytes, however, some members are also pathogenic agents causing Nocardiosis in human and animal lungs, and affecting the central nervous system, brain and skin (Brown *et al*., 1999). The only fully sequenced Nocardial genome, *Nocardia farcinica* IFM10152, includes many

candidate genes for virulence and multidrug resistance as well as secondary metabolism (Ishikawa *et al*., 2004). The group *Corynebacterium* consists of nearly 70 species, including many new species recently isolated from human clinical samples, wild animals or soil. The well studied *Corynebacterium glutamicum* is mainly found in soil contaminated with bird faeces, sewage and manure and was sequenced in 2003 (Kalinowski *et al*., 2003). *C. glutamicum* is widely used in the industrial production of amino-acids especially L-glutamic acid and L lysine because of their importance in human and animal nutrition respectively (Leuchtenberger *et al*., 2005). The genome of the human pathogen *Corynebacterium diphtheriae* which causes diphtheria was sequenced in 2003 (Cerdeno-Tarraga *et al*., 2003). The disease is caused by the production of an endotoxin in the pharynx, which inhibits protein synthesis thereby killing susceptible host cells (Holmes, 2000). The examples listed are suggestive of a link between sequence specificity and virulence in members of the CMN sub-group of bacteria.

Features	***Corynebacterium***	***Mycobacterium***	***Nocardia***
GC content	51% - 60%	62% - 70%	64% - 72%
Cell shape	Pleomorphic rods	Mainly rods, sometimes branching	Mycelium fragmenting into rods or cocci
Mycolic acids	22 - 66 carbon atoms	60 - 90 carbon atoms	46 - 60 carbon atoms
Time taken for visible colonies to appear (Growth rate)	1 - 2 days	2 - 40 days	1 - 5 days

Table 1. 1: Comparison of some of the features between the CMN sub-group
(table adapted from Cook *et al*., 2009).

1.1.2 Mycobacteria

The mycobacteria contain more than 100 different species which have been identified based on a comparison of their 16S rRNA sequences (Tortoli, 2006). While this genus is well known for its successful human pathogens, *M. tuberculosis*, *M. leprae* and *M. ulcerans* which cause tuberculosis, leprosy and Buruli ulcers respectively, the majority of its members are non-pathogenic environmental bacteria. Tuberculosis was declared a global emergency by the World Health Organisation (WHO) in 1993. However, today it is still one of the leading causes of human death due to bacterial infection with as many as 2 million deaths reported each year (Glaziou *et al*., 2009). The bacteria latently infects nearly a third of the World's population. A remarkably feature of *M. tuberculosis* is its ability to remain in a dormant state until reactivation and manifestation of active tuberculosis. The treatment of the disease has been a major concern due to a number of factors such as cost of treatment, number of people infected and rising antibiotic resistance. The complex nature of the cell wall makes treatment difficult, contributing to the persistence and survival of the bacteria.

1.1.2.1 Mycobacterial Cell Wall

One of the areas that has received a great deal of attention is the mycobacterial cell wall. The reason for this interest is due to the complexity of the cell wall structure which is responsible for many distinctive features, such as its acid-fast nature, distinct colony morphology, its resistance to chemicals and impermeability to therapeutic agents. The mycobacterial cell wall is composed of an inner and outer layer surrounding the plasma membrane (figure 1.1) (Brennan, 2003). The inner layer comprising peptidoglycan (PG) covalently attached to arabinogalactan (AG) which in turn is attached to mycolic acids,

is known as the essential cell wall core or the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex. The outer layer is composed of free lipids, some with long fatty acids complementing the short α chains and short fatty acids complementing the long α' chains. The lipid-linked cell wall sugars consist of phosphatidylinositol mannosides (PIMs), the phthiocerol containing lipids, lipomannan (LM) and lipoarabinomannan (LAM). During cell wall disruption to extract the various components, due to the action of detergents, free lipids, proteins, LAMs and PIMs are solubilised, while the mAGP complex remains insoluble (Brennan, 2003). This in essence indicates that the soluble outer layer plays a role in signalling and acts as effector molecules in the virulence process. Hence, the essential insoluble core, which is responsible for cell viability could be used as possible drug targets (Brennan, 2003). The mAGP complex will be explained in some detail below.

Peptidoglycan: This rigid murein layer is responsible for maintaining the shape of the cell and protecting it from osmotic turgor pressure but yet is flexible enough to allow for cell growth and expansion (Hett and Rubin, 2008). With the exception of *Mycoplasmas* and possibly *Chlamydia* (Pavelka, 2007), all eubacterial cell walls contain peptidoglycan. The glycans consist of repeating units of N-acetyl glucosamine (NAG) linked to N-acetylmuramic acid (NAM). These glycan chains are cross-linked by peptides bound to the lactyl groups on NAM for different glycan strands. The peptides that crosslink the glycan chains are L-alanyl-D-*iso*-glutaminyl-*meso*-diaminopimelic acid (DAP) from one strand which is linked to the terminal D-alanine residue from L-alanyl-D-*iso*-glutaminyl-*meso*-DAP-D-alanine (Brennan, 2003). The *Mycobacteria* along with the *Corynebacterium* and *Nocardia* groups of bacteria form a Chemotype IV

cell wall (Schleifer and Kandler, 1972), which differs from other bacterial cell walls in two ways. First, by the presence of some N-glycolylated muramic acid units along with NAM and secondly, while the majority of peptide linkages are the typical DAP-alanine type, a small percentage (25%) of these linkages are between two DAP molecules (Adam *et al*., 1969; Lederer *et al*., 1975).

Arabinogalactan: This polysaccharide anchors the impermeable mycolic acids to the peptidoglycan layer. It is composed of the unusual D-arabinofuranose and D-galactofuranose sugars that have been found to be serologically active (Lee *et al*., 1996).

Mycolic acids: The mycolic acids are high molecular weight α-alkyl and long β-hydroxyl fatty acids that are attached to arabinogalactan by ester bonds and are also present as free extractable lipids. The mycolic acids can be separated into the long meromycolate moiety (comprising 60-90 carbon atoms per chain) and the short α branch lipids (Minnikin, 1982). The mycobacterial mycolic acids differ from the mycolic acids found in *Corynebacteria*, *Norcadia* and *Rhodococcus* in a number of aspects. Firstly they contain the largest number of carbon atoms (C_{70}- C_{90}), secondly, they have the largest α branch (C_{20}-C_{25}) and thirdly, they contain double bonds, cyclopropane rings or methyl groups as non-oxygen functional groups in the main chain (meromycolate moiety) (Brennan and Nikaido, 1995). In Mycobacteria, the short chain fatty acids are synthesised by the Fatty Acid Synthase I (FASI) complex, while the long chain fatty acids are synthesised by the extension of the mercomycolate precursors from FASI by the FASII complex (Barry *et al*., 1998). In *M. tuberculosis*, the genes encoding FASII enzymes are located in 2 loci, an operon containing four genes, *fabD-acpM-kasA-kasB*, and a second operon containing the reductases *mabA* and *inhA* (Cole *et al*.,

1998). Due to this indispensable nature of the mycolic acid biosynthetic enzymes, they have been tested as potential drug targets for TB.

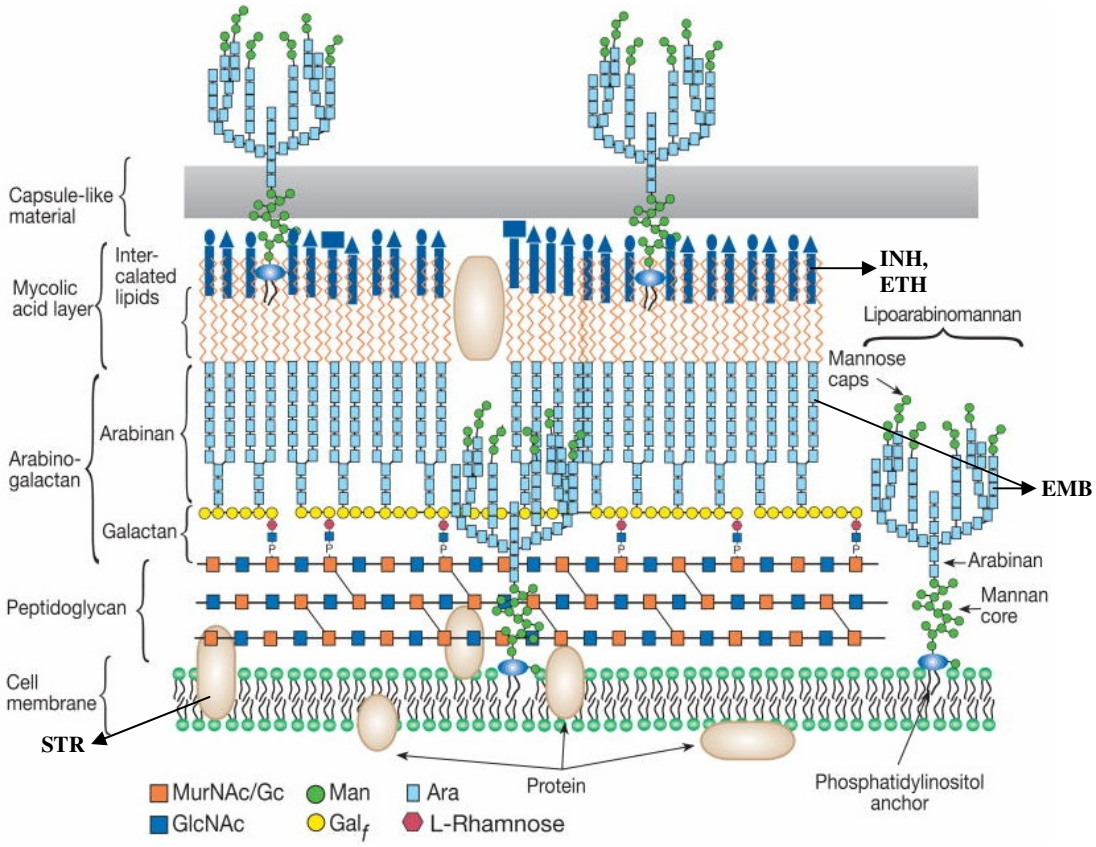


Figure 1. 1: Structure of a mycobacterial cell wall

(Image taken from Esko *et al*., 2009)

This complex cell wall architecture is responsible for the bacterium's growth and survival in an infected host (Daffe and Draper, 1998). The biosynthesis of the cell envelope especially the synthesis of the mycolic acids have been exploited as drug targets by drugs such as Ethionamide (ETH) and isoniazid (INH) which target the *inhA* protein (Banerjee *et al*., 1994) and Ethambutol (EMB) and Streptomycin (STR) which work by inhibiting synthesis arabinan or by inhibition of protein synthesis respectively (Belanger *et al*., 1996; Zhang *et al*., 2006). A schematic representation of the

Mycobacterial cell wall is shown in figure 1.1, where the arrows indicate the sites of action of the above mentioned drugs (Alsteens *et al*., 2008).

1.2.1.2 *Mycobacterium smegmatis* as a model organism

Mycobacterium smegmatis was first reported in November 1884 by Lustgarten and then isolated from human smegma the following year (Alvarez, 1885). This saprophytic, free-living, acid-fast, soil bacterium can be cultured in simple defined media and grows ten times faster than *M. tuberculosis* (Jacobs, 2000). Although the non- pathogenic *M. smegmatis* doesn't enter epithelial cells or persist in professional phagocytes, it has been suggested as an ideal model organism for studying mycobacteria in general (Reyrat and Kahn, 2001). However, it can be argued that due to a number of reasons it can be used to understand the more pathogenic mycobacteria. Firstly, it is easily transformable and can be used in Containment Level II facilities (Snapper *et al*., 1990), secondly, it shares over 60% identity with virulence genes, some of which were previously reported to have been unique to *M. tuberculosis* (Berthet *et al*., 1998; Glickman *et al*., 2000) and thirdly, it shares a high similarity in cell wall architecture with the *Mycobacterium avium* complex (MAC), making studies of cell wall associated virulence lipids easy (Belisle, 1991).

1.1.3 Protein Folding

1.1.3.1 Background on protein folding

In order to be functional, proteins must fold and reach their stable state. The environment of a cell is a complex mixture of macromolecules. So how does a protein fold in order to attain its native configuration and become biologically active? Initial

studies indicate that all the information required for the protein to achieve this native conformation is present in its amino acid sequence (Haber and Anfinsen, 1962). This was demonstrated by Anfinsen and his co-workers in the late 1950s and early 1960s. These studies involved the reduction and unfolding of bovine pancreatic ribonuclease by treatment with mercaptoethanol in the presence of 8M urea (chaotropic agent). It was observed that upon reduction and protein unfolding, eight cysteine residues were generated which produced a non functional randomly coiled structure. However, upon removal of the urea and re-oxidation by molecular oxygen, reformation of the disulphide bonds occurred and helped in restoration of the native configuration of the protein which was found to be fully active (Anfinsen *et al*., 1961). This series of experiments gave rise to the 'Thermodynamic Hypothesis' which states that the native conformation of a protein under its normal physiological conditions is attained when the total Gibbs free energy of the system is lowest. Since the free energy of a macromolecule is determined by its inter-atomic interactions which in the case of a polypeptide is its amino acid sequence, it can be conclusively stated that all the information required by a protein to attain its native configuration is present in its amino acid sequence (Anfinsen, 1973).

Towards the late 1960s another question arose as to the amount of time taken before a protein achieves its native state. In 1968, Levinthal proposed that an unfolded protein (for example consisting of 100 amino acid residues) would have to go through at least 10^{30} different conformations (i.e. assuming there are only two possible conformations per residue) before attaining its native conformation (Levinthal, 1968). Assuming that it takes 10^{-11} of a second (this is the rate of atomic bond rotation) to change from one

conformation to another, it would take nearly 10^{19} second or 10^{11} year for the polypeptide chain to go through all the possible conformations. However, it takes less than a minute for proteins to fold *in vitro* to its native state. From this we infer that a protein cannot sample all the possible conformations before attaining its native state and this conflict between the theoretical estimate and the actual time taken for protein folding to take place has came to be known as the 'Levinthal Paradox'. This leads to the conclusion that protein folding occurs along a specific pathway or pathways and not in a random manner.

1.1.3.2 Models of protein folding

According to the Levinthal Paradox, protein folding would be a very time consuming process, however, the fact that proteins do indeed fold in a few seconds to minutes has been established. Over the years, various groups have put forward their ideas on the different models for protein folding. These various models are discussed in brief below.

The 'Nucleation-Growth Model' proposes that a set of 8-18 amino acid residues forms a nucleus. The process of protein folding occurs by a rapid 'growth' or addition of residues closest to the nucleus initially and then followed by residues that are located further away (Wetlaufer, 1973). According to the 'Diffusion-Collision Model' put forth by Karplus and Weaver in 1976, a protein is composed of several parts or microdomains (Karplus and Weaver, 1976). Each of these microdomains is small enough for all the possible conformations to be searched through rapidly. These microdomains are unstable and during diffusive encounters in the cell, they collide and coalesce with each other to form more stable structures as seen in figure 1.2. In 1982, another model called

the 'Framework Model' was proposed by Kim and Baldwin (Kim and Baldwin, 1982). Their model states that individual H-bonded transient secondary structures are stabilised by packing against each other. These simple structures that are formed by packing interact further with each other to produce more complex stable structures. This model is very similar to the previously mentioned diffusion-collision model. The 'Hydrophobic Collapse Model' proposes that protein folding is initiated by the uniform collapse of the protein molecule by hydrophobic effects which is then followed by the formation of the more stable secondary structures (reviewed in Nolting and Andert, 2000).

The 'New View' of protein folding puts forward The 'Energy Landscape Theory'. This theory uses a folding funnel to explain its model of protein folding. A folding funnel is an energy diagram representing the thermodynamics of a population of folding proteins as seen in figures 1.3 and 1.4 (Leopold *et al*., 1992). It has a very broad mouth which corresponds to a state of favourable high entropy and unfavourable high energy unfolded state. It then follows the thermodynamically easy route to form the active state of the protein. This is represented as the funnel narrows, and the protein begins to fold and become more stable and the number of possible conformations reduces significantly. It finally reaches a state of unfavourable low entropy and favourable low energy that represents the native conformation of the protein (Fitzkee *et al*., 2005; Wolynes *et al*., 1995). The transition from the unfolded to the folded state is interspersed with intermediate folded forms that are generated by local energy barriers created along the energy funnel. The 'molten globule' states which are defined as the states of a polymer in which there is negligible to partial amount of native structure and it is also characterised by a high level of entropy (Plotkin and Onuchic, 2002). This

view seems to give a good resolution to the Levinthal paradox as it explains how a protein with a huge number of possible conformations, does not need to sample all the possible forms but instead can adopt those selected states which entropically favour the formation of the active form while making its way down the funnel.

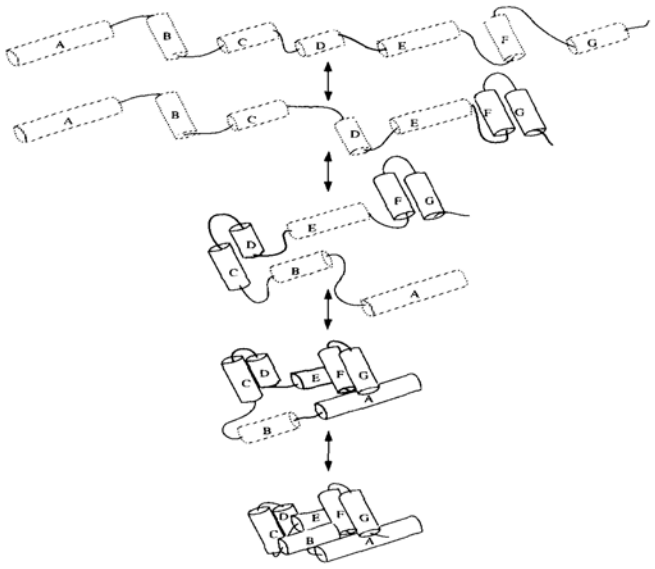


Figure 1. 2: Illustration depicting diffusion collision model of protein folding

An illustration of the diffusion collision model where A-G are the microdomains. The dashed outlines indicate instability and the solid lines indicate stability. The process starts with the random coil set of conformations. The microdomains transiently occupy folded secondary structure conformations and during diffusive encounters, they collide and coalesce to form more stable intermediates which are held by hydrophobic interactions. These multimicrodomain intermediates in turn collide to form more stable structures. Image taken from Karplus and Weaver (1994).

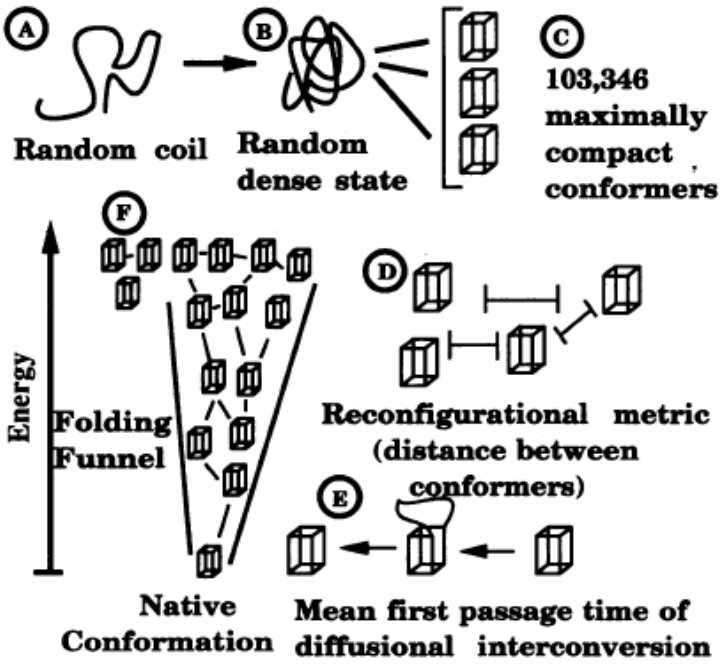


Figure 1. 3: Illustration depicting protein folding pathways

The image is a schematic representation of the folding process where the protein initially in a denatured coil (A), collapses to form a more dense structure (B) and eventually forms compact conformer (C). The conformers then arrange themselves according to their configurations and then enter the 'folding funnel' where they eventually form the active state of the protein. Image taken from Leopold *et al*., (1992).

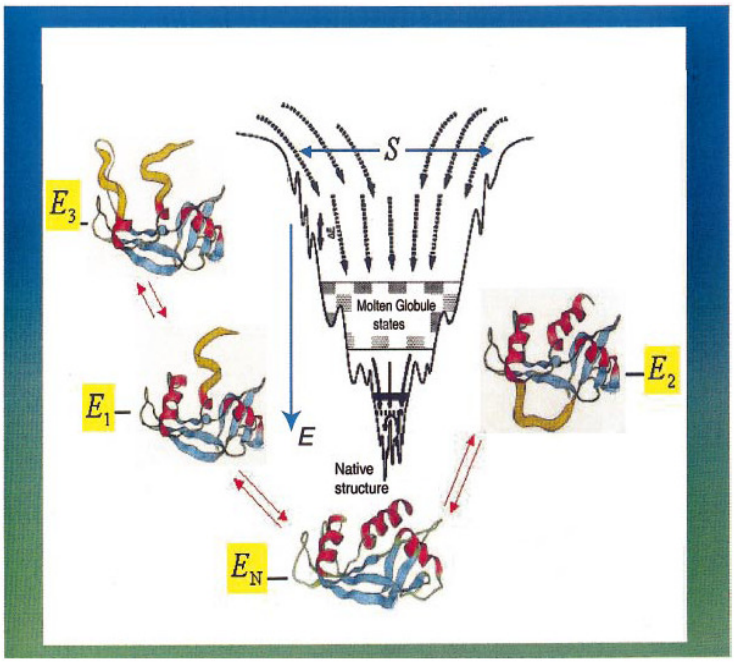


Figure 1. 4: Illustration depicting energy landscape theory of protein folding

The image depicts a typical folding funnel which represents the path a protein takes while it folds and finally achieves its stable native state. The depth of the funnel represents the free energy (E) of the conformational states while the width is a measure of the configurational entropy (S). As can be seen in the image, the protein is initially in a state of high energy (E_3) and high entropy. However as the protein starts folding, its energy levels decrease (E_2 and E_1) before it finally attains its stable native configuration (E_N). Also seen in the image is the 'molten globule states', transient compact intermediate structures with a certain degree of secondary structure. Image taken from Plotkin and Onuchic (2002).

1.1.3.3 Protein Aggregation

Protein aggregation can be described as the interaction of polypeptide chains with each other that gives rise to non-functional structures. While the fundamental cause of protein misfolding is the presence of unfavourable conditions, the two main things that contribute to protein aggregation in the cell are probably the presence of polyribosomes and the phenomenon of macromolecular crowding. Under unfavourable growth

conditions, such as a rise in the optimal growth temperature, proteins tend to unfold en masse leading to protein aggregation (Gragerov *et al*., 1991). The intracellular environment of a cell is very crowded due to the presence of a high concentration of macromolecules, such as proteins, nucleic acids and complex sugars which are present in the range of 200-400 mg ml^{-1} (Ellis, 1997). The presence of these macromolecules has high energetic consequences for the cell as they occupy nearly 40% of the total volume within a cell. Thus, reactions that increase the available volume are theoretically stimulated by crowded conditions (Ellis and Minton, 2003). The binding of macromolecules to each other, folding of proteins and nucleic acids into compact structures and the formation of aggregates are thus favoured as they all contribute to increasing the available volume within the cell. Thus it can be concluded that macromolecular crowding within a cell leads to protein aggregation which is deleterious to the cell and could result in disease and cell death (Ellis, 2001).

During transcription, a single strand of mRNA can become associated with more than one ribosomal complex as seen in figure 1.5. This results in the formation of a polyribosome where newly emerging polypeptide chains are located within a small distance of each other. This close proximity could lead to interaction between the unfolded chains causing aggregation (Frydman, 2001).

Previous studies have found that the rate of peptide-chain elongation in *E. coli* in fast and slow growing cells is 16 amino acids/second and 13 amino acids/second respectively (Dalbow and Young, 1975). Thus the rate of synthesis from the ribosome implies that a nascent protein will be incomplete for many seconds and may begin to

fold before all the information required is available, leading to the formation of misfolded proteins. Furthermore, it has been observed that the exit channel of the ribosome is approximately 15Å wide which only allows the polypeptide to attain a helical structure (Nissen *et al*., 2000). It has been demonstrated that the polypeptide chain can attain its native state only when its entire chain has emerged from the ribosome and this could take up to a few minutes for a 300 residue protein. Since protein folding is mainly driven by hydrophobic interactions, the time lag between the emergence of the chain from the ribosome and its folding to the native conformation, means that there is a high possibility of the interaction of two nascent proteins with each other, which could contribute to the misfolding and aggregation of the protein (Hartl and Hayer-Hartl, 2002).

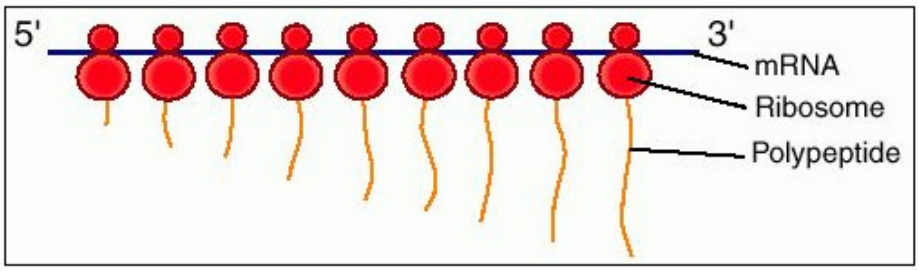


Figure 1. 5: **Illustration of mRNA strand with poly ribosomes**

A strand of mRNA along which a number of ribosomes are attached. Newly synthesised polypeptide chains can be seen emerging from the ribosomal complex.

1.1.4 Molecular Chaperones

1.1.4.1 Definition of molecular chaperones

The term 'molecular chaperone' was first used by Laskey *et al*. in 1978 to describe a nuclear protein (nucleoplasmin) that prevented the misassembly of proteins during the formation of nucleosomes in amphibian eggs (Laskey *et al*., 1978). Molecular

chaperones have been described by Ellis, as a group of unrelated protein families that assist in the proper folding and assembly of other proteins *in vivo*, which however, do not form a part of the final functional protein (Ellis, 1987; Ellis, 1993). There are a few important points to note from this definition of molecular chaperons (reviewed in Lund, 2001). First, protein assembly in the above context refers to de novo protein folding, association of protomers into oligomers, transport of proteins across and into membranes and also refolding of partially unfolded proteins or aggregates after stresses such as heat shock. Second, the mechanism of action of some classes of molecular chaperones has shown that they function as complexes rather than as individual proteins and the above definition of molecular chaperons refers only to the functional complex and not to the individual components. Finally, although some molecular chaperones have been classed as heat shock proteins due to their upregulation during higher temperatures, there are some molecular chaperones that are not induced by heat shock and similarly some heat shock proteins that are not molecular chaperones.

1.1.4.2 Necessity for molecular chaperones

The general view proposed by Anfinsen was that proteins were capable of folding on their own (without the assistance of other molecules) by a thermodynamically spontaneous process (Anfinsen, 1973). However, this view was disproved when evidence of certain specialised proteins (molecular chaperones) required to help in the proper folding of other proteins began to emerge.

Protein folding *de novo* is not from a denatured chain and the rate of synthesis of the polypeptide chain from the ribosome means that the nascent protein will be incomplete

for many seconds. This poses two potential problems, first the nascent chain may begin to fold before all the information required for correct folding is available and second the interaction of two nascent chains could lead to misfolding of the protein. Thus one of the functions of molecular chaperones is in protecting these nascent protein chains from misfolding.

There are a number of reasons which lead to protein aggregation (section 1.1.3.3). Due to the roles of molecular chaperones in post-translational folding, assembly and transport of other proteins, protein aggregation is greatly reduced.

Furthermore, in terms of free energy, the unfolded state of a protein is favoured compared to the native state. This makes proteins highly susceptible to increases in temperature. Although small changes may not cause complete denaturation, the unfolding of a few essential proteins can lead to the formation of aggregates. Molecular chaperones prevent this aggregation of proteins and also reverse it if it happens by binding and stabilising cellular proteins (Kusukawa and Yura, 1988). Hence, many of the molecular chaperones are also known as heat shock proteins (HSPs). The genes encoding heat shock proteins are highly conserved in nature and are found in all three domains of life (bacteria, archaea and eukaryotes) (Lindquist and Craig, 1988). Some chaperones require a co-chaperone to carry out their function. The study of the individual chaperones as well as their interactions with each other is vital as they play such a crucial role in preventing the misfolding and hence aggregation of proteins and accumulation of toxic products.

1.1.4.3 Classes of Molecular Chaperones

Molecular chaperones can be classified based on their size, mode of action, substrates they bind or their phylogeny. There are a number of classes of molecular chaperones which utilise different mechanisms to prevent protein folding and aggregation. Some groups of chaperones require ATP for binding and release of the protein (Hartl and Hayer-Hartl, 2002), while other groups play a major role in protecting against hydrophobic interaction of proteins with themselves during protein biosynthesis, as this will lead to aggregation. Moreover, they prevent partially folded proteins from exposing their interactive surfaces and forming small non-functional complexes (Lund, 2001). Some of the various classes of molecular chaperones with a few examples are listed in the table below (table 1.2).

Chaperone family	**Examples**	**Function**
HSP100	ClpA, ClpB, ClpX } Bacteria HSP78 } Mitochondria ClpC, ClpB } Chloroplasts	Disassembly of oligomers and aggregates
HSP90	HtpG } Bacteria TRAP1 } Mitochondria HSP90 } Chloroplasts HSP90, HSP82, HSP83 } Eukaryotic cytosol GRP94 } ER	Protein folding and degradation. In eukaryotes, regulation of assembly of steroid receptors and signal transduction proteins
HSP70	DnaK } Bacteria SSC1 } Mitochondria HSP70, HSC70, SSB, SSA } Eukaryotic cytosol BiP } ER	Prevention of aggregation of unfolded or newly translated proteins; regulation of heat shock response
HSP60	GroEL (Cpn60) } Bacteria HSP60 } Mitochondria Rubisco sub-unit binding protein } Chloroplasts Thermosome, TF55, CCT } Archaea TRiC, CCT } Eukaryotic cytosol	Folding of newly translated and newly folded proteins
HSP40	DnaJ, CbpA, DjlA } Bacteria Ydj1 } Mitochondria	Binding of unfolded proteins, stimulation of ATPase activity of HSP70 (HSP70 cochaperone). Very

	Cdj1 } Chloroplasts HSP40, Sis1, Zuo1 } Eukaryotic cytosol Sec63 } ER	heterologous group as many proteins possess "DnaJ-domain"
sHSP (small heat shock proteins)	IbpA, IbpB } Bacteria HSP22 } Mitochondria HSP21 } Chloroplasts HSP16.5 } Archaea HSP26, Alpha-crystallin } Eukaryotic cytosol	Stabilisation of unfolded proteins, prevention of protein aggregation
Co-chaperonin GroES	GroES (Cpn10) } Bacteria	Essential for function of group I chaperonins
Trigger factor	Tig } Bacteria	Assists folding of nascent ribosome-bound polypeptide
Nascent polypeptide-associated complex	MTH177 } Archaea NAC } Eukaryotic cytosol	Interacts with and probably assists folding of nascent ribosome-bound polypeptide

Table 1. 2: **Classes of molecular chaperones**

This table lists the classes of molecular chaperones and gives some examples of the proteins in the different domains (Lund and Ellis, 2008).

1.1.4.4 Heat shock proteins and their regulation in *E. coli*

E. coli has served as the paradigm species for the study of heat shock proteins (HSP) in Gram-negative bacteria. The heat shock response in *E. coli* is regulated by an alternative sigma factor, the σ^{32} subunit of the RNA polymerase that is encoded by the *rpoH* gene (Grossman *et al*., 1984; Bukau, 1993). This sigma factor is required by the RNA polymerase to direct it specifically to the promoters of the heat shock genes as these

differ from other housekeeping promoters (σ^{70}-dependent promoters) (Narberhaus, 1999). Upon heat shock, the expression of a set of highly conserved proteins (heat shock proteins) is increased. It was demonstrated that some of these heat shock proteins were also required at normal temperatures and that they were at least partially regulated by σ^{32} as deletion of the *rpoH* gene resulted in a temperature sensitive phenotype (Zhou *et al*., 1988). Compensatory mutants that allowed for growth up to 40°C were screened and mutations were found to be in the region of the *groE* operon (Kusukawa and Yura, 1988). The genes of the *groE* operon were previously shown to play a role in bacteriophage lambda assembly (Georgopoulos *et al*., 1973). When growth of the *rpoH* mutant above 40°C was analysed a new set of up-promoter mutations were observed in the *dnaK* operon. Although the genes of the *dnaK* operon, *dnaK* and *dnaJ* were originally identified for their role in host DNA replication, they were also been found to play a role bacteriophage lambda replication (Saito and Uchida, 1977, 1978). The *dnaK* and *dnaJ* genes were also found to play a role in the replication of plasmids mini P1 and mini F1 (Kawasaki *et al*., 1990; Tilly and Yarmolinsky, 1989).

Under normal physiological conditions, the concentration of σ^{32} in the cell is low due to weak transcription of the *rpoH* gene and due to its association with DnaK, which makes it more susceptible to degradation by proteases (Herman *et al*., 1995; Straus *et al*., 1987). However, on exposure to stress, the HSPs (such as DnaK) become associated with the non-native proteins, thus allowing the immediate release of a small amount of already present active protein and simultaneously the translation of the *rpoH* mRNA results in increased expression of the heat shock genes. HSPs act either independently or in association with one another to protect newly synthesised proteins or repair or

degrade damaged proteins. Once the non-native proteins are repaired or degraded an accumulation of the heat shock proteins cause repression of σ^{32} factor expression by binding to it and thus serves as a negative feedback loop regulator (Bukau, 1993; Guisbert *et al*., 2004; Schroder *et al*., 1993).

1.1.4.5 Regulation of the heat shock response in Actinobacteria

Even though the heat shock regulation by the *rpoH* gene has been extensively studied in *E. coli*, it is not the most commonly used mechanism. This may be due to the absence of the *rpoH* gene as is the case in *M. smegmatis*. The regulation of the heat shock response by a transcriptional repressor protein has been studied in other bacteria such as *Bacillus subtilis* (Yura and Nakahigashi, 1999). One of the well characterised systems is the CIRCE/HrcA system that regulates the expression of the DnaK and GroE operons in *Bacillus subtilis* (Zuber and Schumann, 1994). In this system, the repressor protein HrcA binds to an operator sequence located in the promoter regions of the genes it regulates (Roberts *et al*., 1996). This inverted repeat sequence, designated as the CIRCE sequence (controlling inverted repeats of chaperone expression) was originally identified upstream of the Mycobacterial *groE* genes (Baird *et al*., 1989), implying that the *groE* genes may be regulated by the CIRCE/HrcA system.

Another well characterised regulation system is the HspR/HAIR system, in which a repressor protein HspR binds to an operator sequence known as the HAIR sequence (HspR-associated inverted repeats) (Bucca *et al*., 1995; Grandvalet, 1999). Studies conducted in *M. tuberculosis*, have demonstrated that only the DnaK and ClpB chaperone systems are regulated by the HspR/HAIR system (Stewart *et al*., 2002).

1.1.5 HSP60 family of chaperones

This family of chaperones belongs to the group of larger chaperones that are barrel-shaped and has been well characterised. The term 'chaperonin' was coined by Hemmingsen in 1988 to represent this subclass of molecular chaperones (Hemmingsen *et al.*, 1988). Over the years, the molecular chaperones have been called by various names (GroE, heat shock 58, P1, and HSP60 and 65kDa antigen). In this study we adopt the nomenclature proposed by Ellis and Coates in 1993 (Coates *et al.*, 1993). This system states that they will be known as 'GroEL chaperonin 60 gene' and that the duplicate copies of the gene will be known as chaperonin 60.1 and chaperonin 10.1 (in the case of GroES). Hence, while the *E. coli* chaperonins will continue to be referred to as *groEL* and *groES* genes in this study, the homologues from other bacteria will henceforth be referred to as chaperonin 60 (*cpn60*) and chaperonin 10 (*cpn10*) respectively. The chaperonin genes in *M. tuberculosis* are *cpn60.1* (Rv3417c), *cpn60.2* (Rv0440c) and *cpn10* (Rv3418c).

The chaperonins can be divided into two groups based on their sequence similarity (Gutsche *et al.*, 1999; Lund *et al.*, 2003):

- Group I chaperonins: found in bacteria (e.g. GroEL in *E. coli*), chloroplasts (rubisco binding protein) and eukaryotic organelles like mitochondria (Hsp60), require the presence of co-chaperonin (GroES) for their function and form a seven membered ring structure.
- Group II chaperonins: are found in archaea (thermosome) and eukaryotic cytosols, do not require the presence of a co-chaperonin and form an eight or nine membered ring structure (Macario *et al.*, 1999). An example of this group

is the TRiC (TCP-1 ring complex, which is also known as CCT for chaperonin-containing TCP-1) in eukaryotic cytosols and thermosome in the archaeal cytoplasm (Gutsche *et al*., 1999; Hartl and Hayer-Hartl, 2002).

Both groups of chaperonins are similar in that they require ATP (Hartl and Hayer-Hartl, 2002). While GroEL and the thermosome are induced by heat shock, CCT in eukaryotes is not induced by stress. Some of the differences between the two groups of chaperonins is the presence of 7 (Group I) and 8 or more (Group II) subunits per ring and the presence of a detachable 'lid' like structure (GroES) in group I chaperonins and a built-in protrusion in group II chaperones (reviewed in Horwich *et al*., 2007).

1.1.5.1 Significance of the GroE system

The GroEL protein is conserved across the kingdoms not only in sequence but in function as well (Hemmingsen *et al*., 1988). The *groE* genes were first identified when mutations in them prevented the proper assembly and growth of bacteriophage (Georgopoulos *et al*., 1972; Takano and Kakefuda, 1972). It has also been demonstrated that both GroEL and GroES are essential for growth at all temperatures in *E. coli* (Fayet *et al*., 1989). A homologue of the GroEL protein in chloroplasts was found to be involved in the assembly of rubisco (Hemmingsen *et al*., 1988). In *E. coli*, over expression of GroES and GroEL assisted in the folding of denatured Rubisco (Goloubinoff *et al*., 1989). These studies and many more indicate that the GroE family of chaperonins is essential for growth not only at normal temperatures but under stress conditions as well, where they help in folding of unfolded proteins.

1.1.5.2 Structure of GroEL and GroES

Initial purification studies conducted in 1979, revealed that GroEL was a soluble protein consisting of 14 subunits each having a molecular weight of approximately 65kDa (Hendrix, 1979). GroEL has a characteristic arrangement, in which there are two heptameric rings stacked back to back; the entire structure is approximately 147 Å in length and 137 Å in diameter (figures 1.6 and 1.7). Each ring encloses a central hydrophobic cavity which is 47 Å in diameter, where the unfolded proteins (<70kDa in size) can attain their native forms before being released into the external medium (Braig *et al*., 1994). The 10kDa co-chaperonin GroES forms a heptameric ring that caps one of the two GroEL rings (*cis* ring) forming an asymmetric complex (Xu *et al*., 1997). This cavity provides an environment where the unfolded polypeptide can fold correctly without the presence of other folding chains (Fenton and Horwich, 2003). This enclosed cavity is known as the 'Anfinsen cage'.

Each subunit can be further divided into three domains:

- Apical domain: this domain is present at the top of the complex. Since the binding sites for substrates located in this domain are also involved in GroES binding (Fenton *et al*., 1994), binding of GroES displaces the bound substrate into the central cavity.
- Intermediate domain: this slender layer connects the apical and equatorial domains. Upon nucleotide binding, this domain is responsible for the transfer of allosteric signals from equatorial domain to the apical domain, which helps in binding and encapsulation of substrates (Horwich *et al*., 2007).

- Equatorial domain: this is the part that contains the ATP binding sites and which connects the two rings to each other (Fenton *et al*., 1994).

GroES

The co-chaperonin consists of a single heptameric ring that is made up of identical subunits that are 10kDa in size. Each subunit has a β-barrel structure with two protrusions extending from it. One of these loops forms the top of the dome and is called the dome loop, while the other loop interacts with the apical domain of the GroEL subunit and is called the mobile loop. This GroES in the presence of ATP/ADP binds to the chaperonin GroEL to form an asymmetric complex that aids in protein folding (Hunt *et al*., 1996; Xu *et al*., 1997).

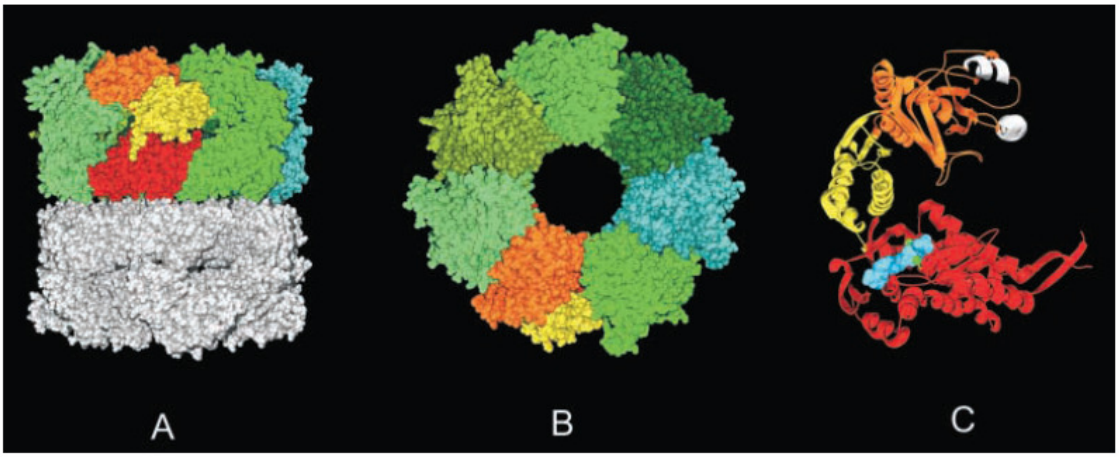


Figure 1. 6: Structure of *E. coli* GroEL complex

This figure depicts the structure of GroEL. The first image (A) represents the side view showing the two rings. The top ring shows the different subunits (shades of green) and one subunit has further been divided into the apical domain (orange), intermediate domain (yellow) and the equatorial domain (red). Image (B) is the top view of the molecule in which one subunit has been divided into its domains, while image (C) is a ribbon representation of a single subunit showing the three domains. Also seen in this image is the ATP (blue) bound to the intermediate domain and the hydrophobic regions (white) in the apical domain where GroES binding takes place. Image taken from Walter (2002).

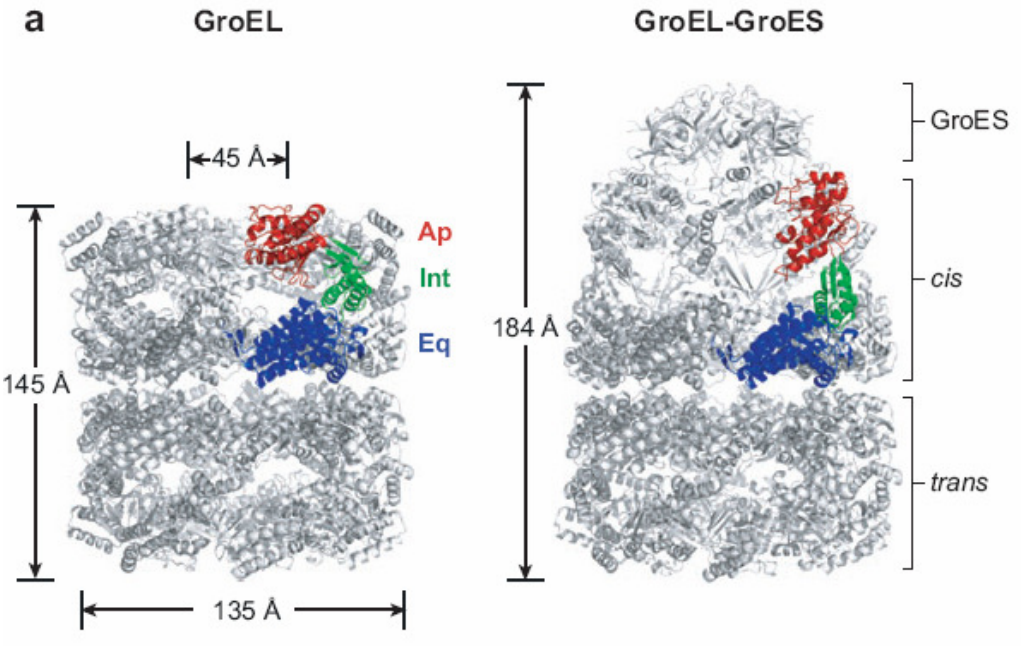


Figure 1. 7: **Structure of *E. coli* GroEL-GroES complex**

Crystallographic models of the GroEL complex (left) and the GroES-GroEL complex (right). Each subunit of tetradecameric GroEL is composed of an apical (red), intermediate (green) and equatorial (blue) domain. Image taken from Horwich *et al*. (2007).

1.1.5.3 Anfinsen Cage

In order to explain the mechanism of action of GroEL, a simple model was developed to explain the manner in which GroEL stimulated protein folding (Ellis, 1994). ATP hydrolysis and transient binding of GroES are thought to contribute to the alternation of high and low affinity states of GroEL to its non-native substrates. The non-native protein binding sites located in the apical domain of the GroEL molecule, shift away from the lumen of the cavity to be sequestered by GroES. This provides the folding protein with an isolated, diluted, hydrophilic environment in which protein folding can

take place unhindered. This model of GroEL mediated protein folding came to be known as the Anfinsen cage model (Ellis, 1994; Saibil *et al*., 1993).

1.1.5.4 Iterative annealing mechanism

Another model for GroEL mediated protein folding was put forth by Todd *et al.,* in 1996 and is known as the iterative annealing mechanism (Todd *et al*., 1996). This model makes use of the energy landscape theory to explain the unfolding and folding of the substrate protein by GroEL. According to this model, binding of GroES to GroEL causes a conformational change in GroEL. A substrate that is trapped in a kinetically unfavourable state then comes in contact with GroEL which causes the unfolding of the substrate as seen in figure 1.8. This unfolding is followed by its release from GroEL. The substrate protein is now in a position to adopt two different routes; it can either fall back into the kinetic trap or it can follow its productive route and fold to its native state. The portion of substrates that fall back into the kinetic trap once again come in association with GroEL and the cycle continues (Thirumalai and Lorimer, 2001).

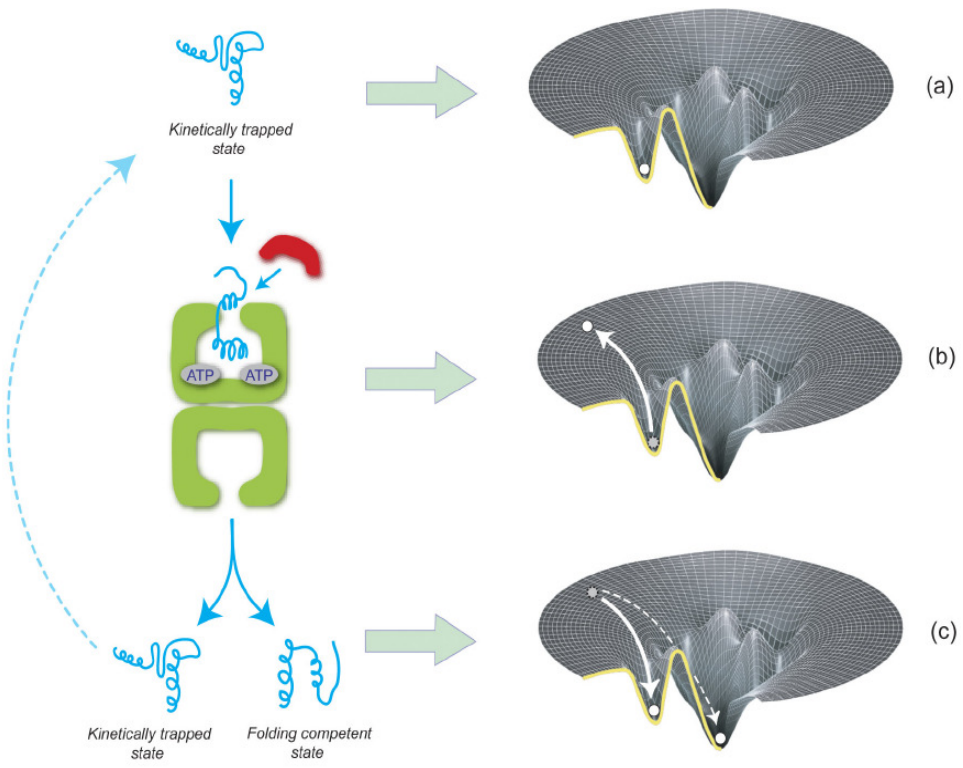


Figure 1. 8: **Illustration depicting the iterative annealing mechanism**

This image represents the Iterative mechanism of annealing. The position of the kinetically trapped substrate can be seen in the folding funnel as well (a). This substrate then comes into contact with GroEL and unfolding takes place which causes the substrate to be released from the kinetic trap (b). the substrate is now free to follow a route leading to proper folding to the native state or may once again become trapped in which case it will interact with GroEL once again (c). Image taken from Lin and Rye (2006).

1.1.5.5 Substrates of GroEL

Recently, various groups have been carrying out studies to determine the substrates of the GroEL/GroES system. This work that has been conducted in *E. coli* provides insight into the various proteins that require GroEL for their folding. Initially an approach where temperature sensitive mutants of *groES* and *groEL* were screened for their *in vivo* properties showed that defects in DNA and RNA synthesis (Wada and Itikawa, 1984),

cell division (Georgopoulos and Eisen, 1974) and proteolysis (Straus *et al*., 1988). However that kind of approach has it limitations as it is difficult to determine whether the observed effects are a direct or indirect consequence of the *groE* genes. Diaminopimelic acid (DAP) is a precursor in cell wall synthesis and one of the enzymes (DapA) required for its synthesis was shown to be a substrate of GroEL (McLennan and Masters, 1998). When levels of GroEL were depleted in a modified strain of *E. coli* (*E. coli* MGM100), cell lysis occurred which was prevented by supplementing the media with DAP, thus establishing a link between DapA and GroEL. More recently a direct approach involving the complete proteome analysis of *E. coli* by biochemical and quantitative proteomics showed that while nearly 350 proteins interact with GroEL, only about 85 of these proteins were found to be obligate GroEL substrates (Kerner *et al*., 2005). This set of obligate substrates was identified by monitoring their refolding in the presence of GroEL. Furthermore, 13 of the above mentioned proteins were found to be essential proteins. This explains the fact that GroEL is a vital factor in the growth and survival of the bacteria. The various substrates were divided on the basis of their requirement for GroEL as:

- Obligate GroEL proteins: e.g., MetF, MetK, GatY, DapA
- GroEL/ATP associated proteins: e.g., EF-G, FtsZ, AlaRS, Acyl CoA synthetase
- Reversibly GroEL interacting proteins: e.g., LysA

Of particular interest was the MetE protein (serves as a marker of GroEL function), whose levels appeared to increase after a temperature shift. This was observed in a temperature sensitive GroEL mutant, which was grown in LB. The *metE* gene is repressed by MetF and MetK both of which have been found to be obligate GroEL

substrates. Hence, in accordance with findings, when GroEL is non-functional, derepression of MetE occurs and its levels increase (Chapman *et al*., 2006).

1.1.5.6 Mechanism of GroES-GroEL mediated protein folding

The mechanism of protein folding using the GroESL (GroES and GroEL complex) machinery is complex. There are two kinds of allosteric interactions taking place that facilitate the folding of a substrate protein by the GroESL machinery. These are intra-ring positive co-operativity by which the seven ATP molecules bind co-operatively to each of the seven subunits of each ring and the inter-ring negative co-operativity where the binding of ATP molecules to one ring results in the decrease in ATP affinity to the other ring. The steps involved in the GroESL reaction cycle will be summarized below.

- The chaperonin reaction cycle begins by the binding of the unfolded protein to the hydrophobic domain of the free *cis* ring of the GroEL protein (figure 1.9a). This binding causes the dissociation of the GroES molecule from the *trans* ring. The hydrophobic apical surface of the GroEL ring acts as the binding sites not only for the co-chaperonin GroES but also for the non-native substrate proteins (Fenton *et al*., 1994).
- In the next step, seven molecules of ATP bind to the equatorial domain of the *cis* ring (intra-ring positive co-operativity) followed by the binding of GroES (through the mobile loops) to the apical domain of the GroEL *cis* ring. This binding of GroES results in the enlargement of the cavity, and conversion of the inner surface from hydrophobic to hydrophilic (Weissman *et al*., 1995; Xu *et al*., 1997).

- The binding of ATP and GroES to the GroEL protein results in a number of domain movements to enable the central cavity to switch from the peptide accepting state to the folding active state. These movements consist of a 25° downward rotation of the apical domain results in lowering of the intermediate domain onto the equatorial domain locking the nucleotide into the ATP site as seen in figure 1.10 (Ranson *et al*., 2001). A 65° upward rotation of the apical domain then causes an upward elevation of the apical domain which is followed by a 90° clockwise twist of the domain along its long axis (Xu *et al*., 1997; Ranson *et al*., 2006).
- These rigid movements in the *cis* ring of GroEL result in the burying of the hydrophobic residues originally facing the inner cavity thus displacing the bound polypeptide into the central cavity (Weissman *et al*., 1995, 1996).
- The non-native protein is now enclosed in this cavity and is free to fold in the absence of other folding polypeptides or crowding agents (figure 1.9b). Folding of the polypeptide proceeds until the bound ATP is hydrolysed (about 10 seconds). This hydrolysis of ATP causes a decrease in the affinity of GroES by GroEL (figure 1.10c) (Rye *et al*., 1997).
- This is followed by the binding of seven molecules of ATP to the opposite *trans* ring (figure 1.9d), which triggers the dissociation of GroES and substrate protein (completely folded or partially folded) from the *cis* ring (inter-ring negative co-operativity) (Todd *et al*., 1994; Weissman *et al*., 1994).

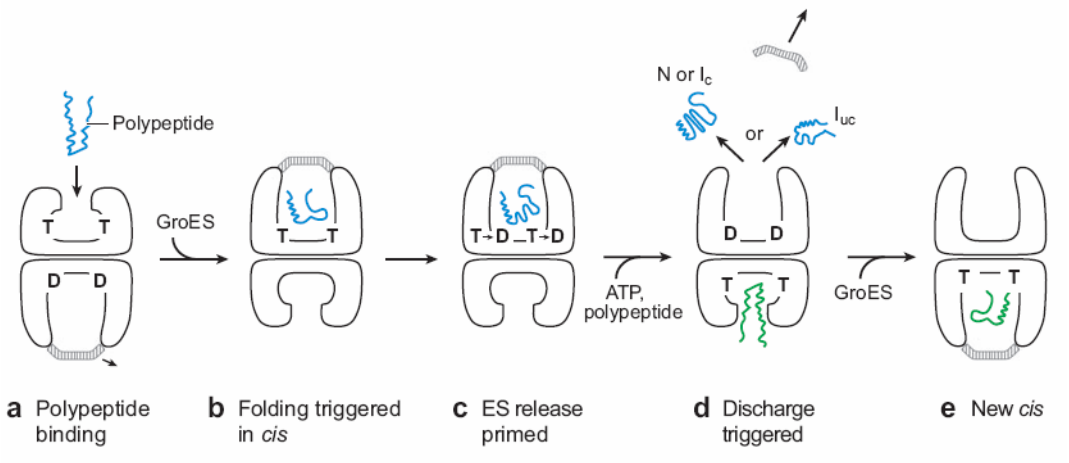


Figure 1. 9: The reaction cycle of GroES-GroEL.

The unfolded polypeptide binds to the *cis* ring causing the dissociation of GroES from the *trans* ring (a), which is followed by the binding of ATP (T) and GroES to the *cis* ring where folding takes place within the cavity (b), ATP hydrolysis (D) in the *cis* ring causes the release of GroES (c) and the discharge of the completely folded (N) or partially folded intermediates (I_c and I_{uc}) (d). The new cycle then starts by the binding of unfolded polypeptide and ATP to the *trans* ring (d and e) which now comes to be known as the *cis* ring. Image taken from Horwich *et al.* (2007).

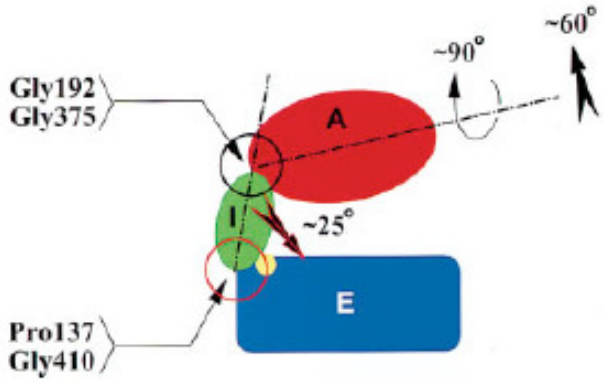


Figure 1. 10: **Illustration of domain movements of GroEL during protein folding cycle**

Figure shows the domain movements of the GroEL ring which enables the protein to switch from a peptide accepting state to a folding-active state. The Apical domain is represented in red, intermediate in green, equatorial in blue and the ATP nucleotide in yellow. Image taken from Bukau and Horwich (1998).

1.1.5.7 Group II chaperonins

This group of chaperones is found in the eukaryotic cytosol TRiC (TCP-1 ring complex, which is also known as CCT for chaperonin-containing TCP-1) and in the archaeal cytoplasm (thermosome) (Gutsche *et al*., 1999; Hartl and Hayer-Hartl, 2002). They are hexadecameric or nonadecameric oligomers consisting of one to eight subunits. They are similar to the group I chaperonins in that they are double ringed structures with each sub unit having the same three domains (apical, equatorial and intermediate) and have been shown to use the same mechanism of action, i.e., the binding of the substrate to a central cavity and followed by its ATP-dependant release. However, the Group II chaperonins differ in that they are hetero-oligomeric structures which do not require the function of a co-chaperone for folding.

1.1.6 Multiple chaperonins in bacteria

The presence of more than one copy of the *cpn60* gene is common and this occurrence has been observed in a number of groups of bacteria (figure 1.11) (Lund, 2009). From the table in figure 1.12, it is clear that the presence of multiple chaperonins in bacteria is not randomly distributed but is limited to certain groups. Although, the presence of multiple copies of *cpn60* genes was initially reported in the *Actinobacteria* (Rinke de Wit *et al*., 1992; Kong *et al*., 1993), it has since been reported in members of Cyanobacteria (Lehel *et al*., 1993), Rhizobia (Fischer *et al*., 1993; Rusanganwa and Gupta, 1993; Wallington and Lund, 1994) and Chalamydia (Karunakaran *et al*., 2003). Furthermore, the number of copies of the *cpn60* genes in a group appears to follow a particular pattern, for example, all the members Chlamydiae possess three homologues of *cpn60*, while the number of homologues in the Rhizobia range from one to seven.

This strongly suggests that selective pressure maintains the presence of multiple copies and that the presence of these multiple copies is not due to a random event.

It has already been demonstrated that a single copy of GroEL in *E. coli* is capable of folding a range of substrates, so what is the purpose of these multiple copies? One possibility is that the extra copies are present to maximise chaperoning activity within the cell, while the other possibility is that the duplicated copies have evolved to take on a more specialised functions in the cell. However, as results from a number of studies have demonstrated, the situation is probably in between the two hypotheses, where some copies of genes are dispensable for growth while others are not due to housekeeping or specialised roles in the cell (Fischer, 1999; Rodriguez-Quinones *et al.*, 2005; Kapatai *et al*., 2006).

A phylogenetic analysis of 249 multiple chaperonins in bacteria found that the phenomenon of multiple chaperonins was due to independent gene duplications in different lineages leading to distinct histories of adaptation and selection in each group (Goyal *et al*., 2006). Thus the characteristics of these multiple chaperonins may differ between two organisms or taxonomic groups.

1.1.6.1 Multiple chaperonins in Actinobacteria

While most members of this group contain at least two copies of the *cpn60* gene, in all cases only a single copy of the *cpn10* gene is found linked to one *cpn60* gene. In cases where only one copy of the *cpn60* is present (*Tropheryma whipplei*), it is located elsewhere on the genome and linked to the *cpn10* gene (Maiwald *et al*., 2003). This

implies that a common ancestor of all the *Actinobacteria* had two copies of the *cpn60* genes one with *cpn10* and one without and due to functional redundancy, the copy located with the *cpn10* gene was lost (Goyal *et al*., 2006). Studies on the evolutionary rates between the *cpn60* genes in *M. tuberculosis* and *M. leprae*, suggest that the *cpn60.1* genes have evolved at a much faster rate than the *cpn60.2* genes (Hughes, 1993). This strongly suggests that the *cpn60.2* genes could be the essential housekeeping genes while the *cpn60.1* genes could have evolved to play a more specialised role in the cell. This will be discussed shortly.

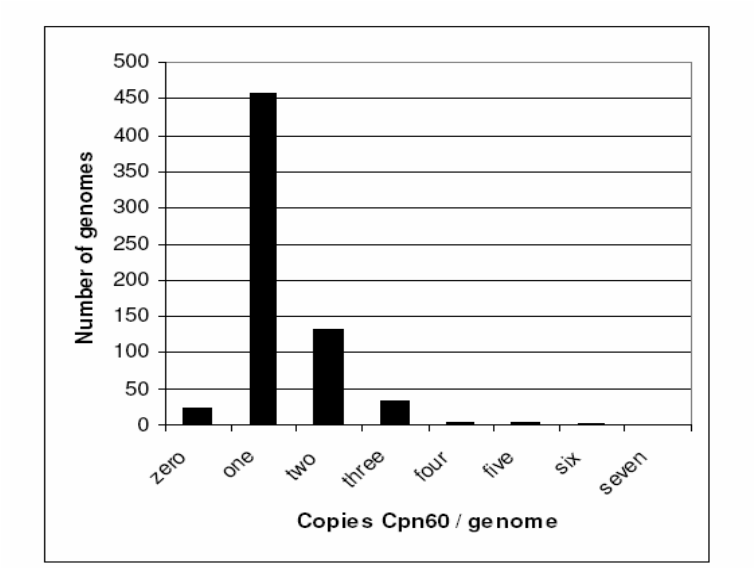


Figure 1. 11: Distribution of the copies of *cpn60* genes amongst all the sequenced bacterial genomes

(Taken from Lund, 2009)

		Number of Cpn60 homologues							
NCBI group	Total	0	1	2	3	4	5	6	7
Actinobacteria	54	0	9	40	2	3	0	0	0
Alpha proteobacteria	90	0	55	18	8	2	5	1	1
Aquificae	1	0	1	0	0	0	0	0	0
Bacteroidetes/Chlorobi	23	0	22	1	0	0	0	0	0
Beta proteobacteria	60	0	35	17	6	0	1	1	0
Chlamydiae/Verrucomicrobia	13	0	0	0	13	0	0	0	0
Chloroflexi	7	0	4	3	0	0	0	0	0
Cyanobacteia	33	0	0	30	3	0	0	0	0
Deinococcus-Thermus	4	0	4	0	0	0	0	0	0
Delta proteobacteria	19	0	12	6	1	0	0	0	0
Epsilon proteobacteria	20	0	20	0	0	0	0	0	0
Fibrobacter/Acidobacteria	2	0	2	0	0	0	0	0	0
Firmicutes	148	13	125	10	0	0	0	0	0
Fusobacteria	1	0	1	0	0	0	0	0	0
Planctomycetes	1	0	0	0	1	0	0	0	0
Gamma proteobacteria	172	0	158	12	2	0	0	0	0
Spirochaetes	14	0	14	0	0	0	0	0	0
Thermotogae	7	0	5	2	0	0	0	0	0
TOTAL	669	13	467	139	36	5	6	2	1

Figure 1. 12: Distribution of the multiple *cpn60* homologues in all sequenced bacterial genomes.
(Taken from Lund, 2009).

1.1.7 The unusual chaperonins of Mycobacteria

The chaperonins from mycobacteria are interesting for a number of reasons ranging from their oligomeric states to their immunomodulatory roles to even more specialised functions within the cell. Each of these unique characteristics of these chaperonins will be discussed below.

1.1.7.1 Sequence specificity

All the mycobacteria contain at least two copies of the *cpn60* gene while *M. smegmatis* contains three copies of the *cpn60* gene (Ojha *et al*., 2005). Sequence analyses and phylogenetic studies have demonstrated that all the *cpn60.1* genes cluster and form a clade while the *cpn60.2* genes form a separate clade (Goyal *et al*., 2006). This indicates divergence of function. The C-terminal ends of all the mycobacterial Cpn60 proteins

have characteristic repeating motifs (Lund, 2009), with Cpn60.1 proteins displaying a histidine rich C-terminus and Cpn60.2 proteins containing characteristic glycine-glycine-methionine (GGM) repeats. Although the significance of these repeating units is not known, the *E. coli* GroEL protein also has the GGM tail which is very flexible in the crystal structure (Braig *et al*., 1994). One study showed that this GGM tail in *E. coli* GroEL complex was required for association with the lipid membranes, implying a role in folding of soluble and membrane proteins (Torok *et al*., 1997).

1.1.7.2 Oligomeric state

As explained previously, the well characterised *E. coli* GroEL protein has been shown to form a tetradecamer, and along with co-chaperonin GroES (which forms a lid), encloses a hydrophobic cavity within which protein folding takes place. Thus, the double ring structure is of utmost importance in the process of protein folding. In this context, it is intriguing that the chaperonins from *M. tuberculosis* have been reported to only form lower oligomers (Qamra *et al*., 2004). Even when oligomer formation was tested under a range of buffer conditions (ATP, GroES, Cpn10, $(NH_4)_2SO_4$, Mg^{2+}, K^+, glycerol) the formation of higher oligomers was not observed. Since the folding cycle involves the hydrolysis of ATP, the chaperonins from *M. tuberculosis* were tested for their ATPase activity and found to have very low activity (Qamra *et al*., 2004). Even more significant was the finding that the Cpn60.2 protein crystallised as a dimer, further supporting the hypothesis that these proteins exist as unusually low oligomers (Qamra and Mande, 2004). However, contradicting these experiments are the findings that Cpn60.2 can function together with the heptameric GroES and complement for loss of GroEL in *E. coli* (Hu *et al*., 2008). As explained earlier (1.1.5.6), the presence of the

double ring tetrameric structure of GroEL in *E. coli* is crucial for the binding of ATP and heptameric GroES during the protein folding cycle. Thus the strong evidence proving that *M. tuberculosis* Cpn60.2 complements for loss of GroEL in *E. coli* suggests that due to possible instability of the oligomers, the conditions being tested for oligomer formation need to be altered. A part of this study (chapter 6) aims to address if this is also the case with the *M. smegmatis* chaperonins.

1.1.7.3 Role in Immunomodulation

The mycobacterial chaperonins have been shown to play a role in eliciting an immune response. The 65kDa protein isolated from *M. bovis* BCG was found to be highly immunogenic (De Bruyn *et al*., 1987). The Cpn10 protein from *M. tuberculosis* was shown to have T-cell and B-cell epitopes (Baird *et al*., 1989). Further studies conducted on the Cpn10 protein showed that it stimulated bone resorption (Meghji *et al*., 1997) and was secreted in the macrophage phagosome (Fossati *et al*., 2003). The Cpn60.1 protein from *M. tuberculosis* was shown to have antigenic properties (Shinnick *et al*., 1988) and also stimulate human monocytes (Tormay *et al*., 2005). Interestingly, deletion of *cpn60.1* prevented the formation of granulomas in lungs of infected mice (Hu *et al*., 2008). In summary, the Cpn60 proteins from *M. tuberculosis* have shown to elicit pro-inflammatory and anti-inflammatory responses (Qamra *et al*., 2005; Henderson *et al*., 2006; Lewthwaite *et al*., 2007).

1.1.7.4 Role in heat shock response

The role of molecular chaperones in the heat shock response has previously been discussed (1.1.4.4). The heat shock regulation in mycobacteria has previously been explained (1.1.4.5). The expression of the *cpn10*, *cpn60.1* and *cpn60.2* genes from *M. tuberculosis* were shown to be upregulated in a HrcA mutant, confirming its regulatory role in the heat shock response (Stewart *et al*., 2002). Although *M. leprae* posses *dnaK* and *groE* operons which have candidate binding sequences for the HspR and HrcA repressors respectively, neither the DnaK or GroE proteins appeared to be induced upon heat shock (Williams *et al*., 2007). Since the role of extra-cytoplasmic sigma factors (*sigH*) in heat shock and other stress regulation has previously been demonstrated in *M. tuberculosis* and *M. smegmatis* (Fernandes *et al*., 1999; Raman *et al*., 2001), the reason for the lack of heat shock responses in *M. leprae* has been attributed to the fact that the regulatory *sigH* gene is a pseudogene and could explain its poor growth even at temperatures of 37°C (Williams *et al*., 2007). The response of the *M. smegmatis* chaperonins in the heat shock response will be analysed in this study (chapter 3).

1.1.7.5 Specialised functions

***Biofilm formation*:** Biofilm formation was observed in mycobacterial species nearly ten years ago (Hall-Stoodley and Lappin-Scott, 1998). The formation of biofilms by bacteria has been shown to contribute to protection against environmental conditions for example changes in temperature and nutrient availability and also to drug tolerance (Hall-Stoodley and Stoodley, 2009). The *M. tuberculosis* cells present in biofilms have been shown to be drug tolerant leading to the persistent nature of the disease (Ojha *et al*., 2008). In *M. smegmatis*, the absence of one of the chaperonin genes, *cpn60.1*, has

been shown to be responsible for lack of biofilm maturation (Ojha *et al*., 2005), which will be further investigated as part of this study. Furthermore, the reason for the absence of the mature biofilm has been attributed to the association of the Cpn60.1 protein with the KasA protein. As mentioned previously, KasA is a part of the FASII complex, which is responsible for the elongation of the short chain meromycolates during cell wall synthesis. This suggests a link between cell wall structure and chaperonin function. Studies in *M. smegmatis* have shown that *kasA* is essential, while *kasB* is non-essential (Bhatt *et al*., 2005). Further work has since been done and one study demonstrated that isoniazid resistant clinical isolates were found to have mutations in *kasA* (Mdluli *et al*., 1998). Conditional mutants of *kasA* have a dramatic effect on cell survival, and depletion of this protein results in changes in cell morphology leading to cell lysis (Bhatt *et al*., 2005). In *M. tuberculosis*, deletion of the *kasB* gene has been shown to affect mycolic acid synthesis leading to a loss of acid-fast nature of the cells, and more importantly, increased the ability of the cells to persist in immuno-compromised mice (Bhatt *et al*., 2007). The elucidation of the structures of KasA (Lucker *et al*., 2009) and KasB (Sridharan *et al*., 2007) is hoped to aid in the search for potential drug targets due to their critical role in mycolic acid formation during cell wall synthesis (Swanson *et al*., 2009).

***Role of phosphorylation*:** Phosphorylation of proteins has been shown to be important for sensing environmental conditions and in signal transduction. In bacteria, this function is accomplished by a two-component system composed of histidine kinase sensors and their response regulators (Stock *et al*., 1989), however, in eukaryotes, the phosphorylation is mainly observed on serine, threonine or tyrosine residues. Recently,

however, these eukaryotic type serine threonine protein kinases (STPKs) have been shown to play a role in bacterial development, stress responses and pathogenicity (Zhang, 1996). Furthermore, 11 such STPKs were found in the *M. tuberculosis* genome, where their roles ranged from normal cell growth (PknB), to phosphate transport (PknD) to septum formation and glucose transport (PknF), to pathogenesis and survival within the host (PknG and PknH) (Cole *et al*., 1998; Av-Gay and Everett, 2000; Walburger *et al*., 2004; Deol *et al*., 2005; Papavinsasundaram *et al*., 2005). A recent study reported the presence of 18 STPKs in *M. smegmatis* and demonstrated that the over expression of one such STPK, PknF, affected cell structure, sliding motility and biofilm formation (Gopalaswamy *et al*., 2008). As disruption of the *cpn60.1* gene in *M. smegmatis* was shown to abolish biofilm maturation, a possible link between phosphorylation and chaperonin function could be argued. Interestingly, a recent study demonstrated that *M. tuberculosis* Cpn60.1 is a substrate of the STPK PknF, which phosphorylates residues Thr25 and Thr54 (Canova *et al*., 2009). Although the Thr25 residue appears to be conserved among the pathogenic *Mycobacteria* (*M. tuberculosis*, *M. leprae* and *M. ulcerans*), its replacement by an alanine in *M. smegmatis* was responsible for the lack of phosphorylation by PknF (Canova *et al*., 2009). This suggests a possible link between pathogenesis and phosphorylation of Cpn60.1.

***Other functions*:** The *M. tuberculosis* Cpn60.1 protein has recently been shown to bind to single stranded DNA (ssDNA) with high affinity but without any specificity (Basu *et al*., 2009). Furthermore, it has been shown to condense DNA by binding in the minor groove, implying that it may have switched roles of polypeptide chaperoning to polynucleotide chaperoning (Basu *et al*., 2009). While this is a novel function of a

chaperonin protein in bacteria, previous studies have shown that the mitochondrial chaperonin (Hsp60) in the yeast *Saccharomyces cerevisiae* binds to ssDNA with strand specificity (Kaufman *et al*., 2000). This finding was further confirmed when the morphology of the mitochondrial DNA nucleoids was found to be altered in a temperature sensitive *hsp60* mutant (Kaufman *et al*., 2003). An *M. smegmatis lsR2* mutant was found to be defective in biofilm formation (Chen *et al*., 2006), and more importantly, this protein was found to play a role in nucleoid organisation and compactness in *M. tuberculosis* (Chen *et al*., 2008). These findings are very interesting as they suggest a possible link between biofilm formation and nucleoid association for which the Cpn60.1 protein has been shown to play similar roles in *M. smegmatis* and *M. tuberculosis*.

A recent report has also demonstrated that the *M. tuberculosis* Cpn60.2 protein is located in the capsular material where it is responsible for the interaction of the bacteria with the macrophages (Hickey *et al*., 2009).

1.1.8 Aims of the project

The ultimate aim of this project is to lead to a better understanding of the multiple chaperonins in *Mycobacterium smegmatis*. This study is a functional characterisation of the chaperonins of *M. smegmatis*. In this study, the hypothesis that any sequence differences between the three homologues may contribute to a specialisation of function, was investigated. Furthermore, experiments distinguishing between functional specialisation and redundancy were performed. Since large oligomer formation is known to be essential for chaperonin function, it is hypothesised that the chaperonins of *M. smegmatis* that function in *E. coli* would form large oligomers. The various topics addressed in the study are listed below:

- Determine the relationship of multiple *cpn60* genes in *M. smegmatis* to other *cpn60* genes by phylogenetic analysis
- Determine any sequence properties that contribute to the roles of each copy of the *cpn60* homologue in the cell (sequence analyses, homology and pattern searches)
- Analyse their chaperonin function in *M. smegmatis* by monitoring expression levels of the genes under stress conditions (Quantitative Real Time PCR)
- Define the transcription of the *cpn10-cpn60.1* operon (RACE and promoter probe analysis)
- Characterise further the role of Cpn60.1 in biofilm formation and test other Cpn60 homologues for their role in biofilm formation (biofilm assays)
- Characterise the *M. smegmatis* chaperonins in the well studied *E. coli* (cloning, protein expression and complementation studies)

- Analyse oligomer formation of purified *M. smegmatis* Cpn60.1 and Cpn60.2 proteins (recombinant protein purification, native gels, AUC).

2.1 Bacterial strains and plasmids

Table 2. 1: Bacterial strains

Strain	Genotype	Source or Reference
E. coli		
DH5α	F^-, *endA1*, *glnV44*, *thi-1*, *recA1*, *relA1*, *gyrA96*, *deoR*, *nupG*, Φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)*U169*, *hsdR17*(r_K^- m_K^+), λ–	Lab stock (Hanahan, 1983)
MGM100	Modified MG1655 (F^- λ^- ilvG- rfb-50 rph-1) strain in which *groESL* promoter has been replaced by pBAD promoter	(McLennan and Masters, 1998)
SF103	Modified TG1 (*supE*, *hsd*Δ5, *thi*Δ(*lac-proAB*) F^- (*traD36proAB*$^+$, *lacI*^{q} *lacZ*Δ15) which has a *groEL* 44ts mutation	Lab stock
Mycobacterium smegmatis		
mc^2155		(Snapper *et al*., 1990)
mc^2155 Δ*cpn60.1*		(Ojha *et al.*, 2005)

Table 2. 2: **Plasmids**

Plasmid	Description	Source or Reference
pUC18	General purpose cloning vector, Amp^r, *lacZ*	Lab stock
p*trc*99A	General purpose cloning vector with strong *trp-lac* fusion promoter, amp^r, colE1 ori	Amman, 1985
p*trc*ESL	Derived from p*trc*99A, containing *groESL* from *E. coli*	Lab stock
p*trc*ES-SpeI	Derived from p*trc*ES-60.1 tb,	This study

	where the *cpn60.1* gene has been deleted, amp^r, colE1 ori	
pGEMT-easy	Cloning vector, amp^r, f1 ori	Promega
pGEMT-cpn60.1	pGEMT-easy vector containing *cpn60.1* gene	This study
pGEMT-cpn60.2	pGEMT-easy vector containing *cpn60.2* gene	This study
pGEMT-cpn60.3	pGEMT-easy vector containing *cpn60.3* gene	This study
pGEMT-cpn10	pGEMT-easy vector containing *cpn10* gene	Kausar, S
p*trc*ES-cpn60.2	Derived from p*trc*ES-SpeI, containing *cpn60.2* gene	This study
p*trc*ES-cpn60.3	Derived from p*trc*ES-SpeI, containing *cpn60.3* gene	This study
p*trc*10-cpn60.2	Derived from p*trc*ES-cpn60.2, containing *cpn10*	Kausar, S
p*trc*10-groEL	Derived from p*trc*10-cpn60.2, containing *E. coli groEL*	This study
p*trc*10-cpn60.1	Derived from p*trc*10-groEL, containing *cpn60.1*	This study
p*trc*10-cpn60.3	Derived from p*trc*ES-cpn60.3, containing *cpn10*	This study
pBAD24	General cloning vector with a pBAD promoter, amp^r, colE1 ori	(Guzman *et al.*, 1995)
pBAD24-cpn60.1	Derivative of pBAD24 expressing *cpn60.1* from pBAD promoter	This study
pOS2	pBR322 derivative, *E. coli groE* promoter expressing *E. coli groES* and *groEL*	Lab stock
p*groE*-EScpn60.1	Derived from pOS2, *cpn60.1* and *groES* expressed from *groE* promoter	This study

pSD5B	*E. coli - Mycobacterium* shuttle vector containing promoterless lacZ gene, kan^r, P15A ori, oriM	(Jain *et al*., 1997)
pSD5B-SF	Derived from pSD5B, containing 204bp upstream region of *cpn60.1*	This study
pSD5B-LF	Derived from pSD5B, containing 393bp upstream region of *cpn60.1*	This study
pSD5B-cpn10	Derived from pSD5B, containing 369bp upstream region of *cpn10*	This study
pSD5B-cpn60.2	Derived from pSD5B, containing 335bp upstream region of *cpn60.2*	This study
pSD5B-cpn60.3	Derived from pSD5B, containing 335bp upstream region of *cpn60.3*	This study
pMSMEGgroEL1	pMV361 backbone with *cpn10* and *cpn60.1*, kan^r	(Ojha *et al*., 2005)
pMs10-EcgroEL	Derived from pMSMEGgroEL1, containing *E. coli groEL*	This study
pMs10-cpn60.2	Derived from pMSMEGgroEL1, containing *cpn60.2*	This study
pMs10-cpn60.3	Derived from pMSMEGgroEL1, containing *cpn60.3*	This study

Table 2. 3: **Oligonucleotides**

Name	**Sequence**(5'→ 3')	**Description**
Cpn60.1 fwd SpeI	AATACTAGTATGAGCAAGCA GATTGAATTC	Fwd primer to amplify *cpn60.1* from *M. smegmatis*
Cpn60.1 rev SpeI	AATACTAGTCATATGGCGGA TCTTGTC	Rev primer to amplify *cpn60.1* from *M. smegmatis*

Cpn60.2 fwd	ACGACTAGTATGGCTAAGAC AATTGCGTA	Fwd primer to amplify *cpn60.2* from *M. smegmatis*
Cpn60.2 rev	AATACTAGTCAGAGCCACCG ACGAACG	Rev primer to amplify *cpn60.2* from *M. smegmatis*
Cpn60.3 fwd	AATACTAGTATGGCCAAAGA ACTGCGGTTC	Fwd primer to amplify *cpn60.3* from *M. smegmatis*
Cpn60.3 rev	AATACTAGTCCTCCGAGTCCC GGAACG	Rev primer to amplify *cpn60.3* from *M. smegmatis*
p*trc*ES SpeI fwd	CGAAAGCGACATTCTGGCA	Forward sequencing primer for p*Trc*ES-SpeI
pt*rc*ES SpeI Rev	CAC TTC TGA GTT CGG CAT GG	Reverse sequencing primer for p*Trc*ES-SpeI
SP6	CAA GCT ATT TAG GTG ACA CTA TAG	Primer to sequence pGEMTeasy
T7promoter	TAA TAC GAC TCA CTA TAG GG	Primer to sequence T7 promoter in pGEMTeasy
pBAD seq Fwd	GCG GAT CCT ACC TGA CGC	Forward sequencing primer for pBAD24
pBAD seq Rev	GAC CGC TTC TGC GTT CTG	Reverse sequencing primer for pBAD24
Cpn60.1 gsp1	GACGTCGTTGGTCTTGGT G	Reverse primer for 5'RACE of cpn60.1
Cpn60.1 gsp2	GACGGACTTCACCAGCTG	Nested reverse primer for 5'RACE of cpn60.1
Cpn10 gsp2	GTTGTACTTGATCTCGGTG	Nested reverse primer for 5'RACE of cpn10
SF fwd XbaI	GACGTTCTAGAGGGTGACAC CGT	Forward primer to amplify shorter *cpn60.1* upstream fragment
SF rev SphI	TAAGTCTCTTGCATGCGCCTA CG	Reverse primer to amplify *cpn60.1* upstream fragment
LF fwd XbaI	TCCATCGTCTAGAGCGTGAA C	Forward primer to amplify longer *cpn60.1* upstream fragment
Cpn10 fwd XbaI	CGTCGTCATCTAGAACACCG AG	Forward primer to amplify *cpn10* upstream fragment

Cpn10 rev XbaI	CACGCTCGTCTAGATGGAGCCC	Reverse primer to amplify *cpn10* upstream fragment
Cpn60.2 fwd XbaI	GCCATGCAGGTCTAGAACGCCA	Forward primer to amplify *cpn60.2* upstream region
Cpn60.2 rev SphI	TGTCTTAGCGCATGCGAAGTGT	Reverse primer to amplify *cpn60.2* upstream region
Cpn60.3 fwd XbaI	CGTCGGTCTAGAAGGCGCGACC	Forward primer to amplify *cpn60.3* upstream region
Cpn60.3 rev SphI	CTTTGGGCATGCTCGGAGTCC	Reverse primer to amplify *cpn60.3* upstream region

- Underlined bases denote restriction site

Table 2. 4: Primers and probe combinations for real time PCR reactions

Target mRNA	Gene expression assay component	Sequence (5' to 3')
cpn10	probe	6FAM-CTCGCCGTCCTCATCC
	forward primer	CACCGTCGTCGCAGTTG
	reverse primer	GTCACCCTCGGCAACGT
cpn60.1	probe	6FAM-ATCCGGCAGCGAGCTG
	forward primer	GCTGCACCGCGACAAG
	reverse primer	GCCTCGGCGACCTTCTC
cpn60.2	probe	6FAM-TCGCTGCCACCGCCG
	forward primer	AAGGAGGTGGAGACCAAGGA
	reverse primer	GGTCACCGGCGGAGATAC
cpn60.3	probe	6FAM-CATGGCGGTCGACGTC
	forward primer	ACGACCGAGGAGAGTGACA
	reverse primer	GTATCCGTGGTCGAACTCGAT
gapdh	probe	6FAM-ACGCCTCGTGCACCAC
	forward primer	CGGCAGCCAGAACATCATCT
	reverse primer	CGGGCCCAGGCAGTT

2.2 Bacterial Media and Growth Conditions:

2.2.1 Media:

LB (Luria-Bertani) medium (Sambrook, 2001): Tryptone 10g/L, yeast extract 5g/L and NaCl 10g/L in deionised water. For LB agar plates 15g/L of agar was added to the liquid media and then sterilised by autoclaving.

Difco Middlebrook 7H9 broth (BD biosciences): 4.7g was added to 900ml deionised water and sterilised by autoclaving, 100ml Middlebrook ADC enrichment (BBL) and 0.05% Tween80 was added before use.

Difco Middlebrook 7H10 agar (BD biosciences): 19g was added to 900ml deionised water and sterilised by autoclaving, add 100ml Middlebrook ADC enrichment (BBL) and 0.05% Tween80 was added before use.

2.2.2 Media supplements

The media supplements that were used are listed with their abbreviation and working concentration as follows: ampicillin (Amp) at 100μg/ml, kanamycin (Kan) at 50μg/ml (*E. coli*) or 20μg/ml (*M. smegmatis*), hygromycin (Hyg) at 150μg/ml, chloramphenicol (Cmp) at 20μg/ml and tetracycline (Tet) at 10μg/ml. All supplements were supplied by Sigma and filter sterilised using a 0.2μm diameter Acrodisc® syringe filter (Pall Life Sciences). 5-bromo-4chloro-3-indolyl-β-D-galactoside (XGAL) at 20mg/ml in DMF (working concentration 40mg/ml) and isopropyl-β-D-thiogalactopyranoside (IPTG) at 100μg/ml or 0.1M (working concentration 0.1mM) were added to broth or agar where

necessary. Stock solutions of 20% (w/v) arabinose and 20% (w/v) glucose were made and added to LB broth or agar at a final concentration of 0.2% where required.

2.2.3 Growth of bacterial cultures

Small scale (5-10ml) bacterial cultures were prepared in glass universal tubes and larger volumes of culture were prepared in glass conical flasks (up to 1 litre).

E. coli strains: Overnight liquid cultures were grown in LB broth with appropriate media supplements with aeration at 180-200 rpm in an Infors AG CH-4103 shaker incubator (Infors AG, Bottmingen, Switzerland) at 37°C. Plate cultures were incubated overnight at 37°C on LB agar plates containing the suitable supplements.

M. smegmatis: Liquid cultures were grown in Difco Middlebrook 7H9 broth with Middlebrook ADC enrichment and 0.05% Tween80 (BBL) at 37°C for 36 hours with appropriate media supplements with aeration at 180-200 rpm in an Infors AG CH-4103 shaker incubator. Plate cultures were incubated for 48 hours at 37°C on Difco Middlebrook 7H10 with Middlebrook ADC enrichment (BBL) and Tween80 plates containing the suitable supplements.

2.2.4 Monitoring of bacterial growth

Growth of liquid bacterial cultures was monitored by measuring the optical density at a wavelength of 600nm (OD_{600}) using an Ultrospec 2100 pro UV/Visible spectrophotometer.

2.2.5 Strain maintenance

Long-term storage of bacterial strains was carried out by adding 1ml of 80% v/v sterile glycerol to 1 ml aliquots of over night culture and stored at -80°C. Cultures were stored for short-term periods on agar plates at 4°C for up to 2 weeks.

2.3 DNA manipulation and analysis

All enzymes and buffers were supplied by New England Biolabs unless otherwise stated.

2.3.1 Preparation of plasmid DNA

The preparation of plasmid DNA described by (Birnboim and Doly, 1979) from *E. coli* cultures was performed using the Qiagen QiAprep Spin Plasmid Miniprep kit and Midiprep kit for small scale and large-scale preparations respectively, in accordance with the manufacturer's instructions.

2.3.1.1Alkaline lysis method

This is a modified method of plasmid extraction that is carried out in the lab.

Materials required:

Sterile GTE (50mM glucose, 25mM TRIS pH 8.0, 10mM EDTA), 1% SDS/0.2M NaOH (from stock of 10% SDS and 10M NaOH), 3M Na/5M acetate, sterile TE (10mM TRIS pH8.0, 1mM EDTA), Isopropanol, 70% ethanol (in TE) and 100% ethanol.

Method: Cells were harvested from 1-5mls of overnight culture by centrifuging at 13000rpm for 2 minutes and the supernatant discarded. The pellet was then spun again

for 30 seconds and any remaining LB removed using a Gilson pipette. This is an important step as any LB present can inhibit action of restriction enzymes and ligases. The pellet was then resuspended in 200µl GTE by pipetting and 400µl 1%SDS/0.2M NaOH was added to lyse the cells (by inverting the tube several times). Three molar Sodium Acetate (300µl) was then added to the solution which was then inverted several times until a white precipitate was formed. This solution was then incubated for 5 minutes at room temperature and then centrifuged at 13000rpm for a further 5 minutes. The resulting supernatant, which contains the plasmid DNA, was purified by isopropanol precipitation. The supernatant was poured into a clean labelled eppendorf and 600µl isopropanol was added to precipitate the nucleic acids (mixed very thoroughly). This was then spun at 13000rpm for 10 minutes to form a pellet of nucleic acid at the bottom of the eppendorf. The pellet was then washed with 500µl of 70% ethanol and then again with 500µl of 100% ethanol. The pellet was then dried at 37°C for 5-10 minutes and resuspended in 50µl TE to obtain the plasmid DNA.

2.3.2 DNA amplification using the polymerase chain reaction (PCR)

The specific amplification of DNA regions was performed using the polymerase chain reaction. Oligonucleotide primers were designed using an online primer design programme (Intergrated technologies DNA primer design, http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx).

Standard protocol

Reactions were performed in 50µL volumes containing 1.25U *Taq* DNA polymerase (Invitrogen, Bioline or Finnzymes), 0.4ul of dNTP mix at a final concentration of 2.5mM each dNTP (Bioline dNTP mix), 10 x PCR buffer from manufacturer (0.5M

KCl) 25mM $MgCl_2$ (Bioline or Invitrogen), primers were used at a final concentration of 25pmol and 20ng of DNA template or a purified bacterial colony. The PCR cycle was carried out in a Hybaid Gradient Thermocycler (Thermo Hybaid) or the GeneAmp PCr system 9700 (Applied Biosystems) and the cycle is given below:

Step	Temperature	Time	
1. Initial denaturation	**96°C**	**3 minutes**	
2. Denaturation	**96°C**	**30 seconds**	**30 cycles**
3. Annealing	**'x'°C**	**'x' minutes (1min/kb)**	
4. Extension	**72°C**	**1 minute**	
5. Final extension	**72°C**	**5 minutes**	

Colony PCR

The reaction was carried out in 50µl volumes and the same PCR cycle was also maintained. The colony was resuspended in the PCR mix and the PCR performed on it.

2.3.3 Agarose gel electrophoresis

Analysis of DNA samples was done by agarose gel electrophoresis according to (Sambrook, 2001). Agarose gels were prepared by the addition of agarose (SeaKem LE, Flowgen) to 1 X TBE buffer (90mM Tris-Borate pH 8.3, 1mM EDTA) to a final concentration of 1% w/v and heated (Sambrook, 2001). Prior to casting, ethidium bromide was added to a final concentration of 0.5µg/ml or Sybr Safe Green (Invitrogen) at 1µl/100 ml agarose to allow staining of DNA fragments. Typically 10µl samples

were mixed with 1.6µl 6X loading buffer (0.25% bromophenol blue, 0.25% Xylene cyanol FF, 15% Ficoll type 400 [Pharmacia], made up using 10ml sterile water) (Sambrook, 2001) and electrophoresed in 1 X TBE buffer at 95V for 30-45 minutes. Bioline Hyperladder I or IV was used as the marker to size the DNA. UV visualisation of DNA fragments was recorded using a Bio Rad UV GelDoc system (BioRad).

2.3.4 Purification of DNA

The purification of PCR products and restriction digests was carried out using the Qiagen QiAquick PCR Purification Kit in accordance with the manufacturer's instructions. For the purification of DNA from an agarose gel slice the Qiagen QiAquick Gel Extraction kit was used according to the manufacturer's instructions.

2.3.5 Quantification of DNA

DNA samples were quantified using a Nanodrop ND-1000 spectrophotometer at 260/280nm wavelength according to the manufacturer's instructions or estimated by running 1µl of sample on an agarose gel alongside a known concentration of marker DNA.

2.3.6 DNA sequencing and oligonucleotide synthesis

DNA samples were sequenced using the modified dideoxy-chain termination method described by Sanger *et al*. (1977) at the Genomics Lab, University of Birmingham, Birmingham, UK. Oligonucleotides were synthesised by Alta Biosciences, University of Birmingham, Birmingham, UK.

2.3.7 Restriction digestion of DNA

Restriction digestion of DNA was performed in 20µl reaction volumes in the appropriate reaction buffer for 2-3 hours (~ 400ng of plasmid DNA) at the appropriate temperature. Bovine Serum Albumin was added where necessary.

2.3.8 Ligation of DNA fragments

For standard ligations DNA fragments were digested with the appropriate restriction enzymes in order to generate site-specific cohesive ends prior to ligation. DNA fragments were ligated in a reaction volume of 20µl containing approximately a 1:3 molar concentration of vector:insert DNA using 1U T4 DNA Ligase (Bioline Quick Stick Ligase) in 4X Ligase buffer at room temperature for 30 min - 60 min or at 4°C overnight.

2.4 Bacterial Transformations

2.4.1 Preparation and transformation of chemically competent *E. coli*

Reagents required

TfbI:

294.5mg KAc, 121mg $RbCl_2$, 220mg $CaCl_2$, 1g $MnCl_2$ was dissolved in 65 ml sterile water. To the resultant solution, 15 ml of glycerol was added. The pH was then corrected to 5.8 using 0.2M acetic acid. The solution was then made up to 100ml with sterile water and filter sterilised to be stored at -20°C.

TfbII:

209mg MOPS, 1.65g $CaCl_2$, 121mg $RbCl_2$ was dissolved in 60ml sterile water. To this solution 15ml of glycerol was added and the pH fixed to 6.5 using 1M KOH. The solution was then made up to 100ml and filter sterilised to be stored at -20°C.

Method: Chemically competent *E. coli* cells were prepared according to the method performed routinely in the lab. A single colony of *E. coli* was inoculated into 5 ml of LB broth with no antibiotics and incubated at 37°C with shaking overnight. One ml of the overnight culture was added to 100ml of fresh LB broth in a 400ml conical flask and incubated at 37°C with shaking. The growth of the culture was monitored at several time points until the OD_{600} reached 0.4-0.5. The culture was chilled on ice for 5 minutes following which the cells were harvested by centrifuging at 4000rpm for 15 minutes at 4°C. The supernatant was decanted and the cell pellet resuspended in 25ml of TfbI. This was then incubated on ice for 10 minutes and the solution spun down at 5000 rpm for 5 minutes. The resultant pellet was then resuspended in 2ml of TfbII and incubated on ice for 2 hours. Aliquots (100µl) of the cells were made and stored at -80°C.

2.4.2 Heat Shock Transformation

The DNA, competent cells and Eppendorf tubes were pre-chilled on ice. About 1-10ng of plasmid DNA or 5-10µl of ligation reactions were added to the competent cells and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 45 seconds only and immediately placed on ice for 2 minutes. Eight hundred ml of LB media was added to the cells and this was incubated at 37°C with shaking for one hour. The cells were then plated on the appropriate antibiotic plates.

2.4.3 Transformation of *M. smegmatis*

A mid-log culture of *M. smegmatis* was chilled on ice for 10min and the cells harvested by centrifuging at 5000rpm for 15min. The cell pellet was washed thrice with ice cold sterile 10% (v/v) glycerol. If a 100ml starting volume of culture was used, for the first wash, 25ml of 10% glycerol was used, 10ml for the second wash and 5ml for the third wash. Finally the pellet was resuspended in 1ml ice cold 10% glycerol for use immediately or aliquoted and stored at -70°C.

For the transformation, approximately 1µg salt free DNA was added to 400µL of electro-competent cells and kept on ice for 10 min. The cells and DNA were transferred to a 0.2cm electrode gap electroporation cuvet (pre-chilled on ice), placed in an electroporation chamber (Equibio, Geneflow Ltd, UK) and subjected to one pulse of 1.8kV. 1ml of Middlebrook 7H9 broth +ADC was added immediately and the culture incubated at 37°C with shaking for 3 hours. The cells were harvested after incubation and plated on suitable media.

2.5 Protein analysis

The proteins were observed on denaturing Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS PAGE) gels on which they could be separated according to their molecular mass.

Gel composition for a single gel:

Material	10% resolving gel (ml)	5% stacking gel (ml)
Sterile water	4	2.7
30% acrylamide solution	3.3	0.5
1.5M Tris (pH8.8)	2.5	-
1M Tris (pH 6.8)	-	0.67
10% SDS	0.1	0.04
10% ammonium persulphate	0.1	0.04
TEMED	0.004	0.004
Total	10	4

2.5.1 SDS PAGE sample preparation and electrophoresis

Reagents required:

Resuspension buffer: 50mM Tris-Cl, 50mM NaCl, 1mM dithiothreitol (DTT), 1mM EDTA and 50µg/ml protein inhibitor cocktail (PIC). The DTT and PIC are added just before use.

1X Loading buffer (SDS-LB): 500µl of 0.05 M Tris-Cl (pH 6.8) was mixed with 2 ml of 2% SDS, 1 ml of 10% Glycerol and 0.05g of bromophenol blue. The solution was then made up to 10ml in sterile water and 500µl of β-Mercaptoethanol was added.

Electrophoresis buffer: A stock solution of 5X electrophoresis buffer (1L) was made by dissolving 15.1g of Tris Base, 94g of glycine, and 5g of SDS in 1L of sterile water. When running the gels a solution of 1X electrophoresis buffer was used.

Staining solution: A 1 litre stock of staining solution was made with 2.5g of coomassie brilliant blue R250, 450ml of ethanol, 450ml sterile water, and 100ml of glacial acetic acid mixed thoroughly to dissolve the coomassie powder.

Destaining solutions: 300ml Methanol and 100ml glacial acetic acid made up to 1L in sterile water.

Method: 1ml of an overnight culture at OD_{600} of 1 was centrifuged for 5 minutes at 13000rpm. The pellet was washed twice in 1X PBS. The resultant pellet was then resuspended in 75-150µl SDS loading buffer and frozen at -20°C. Before the protein samples were loaded on a gel they were heated in a steamer for 10 minutes and centrifuged for 5 minutes at 13000rpm. The other method for making protein extracts involved resuspending the cell pellet in 1ml of resuspension buffer and sonication (3x 1min, cooling intervals of 10sec) at a setting of 5. Three rounds of sonication were done for each sample or until the culture appeared to clear (indicating cell lysis). A sample of the sonicated cells was loaded on a gel as total lysate, while the rest was spun down at 13000rpm for 10 minutes to give the soluble fraction (supernatant) or the insoluble fraction (pellet). The gel was run at a voltage of 150V until the samples reached the interphase of the two layers; the voltage was then increased to 200V and the gel run for 60-90 minutes. After completion of electrophoresis, the gel was stained over night and then destained twice with destaining solution for 20 minutes. The gel was then rehydrated with water and dried using the Slab gel dryer SGD4050 (Savant) at a temperature for 50°C for 2 hours.

2.5.2 Native PAGE sample preparation and electrophoresis

Reagents required:

Resuspension buffer: 50mM Tris-Cl (pH 7.5), 5mM EDTA, 1mM PMSF, 100µg/ml DNase (from 10mg/ml stock), 500µg/ml lysozyme (from 10mg/ml stock) and 50µg/ml PIC.

5X Native sample buffer: 15ml of 1M Tris-Cl (pH 6.8), 2.5ml of 1% solution of bromophenol blue, 7ml of distilled water and 25ml of glycerol.

5X Native gel running buffer: 15.1g Tris base, 94g glycine. The components were dissolved and the volume made up to 1 litre.

Gel composition for a single gel:

Material	10% resolving gel (ml)	5% stacking gel (ml)
Sterile water	1.4	2.15
30% acrylamide solution	1.25	.85
1.5M Tris (pH8.8)	1.25	-
1M Tris (pH 6.8)	-	0.625
80% glycerol	1	1.25
10% ammonium persulphate	0.05	0.05
TEMED	0.004	0.005
Total	5	5

Method: Cells from 1ml of overnight culture were harvested by centrifugation at 13000rpm for 5 minutes. The resulting cell pellet was resuspended in 120µL resuspension buffer and allowed to stand at room temperature for 10 minutes. The cell solution was then sonicated (3x 1min, cooling intervals of 10sec). The sonicated

samples were then centrifuged at 13000rpm for 10 minutes and the supernatant transferred to a new eppendorf tube and a calculated volume of sample buffer was then added. The gel was run at a voltage of 220V for 5-6 hours. After completion of electrophoresis, the gel was stained over night and then destained twice with destaining solution for 20 minutes.

2.5.3 Western blotting

Reagents required:

Transfer Buffer (1L): 2.21g (10mM) of CAPS in 10% methanol made up to 1L with sterile water. The pH was adjusted to 11 using NaOH prior to the addition of methanol. The solution was stored in the fridge.

10X PBS: 2g of KH_2PO_4 was added to 14.48g of Na_2HPO_4, 80g of NaCl, and 2g of KCl before being dissolved in 1L of sterile water. The solution was then autoclaved and stored at room temperature.

TPBS: 0.1% of Tween 20 was added to 1X PBS to make TPBS.

TBS: 10mM Tris-HCl and 0.9% NaCl were made up to 1L in sterile water and sterilised before storing it at room temperature.

Blocking solution: 5% non fat milk powder was mixed with TPBS.

Diluted primary antibody solution: A dilution of 1:1000 of primary antibody in blocking solution was made and stored in the fridge until required.

Diluted secondary antibody: A dilution of 1:10000 of secondary antibody (anti-mouse HRP conjugate from Amersham ECL plus Western Blotting reagent pack, [GE Healthcare]) was added to the blocking solution.

Method: Four sheets of Whatman MM3 paper were cut slightly larger than the size of the protein gel. The PVDF membrane was also cut to the same size as the protein gel, and the upper right corner marked. The PVDF membranes were prewet in absolute methanol and incubated in the transfer buffer. Two sheets of Whatman paper were wet in the transfer buffer and laid on one of the fiber pads, followed by the wet membrane on it, and the gel rinsed in transfer buffer was placed on it. The other two sheets of wetted Whatman paper were laid over the gel and the remaining fiber pad finally placed on the top. The whole sandwich was put between the gel holder cassette with the membrane to the white plastic cover and the gel to the black cover. Pre-chilled transfer buffer was poured into the tank and put in the cooling unit. Finally the gel holder with the sandwich was placed in the tank with the white cover facing the white side of the inner frame containing the electrodes. In this case the membrane is facing the positive electrode (red) and the proteins are running towards that direction. The tank is connected to the power supply and the parameters set. For fast transfer of higher molecular weight region 200mA current was used for 70 min with the mini apparatus and 200mA for 5 hours with the big apparatus. When finished, the membrane was removed.

Antibody binding and detection

The membrane was placed in a tray with the protein containing side facing up and rinsed with TPBS for 30 minutes. The membrane was then incubated in the blocking solution (5% milk powder in TPBS) overnight. The next morning, the membrane was washed thrice with TPBS and incubated in the primary antibody solution (diluted in the blocking solution) for one hour room temperature. The membrane was then washed 3 x

15 minutes in TPBS and incubated in the secondary antibody solution diluted 1:10000 (in blocking solution) for one hour. The membrane was then washed 2 x 15 min in TPBS followed by washing in TBS for 5 minutes. The ECL developing solution was prepared by mixing solutions A and B in a 40:1 ratio (3ml solution A and 0.075ml solution B). The excess liquid was removed from the membrane and laid on a piece of aluminium foil, protein containing side up. The developing solution was then evenly laid on the surface of the membrane and this was incubated for 5 minutes. The excess liquid was removed and the membrane wrapped in cling film. The membrane was exposed to X-ray film for one min and the film developed. The exposure times are adjusted to improve the strength of the signal.

2.6 Complementation Analysis (Spot Test)

The required strains were grown in 5ml LB broth (either with only 0.2% arabinose or 0.2% arabinose and 0.1mM IPTG) overnight at 37°C. Three ml of the over night culture was taken and centrifuged at 13000rpm for 10 minutes. The pellet was then washed twice with LB broth before resuspending in 1ml of the same media. The OD_{600} of this culture was measured and then adjusted to 1 by diluting cultures with the LB media. A 96 well plate was used to make the serial dilutions. In the first well, 200µl of each strain at OD_{600} 1 was added and ten fold serial dilutions made up to a dilution of 10^{-6} or 10^{-7}. A volume of 5-10µl was then spotted onto a labelled square plate containing LB agar with agarose or glucose with and without IPTG and allowed to dry. Once the spots had dried, the plates were incubated at various temperatures (30°C, 37°C, and 42°C) to test for complementation at those temperatures.

2.7 RNA Extraction of *M. smegmatis*

Reagents required

5M GTC solution: 5M guanidinium isothiocyanate [Fluka; ultrapure grade], 0.5% sodium N-lauryl sarcosine, 25mM tri-sodium citrate, pH 7.0, 0.1M 2-mercaptoethanol, 0.5% Tween 80.

Method: A 10ml mid-log culture of *M. smegmatis* (grown in LB broth with 0.05% Tween80) was mixed with 4 volumes of 5M GTC lysis solution and mixed rapidly by swirling. This produces a final concentration of 4M GTC. This solution was then centrifuged at 3000rpm for 30 minutes to pellet the cells. The supernatant was discarded carefully and the pellet resuspended in 1 ml of 4M GTC solution. This 1ml of culture in GTC solution was transferred to a 2ml tube and centrifuged for a minute at 13000rpm. The excess GTC solution was poured off and the pellet resuspended in 1.2mL of TRI reagent (Fluka Biochemicals) and added to 2 ml screw-cap skirted microcentrifuge tubes containing 0.5ml of 0.1mm ceramic beads. The tubes were spun using a reciprocal shaker (Hybaid Ribolyser) at maximum speed setting (6.5) for 45 seconds. The tubes were removed from the Ribolyser and left at room temperature for 10 minutes. Chloroform (200µl) was added to the tube and vortexed for 30 seconds. The tubes were then left at room temperature for 10 minutes to partition the aqueous and phenolic phases and then centrifuged at 13000rpm at 4°C for 15 minutes. The upper phase was transferred to a fresh tube and an equal volume of chloroform added to it. The tubes were vortexed again and incubated at room temperature for 10 minutes before centrifuging at 13000rpm at 4°C 15 minutes. The upper aqueous phase was transferred to a new tube and 0.8 volume of isopropanol added to precipitate the nucleic acid. This

was incubated at -20°C overnight. The next day, the RNA was pelleted at 13,000rpm in a microcentrifuge for 20 minutes at 4°C. The supernatant was discarded and the pellet washed with 1ml of 70% ethanol and re-pelleted. The residual ethanol is removed with a pipette and the pellet briefly air-dried before being resuspended in 100µl of RNase free water. The resuspended RNA was then purified using the Qiagen RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. The pure RNA was stored at -80°C.

The RNA from above was once again DNase treated using the Ambion turbo free DNA kit according to manufacturer's instructions.

2.8 cDNA synthesis

The DNase treated RNA samples were used to synthesise cDNA using the High capacity cDNA reverse transcription kit (Applied Biosystems) according to manufacturer's instructions. A total of ~1.2 - 1.5µg of RNA was used in a 20µL reaction in all cases and the cDNA synthesised. The PCR cycle was carried out in a Hybaid Gradient Thermocycler (Thermo Hybaid) or the GeneAmp PCR system 9700 (Applied Biosystems) and the cycle is given below:

Temperature	Time
25°C	**10 minutes**
37°C	**120 minutes**
85°C	**5 seconds**
4°C	**store**

2.9 Quantitative Real Time PCR

Quantitative real time PCR was performed using the custom made Taqman gene expression assays (Applied biosystems). A total of 60ng of cDNA was used in each 25µL reaction.

Standard protocol

Reactions were performed in 20µL volumes containing 10µL 2x Taqman gene expression mastermix (AmpliTaq Gold DNA Polymerase, Ultra Pure, UDG, dNTPs with dUTP, ROX^{TM} passive reference and optimised buffer components), 1µL target specific Custom Gene Expression Assay (250nm probe and 900nM each primer) and 9µL cDNA (60ng). The real time PCR cycle was carried out in an ABI Prism 7000 Sequence Detection System (Applied Biosystems) and the cycle is given below:

Step	Temperature	Time	
1. Incubation	50°C	2 minutes	
2. Enzyme activation	95°C	10 minutes	
3. PCR-denaturation	95°C	15 seconds	40 cycles
4. PCR-annealing/extension	60°C	1 minute	

2.10 Rapid Amplification of cDNA ends (5'RACE)

RNA was extracted from mid-log cultures of *M. smegmatis* as described above (2.7) and the 5'RACE system for rapid amplification of cDNA ends (Version 2.0, Invitrogen) was done according to manufacturers instructions, using primers cpn60.1 gsp1, cpn60.1

gsp2 and cpn10 gsp2. cDNA was tailed at the 5' ends using poly-cytosine and transcriptional start sites were identified by detection of the junction of this poly-C tail in the sequenced cDNA.

2.11 Purification of Cpn60.1 protein

The presence of a histidine rich region at the C-terminal end of the Cpn60.2 protein, made it useful to purify the protein using a Nickel affinity purification column.

Buffers

Lysis buffer: 50mM Tris (pH6.5), 50mM NaCl, 1mM DTT, 1mM EDTA, PIC 1ml /200ml cell volume.

1X Binding buffer: 50mM NaCl, 50mM Tris pH 7.9

1X Wash buffer: 150mM NaCl, 50mM Tris pH 7.9, 20mM imidazole

1X Elution buffer: 150mM NaCl, 50mM Tris pH 7.9, 200mM imidazole

Cell lysate preparation

A single colony of *E. coli* containing pBAD24-60.1 was inoculated into 5 ml of LB broth with Amp and incubated at 37°C with shaking overnight. An overnight culture (4ml) was added to 400ml of fresh LB broth in a 1000ml conical flask and incubated at 37°C with shaking (180rpm). The growth of the culture was monitored at several time points until the OD_{600} reached 0.4-0.5. The culture was induced with 0.2% arabinose for 4 hours. The cells were harvested by centrifuging at 4000rpm for 15 minutes at 4°C. The supernatant was decanted and the cell pellet resuspended in 2ml of lysis buffer. For cell lysate preparation, the resuspended pellet was sonicated (3x 1min, cooling intervals of 10sec). The samples were then centrifuged at 12000g for 10mins and the

supernatants centrifuged once again at 100000g for 1 hour (Beckman J2-21M/E centrifuge, JLA20 rotor).

Affinity purification chromatography

The His-Bind resin (Novagen, 3ml) was connected to Bio-Rad Biologic LP HPLC system and equilibrated with 1x binding buffer (10 fold bed volume). The crude cell extract was applied to the column and allowed to flow through by gravitational force. The column was then washed with 5 fold 1x binding buffer followed by a wash with 1x wash buffer (15ml). Proteins were then eluted with 1x elution buffer (15ml). The collected samples were checked on a 12% SDS PAGE.

Hydrophobic column

Buffers

Buffer 1 (high salt buffer): 0.5M $(NH_4)_2SO_4$, 50mM KPi (KH_2PO_4 and K_2HPO_4)

Buffer 2 (eluting buffer): 50mM KPi

The hydrophobic column (butyl agarose, 25ml, GE Healthcare) was connected to Bio-Rad Biologic LP HPLC system and the column equilibrated with 10 fold bed volume (250ml) buffer 1. The pooled samples from the last step were loaded onto the column and allowed to flow through the column by gravitational force. The column was washed with 5 fold bed volume (125ml) buffer 1. Proteins were eluted by using a linear gradient (10 fold bed volume) with decreasing ionic strength (100% 0.5M $(NH_4)_2SO_4$ to 0% $(NH_4)_2SO_4$). The column was then cleaned with water. Protein samples were checked on 12% SDS PAGE and collected. The purified proteins were concentrated using Vivaspin 5K cut concentrators (Vivascience) and stored at -80°C for further use.

2.12 Purification of Cpn60.2 protein

The purification of the Cpn60.2 protein involved a number of stages which will be discussed below.

DEAE sepharose ion exchange chromatography

Buffers

Buffer 1 (low salt buffer): 50mM Tris (pH 6.5), 50mM NaCl, and 1mM EDTA

Buffer 2 (high salt buffer): 50mM Tris (pH 6.5), 500mM NaCl and 1mM EDTA

Cell lysate preparation

A single colony of *E. coli* containing p*trc*ES-60.2 was inoculated into 5 ml of LB broth with Amp and incubated at 37°C with shaking overnight (180rpm). An overnight culture (4ml) was added to 400ml of fresh LB broth in a 1000ml conical flask and incubated at 37°C with shaking. The growth of the culture was monitored at several time points until the OD_{600} reached 0.4-0.5. The culture was induced with 0.1mM IPTG for 3 hours. The cells were harvested by centrifuging at 4000rpm for 15 minutes at 4°C. The supernatant was decanted and the cell pellet resuspended in 2ml of lysis buffer. For cell lysate preparation, the resuspended pellet was sonicated (3x 1min, cooling intervals of 10sec). The samples were then centrifuged at 12000g for 10mins and the supernatants centrifuged once again at 100000g for 1 hour (Beckman J2-21M/E centrifuge, JLA20 rotor).

Chromatography

The DEAE sepharose column (GE Healthcare, 1cm x 20cm, 12ml) was connected to Bio-Rad Biologic LP HPLC system and equilibrated with 250ml buffer 1 before use.

The crude cell extract was applied to the column followed by 5 fold bed volume (125ml) of buffer 1. Proteins were eluted by using a linear gradient (10 fold bed volume) with increasing ionic concentration from buffer 1 to buffer 2 (0% - 100%). The column then was washed with ~72ml buffer 2. The column was cleaned using 1M NaCl buffer. Protein samples were checked on 12% SDS PAGE and collected. The purified proteins were concentrated using Vivaspin 5K cut concentrators (Vivascience).

Hydrophobic column

Buffers

Buffer 1 (high salt buffer): 0.5M $(NH_4)_2SO_4$, 50mM KPi

Buffer 2 (eluting buffer): 50mM KPi

The hydrophobic column (butyl agarose, 25ml, GE Healthcare) was connected to Bio-Rad Biologic LP HPLC system and the column equilibrated with 10 fold bed volume (250ml) buffer 1. The pooled samples from the last step were loaded onto the column and allowed to flow through the column by gravitational force. The column was washed with 5 fold bed volume (125ml) buffer 1. Proteins were eluted by using a linear gradient (10 fold bed volume) with decreasing ionic strength (100% 0.5M $(NH_4)_2SO_4$ to 0% $(NH_4)_2SO_4$). The column was then cleaned with water. Protein samples were checked on 12% SDS PAGE and collected. The purified proteins were concentrated using Vivaspin 5K cut concentrators (Vivascience).

Size exclusion chromatography

Buffers

Buffer 1: 50mM Tris, 50mM KCl, 10mM $MgCl_2$ pH7.5

Chromatography

The size exclusion column (Superdex 200, 35ml, GE Healthcare) was connected to a Bio-Rad Biologic LP HPLC system and equilibrated with 250ml buffer 1 before use. Protein samples obtained from the previous step were applied to the column and eluted using buffer 1 at a flow rate of 1ml/min. The collected sample was analysed on 12% SDS PAGE and concentrated using Vivaspin 5K cut concentrators (Vivascience).

2.13 Cell free extracts from *M. smegmatis*

A 3ml culture of LB was inoculated with a loop full of *M. smegmatis* and grown at 37°C in a shaker (170rpm) for 48 hours. This was then sub-cultured into 50ml LB with 0.05% Tween80 and grown at the same conditions overnight. Cell-free extracts were then made according to the protocol by Papavinasasundaram *et al*. (2001).

The overnight 50ml culture was harvested by centrifuging at 4°C at 4000rpm for 15 minutes. The resulting pellet was then washed with 20ml of 1X PBS with protease inhibitor cocktail (50µg/ml) by centrifuging at 4°C at 4000rpm for 5 minutes. The supernatant was discarded and the pellet washed again with 10ml of 1X PBS for 3 minutes and the resulting pellet was then resuspended in 400µl 1X PBS. Eppendorf tubes were filled 1/3 with glass beads (unwashed, 212-300µm, Sigma) and the 400µl culture from the previous step was put into the tubes. The tubes were incubated on ice for 5 minutes and the cells then lysed in the reciprocal shaker (Hybaid Ribolyser) at a speed of 6.5 for 30 seconds after which the tubes were placed on ice for a minute before being disrupted in the ribolyser once again for 30 seconds. The tubes were then centrifuged at 13000rpm for 2 minutes and the supernatant removed and put into a new

tube. This was then centrifuged once again for 2 minutes and the cell free extracts (supernatant) moved to a new tube. The cell free extracts were then stored at -80°C.

2.14 Bio-Rad protein assay

This assay to quantify protein samples is based on the Bradford assay (Bradford, 1976). For this experiment, a standard protein stock was made from BSA (10mg/ml) supplied by NEB. Ten-fold dilutions (using 1X PBS) were made to obtain final working concentrations of 100µg/ml, 10µg/ml and 1µg/ml of BSA. A series of known concentrations of BSA were then made ranging from 0.1µg/ml - 100µg/ml by adding appropriate volumes of the BSA stock solution and 1X PBS. The assay was performed in a 96-well microtitre plate. A volume of 160µl of each BSA standard was put into a well along with 40µl of Dye reagent (BioRad protein assay kit). These samples were made in triplicate. The plate was then incubated at 37°C for 30minutes and the absorbance at 595nm then measured in a plate reader (Labsystems Multiscan MS). A graph was then plotted to produce a standard curve. When measuring concentration of samples, 10µl of the sample and 200µl of the dye reagent was added to each well and the absorbance measured. The samples were measured in triplicate. The average value for each sample was then plotted on the standard graph to determine its concentration.

2.15 Biofilm assay

Reagents required

Biofilm base media:

13.6g KH_2PO_4, 2g $(NH_4)_2SO_4$ adjust the pH to 7.2 using KOH pellets and add 0.5mg $FeSO_4.7H_2O$, 5g Casamino acid. Autoclave for 20min at 121psi, and store at room temperature till required.

Biofilm media:

94ml Biofilm base media, 5ml of 40% glucose, 1ml 0.1M $CaCl_2$, 100µl of 1M $MgSO_4$.

A single colony of each of the strains of *M. smegmatis* was inoculated into 3ml Difco Middlebrook 7H9 broth + ADC + 0.05% Tween80 + appropriate antibiotics and incubated at 37°C with shaking for 24-36 hours. 10ml of Biofilm media was aliquoted into small Petri dishes (60mm X 15mm). To the aliquoted biofilm media, 10µl of the culture was added and kept stationary at 30°C.

2.16 β-galactosidase Assay

Reagents required

Z buffer:

12ml 0.5M Na_2HPO_4, 8ml 0.5M NaH_2PO_4, 1ml 1M KCl, 100µL 1M $MgSO_4$, 270µL β-mercaptoethanol made up to 100ml with sterile distilled water. This buffer should be made up fresh every week.

ONPG: 4mg/ml stock in Z buffer

1M Na_2CO_3, 0.1% SDS and chloroform.

A single colony of each strain of *M. smegmatis* harboring the pSD5B construct was grown in 3ml 7H9 Middlebrook 7H9 broth + ADC + 0.05% Tween80 + appropriate antibiotics and incubated at 37°C with shaking for 24-36 hours. Tubes containing 900µL Z buffer were incubated at 30°C. The OD_{600} of the cultures was measured and 100µL of culture was added to the tubes containing Z buffer. Blanks containing only Z buffer or Z buffer + media were also set up. One drop each of 0.1% SDS and chloroform was added to the tubes and tubes vortexed well to lyse the cells. The reaction was started by adding 200µL of ONPG and the tubes were shaken. A timer was started at this point. When a significant yellow colour had developed, the reactions were stopped by adding 500µL of 1M Na_2CO_3. One ml of solution from each tube was taken and centrifuged in eppendorf tubes at maximum speed for 5 minutes to pellet any cells. The OD_{420} of the supernatant was measured after blanking with Z buffer.

The promoter activity was calculated using the formula:

$$\frac{OD_{420}}{OD_{600} \text{ x culture vol (ml) x time of incubation (min)}}$$

2.17 Colony morphology of *Mycobacterium smegmatis*

The *M. smegmatis* strains to be tested were grown in Middlebrook 7H9 broth with 10% ADC and 0.05% Tween80 and required antibiotics and at 37°C in a shaking incubator for 36 hours. The absorbance of the cultures was measured at OD_{600} and the cultures were normalised to an OD_{600} of 1. 5µL of the test cultures were then spotted onto media (7H10 agar, 7H11 agar, Tryptic Soy agar and LB agar) with or without Tween80 and the plates incubated at 37°C for five days. Colony morphology was analysed using a Nikon SMZ-U microscope with a magnification of 0.75X.

3.1 Introduction

A survey of nearly 669 sequenced genomes revealed that nearly 30% contain two or more chaperonin genes (Lund, 2009). So why is the presence of multiple chaperonins interesting? As described earlier in chapter 1, the *E. coli* GroEL protein is capable of folding a wide range of substrates (Houry *et al*., 1999; Kerner *et al*., 2005; Chapman *et al*., 2006). Hence, if a single copy of GroEL together with GroES can fold partially folded or unfolded proteins, then what is the purpose of the multiple copies of chaperonins in certain bacteria? Two possibilities exist, one of which is that the reason for multiple chaperonins is to increase chaperoning activity in a cell, while the second reason is that each of the duplicated genes may have evolved to take on a more specialised function in the cell. However, the actual situation might be intermediate of the two possibilities. A number of previous studies in bacteria which have multiple copies of *cpn60* genes have shown that at least one copy is essential. Studies analysing the multiple chaperonins of *Rhizobium leguminosarum* have shown that out of the three copies of *cpn10-cpn60*, the presence of one copy in the cell is essential (Rodriguez-Quinones *et al.*, 2005; Wallington and Lund, 1994). In *Bradyrhizobium japonicum*, while all of the five *cpn60* genes are individually indispensable (along with either *cpn60.1* or *cpn60.2*), double knock-outs cause a reduction in nitrogen fixation (Fischer, 1999). Moreover, even in the archaeon *Haloferax volcanii*, while two of the *cpn60* genes (*cct1* or *cct2*) are individually dispensable for growth, the presence of at least one of them in the cell is essential and the third copy (*cct3*) is completely dispensable and does not support growth on its own (Kapatai *et al*., 2006). These examples imply that chaperonin function is essential in these organisms which have multiple *cpn60* genes (since at least one of the multiple *cpn60* gene is always present in the cell) and also

show that in most cases, the presence of any of the multiple *cpn60* genes on their own cannot support growth, only a few of them are capable of doing so.

Most members from the Actinomycete group of bacteria have been found to contain at least two copies of *cpn60* genes (Lund, 2009). Almost all mycobacteria are known to have at least two copies of the *cpn60* genes, with one serving as the housekeeping chaperonin, while the other chaperonin appears to have a specialised function (Lund, 2009, Hu *et al*., 2008, Ojha *et al*., 2005). *M. smegmatis* is unique among mycobacteria for which a complete genome sequence is available to date, in that it contains three copies of chaperonin genes (Ojha *et al*., 2005). Previous studies have shown that *M. smegmatis cpn60.1* is not essential, but plays a role in biofilm maturation, indicating that each gene probably has a specific role in the cell (Ojha *et al*., 2005).

Previous phylogenetic studies have shown that when there are two copies of the genes, namely *cpn60.1* and *cpn60.2*, only one copy has a genetically linked co-chaperonin *cpn10* present. Only one *cpn10* gene has been reported in all Actinomycetes. In cases where there is only a single copy of the *cpn60* gene, it has always been noticed that the *cpn10* gene is present elsewhere on the genome, as is the case in *Tropheryma whipplei* (Maiwald *et al*., 2003). This finding implies that a common ancestor containing both copies existed, and due to functional redundancy, one copy was lost while the essential one (not found with *cpn10*) remained.

While the Cpn60.1 and Cpn60.2 proteins from *M. tuberculosis* and *M. smegmatis* share a similarity of around 90%, the Cpn60.3 protein is no more than 50% similar to any of the *M. smegmatis* or *M. tuberculosis* homologues. Furthermore, none of the three homologues share more than 55% identity with the well characterised *E. coli* GroEL protein.

In order to characterise these chaperonins based on their expression, studies were conducted in *M. smegmatis* as well as in *E. coli*. The first two chapters of this thesis will focus on the studies conducted in *M. smegmatis*, while the later two chapters will discuss the results obtained from experiments conducted in *E. coli*.

In this chapter, a few questions will be answered, which will lead to a better understanding of the *M. smegmatis* chaperonins. First, what are the sequence differences between the three homologues and second, do these sequences differences contribute to a specialisation of function of the proteins they encode? The first question was answered by analysing gene sequences, searching for regulatory regions and looking at the phylogeny as well. To begin to test the hypothesis that any sequence differences between the *cpn60* genes might indicate a specialisation in function, the expression of the *cpn60* and *cpn10* under a range of different stress conditions was monitored. By answering these questions we aim to establish the reason for the presence of three copies of the same gene.

The heat shock response is characterised by global transcriptional changes with the increase in expression of certain classes of genes such as the molecular chaperones

(Hartl *et al*., 1996). All bacteria require molecular chaperones to fold unfolded or misfolded proteins. The *E. coli* chaperonin GroEL serves as the paradigm for the study of these proteins. Not only is the *E. coli* chaperonin GroEL essential for growth at higher temperatures, it has also been shown to be required for growth at all temperatures (Fayet *et al*., 1989). Bacterial stress responses are regulated by various systems. In *E. coli* the response is induced by the transcriptional activation of the alternative sigma factor, the σ^{32} subunit of RNA polymerase that is encoded by the *rpoH* gene (Bukau, 1993), while in *B. subtilis*, the heat shock response is regulated by transcriptional repression (Yura and Nakahigashi, 1999). In particular, the heat shock response of Gram-positive bacteria can be broadly divided into three classes, Class I genes include genes whose expression is negatively regulated by the repressor protein HrcA, Class II genes comprise genes whose expression is positively regulated by the general stress sigma factor, σ^{B} and Class III genes include all the remaining heat shock inducible genes by some other regulatory mechanism (Yura and Nakahigashi, 1999).

Baird *et al*. (1989) were the first to identify an inverted repeat sequence upstream of the mycobacterial *groE* genes. This element has since been identified in a number of bacterial species and has a consensus sequence, CTAGCACTC-N9-GAGTGCTAG, named CIRCE (controlling inverted repeats of chaperone expression). This element has been characterised in *B. subtilis* by Zuber and Schumann and is associated with the *groE* and *dnaK* operons (Zuber and Schumann, 1994). Subsequently, HrcA (heat regulation at CIRCE) was identified as the repressor protein (Roberts *et al*., 1996). Studies of this regulation in *B. subtilis* have shown that newly synthesised HrcA is inactive and must interact with the GroESL chaperonin system to bring about

conformational changes so that it can bind to the CIRCE region. Under stress conditions, the chaperonins are titrated by non-native proteins and the inactive HrcA cannot bind to the CIRCE regions located upstream of the *groE* operon resulting in transcription of these genes. Once the cell has recovered, HrcA associates with the promoter regions thus preventing the expression of the two heat shock proteins (Hecker *et al*., 1996; Mogk *et al*., 1997). A search for CIRCE sequences in the *M. tuberculosis* genome revealed the presence of two such sequences upstream of *cpn60.2* and one sequence upstream of *cpn10*. Furthermore, the levels of the *cpn10*, *cpn60.1* and *cpn60.2* genes have been shown to be upregulated in a HrcA mutant (Stewart *et al*., 2002).

Another well known regulator of the stress response is the regulatory protein HspR, which binds to a sequence known as HAIR (HspR-associated inverted repeats) (Bucca *et al*., 1995; Grandvalet, 1999). Extensive studies of this regulatory system have been conducted in *Streptomyces coelicolor* and found to mainly involve regulation of the DnaJ/DnaK chaperones (Bucca *et al*., 2000; Bucca *et al*., 2003). Furthermore, from studies of an *M. tuberculosis* HspR mutant, it appeared that only the DnaJ/DnaK genes and ATPase ClpB were controlled by this protein, while the chaperonin genes appeared unaffected in the mutant (Stewart *et al*., 2001). As there is no homologue of the *rpoH* gene in *M. smegmatis*, it is likely that the heat shock response is regulated by other systems, possibly the HrcA/controlling inverted repeat of chaperone expression (CIRCE) or the HspR/HspR-associated inverted repeat (HAIR) systems as is the case in *M. tuberculosis* (Stewart *et al*., 2001; Stewart *et al*., 2002).

To begin the study, simple bioinformatic analyses to trace the phylogeny of these three *cpn60* genes, along with a search of regulatory sequences was performed. This was then followed by a study aimed at characterising gene expression under stress conditions. Finally, preliminary experiments to determine the 5' ends of the *cpn60.1* and *cpn10* mRNA were performed.

3.2 Experimental approach

To look at changes in expression levels of the three *cpn60* genes and the co-chaperonin *cpn10*, quantitative real time PCR (qRT-PCR) was used. Briefly, cells were grown up to mid-log phase and stressed for different periods of time using various stress treatments. Samples were collected and RNA extracted for further analysis. Since equal amounts of cDNA was used in each reaction to calculate the fold change of expression levels of each gene, $2^{-\Delta\Delta C}{}_T$ method was used to calculate fold change in expression levels of each gene (Livak and Schmittgen, 2001). The control gene chosen was GAPDH. A sample of the stressed culture was plated to check for cell viability after the maximum period of stress in each case.

Quantitative real time PCR works on the principle of FRET (Förster Resonance Energy Transfer or Fluorescence Resonance Energy Transfer), where fluorescence is generated due to the coupling of a fluorescent reporter dye and a quencher molecule present on the same or separate oligonucleotides. The probes chosen for this study were the TaqMan probes, which are oligonucleotides designed to bind specifically to the minor groove region on the target gene. At the 5' end of each probe is a FAM™ fluorescent reporter dye, while at the 3' end is a quencher molecule. When the probe is bound to the target

strand, the fluorescence is quenched due to the proximity of the reporter dye to the quencher. However, during amplification of the target strand, the 5' nuclease activity of DNA polymerase causes a cleavage of the probe. This cleavage results in an increase in fluorescence in each cycle, which is proportional to the amount of probe cleaved. The relative levels of target nucleic acid present in the sample at the start of the reaction are inferred by measuring the number of PCR cycles it takes for fluorescence levels to accumulate above a critical threshold.

3.3 Results

3.3.1 Bioinformatic analysis

To begin the analysis, *cpn60* genes from all the available sequenced mycobacterial genomes were aligned. For this purpose, sequences were obtained from xBASE (http://xbase.bham.ac.uk/) and aligned using the default settings on clustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The nucleotide sequences were translated using the translate tool on the Expassy website (http://www.expasy.ch/tools/dna.html). The alignment results gave early indications that there may be a high degree of similarity among all the Cpn60.1 proteins and Cpn60.2 proteins (figure 3.1). Next, a phylogenetic tree of the *cpn60* genes from all the available sequenced mycobacterial genomes was constructed. Nucleotide sequences were extracted from xBASE and aligned using ClustalW with output set to gcg MSF. A phylogenetic tree was drawn using Phylip, using UPGMA for clustering. From the result in figure 3.2, it can be observed that all the *cpn60.1* genes cluster in one clade, while the *cpn60.2* genes cluster together and the *cpn60.3* appears to form an outgroup on its own.

These findings imply that an ancient duplication event occurred and has been retained (Goyal *et al*., 2006) which could further imply that the roles of the Cpn60.1 and Cpn60.2 proteins in the cell might be conserved among the mycobacteria. This is similar to the case with the Cyanobacterial chaperonins where the multiple *cpn60* genes cluster into two distinct clades (Lund, 2009). Furthermore, in the Cyanobacteria *Anabena* L-31, differential expression levels between the two *cpn60* genes have been linked with thermotolerance and nitrogen status in the cell (Rajaram and Apte, 2008). Thus it appears that a functional differentiation between the multiple copies of *cpn60* genes in a cell could be due to subtle sequence variations. Interestingly, when the C-terminal end of the mycobacterial chaperonins was observed, it was apparent that all the Cpn60.1 homologues contained a histidine rich tail, while the Cpn60.2 proteins contained repeats of glycine-glycine-methionine (GGM) (figure 3.1). Although the significance of these specific repeating regions is not known, this repeating sequence is also found in other *cpn60* sequences and in many eukaryotic HSP60 members (Brocchieri and Karlin, 2000). The extensively studied *E. coli* GroEL protein has four repeats of this Gly-Gly-Met tripeptide which has been shown to be very flexible in the crystal structure of the protein (Braig *et al*., 1994). In GroEL, up to 27 residues can be deleted from this C-terminal tail without compromising the function of the protein (Burnett *et al*., 1994). Interestingly, Cpn60.3 from *M. smegmatis* did not have a characteristic sequence at its C-terminal implying that it may have a function very different to that of Cpn60.1 or Cpn60.2. The expression of Cpn60.3 under stress conditions will be investigated in the next set of results.

```
M_gilvum_Cpn601          NAATLEYGDLIGDGVIDPVKVTRSAVVNAASVARMVLTTETAVVEKPAEAEDDGHGHG-----HGHHHH---- 541
M_vanbaalenii_Cpn601     NAATLEYGDLIADGVIDPVKVTRSAVVNAASVARMILTTETAVVEKPEESEDDGHGHG-----HGHHHGHAH- 544
M_sp._JLS_Cpn601         NAATLEFGDLLADGVVDPVKVTRSAVLNAASVARMVLTTETAIVDKPAEEEDHGHGHH-----HGHAH----- 540
M_sp_MCS_Cpn601          NAATLEFGDLLADGVVDPVKVTRSAVLNAASVARMVLTTETAIVDKPAEEEDHGHGHH-----HGHAH----- 540
M_sp_KMS_Cpn601          NAATLEFGDLLADGVVDPVKVTRSAVLNAASVARMVLTTETAIVDKPAEEEDHGHGHH-----HGHAH----- 540
M_smegmatis_Cpn601       NAATLEFGDLVSAGVVDPAKVTRSAVLNAASVARMILTTETAVVDKPADEDEHGHGHH-----HGHAH----- 540
M_bovis_BCG_Cpn601       NVNTLSYGDLAADGVIDPVKVTRSAVLNASSVARMVLTTETVVVDKPAKAEDHDH-H------HGHAH----- 539
M_tuberculosis_1         NVNTLSYGDLAADGVIDPVKVTRSAVLNASSVARMVLTTETVVVDKPAKAEDHDH-H------HGHAH----- 539
M_ulcerans_Cpn601        NAATLTYGDLAADGIVDPVKVTRSAVLNASSVARMVLTTETAIVDKPAEPEDDGHGH------HGHAH----- 539
M_marinum_1_MM1126_      NAATLTYGDLAADGIVDPVKVTRSAVLNASSVARMVLTTETAIVDKPAEPEDDGHGH------HGHAH----- 539
M_avium_paratub_1        NAEKLSYGDLIADGVIDPVKVTRSAVLNSASVARMVLTTETAVVDKPAEEADDHG--------HGHHHH---- 538
M_leprae_1               NASTLGYGDLVADGVVDPVKVTRSAVLNAASVARMMLTTETAVVDKPAKTEEHDH--------HGHAH----- 537
M_gilvum_Cpn602          NAATGEYEDLLKAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_vanbaalenii_Cpn602     NAATGEYEDLLKAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_sp._JLS_Cpn602         NAATGEYEDLLAAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKSAAPAGDPT------GGMGGMDF-- 541
M_sp_MCS_Cpn602          NAATGEYEDLLAAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKSAAPAGDPT------GGMGGMDF-- 541
M_sp_KMS_Cpn602          NAATGEYEDLLAAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKSAAPAGDPT------GGMGGMDF-- 541
M_smegmatis_Cpn602       NAATGEYEDLLAAGVADPVKVTRSALQNAASIAALFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_ulcerans_Cpn602        NAATGVYEDLLKAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_marinum_2_MM0759_      NAATGVYEDLLKAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_avium_paratub_2        NAATGEYEDLLKAGIADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKAAAPAGDPT------GGMGGMDF-- 541
M_bovis_BCG_Cpn602       NAQTGVYEDLLAAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKEKASVPG-G------GDMGGMDF-- 540
M_tuberculosis_2         NAQTGVYEDLLAAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKEKASVPG-G------GDMGGMDF-- 540
M_leprae_2               NAATGEYEDLLKAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKTAAPASDPT------GGMGGMDF-- 541
E_coli_K12_Cpn60         NAATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGG---MGGMGGMGGMM 548
M_smegmatis_Cpn603       NALTGEYGDMFEEGIIDPFKVTRAALESAASIAALLITTETAVVEEILGQPGAIMAPGFGDLAEGMVRPSNIY 549
```

Figure 3. 1: **Multiple alignment of mycobacterial Cpn60 proteins showing C-terminus**

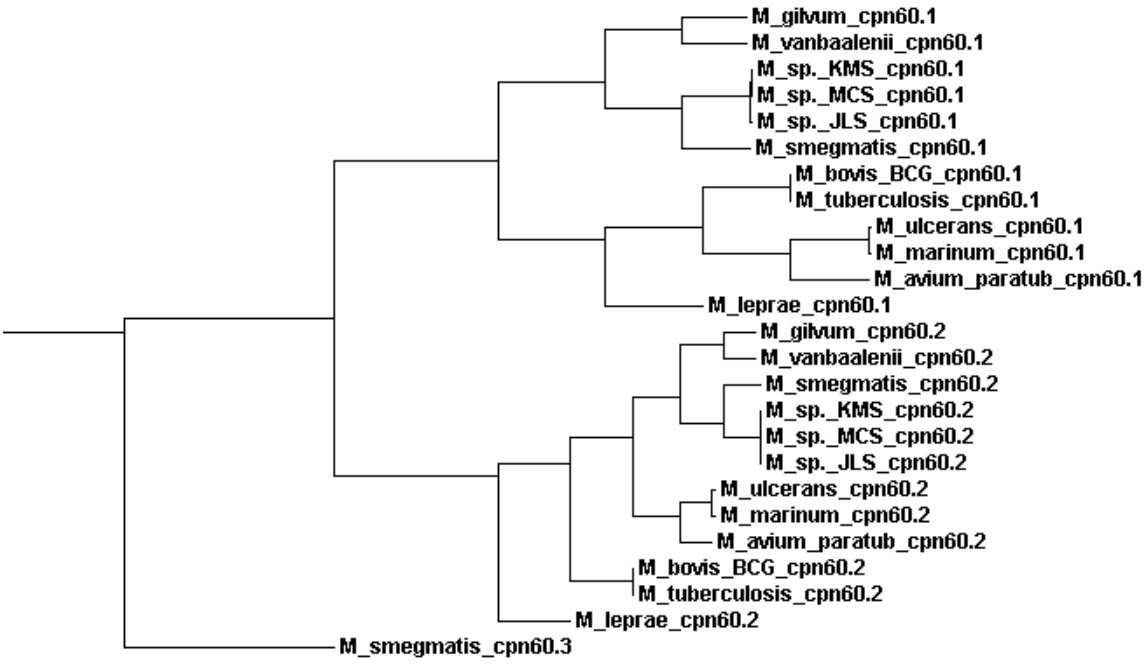


Figure 3. 2: **Phylogenetic tree of mycobacterial chaperonins**

Phylogenetic tree of chaperonin genes from all the sequenced mycobacterial species available

When the nucleotide sequence of *cpn60.3* was compared with other sequenced Actinobacteria, the gene showed nearly 80% similarity to a *cpn60* gene in the soil Actinomycete *Rhodococcus* RHA1. This is significant since this Actinomycete was also found to have three chaperonin genes, one of which is present with the co-chaperonin (McLeod *et al*., 2006), implying that the presence of a third *cpn60* gene is not a random event, but could be due to functional role based on its natural environment. Similar to the other *cpn60* genes, one copy of the *Rhodococcus* RHA1 Cpn60 had a GGM C-terminal tail while the other copy had three histidine residues. Once again, the Cpn60.3 homologue from Rhodococcus did not have a distinguishing C-terminus. An alignment of the region around *cpn60.3* gene showed that a number of neighbouring genes are conserved between the two bacteria. So the presence of the *cpn60.3* gene in *M. smegmatis* could probably be due to a horizontal gene transfer event.

As explained at the beginning of this chapter, two well characterised regulatory systems of the heat shock response exist. The HrcA and HspR repressor proteins have been extensively studied in *M. tuberculosis* (Stewart *et al*., 2001; Stewart *et al*., 2002). In order to identify the presence of these well characterised regulatory sequences, the entire *M. smegmatis* genome was searched for matches to the consensus CIRCE sequence CTAGCACTCNNNNNNNNNGAGTGCTAG and the consensus HAIR sequence CTTGAGTNNNNNNNACTCAAG using the programme PatternSearch implemented on xBASE (Chaudhuri and Pallen, 2006). If no mismatches were allowed, only one occurrence of the CIRCE sequence was found, upstream of *cpn10* (figure 3.3). If up to two mismatches were allowed, a further candidate CIRCE sequence was found upstream of *cpn60.2*, although several other potential matches were also found upstream

of genes that are not usually part of the heat shock regulon. When the genome was searched for the HAIR sequence with up to three mismatches allowed, none of the chaperonin genes were found as candidate sequences. However, during this search, other molecular chaperones such as *dnaJ*, *dnaK* and *grpE* were found to contain HAIR sequences in their upstream regions.

Figure 3. 3: **Illustration of upstream and translated regions of *cpn10* and *cpn60.2* genes**
Upstream region and sequence of *cpn10* (above) and *cpn60.2* (below) genes from *M. smegmatis* showing the CIRCE sequences (bold italics) and translated regions of the genes (red).

3.3.2 Gene expression under stress conditions

Since our previous analyses had shown that as expected each of the *cpn60* genes showed specific sequence differences, the next set of experiments were done to determine if each of the *cpn60* genes played a unique role in the cell. In this next step of characterising the chaperonins from *M. smegmatis*, the expression levels of the *cpn60*

and *cpn10* genes were monitored using quantitative real time PCR (qRT-PCR). For this purpose, wild type *M. smegmatis* cells were grown up to mid log phase (OD_{600} 0.7-0.8, as specified in Belanger and Hatfull, 1999) and cultures were then split into two. While one culture was used as a control, the other was exposed to heat stress, ethanol stress, oxidative stress and osmotic stress for varying amounts of time. The TaqMan probes used in this study for monitoring gene expression are shown in figure 3.4. As can be seen, even though there is a high similarity between the three genes, the probes are located at positions of low similarity to prevent non-specific binding at other regions of homology.

Figure 3. 4: Illustration of Taqman probes in *cpn60* sequences

Sequence alignment of the three chaperonin genes, with the TaqMan probes (black boxes).

3.3.2.1 Heat shock

A culture of *M. smegmatis* was grown up as described in materials and methods. When the OD_{600} was 0.6-0.8, the culture was split into two and one set was grown at 37°C, while the other set was incubated at 42°C. Samples were taken from both cultures at 5min, 10min, 15min and 30min post stress. RNA was extracted from cells according to the method described in section 2.7, cDNA synthesised and equal amounts of cDNA loaded along with ready mix and TAQman probe and qRT-PCR performed. Fold induction for each gene was calculated relative to its level of expression under normal growth at each time point.

Under normal conditions, when expression levels of the genes were monitored, it was observed that *cpn10* was the most highly expressed gene followed by *cpn60.2* and *cpn60.1*, while hardly any expression of *cpn60.3* was observed (figure 3.5).

When the cells were heat shocked at 42°C, expression levels increased drastically even at 5min post stress and continued to increase and peak at 15min, while dropping off a little at 30min (figure 3.6). While all the genes with the exception of *cpn60.3* showed massive increases in expression levels, *cpn60.1* exhibited the highest fold change in expression levels 15min post stress where expression increased more than 180 fold. This massive increase in expression is very significant as previous studies in *M. tuberculosis* have shown increases to be less than 10 fold after 30 minutes of heat shock (Hu *et al.*, 2008).

An *M. smegmatis cpn60.1* knock-out was also analysed for its expression levels under heat shock conditions and the results compared to the wild type. As can be seen in figure 3.7, no expression of *cpn60.1* can be observed, confirming it is indeed a knock-out. When the expression levels of the other genes are compared between the two strains, only a 2 fold difference in expression levels is observed. The significance of the results will be discussed later in the chapter.

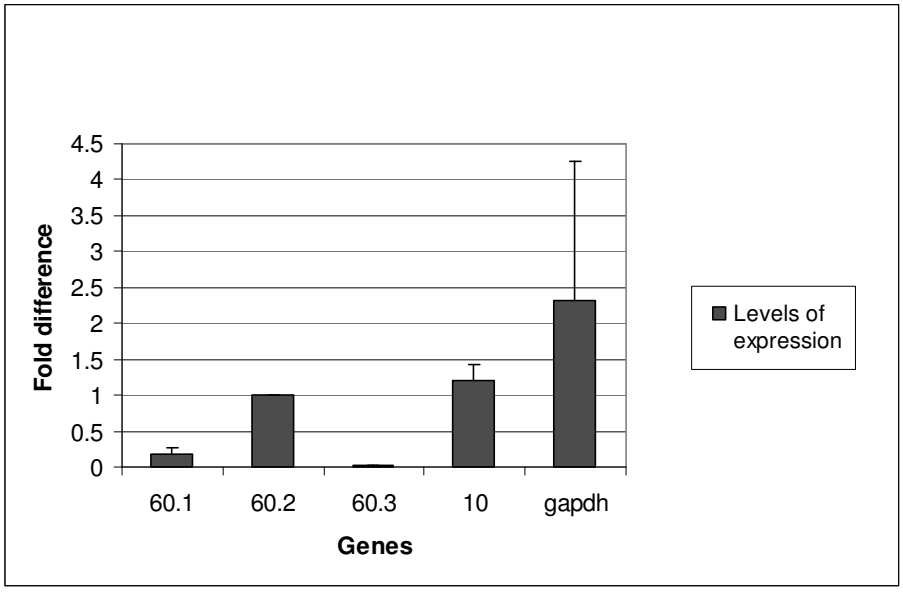


Figure 3. 5: **Relative expression levels of *cpn60* and *cpn10* genes (normal conditions)**
Relative expression levels of the *cpn60* and *cpn10* genes when normalised to *cpn60.2* =1 and grown at 37°C. The results are the average of three independent biological replicates and the error bars show the standard deviations.

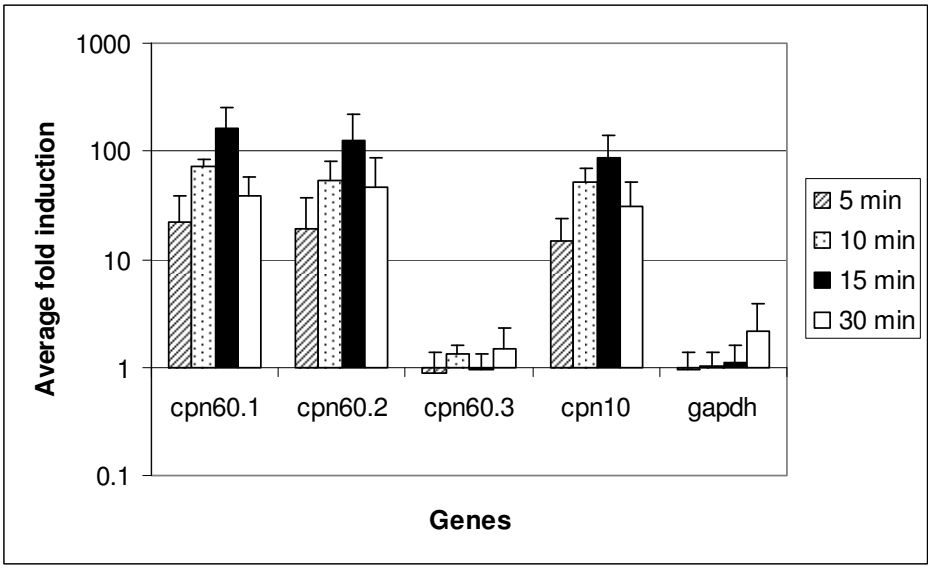


Figure 3. 6: **Average fold change of *cpn60* and *cpn10* genes (heat shock)**

Relative levels of expression of the chaperonin genes after heat shock at 42°C. The results are the average of three independent biological replicates and the error bars show the standard deviations.

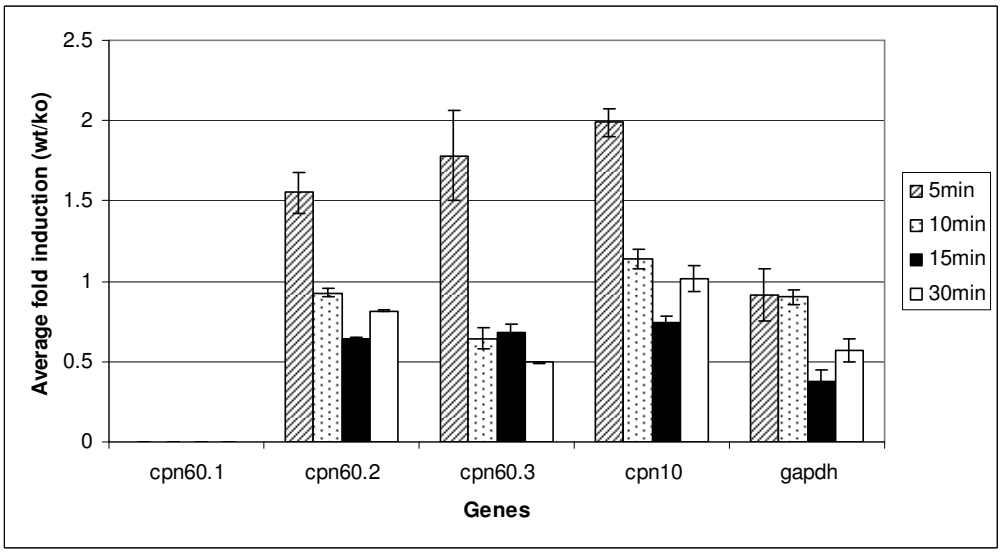


Figure 3. 7: **Average fold change of *cpn60* and *cpn10* genes in *M. smegmatis* strains**

Differences in expression levels of the chaperonin genes in the *cpn60.1* knock-out strain compared to the wild type strain. The results are the average of three independent biological replicates and the error bars show the standard deviations.

3.3.2.2 Ethanol stress

M. smegmatis as grown at 37°C until mid-log phase, and stressed with 5% ethanol. Samples were taken 10 and 30 minutes post stress. qRT-PCR was performed as described in Materials and Methods. Results showed that while there was an increase in expression levels from 10 minutes to 30 minutes (figure 3.8) the levels were not as high as those seen with the heat stress. In this case even *cpn60.3* showed some expression.

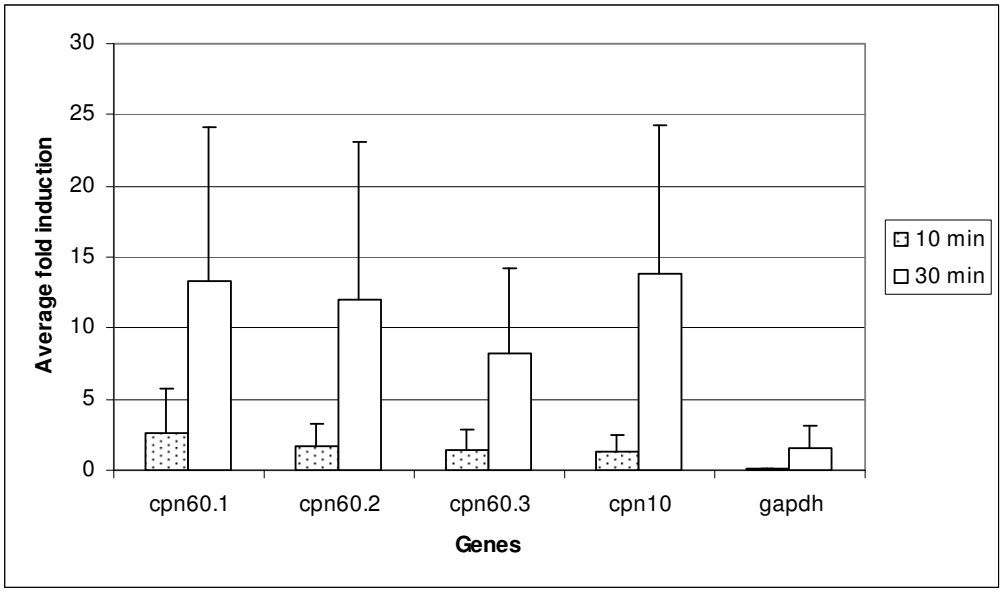
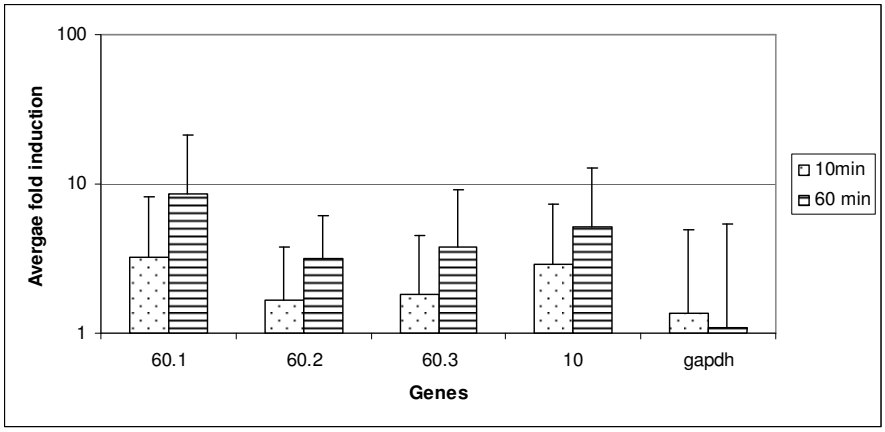


Figure 3. 8: Average fold change of *cpn60* and *cpn10* genes (ethanol stress)

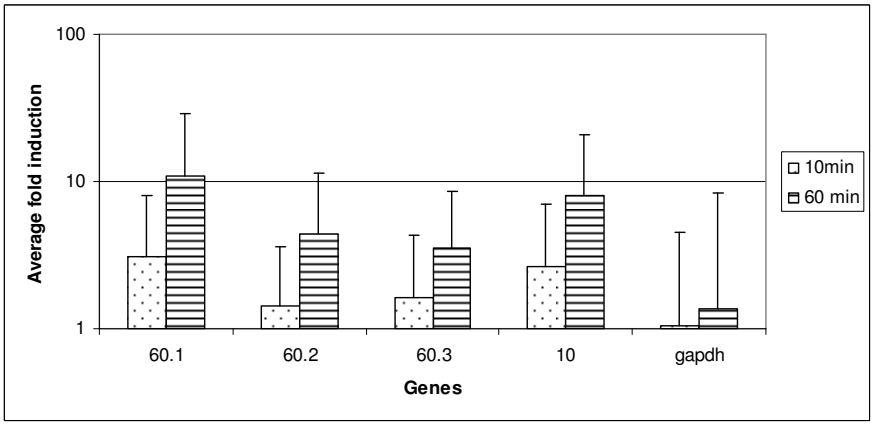
Relative expression levels of chaperonin genes after a 5% ethanol stress. The results are the average of three independent biological replicates and the error bars show the standard deviations.

3.3.2.3 Oxidative stress

An oxidative stress of 10mM or 20mM hydrogen peroxide (H_2O_2) was used to stress *M. smegmatis* wild type cells. The fold change in expression levels of the individual genes was analysed and results can be seen in figure 3.9 a and b. When the difference in the fold induction of all the *cpn60* and *cpn10* genes when exposed to 10mM or 20mM H_2O_2 as compared no large differences were observed.



a



b

Figure 3. 9: Fold change of *cpn60* and *cpn10* genes (oxidative stress)

Fold change in expression levels of genes when exposed to an oxidative stress of 10mM H_2O_2 (a) and 20mM H_2O_2 (b). The results are the average of three independent biological replicates and the error bars show the standard deviations.

3.3.2.4 Osmotic stress

M. smegmatis cells were stressed with 1.5M NaCl to induce an osmotic stress, and results showed that expression levels of the chaperonin genes appeared to decrease upon exposure to the osmotic stress (figure 3.10). A fold change below 1 is considered as not a significant change in expression level, so although expression of *cpn60.3* appeared to increase at 60 minutes post stress, this 1.6 fold increase was not found to be very highly significant.

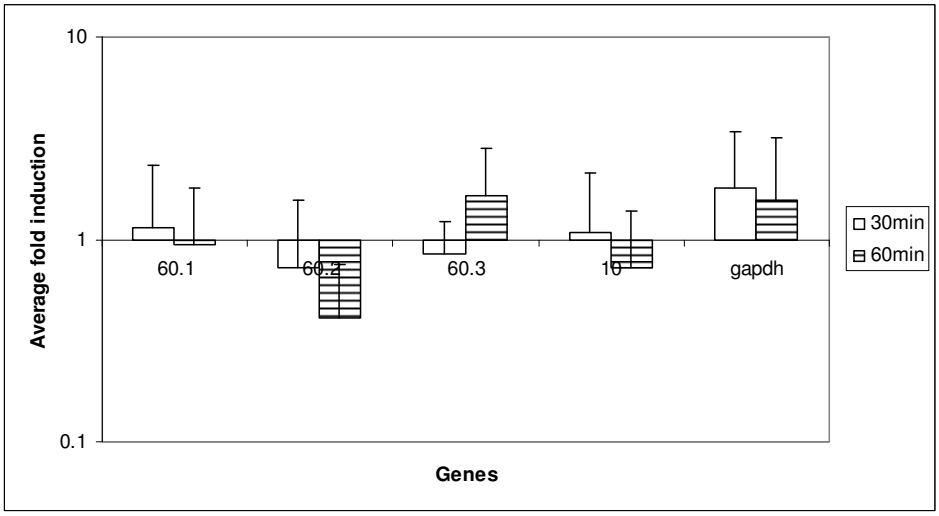


Figure 3. 10: **Relative expression levels of *cpn60* and *cpn10* genes (osmotic stress)**

Decreased levels of expression of the chaperonin genes after exposure to 1.5M NaCl. The results are the average of three independent biological replicates and the error bars show the standard deviations.

3.3.3 Transcript and promoter analysis

In the previous set of experiments, when expression levels of the *cpn10* and *cpn60.1* genes were analysed under unstressed conditions, both genes appeared to be differentially expressed in relation to each other. As the co-chaperonin *cpn10* and chaperonin *cpn60.1* are present in a single operon, the reason for the difference in expression levels was intriguing. One hypothesis was that even though the genes appeared to be present in a single operon, two separate transcriptional sites could exist, one upstream of *cpn10* and one upstream of *cpn60.1* in the intergenic region. The other hypothesis was that even though the genes were co-transcribed, due to a post-transcriptional cleavage event, a smaller less stable fragment of mRNA (containing *cpn60.1*) was produced. Similar studies conducted in *M. tuberculosis* have shown the presence of a second 5' end in the intergenic region, upstream of *cpn60.1* (Kong *et al*., 1993). However, recent reports have shown that there is no promoter present in the intergenic region (consistent with the second hypothesis) (Aravindhan *et al*., 2009). Thus, our prediction was that a similar situation would probably be observed even in *M. smegmatis*.

In order to identify the transcriptional start site (tss), 5' Rapid Amplification of cDNA Ends (5' RACE) was performed. Wild type *M. smegmatis* cells were grown to mid-log phase and RNA extracted as described in chapter 2. The gene specific primer used to synthesise the cDNA was *cpn60.1*gsp1. The synthesised cDNA was then C-tailed at the 5' end using poly-cytosine and transcriptional start sites were identified by detection of the junction between the C-tail in the sequenced DNA using nested primers *cpn10* gsp2 and *cpn60.1*gsp2. Sequencing results identified two separate 5' ends, one upstream of

the *cpn10* gene and one upstream of *cpn60.1*. Two clear 5' ends were found, mapping at 133bp upstream of *cpn10* and in the intergenic region between +29 and +34 upstream of *cpn60.1* (figures 3.11 and 3.12). Due to inconclusive sequencing results (presence of G repeats at that location in the sequence), the precise base pair at the 5' end of the shorter transcript could not be determined. The one upstream of the *cpn10* gene is a good candidate for a genuine transcriptional start site, while the second one could be due to a second start site or to a post-transcriptional cleavage. To distinguish between these possibilities, promoter analysis of the two regions was done.

ACTCTCATGTATAGAGTGCTAGGT*GGCAGGCGGACGCCCCACGTGCCGGCACCCGCGAC

GACGGAGCGAGGGGTCAGGACGCATGCCGAACAACCTGGTAACCGGAATCTGGGGACTT

CCCGGATCTCACCTTTAACTAGTGGAGGGCTCCATC***GTG***GCGAGCGTGAACATCAAGC

CACTCGAGGACAAGATCCTCGTTCAGGCCAACGAGGCCGAGACCACGACCGCTTCGGGT

CTGGTCATCCCCGACACCGCCAAGGAGAAGCCGCAGGAAGGCACCGTCGTCGCAGTTGG

CCCCGGCCGCTGGGATGAGGACGGCGAGAAGCGGATCCCCCTGGACGTTGCCGAGGGTG

ACACCGTCATCTACAGCAAGTACGGCGGCACCGAGATCAAGTACAACGGCGAGGAGTAC

CTGATCCTGTCGGCCCGCGACGTGCTGGCTGTCGTCTCCAAG***TAA***TCACTCGTGTTCA

CCGCCCCGGAGGTCCCCGTCTCA*CACGGGTGATGTCCGGGGCGGTACGCGTAGGCGTCT

GAAAGAGACTT***ATG***AGCAAGCAGATTGAATTCAACGAAACCGCCCGCCGGGCGATGGA

GGCCGGCGTCGACAAGCTCGCCGACGCCGTCAAGGTCACGCTCGGCCCGCGCGGTCGTC

ACGTGGTGCTGGCCAAGTCTTTCGGCGGCCCGCAGGTCACCAACGACGGTGTGA

Figure 3. 11: **Illustration of upstream and translated regions of *cpn10* and *cpn60* genes**

Sequence of *cpn10* (grey) and *cpn60.1* (yellow) showing upstream region with the transcriptional start sites (*), translation start and stop sites (italics) and potential ribosome binding sites upstream of the starts of *cpn10* and *cpn60.1* (underlined).

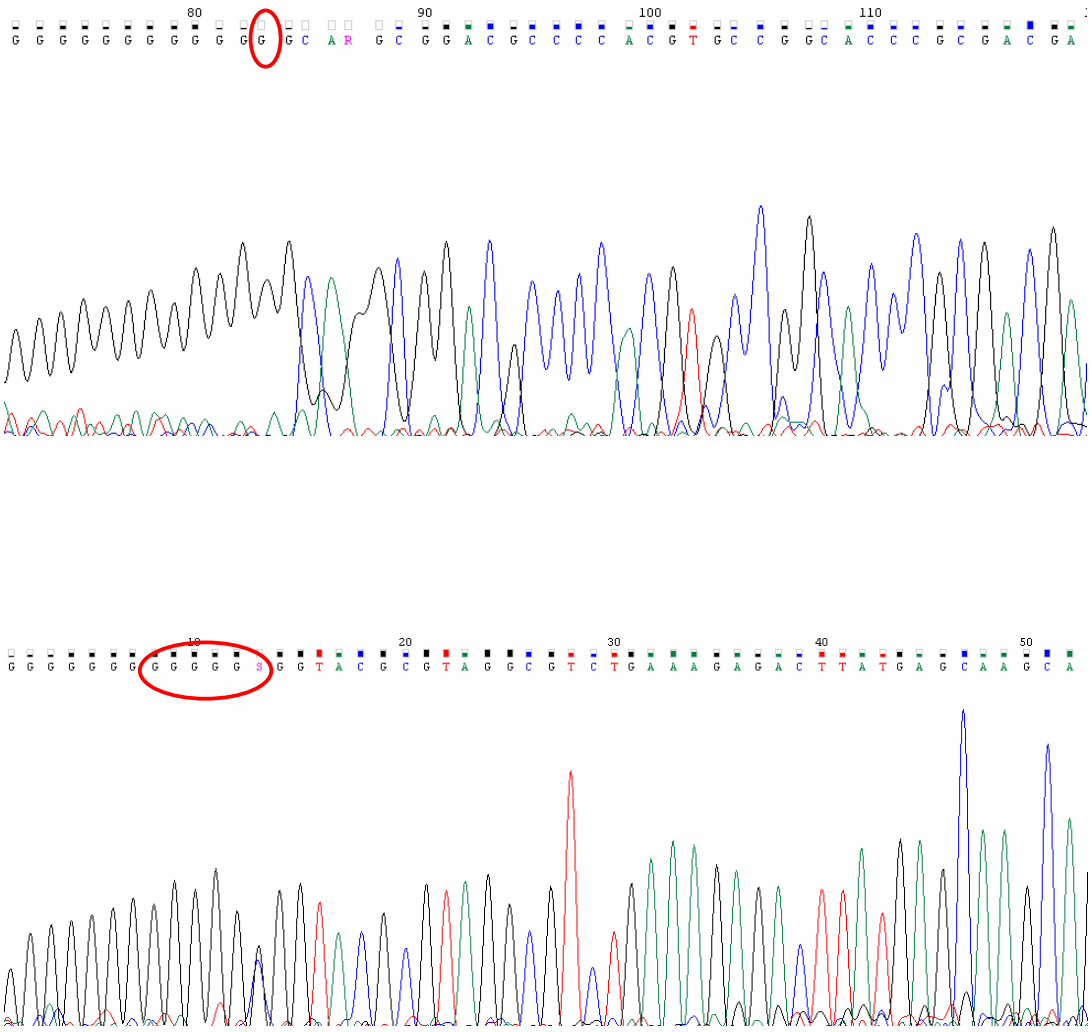


Figure 3. 12: **Chromatogram of sequencing results from 5'RACE experiment**

Chromatogram of sequencing results showing the 5' end of the transcripts (red circles) of *cpn10* (above) and *cpn60.1* (below).

3.3.3.1 Promoter probe studies

Our results from the bioinformatic analyses (section 3.3.1), indicated the presence of regulatory regions, CIRCE sequences, upstream of the *cpn10* but not *cpn60.1* gene. Furthermore, results from the transcript analysis showed the presence a 5' end of the transcript in the intergenic region. Thus, as mentioned previously, one hypothesis was that two separate transcriptional sites could exist, one upstream of *cpn10* and one upstream of *cpn60.1* in the intergenic region. The second hypothesis was that a post-transcriptional cleavage event may result in the presence of a smaller less stable fragment of mRNA (containing *cpn60.1*). To test which of the hypotheses was correct, we decided to investigate the presence of a promoter element in the intergenic region.

To determine the presence of a promoter element in the intergenic region, a promoterless lacZ plasmid, pSD5B which can be used in *E. coli* as well as in mycobacteria was chosen (Jain *et al.*, 1997). Varying lengths of upstream fragments were cloned into plasmid pSD5B on XbaI/SphI sites or on XbaI sites only and β-galactosidase activity measured in strains containing these plasmids. A 369bp upstream region of *cpn10* was amplified and digested with XbaI sites and cloned into pSD5B to give pSD5B-cpn10. For the *cpn10-cpn60.1* intergenic region, two fragment lengths (393bp and 204bp) were amplified and cloned into pSD5B using XbaI and SphI sites to give pSD5B-LF and pSD5B-SF, respectively. As a positive control, a 335bp upstream region of *cpn60.2* was also cloned on XbaI and SphI sites to give pSD5B-cpn60.2. Finally, even though no CIRCE sequences were found in the *cpn60.3* upstream region, a 335bp fragment was cloned into plasmid pSD5B on XbaI and SphI sites to give plasmid pSD5B-cpn60.3 (figure 3.13). Under the various stress conditions, *cpn60.3* showed very

poor or no expression at all, furthermore, bioinformatic analyses revealed no CIRCE region upstream of this gene. Because of these reasons, the upstream region of the *cpn60.3* was investigated for the presence of promoter elements. Although our hypothesis was that no such promoter elements would be found, if a promoter element was to be identified, further experiments involving knocking out the gene or looking at its expression under other conditions would be conducted.

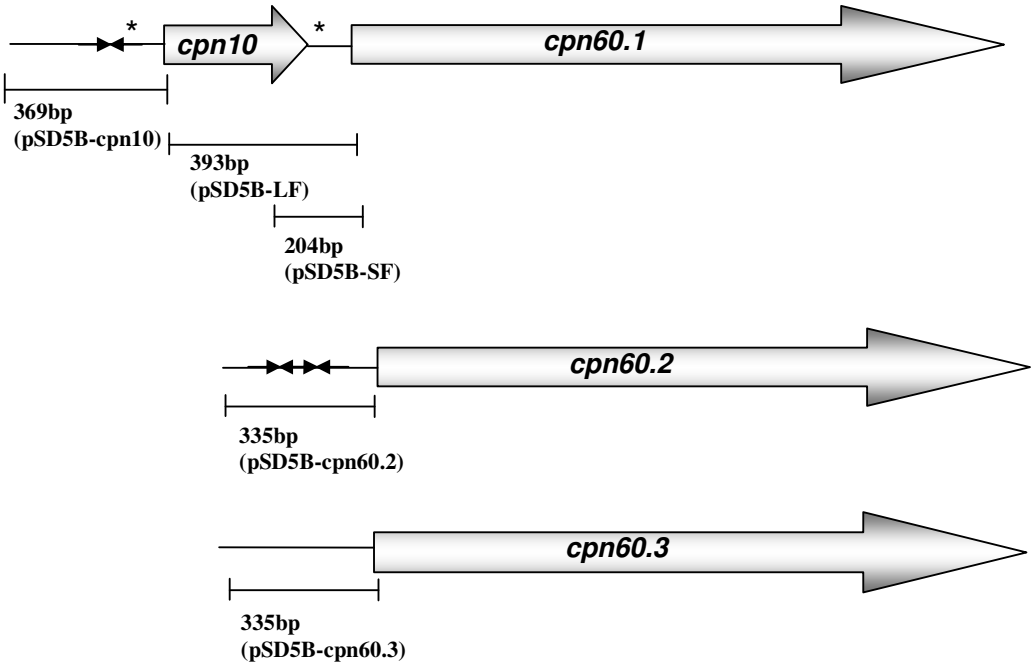
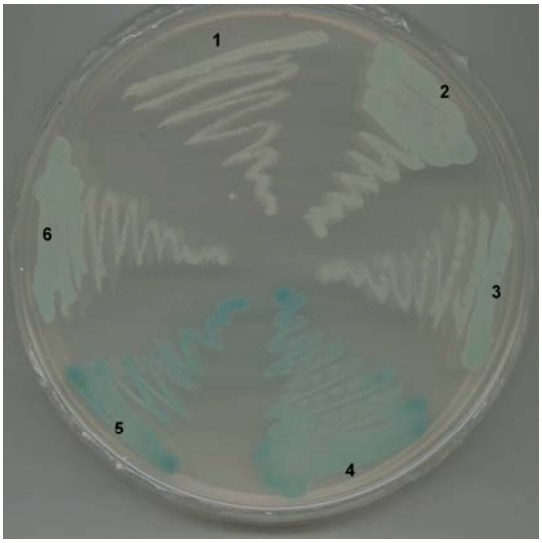
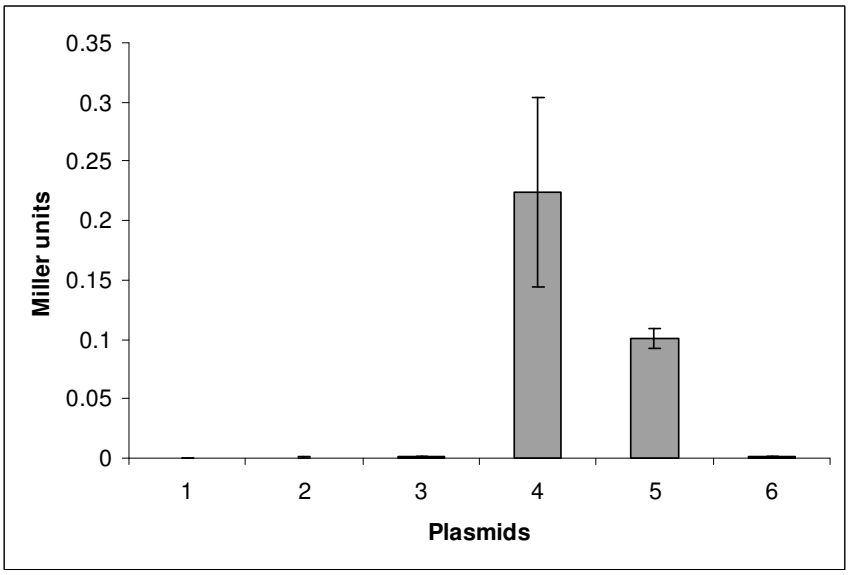


Figure 3. 13: **Diagrammatic representation of regions cloned into plasmid pSD5B**
Diagrammatic representation of the genes and the different fragments cloned into plasmid pSD5B to monitor promoter activity.

Wild type *M. smegmatis* cells were transformed with each of these constructs and the transformants were screened for lacZ activity by conducting a preliminary qualitative plate assay and a quantitative β-galactosidase assay. Previous studies of mycobacterial promoters in *E. coli* have shown that in some cases promoter activity is observed in both *E. coli* and *M. smegmatis* (Aravindhan *et al*., 2009), while in some cases promoter activity is only observed in *M. smegmatis* (Jain *et al*., 2005). Hence this result was not entirely surprising. Results from *M. smegmatis* showed lacZ activity from cells containing plasmids pSD5B-cpn10 and pSD5B-cpn60.2 (figure 3.14). This result was observed on plate assays where the above mentioned strains expressing lacZ breakdown the X-GAL to produce a blue colour (5,5'-dibromo-4,4'-dichloro-indigo) and in the liquid β-galactosidase assay as well (figure 3.14b). These results will be discussed at the end of this chapter.



a



b

Figure 3. 14: Results obtained from β-galactosidase assay

Figure 3.14 a: Media containing Xgal with *M. smegmatis* strains with plasmids pSD5B (1), pSD5B-SF (2), pSD5B-LF (3), pSD5B-cpn10 (4), pSD5B-cpn60.2 (5) and pSD5B-cpn60.3 (6).

Figure 3.14 b: Measurement of promoter activity in *M. smegmatis* transformed with plasmid pSD5B (1), pSD5B-SF (2), pSD5B-LF (3), pSD5B-cpn10 (4), pSD5B-cpn60.2 (5) and pSD5B-cpn60.3 (6).

3.4 Discussion

The chaperonins of mycobacteria are of interest for a number of reasons. Firstly, although many bacteria possess multiple chaperonin genes, *M. smegmatis* is one of only a few where a specific function has been reported for one of the homologues (Ojha *et al*., 2005). This specific role (biofilm maturation) of one homologue, Cpn60.1, will be analysed in detail in the following chapter. Secondly, *M. smegmatis* is unique among the mycobacteria for which a complete genome sequence is available to date, in that it has three chaperonin genes (Ojha *et al*., 2005). As mentioned in the introduction of this chapter, the presence of multiple copies of the *cpn60* genes could be due to either increase chaperoning activity within the cell or it could be that each duplicated copy has taken on a more specialised role within the cell. However, as has been pointed out, the actual situation maybe in between the two hypotheses with some diversification and overlapping of function. Moreover, the fact that one of the *cpn60* genes has been shown to have a specialised function in both *M. tuberculosis* (granuloma formation) and *M. smegmatis* (biofilm maturation), strongly suggests that there may be an overlap of function between these two chaperonins.

To test the above hypotheses, firstly, the multiple *cpn60* sequences from *M. smegmatis* and other mycobacterial members were compared and analysed for regions of conservation and differentiation. Next, as part of a functional characterisation, the multiple *cpn60* and *cpn10* genes from *M. smegmatis* were monitored for their expression levels under various stress conditions.

To this end, the *cpn60* sequences from all complete mycobacterial genomes that are available to date were obtained and aligned. Since most mycobacteria contain at least two Cpn60 proteins, it was not surprising to see that all Cpn60.1 proteins clustered together, while the Cpn60.2 proteins formed a separate cluster. This initial finding suggests that in an ancestral organism, a gene duplication followed by sequence divergence has taken place. Previous studies have shown that amongst the Actinomycetes, in all cases where only a single *cpn60* gene is present (as in *Tropheryma whipplei* [Maiwald *et al*., 2003]), this single copy is not found in an operon with the *cpn10* gene but is located elsewhere on the genome. This can be explained by considering that a common ancestor contained two copies of the *cpn10-cpn60* operon and due to functional redundancy, one copy of *cpn10* was lost (Lund, 2009). The loss of a redundant *cpn60* gene strongly suggests that in bacteria with multiple *cpn60* genes, the copy present away from the *cpn10* gene is most probably the essential housekeeping chaperonin.

Intriguingly, the C-terminus of Cpn60.1 and Cpn60.2 proteins displayed characteristic pattern repeats. The C-terminus of all the Cpn60.1 proteins contained a number of histidine repeats, while the Cpn60.2 tail was composed of glycine-glycine-methionine (GGM) repeats. While the significance of these repeating units is not fully known, the GGM C-terminal of the *E. coli* GroEL protein has been shown to be flexible in the crystal structure of the protein (Briag *et al*., 1994). Moreover, deleting up to 27 residues from the C-terminal does not compromise the function of the protein (Burnett *et al*., 1994). The high conservation of this motif means it almost certainly has some role to play, however, lab experiments have so far failed to detect what this function is. The

Cpn60.3 protein did not have any characteristic repeating motif at the C-terminal end which may imply that its function, if any, in the cell is very different from that of Cpn60.1 or Cpn60.2.

The results obtained from the alignment studies were used to construct a phylogenetic tree. The phylogenetic tree was constructed using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). This is one of the simplest methods of constructing a phylogenetic tree and uses a sequential clustering algorithm in which local topological relationships are identified in order of similarity, and a phylogenetic tree is built in a stepwise manner. However, the drawback of using this method is that it assumes that evolution occurs at the same rate on all tree branches and that the distances are additive. The Neighbor Joining (NJ) clustering method is often considered a better method as it doesn't make the assumption of additive differences. Phylogenetic trees were constructed using both methods and importantly it was observed that both methods produced similar phylogenetic trees which showed all Cpn60.1 genes clustered together and all Cpn60.2 genes forming a separate clade. This result is in agreement with previous phylogenetic studies of Actinobacteria (Goyal *et al.*, 2006). As explained previously, the presence of two separate clades could be due to gene duplication and sequence divergence events in a common ancestor, where the divergence has been conserved. Further, this also supports the hypothesis of conservation of different function in the two clades. Importantly, the Cpn60.3 did not cluster in either of the clades and was present as an outgroup, further suggesting that not only is it different at the sequence level, but probably at a functional level as well. This result also strongly suggests the Cpn60.3 is evolutionarily different from all the other mycobacterial

chaperonins. When a BLAST search was performed with Cpn60.3, it was found to share nearly 80% similarity with a chaperonin from the soil Actinomycete *Rhodococcus* RHA1. Interestingly, this Actinomycete was also found to have three chaperonin proteins (McLeod *et al*., 2006), one containing a GGM tail (most probably the essential housekeeping copy), one containing a histidine rich tail (Cpn60.1 homologue) and the third one which showed most similarity to Cpn60.3. This finding reveals that the presence of Cpn60.3 in *M. smegmatis* could be due to a horizontal gene transfer event between the two bacteria. Further, the presence of this third copy in these two soil Actinomycetes could imply a function related to its natural environment.

Chaperonins are also known as heat shock proteins due to their role in the heat shock response. In *E. coli* the expression of these genes is regulated by the alternative sigma factor, the σ^{32} subunit of RNA polymerase that is encoded by the *rpoH* gene (Bukau, 1993). As there is no homologue of the *rpoH* gene in *M. smegmatis*, it is likely that the heat shock response is regulated by other systems, possibly the HrcA/controlling inverted repeat of chaperone expression (CIRCE) or the HspR/HspR-associated inverted repeat (HAIR) systems (Bucca *et al*., 1995; Zuber and Schumann, 1994).

To address the question as to which regulatory system plays a role in the modulation of the *cpn60* genes, the *M. smegmatis* genome was searched for these regulatory sequences. Furthermore, identification of any of these sequences would provide an insight into the functional role of the *cpn60* gene. A search for CIRCE sequences (CTAGCACTCNNNNNNNNNGAGTGCTAG) in the *M. smegmatis* genome revealed the presence of two such sequences upstream of *cpn60.2* and one sequence upstream of

cpn10. This has previously been shown to be the case in *M. tuberculosis* as well (Stewart *et al*., 2002). This finding is confirmed by the fact that the levels of the *cpn10*, *cpn60.1* and *cpn60.2* genes have been shown to be upregulated in a HrcA mutant (Stewart *et al*., 2002). Thus it would be interesting to do similar studies in *M. smegmatis* and monitor the expression levels of these genes and to also determine if levels of *cpn60.3* were altered. Since no CIRCE sequence was identified upstream of *cpn60.3*, a search for the HspR binding site, the HAIR sequence (CTTGAGTNNNNNNNACTCAAG) was performed. No such sequences were found upstream of *cpn60.3* or any of the other *cpn60* or *cpn10* genes. This strongly suggests, that *cpn60.3* doesn't play a role in the heat shock response in the cell. Interestingly, upstream sequences of the DnaJ/DnaK genes had HAIR binding sites present, which has previously been observed in *M. tuberculosis*. Moreover, studies of an *M. tuberculosis* HspR mutant, showed that only the DnaJ/DnaK genes and ATPase ClpB were controlled by this protein, while the chaperonin genes appeared unaffected in the mutant (Stewart *et al*., 2001). Hence, it is fair to conclude that the *cpn10* and *cpn60* genes in *M. smegmatis* are probably regulated by the HrcA regulon.

Next, to test if the sequence differences observed from the bioinformatic analyses were responsible for functional divergence, a series of experiments monitoring the expression of these genes under various stress conditions was performed. Previous studies looking at expression of the genes under various stress conditions has been performed in related bacteria like *Mycobacterium tuberculosis* and *Corynebacterium glutamicum* (Hu *et al*., 2008; Barreiro *et al*., 2009; Ehira *et al*., 2009). These studies identified the *cpn60* genes to be upregulated among other genes. Studies looking at the response of *M. smegmatis*

to acid stress have previously been conducted (O'Brien *et al*., 1996; Tran *et al*., 2005). There are a number of reports in which the response of *M. smegmatis* to heat and oxidative stress has been investigated (Lundrigan *et al*., 1997; Fernandes *et al*., 1999; Milano *et* al., 2001; O'Toole *et al*., 2003) However, to date, the expression of the chaperonins from *M. smegmatis* under various stresses has not been monitored.

To begin to understand the role of these genes under the various stresses, a change in expression levels of the various genes was determined by monitoring their RNA expression in real time. For this purpose, we chose to use quantitative real time PCR (qRT-PCR). This technique is ideal for quantifying changes in gene expression. For the q-RT-PCR, specific probes, TaqMan were used. To ensure specificity of the probes, regions were chosen with low homology between the three genes and the oligonucleotide sequences were used to search for similarity against the whole genome using the BLAST programme. The change in expression of the target gene in relation to the average of its untreated sample at the same time point was measured. Since the same amounts (60ng) of cDNA was used in each reaction, the data was analysed using the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001). The results shown above reflects the averages of $2^{-\Delta\Delta C_T}$ between three biological repeats at each time point and the error bars show the standard deviation of the mean.

To begin with, the expression levels of the genes under non-stressed conditions were compared to check for differences in expression levels under normal conditions. Results revealed the co-chaperonin *cpn10* was the most highly expressed, followed by *cpn60.2* and *cpn60.1*. The third homologue, *cpn60.3* was not expressed at detectable levels.

These results are interesting for a few reasons. Firstly, there was a difference in the expression levels of the *cpn10* and *cpn60.1* genes which appear to be present in a single operon, secondly, the levels of expression of *cpn60.1* and *cpn60.2* were also considerably different, and finally no expression of *cpn60.3* was observed. There could be a few reasons for the differential expression levels of *cpn10* and *cpn60.1*; one hypothesis is that although both the genes appear to be present in a single operon, they could be transcribed separately and have their own individual promoters. This assumption was further investigated and will be discussed shortly. The next possibility is that since only a single copy of *cpn10* is present in the cell, it would need to be expressed at higher levels to be available to function with all the copies of *cpn60* genes. This hypothesis that high levels of Cpn10 are due to its different roles in the cell could be tested by performing co-immunoprecipitation studies and observing Cpn10 interacting proteins. Next, it has been shown that the housekeeping chaperonin is the one present elsewhere on the genome from the co-chaperonin *cpn10*, which strongly suggests that *cpn60.2* would be the housekeeping chaperonin. Moreover, experiments in *M. tuberculosis* and *M. smegmatis* have failed to knock-out the *cpn60.2* gene implying that it is essential (Ojha *et al*., 2005; Hu *et al*., 2008). Thus, since it is the main chaperoning protein in the cell, it is not entirely surprising that it is expressed at higher levels compared to *cpn60.1* which probably has no essential chaperoning role in the cell. Finally, the lack of expression of *cpn60.3* under the tested conditions of optimal temperature and rich growth media tested may imply that due a specialization of function, it is only expressed under specific conditions such as in poor nutritional conditions (as in its natural environment in the soil). To test these hypotheses, the expression of *cpn60.3* could be monitored under a range of temperatures and in various

media as well. Also, since the expression of all the genes was monitored at mid-log phase, another possibility is that *cpn60.3* could be expressed at a different phase of growth possibly during stationary phase, so expression of the gene could be monitored at different growth phases as well.

Under the heat stress (42°), which is known to elicit a response from the chaperonin genes, as predicted, all the genes except *cpn60.3* showed increased levels of expression, as high as 150 fold in some cases as well. Although expression levels of these genes in *M. tuberculosis* also increased on heat shock, this is the first report to show such high levels of induction. Also, since studies of the heat shock response in *Corynebacterium glutamicum* have shown that at different temperatures, differential expression of the *cpn60* genes is observed (Barriero *et al*., 2009), it would be interesting to investigate the expression patterns of these genes under a range of temperatures. The expression of the *cpn10* and *cpn60* genes was monitored in the *cpn60.1* knock-out strain as well to test if there was a compensatory effect by any of the remaining genes. However as seen from the results, none of the other genes (*cpn10*, *cpn60.2* or *cpn60.3*) are expressed at higher levels when compared to their levels in the wild-type strain. Although *cpn60.1* is highly induced upon heat shock, this result from the knock-out suggests that *cpn60.1* doesn't play an essential role in the heat shock response and the cell does not need to compensate for loss of this gene by increasing levels of expression of the other *cpn60* genes. Thus it could imply that while the presence of this gene can help increase chaperoning activity, due to its more specialised role of biofilm maturation, it is dispensable under normal or heat shocked planktonic growth conditions.

Even though the chaperonins are mainly heat shock proteins, their role in folding unfolded proteins and preventing aggregation could extend to other stress conditions as well. Under a 5% ethanol stress, all the genes showed an increase in expression by more than 10 fold, 30 minutes post stress. This was similar to results observed in *M. tuberculosis*. Although the roles of genes such as *katG* and *furA* in the response to oxidative stress have previously been characterised (Milano *et al*., 2001), the expression of the *cpn10* and *cpn60* genes were investigated in this study. Results showed that expression increased by less than 10 fold, similar to findings in *M. tuberculosis*. In the case of osmotic stress (1.5M NaCl), *cpn10* and *cpn60.2* were down regulated 60 minutes post stress. Importantly, even *cpn60.3* expression increased slightly 60 minutes post stress. In summary, while the highest induction of genes was observed after a heat shock of 42°C, increases in expression levels were observed after ethanol and oxidative stresses. However, the reason for the down regulation of *cpn10* and *cpn60.2* after osmotic stress is unknown. Since *cpn60.3* expression increased upon stresses other than heat, the role of this gene in survival under starvation and physiologically challenging conditions needs to be investigated.

As explained earlier, the levels of expression of the *cpn10* and *cpn60.1* genes varied even under normal conditions, which could be due to the presence of two separate transcriptional start sites and promoter regions, or could be due to cleavage of the mRNA or could also be due to preferential expression of the single copy of *cpn10*. To help answer the question of the differential expression levels between *cpn10* and *cpn60.1*, it was decided that the 5' end of the mRNA would be investigated. 5'RACE was conducted and two 5'ends were identified, one upstream of *cpn10* and the other

upstream of *cpn60.1* in the intergenic region. This result was not surprising as previous reports have identified the presence of two 5' ends at similar positions upstream of these genes in *M. tuberculosis* (Kong *et al*., 1993). However, the presence of the second 5' end in the intergenic region could be due to two possibilities. Firstly, a second promoter region could be present upstream of *cpn60.1* and secondly, a post-transcriptional processing event could result in the cleavage of the single mRNA resulting in the second 5' end. While the possibility of the first scenario is very unlikely due to the absence of regulatory sequences (CIRCE/HAIR sequences) and a conserved ribosome binding site in this intergenic region, to completely rule it out, promoter probe analyses was conducted. Varying lengths of this intergenic region were cloned into a promoterless lacZ plasmid and β-galactosidase activity measured. As expected, no promoter activity was observed from this region, strongly suggesting that the presence of the second 5' end could be due to a cleavage event. The post-transcriptional processing of the *groE* operon in *Agrobacterium tumefaciens* has previously been reported to be heat shock dependant (Segal and Ron, 1995). A recent report analysing the presence of promoter elements in this intergenic region in *M. tuberculosis* also showed no promoter activity and concluded that this was due to mRNA processing (Aravindhan *et al*., 2009). Messenger RNA degradation plays a role in diverse cellular processes and gene expression and is mainly controlled by the activity of a set of proteins known as the degradasome. It has been well characterised in *E. coli*, where the endoribonucleases RNase E binds and cuts the mRNA followed by RNase II and/or PNPase degradation of the fragments generated by the cleavage (reviewed in Coburn and Mackie, 1999). However, this phenomenon has not been well studied in Mycobacteria with only a few known examples of such mRNA processing at the 5' end

being reported (Hu *et al.*, 1999; Unniraman *et al.*, 2002). However, a recent study has reported that the although the mycobacterial bi-cistronic *furA-katG* operon is co-transcribed, due to mRNA processing, two separate transcripts containing *furA* and *katG* are generated (Sala *et al*., 2008). This is very similar to the case with the *cpn10-cpn60.1* operon and the presence of specific processing sites in the intergenic region can be further investigated by analysing this region for polypurine sequences followed by mutagenesis. Furthermore, since the results obtained earlier from the bioinformatic analyses show no CIRCE/HAIR sequences upstream of *cpn60.3*, this upstream region was investigated for the presence of a promoter. However, under the conditions tested, no promoter activity was detected. This does not necessarily mean that there is no promoter activity, and as mentioned earlier, promoter activity could be investigated under various temperatures and growth conditions. The upstream region of *cpn10* showed good promoter activity similar to the positive control (*cpn60.2* upstream region), further confirming that these genes are expressed under normal growth conditions.

In conclusion, in this first chapter, we have looked at the phylogeny of the three homologues in *M. smegmatis* and attempted to assign functions to each of the genes based on their sequences and also on their expression levels under various stresses. Together with previous reports, these studies confirm that Cpn60.2 is the housekeeping chaperonin functioning along with co-chaperonin Cpn10. Cpn60.1 plays some role in the heat shock response, however due to the fact that it is not essential, it would be reasonable to conclude that its main function in the cell is its role in biofilm maturation. The next chapter will focus on uncovering this specialised role of Cpn60.1. Intriguingly,

Cpn60.3 appears to play a role in some stress responses and this role together with other novel functions needs to be investigated further.

4.1 Introduction

Biofilms can be defined as thin coatings of biological material (Karatan and Watnick, 2009). Most bacteria in the environment are found growing as biofilms that are attached to surfaces or at liquid-air interfaces (reviewed in Davey and O'Toole, 2000 and O'Toole *et al*., 2000). Bacteria growing in biofilms have a distinct advantage over planktonically growing cells in that they are protected from environmental conditions and have the capability to survive under low nutrient conditions (Hall-Stoodley *et al*., 2004). Moreover, biofilms help the bacteria evade the effects of antibiotics and antimicrobial agents thus playing a direct role in persistence and virulence (Mah and O'Toole 2001; Stewart and Costerton, 2004; Lewis, 2007). A number of major pathogenic bacteria such as *Pseudomonas aeroginosa, Haemophilus influenza, E. coli*, *Vibrio cholerae* and *Staphylococci* have been shown to form biofilms (reviewed in Hall-Stoodley and Stoodley, 2009). In mycobacteria, the complex cell wall is surrounded by a capsule containing a mixture of proteins, polysaccharides and glycopeptidolipids. This extrapolymeric substance (EPS) which forms the outer layer is responsible for its impermeability to external substances, thus playing an important role in persistence and pathogenesis. Although the link between biofilms and mycobacteria was established many years ago, a study nearly ten years ago was the first to report that the fast growing *M. smegmatis* and slow growing opportunistic pathogen *M. avium*, show sliding motility on a solid surface in a flagella-independent manner (Martínez *et al*., 1999). It was also observed that the mutants defective in sliding motility were defective in biofilm formation and a link between these phenotypes and a deficiency in glycopeptidolipids was proposed (Recht *et al*., 2000).

Mycobacteriophage Bxb1 is a temperate phage of *M. smegmatis* and has been shown to integrate into the 3' end of the *cpn60.1* gene (Kim *et al*., 2003). A study showed that this resulted in loss of biofilm maturation (Kim *et al*., 2003, Ghosh *et al*., 2003). A mutant strain lacking the *cpn60.*1 gene was then created and it was observed that the loss in function of the *cpn60.1* gene did not affect planktonic growth or surface attachment but greatly reduced biofilm maturation (Ojha *et al*., 2005). This loss in maturation was shown to be due to an association of the Cpn60.1 with the KasA protein that is involved with fatty acid synthesis. As explained in chapter 1, KasA which is part of the fatty acid synthase (FAS) II complex is required for the elongation of the meromycolate chain. Thus, the above biofilm results suggest that the composition of the mycolates differ in planktonically growing cells and biofilm cells and for the formation of a mature biofilm, changes in the mycolic acids are necessary. Since Cpn60.1 physically associates with KasA, it was suggested that absence of Cpn60.1 in the cell could indirectly lead to non-assembly of the FASII complex, thereby contributing to a defect in mycolic acid biosynthesis. This biofilm deficient phenotype was complemented when the *cpn60.1* gene was put back into the knock-out strain on a plasmid. According to the published data, a thin pellicle-like layer forms within the first three days and biofilm maturation with the appearance of ridges and troughs on the surface takes four to seven days. According to previous published literature, a biofilm has been described as matrix-enclosed bacterial populations adherent to each other and/or to surfaces or interfaces (Costerton *et al*., 1995). Thus this definition includes biofilm clusters that may be attached to a surface or suspended as flocs in fluids or even present as pellicles at air-liquid interfaces (Hall-Stoodley and Stoodley, 2005). Hence, although the presence of this layer of clustered cells at the air-liquid interface should be referred to as a pellicle,

for reasons of consistency with previous work it will continue to be referred to as a biofilm.

The group in Pittsburgh (Prof. Hatfull's lab, Department of Biological Sciences, University of Pittsburgh, USA) chose to refer to the genes as *groEL1*, *groEL2* and so on. However in keeping with our nomenclature scheme as described by Ellis and Coates in 1993 (Coates *et al*., 1993), we refer to the genes as *cpn60.1*, *cpn60.2* and so on, while the *E. coli* homologue is referred to as *groEL*.

Having determined the expression levels of the *cpn10* and *cpn60* genes under various stress conditions, and confirmed that Cpn60.2 is the housekeeping chaperonin, the next part of this study involved characterising the role of the Cpn60.1 protein in biofilm formation. The aim of this part of the study was to first reproduce the results observed in the paper (Ojha *et al*., 2005) and then determine if this biofilm maturation function was unique to Cpn60.1, or if it could be performed by the other homologues in *M. smegmatis* or homologues from other bacteria. This would be done by cloning various genes (*E. coli groEL, M. smegmatis cpn60.2 and cpn60.3, M. tuberculosis cpn60.1 and cpn60.2*) into the complementing plasmid downstream of *cpn10* and transforming the constructs into the *M. smegmatis Δcpn60.1* strain. Complementing ability was monitored by checking for mature biofilm formation in the complementing strains. The next part of this study would then focus on determining the specific regions of the Cpn60.1 protein responsible for the biofilm maturation. This would be done by making chimeric proteins where regions between Cpn60.1 and any of the complementing homologues would be swapped and complementing ability tested.

4.2 Experimental approach

The strains sent to us were first verified to determine if they were indeed correct and then the biofilm assay was done to determine whether the results obtained previously were reproducible. To begin with, various combinations of genes were cloned onto a plasmid, and a mutant strain of *M. smegmatis* mc^2155 in which the *cpn60.1* has been knocked out was transformed with these constructs. The mutant strain has been reported to be deficient in biofilm maturation and when complemented with a functionally similar protein, biofilm maturation should be restored. Hence, to test for complementation, different genes would have to be cloned into the complementing plasmid and the assay performed as described previously.

However, as will be explained, we encountered a number of difficulties when trying to reproduce the results in our lab. Many variables were altered systematically in an attempt to reproduce the original results, but no consistently reproducible results could be obtained.

4.3 Results

4.3.1 Verification of strains

Before the complementation experiments were started, the strains sent to us by the lab in Pittsburgh were verified using PCR and qRT-PCR. Three independent verifications were done. Firstly, the knock-out strain (which was created by replacing the *cpn60.1* gene with a hygromycin resistance marker) and the complemented strain (containing the complementing plasmid which expressed kanamycin resistance) were grown on media containing the appropriate antibiotics and found to grow as expected. Secondly, PCRs

using various combinations of primers (*cpn60.1* forward and reverse primers, *cpn60.10* forward primer and hygromycin reverse primers) were performed on the wild-type and knock-out strains. Results showed the right sized fragments were obtained with the different primer combinations used. Figure 4.1 shows amplicons of 1.8kb were obtained when *cpn10* forward and hygromycin reverse primers were used on *M. smegmatis* Δ*cpn60.1* (lane 2) and *M. smegmatis* Δ*cpn60.1* with pMsGroEL1 (lane 3). As expected, no product was obtained when wild type *M. smegmatis* was used (lane 1). When *cpn60.1* forward and reverse primers were used, a product can be seen only with *M. smegmatis* (lane 5) and not with *M. smegmatis* Δ*cpn60.1* (lane 6). Thirdly, mRNA was extracted from the knock-out strain and quantitative real time PCR (qRT-PCR) was performed using the various primers described in chapter 3. Results showed that no expression of *cpn60.1* was observed which was consistent with reports that the gene had been deleted (see figure 3.7).

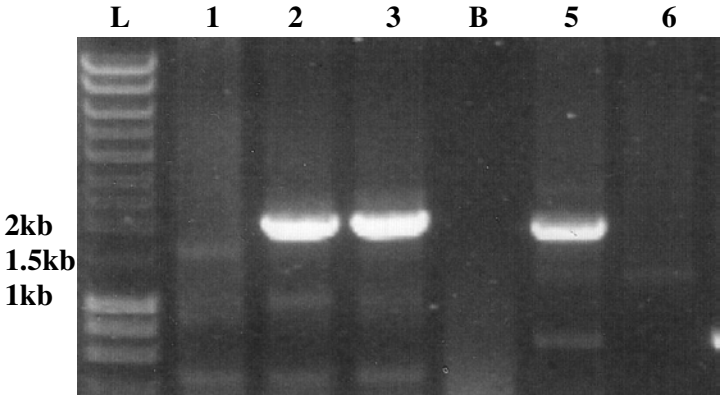


Figure 4. 1: **Confirmation of *M. smegmatis* strains by PCR**
1% agarose gel showing PCR products obtained when using *cpn10* forward and hygromycin reverse primers on *M. smegmatis* (lane 1), *M. smegmatis* Δ*cpn60.1* (lane 2), *M. smegmatis* Δ*cpn60.1* with pMsGroEL1 (lane 3) and water blank (B). When *cpn60.1* forward and reverse primers were used on *M. smegmatis* (lane 5) and *M. smegmatis* Δ*cpn60.1* (lane 6), 1.8kb fragments obtained.

Further, in order to confirm that the lack of biofilm maturation in the mutant strain was not caused due to a slow growing phenotype, a growth curve of the three strains sent to us was performed. As can been observed from the results in figure 4.2, all the three strains grow at similar rates. This confirms that the failure of the mutant to form a mature biofilm is not due to a slow growing phenotype.

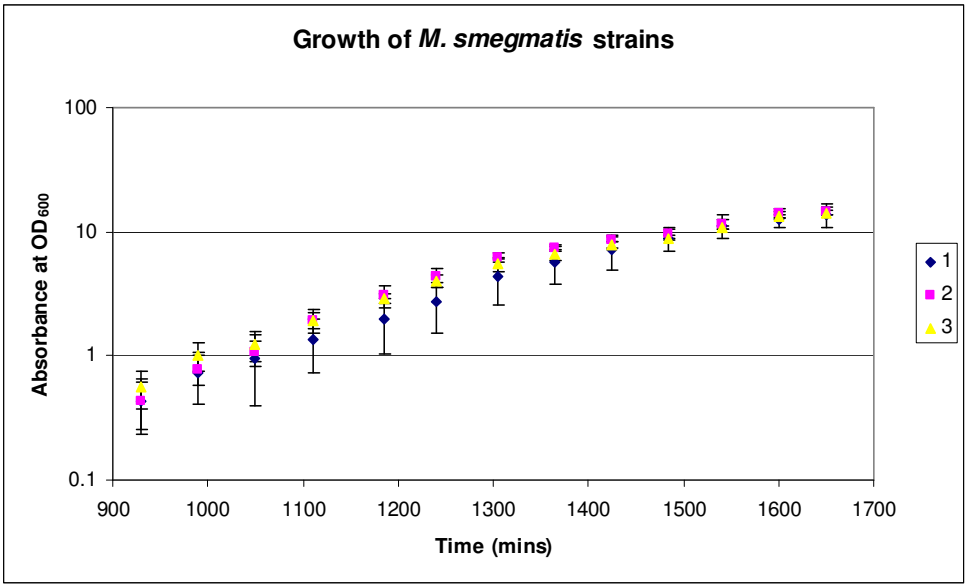


Figure 4. 2: **Comparison of growth between *M. smegmatis* strains**

Growth of wild type *M. smegmatis* (1), *M. smegmatis* Δ*cpn60.1* (2) and complemented strain *M. smegmatis* Δ*cpn60.1* with pMsGroEL1 (3).

4.3.2 Construction of plasmids

Having proved that the strains sent to us were correct, a set of plasmids were constructed to test if any of the homologues of *cpn60.1* were able to complement for loss of biofilm maturation. For this purpose, the original complementing plasmid, pMsGroEL1 used in the previous study was selected as the master plasmid. This extra-

chromosomal shuttle vector is a derivative of plasmid pMV261 and consists of an *E. coli* origin of replication, oriE, a mycobacterial origin of replication, oriM, and a kanamycin resistance marker. Plasmid pMsGroEL1 contains the upstream promoter region of *cpn10*, the *cpn10* gene and *cpn60.1* gene which had been cloned on NheI and KpnI sites.

Since the *cpn60.1* gene was found to be responsible for biofilm maturation, for the purpose of this study, it was decided that only the *cpn60.1* gene would be replaced by the various homologues. A flow diagram briefly outlining the construction strategy is depicted in figures 4.3, 4.4 and 4.5.

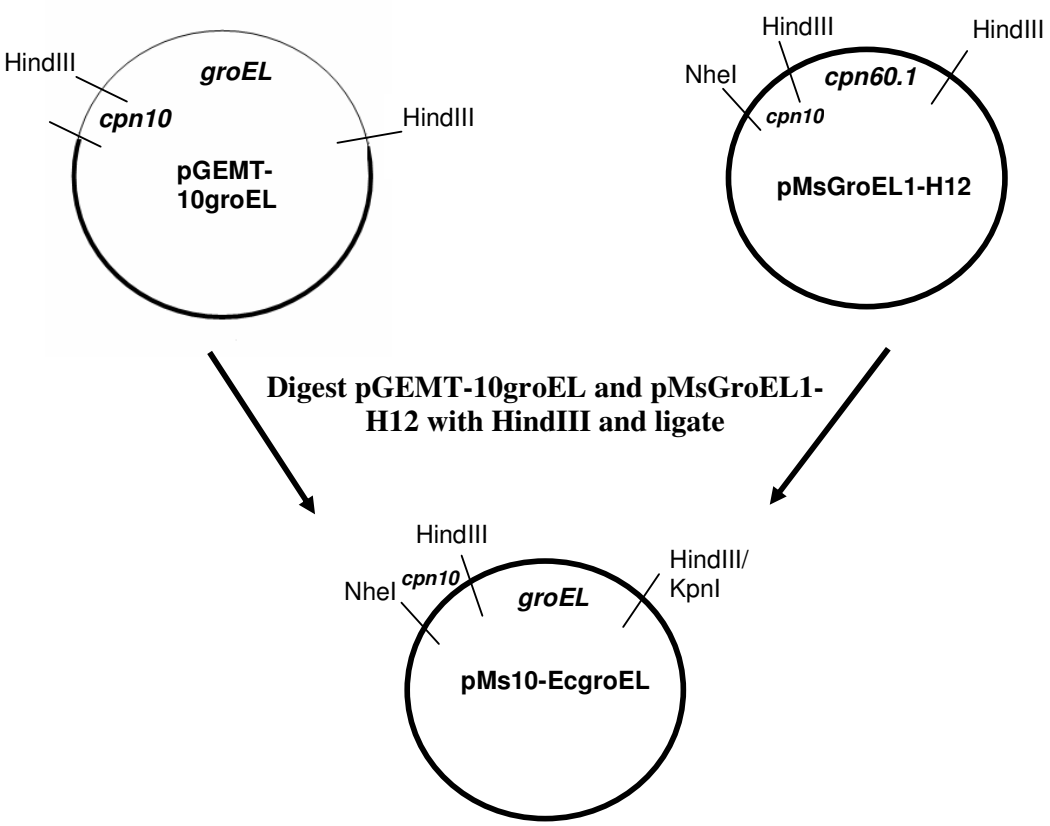


Figure 4. 3: **Schematic representation of construction of pMs10-EcgroEL**

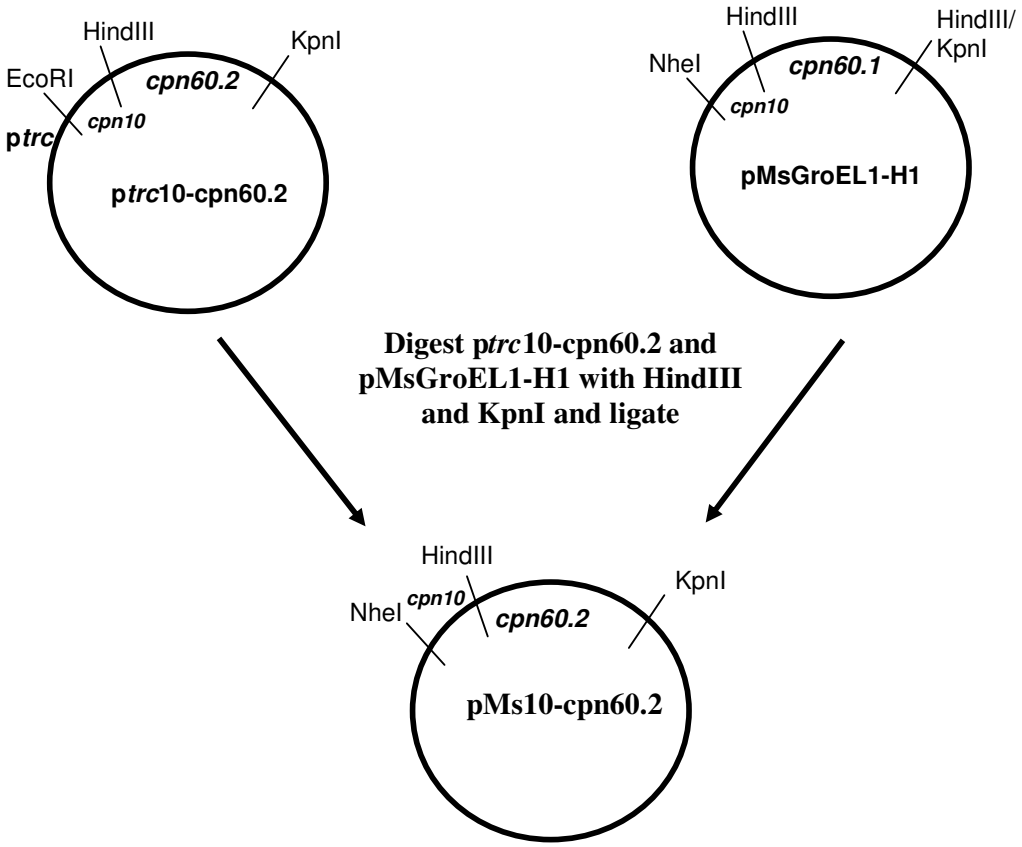


Figure 4. 4: Schematic representation of construction of pMs10-cpn60.2

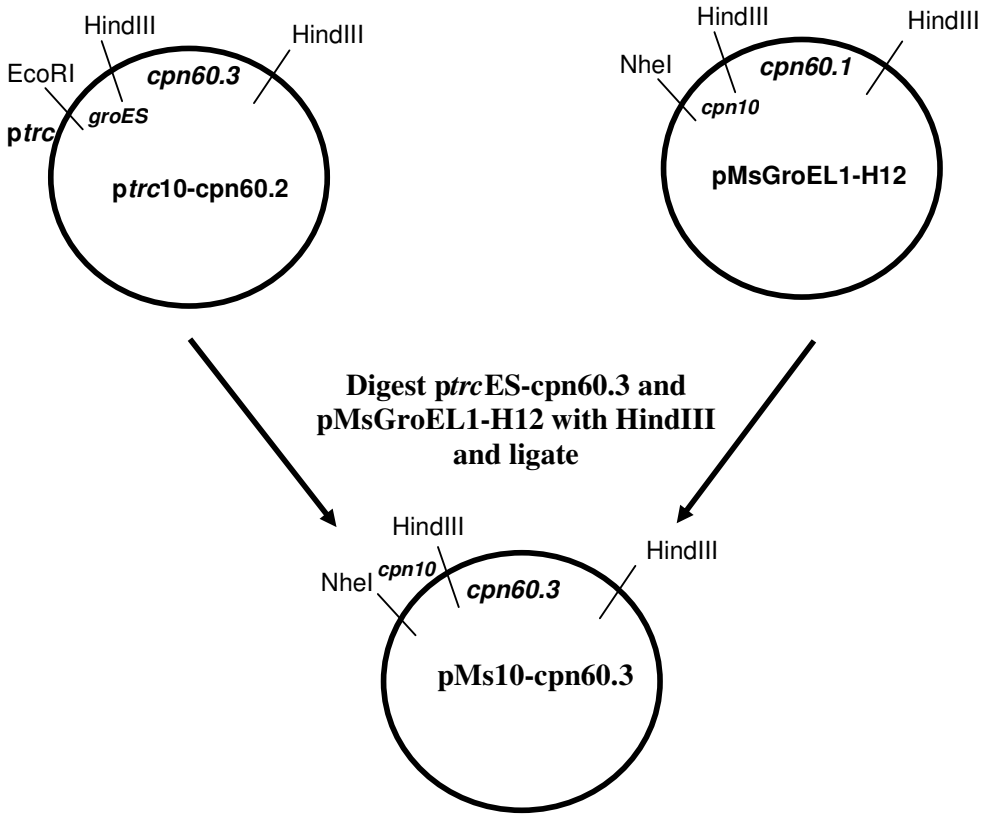


Figure 4. 5: **Schematic representation of construction of pMs10-cpn60.3**

In order to clone the genes, the original plasmid first had to be modified to include restriction sites that could be used to sub-clone the genes. For this purpose, additional restriction sites were added using site directed mutagenesis (SDM). Plasmid pMsGroEL1-H1 contained a HindIII site upstream of the *cpn60.1* gene, while plasmid pMsGroEL1-H2 contained a HindIII site downstream of the *cpn60.1* gene. Plasmid pMsGroEL1-H12 contained HindIII sites upstream and downstream of the *cpn60.1* gene. The *E. coli groEL* gene from plasmid pGEMT-10groEL which was constructed earlier was cloned into pMsGroEL1-H12 on HindIII sites replacing *cpn60.1* to give plasmid pMs10-EcgroEL. The *cpn60.3* gene from plasmid p*trc*ES-cpn60.3 was also cloned into pMsGroEL1-H12 using HindIII sites and the resulting plasmid was named

pMs10-cpn60.3. The *cpn60.1* gene from plasmid pMsGroEL1-H1 was replaced by the *cpn60.2* gene from plasmid p*trc*10-cpn60.2 using HindIII and KpnI sites to give plasmid pMs10-cpn60.2. The mutant strain, in which the *cpn60.1* gene is deleted, was then transformed with the various constructs and the biofilm assay performed as described earlier.

4.3.3 Biofilm assay

For the biofilm assay a single colony of each of the strains of *M. smegmatis* to be used in the biofilm assay was inoculated into 5ml of 7H9 broth containing ADC and antibiotics for 36 hours and then sub-cultured into 20ml of the same medium and grown to mid-log phase. 10µl of the mid-log phase culture was then inoculated into 10ml of sterile re-constituted biofilm media (see chapter 2 for recipe), and incubated in a static 30°C incubator for 7 days. Biofilm formation at the air-liquid interface started between day 2 and 3, while biofilm maturation, characterised by the formation of ridges and troughs on the surface, became apparent by day 4-5. The plates were monitored for growth of the biofilm and pictures were taken at appropriate times. At the end of the 7 day period, in some instances, the biofilm was harvested, and the mass of the biofilm measured to give more quantitative results.

As mentioned earlier, the genes to be tested for complementation were cloned in the original complementing plasmid by replacing the existing *cpn60.1* gene. For reasons of simplicity, the various strains used in the biofilm assay will henceforth be referred to using their assigned numbers as presented in table 4.1.

Strain number	Strain name	Co-chaperonin	Chaperonin
1	*M. smegmatis* mc^2155		
2	*M. smegmatis* Δ*cpn60.1*		
3	*M. smegmatis* Δ*cpn60.1* with pMsGroEL1	MsCpn10	MsCpn60.1
4	*M. smegmatis* Δ*cpn60.1* with pMs10-EcgroEL	MsCpn10	EcGroEL
5	*M. smegmatis* Δ*cpn60.1* with pMs10-cpn60.2	MsCpn10	MsCpn60.2
6	*M. smegmatis* Δ*cpn60.1* with pMs10-cpn60.3	MsCpn10	MsCpn60.3
7	*M. smegmatis* Δ*cpn60.1* with pMs10-tbcpn60.1	MsCpn10	MtbCpn60.1

Table 4. 1: Strains used in biofilm assay along with their corresponding numerical reference

The first aim of the study was to replicate the results produced by the group in Pittsburgh and then test our various constructs for complementation. Unexpectedly, it proved extremely difficult to get the biofilm assay to work in our lab here in Birmingham. A number of various conditions were tested and many variables altered. While we got the assay to work in one instance, there was no consistency in the results. Each of the attempts with the various altered conditions tested along with pictures of the biofilms and masses in some cases will be reported below.

4.3.3.1 Typical biofilm result

According to the results published by Ojha *et al*. (2005) while not much difference in the biofilms of the wild-type and *cpn60.1* knock-out were observed at day 2-3, a marked difference in maturation patterns was observed from day 4-5 onwards. While the wild-type appeared to form a matured biofilm, the knock-out failed to show maturation. Figure 4.6 shows a typical result obtained by them in a routine experiment.

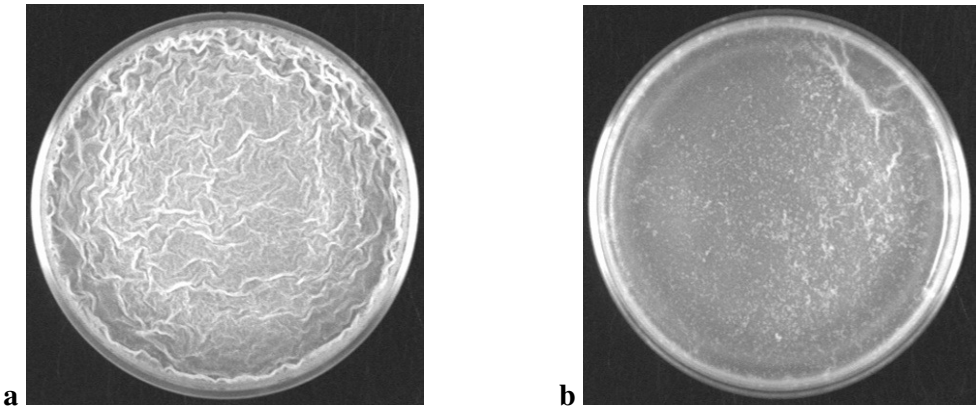


Figure 4. 6: **Typical 5 day biofilm result**

Biofilm assay showing a typical 5 day biofilm of wild-type *M. smegmatis* (a) and *M. smegmatis* Δ*cpn60.1* (b) (Image from Anil Ojha).

4.3.3.2 Initial attempts of biofilm assay

In order to first replicate the results reported in the paper by Ojha *et al*. (2005) the three strains 1, 2 and 3 were grown up as described and inoculated into the biofilm media. The plates were analysed on day 4, and as can be observed from figure 4.7, weak maturation was observed in strains 1 and 3 while strain 2 failed to show any biofilm maturation as expected. However, by day 7, strain 2 showed faint maturation and no striking difference between the three strains could be observed (figure 4.8).

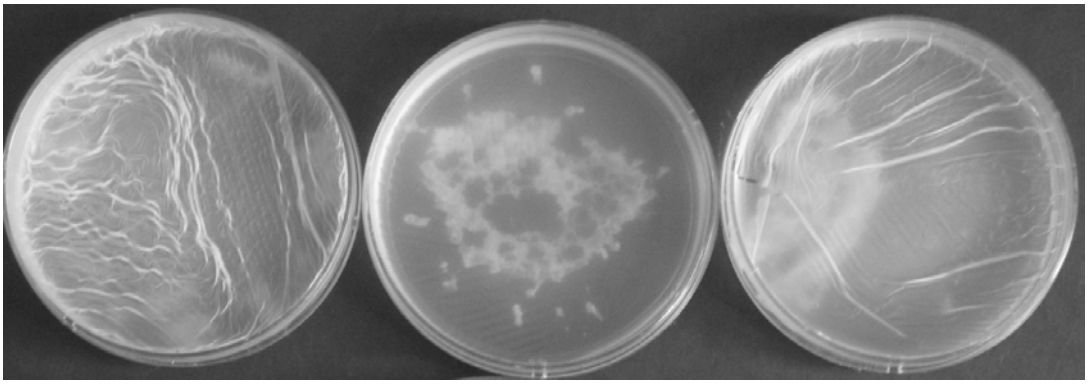


Figure 4. 7: **4 day biofilm result obtained in our lab**
4 day biofilm result showing strains 1 (left), 2 (middle) and 3 (right).

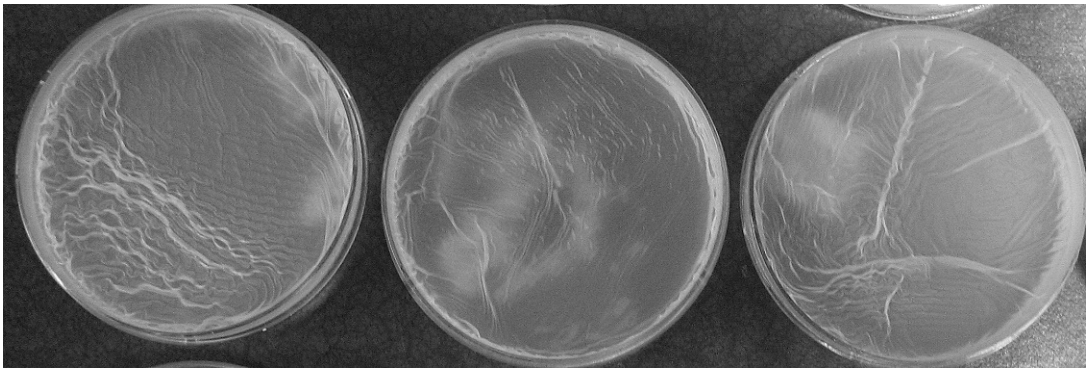


Figure 4. 8: **7 day biofilm result obtained in our lab**
7 day biofilm result showing strains 1 (left), 2 (middle) and 3 (right).

4.3.3.3 Growth stage of inoculum

Since strain 2 also showed slight maturation by day 7 and no clear difference between the three strains could be observed, a series of carefully controlled experiments where parameters that might affect biofilm maturation were systematically altered. Firstly, the biofilm media was inoculated with cultures at different stages of growth. Cultures that were grown up to an OD_{600} of 0.3, 0.5 and 0.7 were used. The inoculating method was also varied and in one case, 10µL of the inoculum was mixed with 10mL of the biofilm media in a sterilin and then added to the Petri dishes, while in another case the 10µL inoculum was added directly to the aliquoted biofilm media in the Petri dish. However,

when the biofilms were analysed on day 5, no differences could be observed between the different innocula (data not shown).

4.3.3.4 Different growth media

The next factor altered was the medium used to grow the strains before inoculation, which we changed from Middlebrook 7H9 + ADC to Tryptic Soy Broth (TSB). However, as can be seen in figure 4.9, a very different phenotype was observed comprising of a distinct pattern of white dots on the surface of the biofilm possibly due to the formation of clumps of cells.

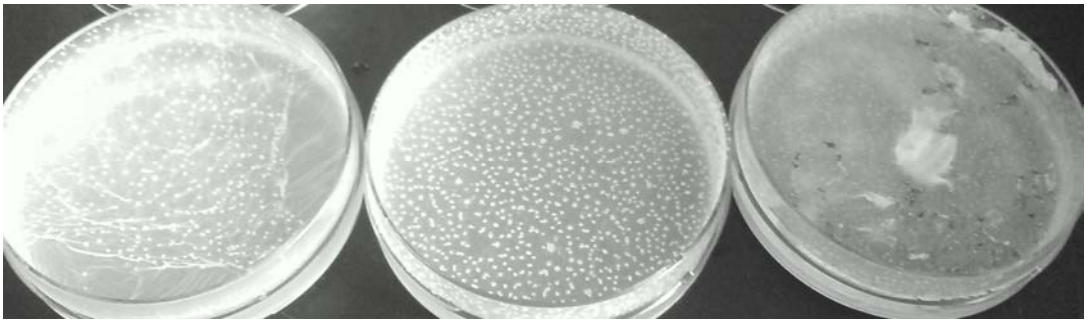


Figure 4. 9: **5 day biofilm result from starter cultures grown in Tryptic Soy broth**
5 day biofilm result showing strains 1 (left), 2 (middle) and 3 (right), where the inoculum was grown in Tryptic Soy Broth.

4.3.3.5 New strain of *M. smegmatis* Δ*cpn60.1*

Although changing the growth media produced a completely different phenotype, it did not produce the expected result of a mature biofilm (strains 1 and 3) or a defective biofilm (strain 2). It was speculated that a suppressor mutation might have arisen in the strain that was able to suppress the original phenotype. To address this issue, we requested the group in Pittsburgh to send us a new aliquot of knockout strain 2. An

aliquot of the strain was sent to us as a liquid glycerol stock and once the culture was streaked out, single colonies were picked and the assay started. The assay was performed simultaneously in our lab and in the lab in Pittsburgh using the same strain. However, when compared, the newly sent strain 2 still formed mature biofilms in our lab, while mature biofilms were not observed in Pittsburgh (figures 4.10 and 4.11).

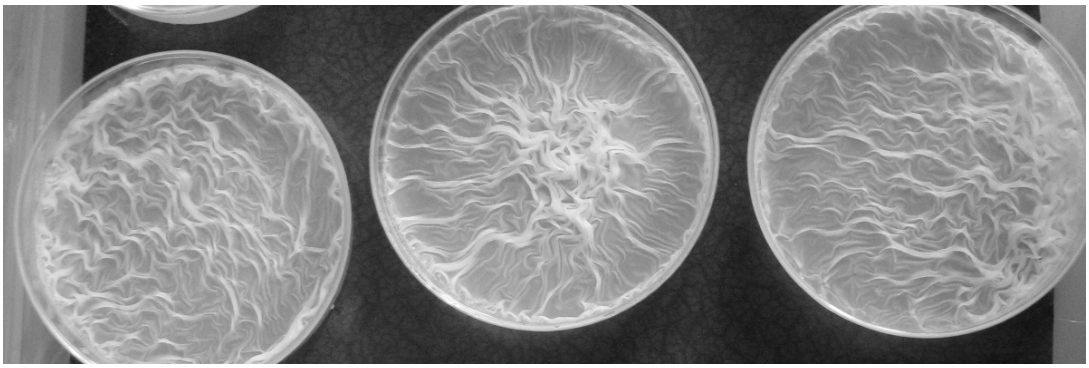


Figure 4. 10: **5 day biofilm result obtained in our lab using new strain**
5day biofilm result obtained in our lab using strain 1 (left), newly sent strain 2 (middle) and strain 3 (right).

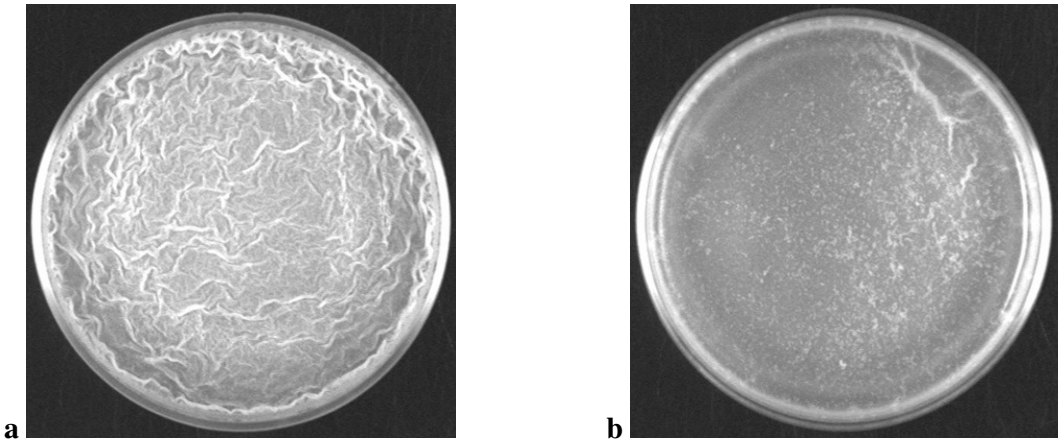


Figure 4. 11: **5 day biofilm result obtained in Pittsburgh lab using new strain**
5 day biofilm result obtained from the lab in Pittsburgh using strain 1 (a) and the same strain 2 as above (b).

4.3.3.6 HPLC grade water

As we were still unable to reproduce the results which were seen in Pittsburgh, it was hypothesized that minor differences in the water used to make up the media might account for the differences. For this purpose, biofilm media and all the various constituents were prepared using HPLC grade water. This was done to rule out any contaminating ingredient that may have been present in the milliQ pure water that was previously being used. However, results obtained showed that even by day 4, strain 2 began to show a matured biofilm. As can be seen in Figure 4.12, even when 2 different colonies of strain 2 were used mature biofilm still resulted from this strain.

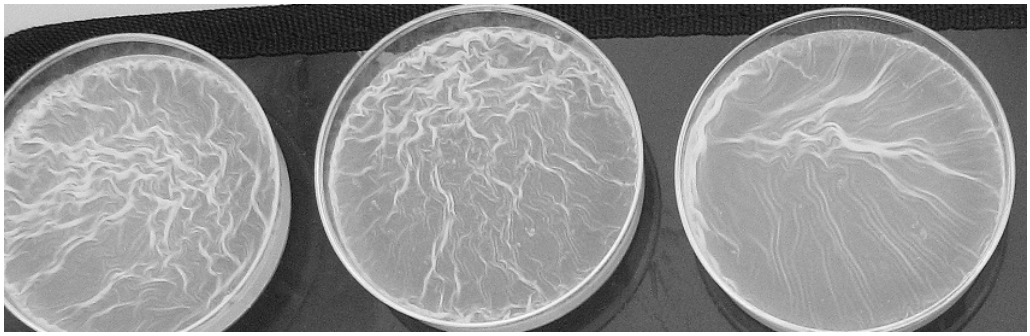
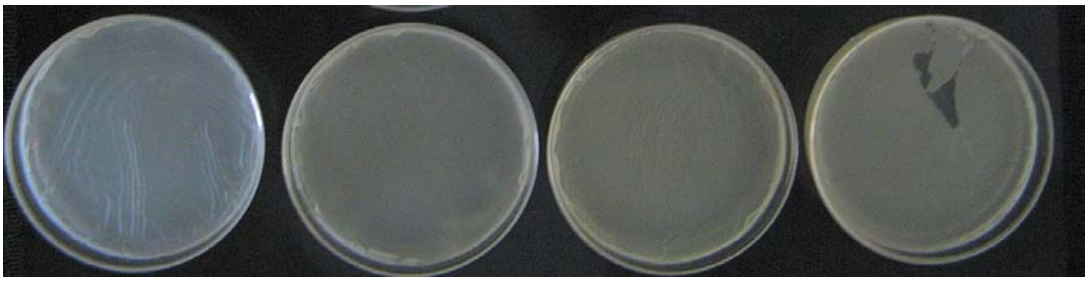


Figure 4. 12: **4 day biofilm result using HPLC water**
4 day biofilm result showing strains 1 (left), 2 (middle) and 3 (right), where the biofilm media and all the various components were prepared using HPLC grade water.

4.3.3.7 Biofilm media from Pittsburgh lab

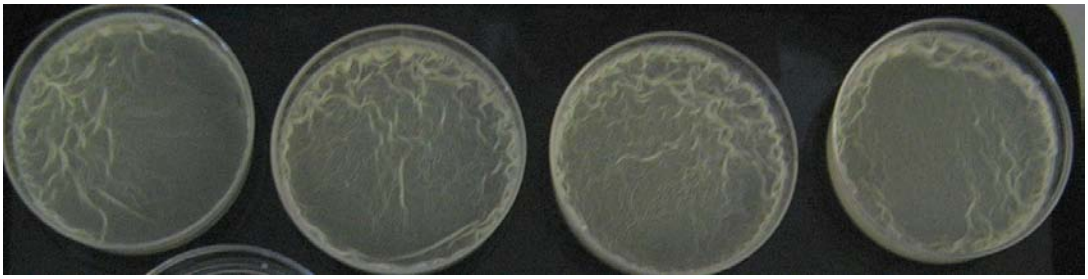
Next, to test the hypothesis that differences in the media between the two labs was contributing to the formation of a mature biofilm by strain 2, we obtained biofilm media from the lab in Pittsburgh. Sterile reconstituted biofilm media was sent to us, and the biofilm assay performed. Since the new aliquot of strain 2 was previously sent as a liquid glycerol stock, we hypothesised that a suppressor mutation could have arisen

during transit. To rule out the fact that the liquid stock culture of strain 2 had acquired a mutation during transit, a new stock of knock-out strain 2 streaked out on a plate was sent to us along with the re-constituted media. At this point, it was also observed that the lab in Pittsburgh did not subculture the inoculum. So, a slight change in our protocol meant that the single colony that was used for the starter culture was grown for about 36 hours till it reached OD_{600} 0.8 and was then directly inoculated into the biofilm media. Simultaneously, an experiment was done using media made in our lab as well. Figure 4.13 shows the results obtained from the assay using media from Pittsburgh, while figure 4.14 shows the results from the assay using media made in our lab. From the results, it can be seen that the assay did not work when media from the other lab was used. This could be due to the media components not withstanding the transit conditions. However, as can be seen for image c in figure 4.14, even the strain sent on a plate showed a mature biofilm by day 5.



a **b** **c** **d**

Figure 4. 13: **5 day biofilm result using Pittsburgh biofilm media**
5 day biofilm result using biofilm media sent by Pittsburgh lab, showing strain 1 (a), strain 2 sent as a culture (b), strain 2 sent on plate (c) and strain 3 (d).

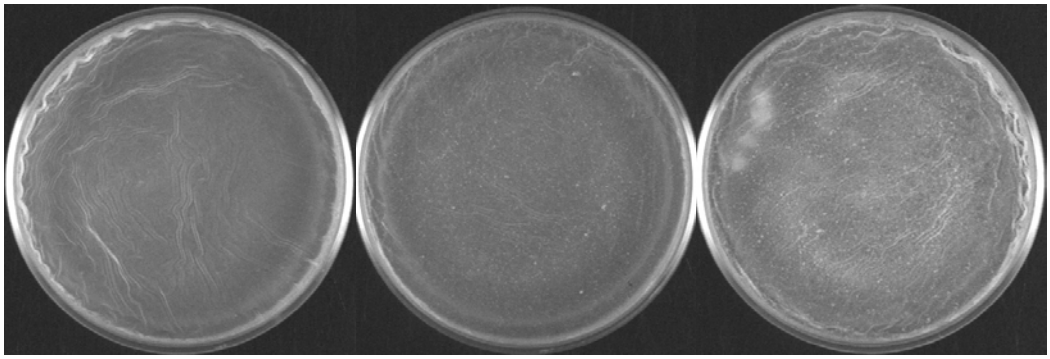


a **b** **c** **d**

Figure 4. 14: **5 day biofilm result using our biofilm media**
5 day biofilm result using biofilm media prepared in our lab, showing strain 1 (a), strain 2 sent as a culture (b), strain 2 sent on plate (c) and strain 3 (d).

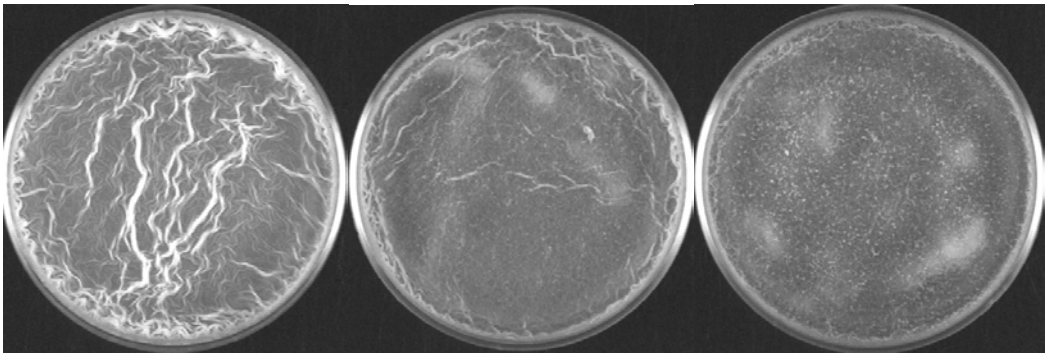
4.3.3.8 Our media to Pittsburgh

Since no biofilm maturation was observed when the assay was done using the media sent to us by the Pittsburgh lab, some of our media as well as the knock-out strain 2 was sent to their lab. The results obtained from the experiment conducted in their lab using our media showed that strain 1 failed to show significant biofilm maturation (figure 4.15a) and resembled their strain 2 (figure 4.15b) as well as the strain 2 sent by us (figure 4.15c). This could be attributed to the fact that the media may have been altered during transit (freezing conditions). Simultaneously, the same set of experiments was conducted using media prepared in their lab. The results obtained were as expected with strain 1 showing maturation (figure 4.16a) and both the strain 2s failing to show maturation (figure 4.16 b and c). These results were also confirmed when the mass of the biofilms was measured (figure 4.17).



a **b** **c**

Figure 4. 15: **5 day biofilm result conducted in Pittsburgh lab using our biofilm media**
4 day biofilm result, conducted in the Pittsburgh lab using our media and strain 1 (a), their strain 2 (b) and the strain 2 sent by us (c).



a **b** **c**

Figure 4. 16: **5 day biofilm result conducted in Pittsburgh lab using their biofilm media**
4 day biofilm result conducted in the Pittsburgh lab using their media and strain 1 (a), their strain 2 (b) and the strain 2 sent by us (c).

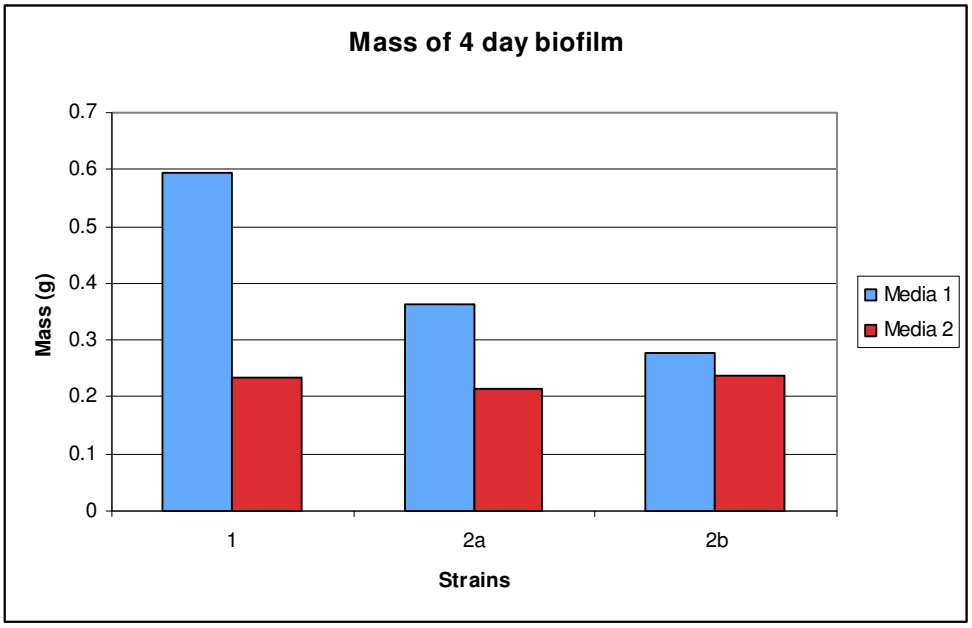
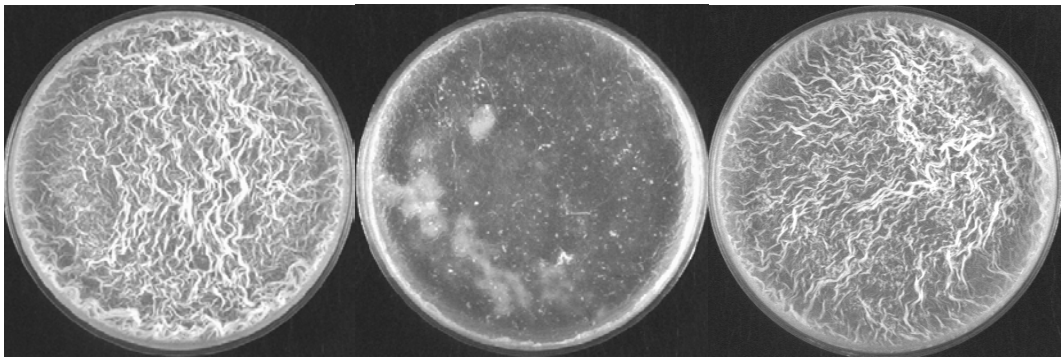


Figure 4. 17: **Comparison of the masses of biofilms using different biofilm media**
Graph showing the mass of 4 day biofilms using their media (blue bars) and our media (red bars) with strain 1 (1), their strain 2 (2a) and our strain 2 (2b).

4.3.3.9 Biofilm assay in Pittsburgh

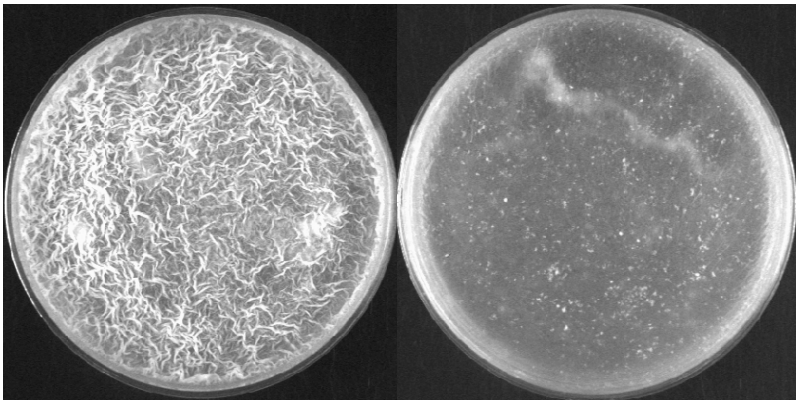
From these experiments, we were unable to identify any variable which, when altered, led to successful reproduction of the results from the Pittsburgh lab. Thus, it was decided that I was to visit the lab in Pittsburgh and perform the assay in the lab there. This was to determine if there were any subtle differences between the constituents or method for the biofilm assay. The assay was performed using their media as well as media made by me in the lab there. As can be observed in figures 4.18 and 4.19, the assay worked with both media used. There were no obvious differences in the method of media preparation. When the mass of 6 day biofilms were measured there was a marked difference between strain 1 and strain 2, irrespective of which media they were grown in (figure 4.20). Since the assay worked with media made by me, we tested one

of our test strains, strain 4 which contains the *E. coli groEL* gene on the complementing plasmid. From figure 4.18c it can be seen that the *groEL* gene can complement the loss in function of *cpn60.1*.



a **b** **c**

Figure 4. 18: **6 day biofilm result conducted in Pittsburgh lab using media made by me**
A day 6 biofilm of strain 1 (a), strain 2 (b) and strain 4 (c). The experiment was conducted in Pittsburgh, with biofilm media made by me.



a **b**

Figure 4. 19: **6 day biofilm result conducted in Pittsburgh lab using their biofilm media**
A day 6 biofilm of strain 1 (a) and strain 2 (b). The experiment was conducted in Pittsburgh, with biofilm media prepared by their lab.

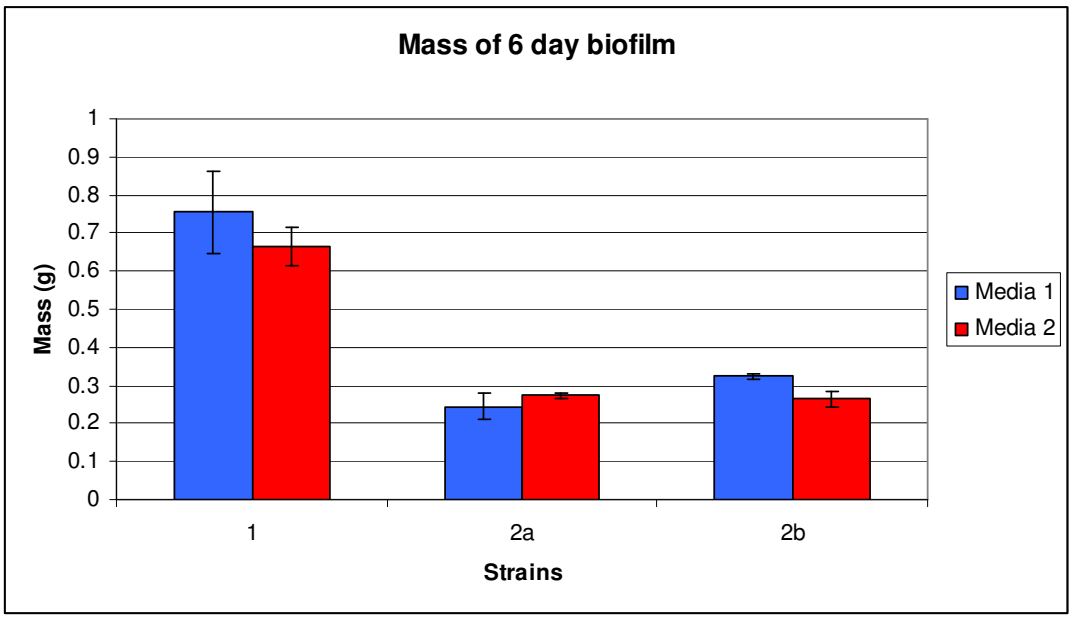


Figure 4. 20: **Comparison of the masses of biofilms using different biofilm media**
Graph showing a comparison of the mass of the biofilms of strain 1, 2a and 2b in their media (blue bars) and media prepared by me in Pittsburgh (red bars). The results represent the average of two biological replicates and the error bars represent the standard deviation.

4.3.3.10 Addition of Tween80

During the trip to their lab, one subtle difference between the methods was noticed. In their lab, cultures that were grown up contained Tween80, while in our lab in Birmingham, the cultures to be used for inoculation did not contain the Tween80 in the media. So an experiment in which Tween80 was left out of the culture media was set up. However as can be seen in figure 4.21, the lack of Tween80 did not make any difference and strain 2 still showed a lack in biofilm maturation.

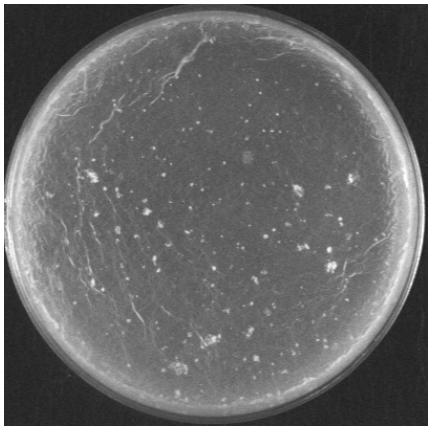
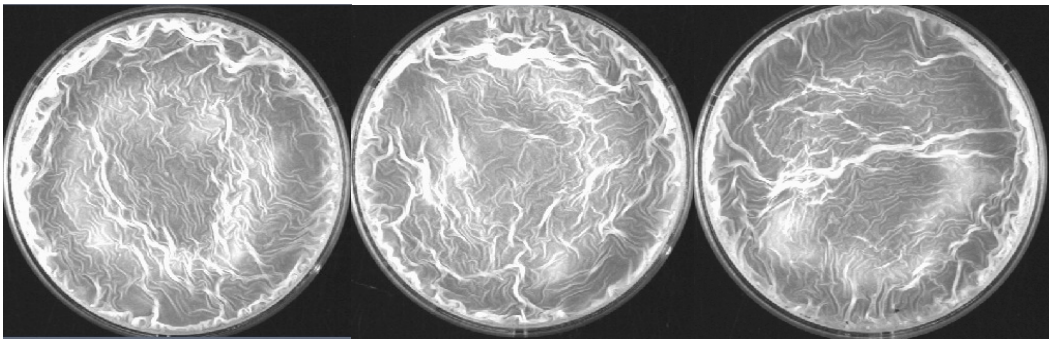


Figure 4. 21: 6day biofilm result of *M. smegmatis*Δ*cpn60.1* strain

A day 6 biofilm result of strain 2, set up from cultures not containing Tween80.

4.3.3.11 On return to Birmingham

However, on return to Birmingham, when the biofilm assay was attempted once again, the same result as before was obtained. Once again, strain 2 formed a mature biofilm similar to strain 1 (figure 4.22) and the three strains could not be visually distinguished. When the mass of the biofilms was measured, no marked difference between the three strains was observed (figure 4.23).



a **b** **c**

Figure 4. 22:7 day biofilm result conducted in our lab

7 day biofilm result with strain 1 (a), strain 2 (b) and strain 3 (c).

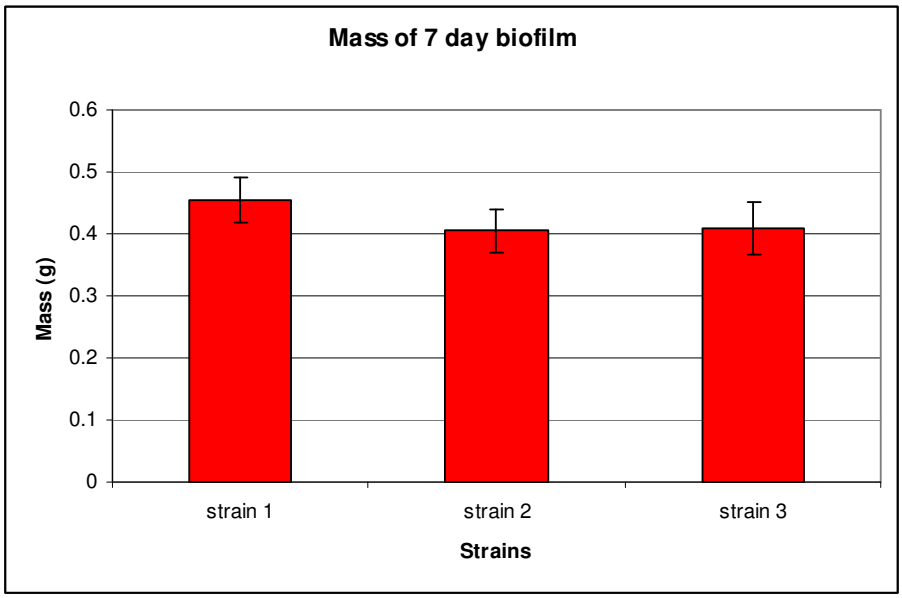


Figure 4. 23: **Masses of 7 day biofilm conducted in our lab**
7 day biofilm result with strain 1, 2 and 3. The assay was conducted in our lab. The results represent the average of three biological replicates and the error bars represent the standard deviation.

4.3.3.12 Iron concentrations

Iron is a crucial component in the biofilm assay (Ojha *et al.*, 2007), so we hypothesised that biofilm formation might be affected by minor differences in iron concentration. Originally the concentration of iron (Fe^{3+}) used was 2µM, however, under instructions from the Pittsburgh lab, the concentration was increased to 4µM. To determine whether alteration of Fe^{3+} concentration has any affect on biofilm maturation, media containing varying concentrations of iron from 0µM to 20µM were prepared and the biofilm assay performed. The typical mass of a 5 day biofilm for strain 1 is between 0.8g and 1.2g, while the mass of a 5 day biofilm of strain 2 is normally in the range of 0.2g - 0.4g (Ojha, Pers. Comm.). Furthermore, it has been observed that the texture of the biofilms

of strain 1 and 2 are different in that strain 1 forms intact biofilms, while strain 2 forms very thin biofilms which disintegrate very easily upon harvesting (Ojha, Pers. Comm.). From the results obtained, it is obvious that the masses of the biofilms from our lab differ greatly from the expected biofilm masses and interestingly strain 1 and 2 formed intact biofilms which did not disintegrate. From the masses of the 7 day biofilms, it can be observed that biofilm maturation is indeed dependant on the concentration of iron in the media, and a minimum concentration of 1μM is essential for complete biofilm maturation (figure 4.24). Thus it can be concluded that while iron is essential for biofilm maturation, it has no effect on the unusual phenotype of the mutant.

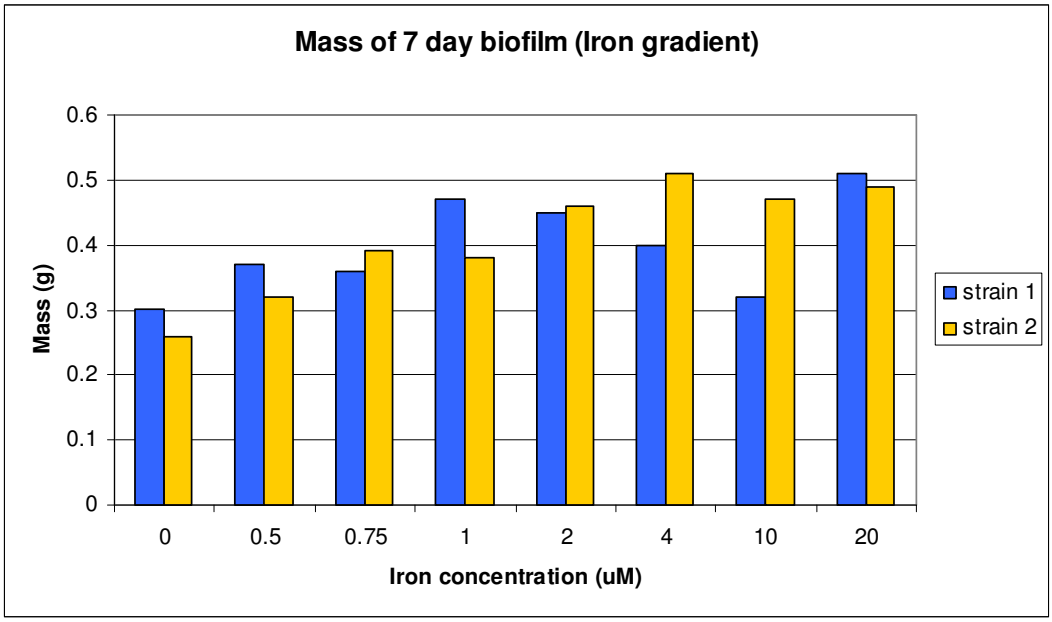


Figure 4. 24: **Effect of iron concentration on biofilm formation**

Mass of 7 day biofilms of strain 1 (blue bars) and strain 2 (yellow bars), grown in biofilm media containing varying concentrations of iron.

4.3.3.13 Calcium is not essential for biofilm maturation

We then decided to determine if Calcium was required for biofilm maturation. Biofilm media containing no calcium was prepared and the biofilm assay performed. As can be observed in figure 4.25, strain 1 forms a mature biofilm in media containing no calcium as does strain 2. This result confirms that calcium is not essential for biofilms formation and maturation.

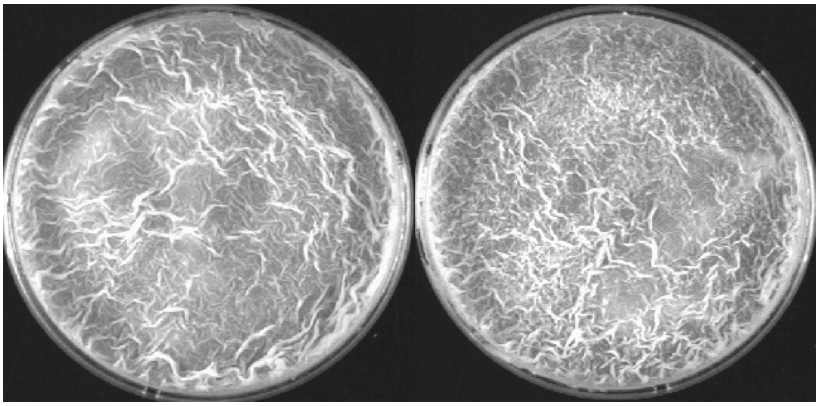


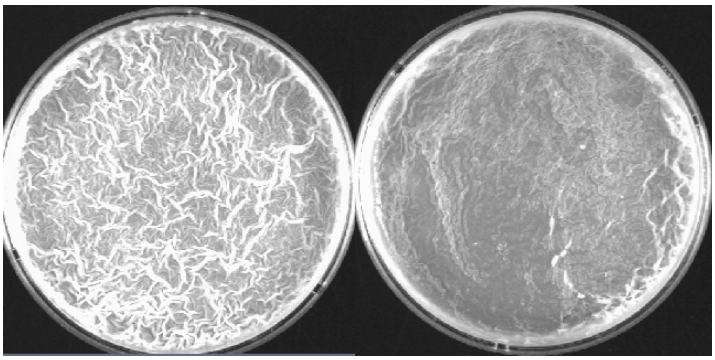
Figure 4. 25: **Calcium is not essential for biofilm formation**

5 day biofilm of strain 1 (left) and strain (right), grown in biofilm media containing no calcium.

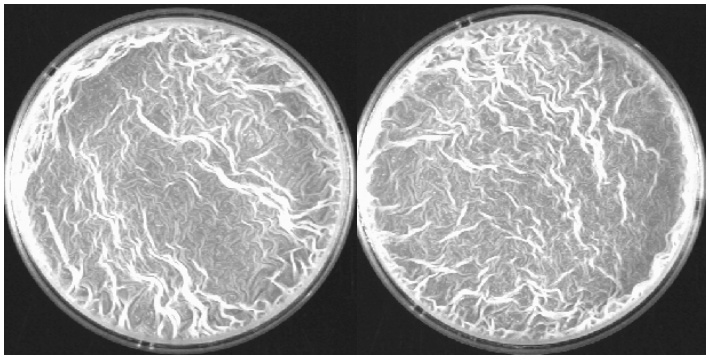
4.3.3.14 Glassware

Up to this point, the glassware used to make the biofilm media and grow the cultures was being washed by the communal washing system in our department. To rule out any possibility of a contaminating substance in the communally washed glassware, new glassware was purchased, washed and autoclaved by us and the biofilm media prepared. Even the cultures to be used in the biofilm assay were grown in plastic sterilins. The initial results were encouraging as the biofilm assay gave the expected result for the first

time (figures 4.26 and 4.27). As we were able to conduct the assay and obtain two independent replicates which worked, we tested strain 4, which previously formed a mature biofilm. As can be observed, strain 4 shows a mature biofilm indicating that the *E. coli groEL* homologue can complement for the loss in function of the *cpn60.1* gene as was seen in the identical experiment done in Pittsburgh (see figure 4.18c).



a **b**



c **d**

Figure 4. 26: **7 day biofilm result using new glassware conducted in our lab**

7 day biofilm result of strain 1 (a), strain 2 (b), strain 3 (c) and strain 4 (d) using biofilm media prepared from newly brought glassware.

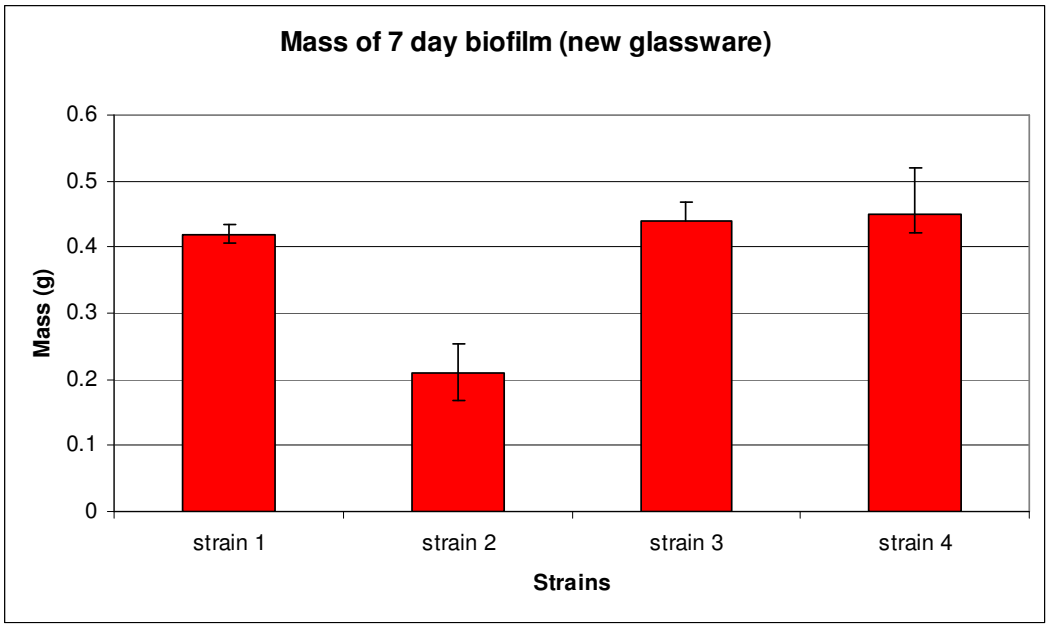


Figure 4. 27: **Masses of 7 day biofilm using new glassware conducted in our lab**

Mass of 7 day biofilms of strains 1, 2, 3 and 4. The assay was conducted with media prepared in new glassware, and the results represent the average of two biological replicates and the error bars the standard deviation.

4.3.3.15 Testing the various strains

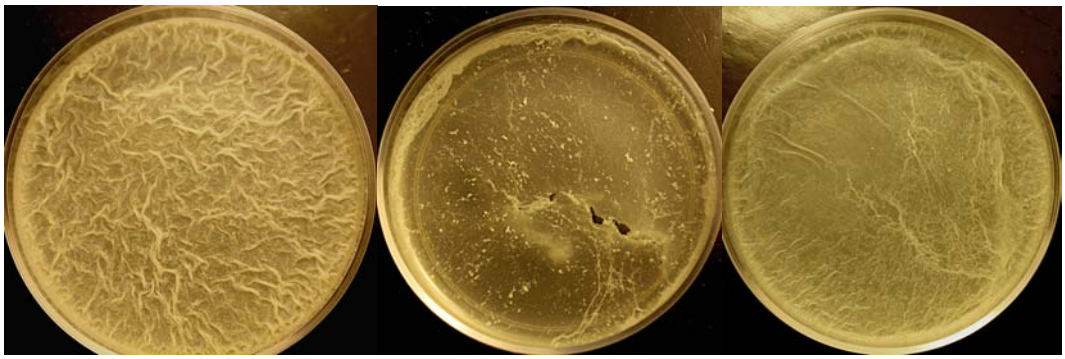
However, after getting the biofilm assay to work in our lab on two separate occasions using new glassware, the assay stopped working once again with strain 2 forming a mature biofilm. At this point it was decided that due to the problems encountered in getting the assay to work reproducibly in Birmingham, future assays would be conducted by the group in Pittsburgh. To check the reliability of this approach, several strains were sent to the Pittsburgh lab blind, i.e., the lab were not aware of the identity of the strains that they were testing. These were the test strains 4, 5 and 6 containing *E. coli groEL*, *cpn60.2* and *cpn60.3* respectively downstream of *cpn10* on the

complementing plasmid. The *M. tuberculosis cpn60.1* gene was also cloned (Tabish Ahmed, unpublished) into the complementing plasmid and the strain 2 transformed with this plasmid to test if the *M. tuberculosis* homologue could functionally complement the *M. smegmatis cpn60.1*. Since the bioinformatic analyses (chapter 3) had shown all the *cpn60.1* genes to be similar based on their sequences, this experiment aims to prove if there is a link between the sequence homology and functional role of these proteins within the cell.

These experiments were conducted in Pittsburgh by our collaborator and results obtained will be presented. Since the strains to be tested for the biofilm assay were sent blind, the first question to be answered was whether the Pittsburgh group could correctly identify strains 1, 2 and 3. From the results sent to us, as expected, they were indeed able to correctly identify strain 1 (figure 4.28a) and strain 3 (figure 4.28c) which showed mature biofilms and also identified strain 2 (figure 4.28b) as growing poorly. Having showed that not only do our strains provide the expected results, but that also that they were able to identify the positive (strain 1 and 3) and negative (strain 2) controls, the results obtained with the test strains were analysed. According to their analyses, strain 5 (containing *cpn60.2*), did not show any complementation and resembled the negative control strain 2 (figure 4.28e), while strains 4 and 6 (*E. coli groEL* and *cpn60.3* respectively) showed partial complementation (biofilms were in between positive and negative control) (figure 4.28 d and f). However, the most significant result is that the strain containing the *M. tuberculosis cpn60.1* gene on the complementing plasmid was able to complement loss of *M. smegmatis cpn60.1* and form a mature biofilm (figure 4.28g). This confirms our hypothesis that the sequence

conservation between the *cpn60.1* genes from *M. tuberculosis* and *M. smegmatis* does play a role in protein function in the cell. However, as can be observed from the results obtained, strains 3 (figure 4.28c) and 7 (figure 4.28g) do not form a luxuriant mature biofilm as observed in the positive control strain 1 (figure 4.28a) and this could be due to the levels of expression of these genes which are present on the plasmid.

In order to confirm that lack of complementation was not due to lack of protein expression, the expression of the Cpn60 homologues from each of the test strains was monitored. Protein was extracted from an equal volume of cells and results showed that there was no observable difference in the expression of a 60kDa protein among the various strains (figure 4.29). However, the expression of the proteins was monitored from planktonically grown cultures and may not necessarily be an accurate indication of protein expression during biofilm growth.



a **b** **c**

d **e** **f**

g

Figure 4. 28: **5 day biofilm result of all test strains conducted in Pittsburgh lab**

5 day biofilm result of assay conducted in Pittsburgh lab with strain 1 (a), strain 2 (b), strain 3 (c), strain 4 (d), strain 5 (e), strain 6 (f) and knock-out strain 2 containing pMs10-tbcpn60.1 (g).

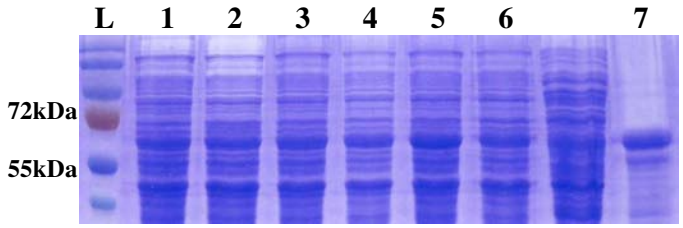


Figure 4. 29: **Expression of Cpn60 proteins from test strains used in biofilm assay**
Expression of a 60kDa protein from strains 1 (1), 2 (2), 3 (3), 4 (4), 5 (5) and 6 (6). The expected size of the protein can be compared to the GroEL protein (7).

4.3.4 Colony Morphology

Since the biofilm assay proved particularly difficult to replicate in our lab, we looked for alternative assays to test the mutant phenotype. Initially, we decided to look at the colony morphology of the various strains to determine if there were any differences and if this could be used as an alternative assay.

For this assay, cultures were grown up as described in chapter 2, the OD_{600} of the cultures were measured and an equal number of cells were spotted onto different media and incubated at 37°C for 5 days. On day 5, the colonies were photographed using a camera attached to a microscope.

The strains were spotted on LB agar, Middlebrook 7H10 agar, Middlebrook 7H11 agar and Tryptic Soy agar with or without Tween80. In some instances, an equal amount of cells were taken and the culture streaked and pictures of single colonies taken.

4.3.4.1 Colony morphology on different media

Since there was no obvious difference between the strains when they were streaked out, we attempted to determine if there were any differences present when cells were spotted onto solid media while looking at morphology on various different types of solid media as well.

The *M. smegmatis* and *M. smegmatis* Δ*cpn60.1* strains were spotted onto Middlebrook 7H11 agar, Middlebrook 7H10 agar, LB agar or Tryptic Soy agar with or without Tween80. Colony morphology was observed after 5 days of incubation at 37°C.

When the colony morphology of the two strains grown on Middlebrook 7H10 agar was observed, both strains showed good growth. The colonies of both strains grown on media without Tween80 showed irregular outlines with rugose surfaces, while colonies grown on media containing Tween80 showed smooth flat edges with raised rugose centres (figure 4.30).

When the growth of the two strains were compared on Middlebrook 7H11 agar, the mutant appears to grow slightly better than the wild-type strain on media without Tween80, while subtle differences in colony morphology were observed on media containing Tween80. As can be observed in figure 4.31b, the wild-type strain appears to have dense growth at the edges while the centre remains flat and smooth. The mutant however appears to show good growth and a raised rugose surface at the centre of the spot (figure 4.31d).

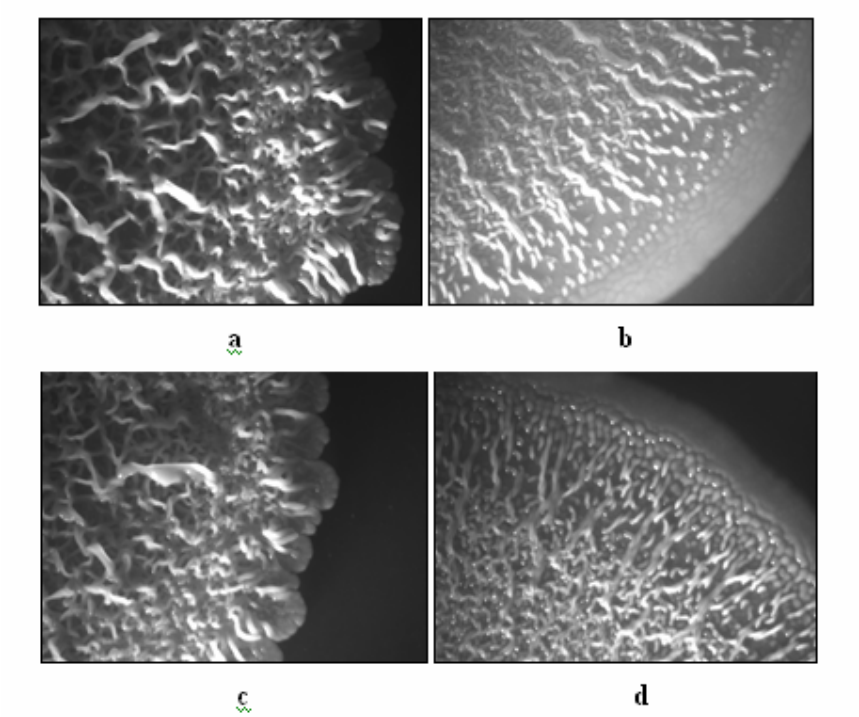


Figure 4. 30: **Colony morphology of *M. smegmatis* strains (Middlebrook 7H10 agar)**

Colony morphology of *M. smegmatis* wild type grown on Middlebrook 7H10 agar (a), Middlebrook 7H10 agar + 0.05% Tween80 (b), *M. smegmatis* Δ*cpn60.1* on Middlebrook 7H10 agar (c) and Middlebrook 7H10 agar + 0.05% Tween80 (d).

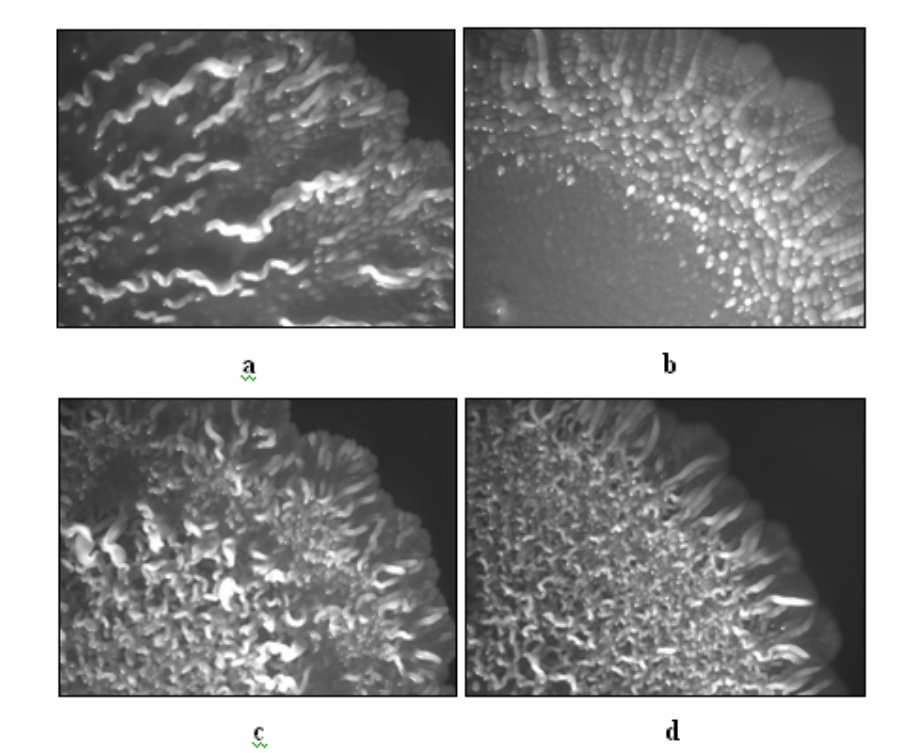


Figure 4. 31: **Colony morphology of *M. smegmatis* strains (Middlebrook 7H1 agar)**

Colony morphology of *M. smegmatis* wild type grown on Middlebrook 7H11 agar (a), Middlebrook 7H11 agar + 0.05% Tween80 (b), *M. smegmatis* Δ*cpn60.1* on Middlebrook 7H11 agar (c) and Middlebrook 7H11 agar + 0.05% Tween80 (d).

The growth of both the strains on LB agar media looked identical and showed no difference (figure 4.32). While colonies grown on media without Tween80 had irregular edges, colonies grown on media containing Tween80 had round edges.

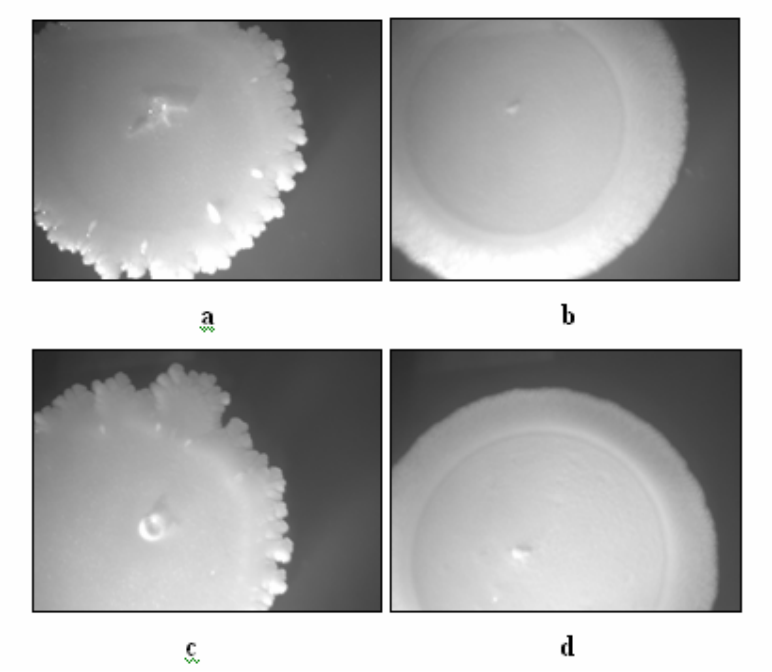


Figure 4. 32: **Colony morphology of *M. smegmatis* strains (agar)**

Colony morphology of *M. smegmatis* wild type grown on LB agar (a), LB agar + 0.05% Tween80 (b), *M. smegmatis* Δ*cpn60.1* on LB agar (c) and LB agar + 0.05% Tween80 (d).

When grown on Tryptic Soy agar, there was no difference between the two strains. Growth on media without Tween80 was good with colonies appearing raised with rugose surface and irregular edges, while growth on media containing Tween80 was poor and colonies had raised rounded edges with a flat centre (figure 4.33).

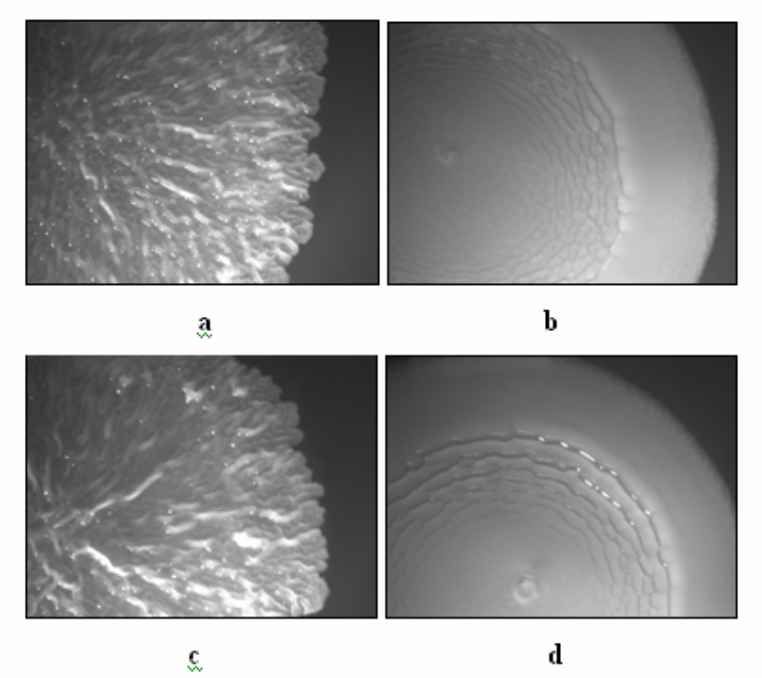


Figure 4. 33: **Colony morphology of *M. smegmatis* strains (Tryptic soy agar)**

Colony morphology of *M. smegmatis* wild type grown on Tryptic Soy agar (a), Tryptic soy agar + 0.05% Tween80 (b), *M. smegmatis* Δ*cpn60.1* on Tryptic soy agar (c) and Tryptic soy agar + 0.05% Tween80 (d).

Since we observed a difference in the morphology of the colonies grown on Middlebrook 7H11 agar with Tween80, we decided to determine if there were any significant morphological differences between all the test strains. However, as can be observed in figure 4.34, *M. smegmatis* Δ*cpn60.1* + pMsGroEL1 (c) doesn't appear to have the same morphology as the wild type strain *M. smegmatis* (a) and resembles the knock-out strain instead (b). The other test strains containing *E. coli groEL* (d), *cpn60.2* (e) and *cpn60.3* (f) appear to resemble the wild type. However, when this experiment was repeated, there was no consistency with the colony morphology of the same strains. Hence, it was decided that this could not be used as an alternative assay.

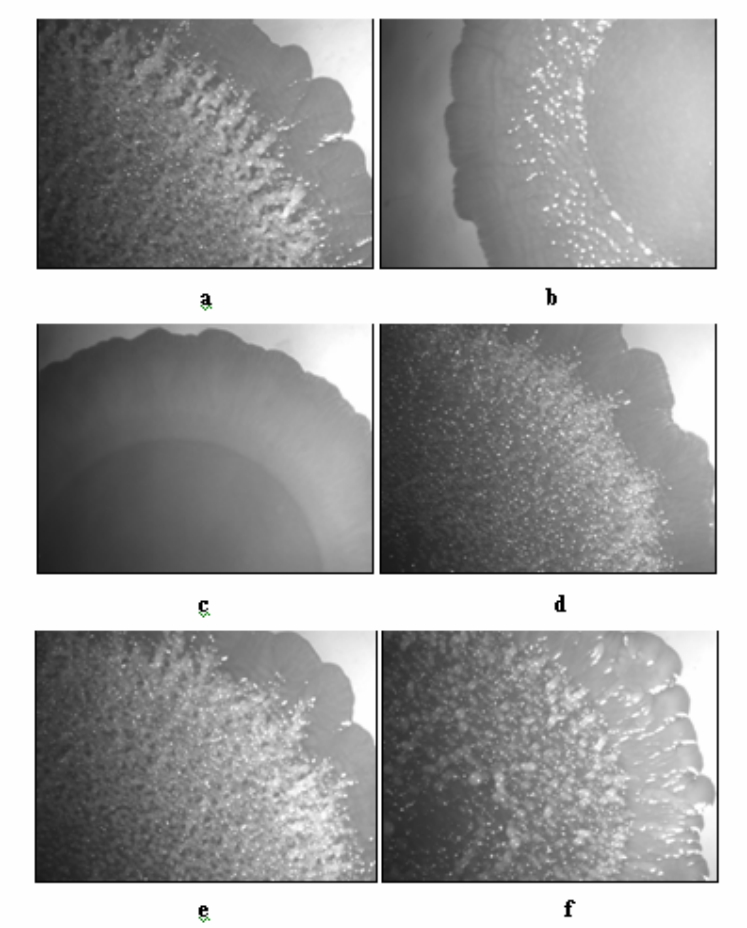


Figure 4. 34: **Colony morphology of *M. smegmatis* test strains (Middlebrook 7H11 agar)**
Colony morphology of all the strains on Middlebrook 7H11 agar with 0.05% Tween80. The strains tested were *M. smegmatis* mc^2155 (a), *M. smegmatis* Δ*cpn60.1* (b), *M. smegmatis* Δ*cpn60.1* + pMsGroEL1 (c), *M. smegmatis* Δ*cpn60.1* + pMs10-EcgroEL (d), *M. smegmatis* Δ*cpn60.1* + pMs10-cpn60.2 (e) and *M. smegmatis* Δ*cpn60.1* + pMs10-cpn60.3 (f).

4.3.4.2 Colony morphology on Middlebrook 7H11 agar

Since morphologically the wild type *M. smegmatis* and *M. smegmatis Δcpn60.1* could be distinguished, we decided to look at the colony morphology of single colonies. For this purpose, an equal number of cells of both strains were spotted onto Middlebrook 7H11 agar, and streaked out to single colonies. After incubation at 37°C for 5 days, the

colonies were photographed. Colonies grown on media without Tween80 have a textured raised rugose surface and an irregular outline, while colonies grown on media containing Tween80 are rounded and have a smooth surface. However, morphologically no differences between the two strains could be observed on media with or without Tween80 (figure 4.35).

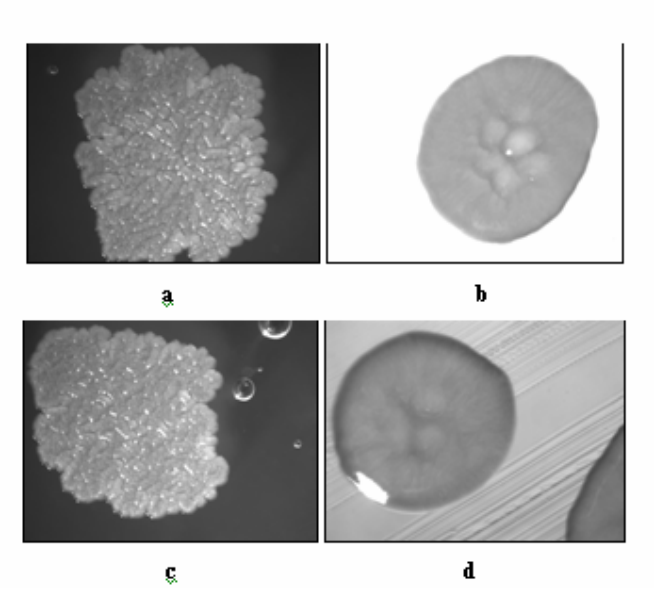


Figure 4. 35: **Single colony morphology of *M. smegmatis* strains (Middlebrook 7H1 agar)**
Single colonies of *M. smegmatis* wild type grown on Middlebrook 7H11 agar (a), Middlebrook 7H11 agar + 0.05% Tween80 (b), *M. smegmatis* Δ*cpn60.1* on Middlebrook 7H11 agar (c) and Middlebrook 7H11 agar + 0.05% Tween80 (d).

4.4 Discussion

Most bacteria are found in the environment as biofilms that are attached to surfaces or present at air-liquid interfaces by the help of an extracellular matrix (Davey and O'Toole, 2000 and O'Toole *et al*., 2000). A number of mycobacterial species such as *M. smegmatis*, *M. avium* (Martínez *et al*., 1999), *M. fortinum* (Hall-Stoodley and Lappin-Scott, 1998) and *M. marinum* (Hall-Stoodley *et al*., 2006) have been shown to form

biofilms. However, it has only recently been shown that even the pathogenic *M. tuberculosis* forms biofilms which play a role in drug tolerance (Ojha *et al*., 2008).

Our previous results from the bioinformatic and expression studies (chapter 3) have confirmed that Cpn60.2 appears to be the housekeeping chaperonin in *M. smegmatis*, while Cpn60.1 appears to have a more specialised role in the cell. Since the role of Cpn60.1 in biofilm maturation has previously been demonstrated (Ojha *et al*., 2005), we wished to exploit this as an assay to determine if there were any structural properties of the proteins that contributed to this function. The questions we aimed to answer were firstly, what makes Cpn60.1 responsible for biofilm maturation and secondly, which regions of the protein are responsible for its association with KasA (part of FASII complex) which plays a role in mycolic acids biosynthesis. To begin to answer these questions, it was necessary to first replicate the results obtained previously, and then test for complementation using the other homologues from *M. smegmatis*, *M. tuberculosis* and *E. coli*. Once complementing genes were determined, more specific analyses would be performed which would include the construction of chimeric proteins to help determine specific regions of the protein responsible for this association with KasA and biofilm maturation.

A literature search established a relationship between mycolic acid synthesis and biofilm formation and more importantly revealed that the *cpn60.1* gene is not the only gene the absence of which causes loss of biofilm formation. An *M. smegmatis* mutant in which the undecaprenyl phosphokinase (*upk*) gene has been deleted has been shown to be defective in biofilm formation (Röse *et al*., 2004). A study in which transposon

insertion mutants were screened for altered colony morphology found a mutant (in the *lsr2* gene) which appeared defective in biofilm formation (Chen *et al*., 2006). The *lsr2* mutant was defective in biofilm formation and also exhibited smooth colony morphology. Though the exact function of this gene remains unknown, it was first identified as an immunodominant T-cell antigen in *M. leprae* (Laal *et al*., 1991). This is interesting as the *M. tuberculosis* Cpn60 proteins have also been shown to be immuno-modulators (as discussed in chapter 1, section 1.1.7.3) Furthermore, as mentioned previously, *M. smegmatis* contains three types of mycolic acids the α-, α'- and epoxy (Barry *et al*., 1998). These mycolic acids are present in two forms, one where they are covalently linked to the arabinogalactan in the cell wall and the other where they are esterified to a trehalose. The study of the *lsr2* mutant identified a third form called mycolyl diacylglycerols (MDAGS) where the mycolic acids are esterified to a glycerol (Chen *et al*., 2006). Even though the dimeric Lsr2 protein has been shown to be cytosolic, the *lsr2* mutant blocked the synthesis of the MDAGS implying that this gene could play a regulatory role in cell wall lipid synthesis (Chen *et al*., 2006). Together with these previous studies, the studies on the *M. smegmatis* Cpn60.1 establish a link between cell wall synthesis and biofilm formation. Furthermore, a link between stress responses and biofilm formation has previously been reported in *E. coli* (Adams and McLean, 1999), further encouraging us to pursue the studies of the chaperonin Cpn60.1 in biofilm formation.

Once the strains sent to us were confirmed to be correct by antibiotic screening, PCR and qRT-PCR, attempts to replicate the previously published results were started. However, surprisingly, a number of difficulties were encountered in this first stage

(4.3.3.2). We therefore systematically altered several parameters in the experiment, starting with changing the growth phase of the inoculating culture (4.3.3.3), methods of inoculation and the nature of the growth medium (4.3.3.4). However, none of these changes were successful at producing the reported phenotype. We then hypothesized that the *M. smegmatis* Δ*cpn60.1* strain might have acquired a suppressor mutation, and asked for new strains to be sent to us. However, the new strains also failed to show a defect in biofilm and instead formed mature biofilms resembling the wild-type strain (4.3.3.5). The next variable changed was the water, however, even the use of high grade HPLC water did not produce the desired result (4.3.3.6). Exchanging biofilm media between the labs did not work either as no biofilms were formed (4.3.3.7 and 4.3.3.8).

It was finally decided that I would visit the lab and learn and conduct the assays in the lab in Pittsburgh. As expected, the assay worked and the *M. smegmatis* Δ*cpn60.1* strain failed to form mature biofilms (4.3.3.9). While a few subtle differences like the addition of Tween80 to cultures was noticed, the assay was performed in the exact same manner as in Birmingham (4.3.3.10). Thus, the reason for the unexpected results was still unknown.

On my return to Birmingham, further attempts were still unsuccessful (4.3.3.11). Since iron has been shown to play an important role in biofilm formation (Ojha *et al*., 2007), we set up an iron gradient to measure biofilm formation (4.3.3.12). Our results revealed that concentrations of less than 0.75μM of iron impaired biofilm formation, confirming that iron is essential for biofilm formation. However, calcium was found to be non-essential for biofilm formation (4.3.3.13). Eventually the only condition where the assay

worked in Birmingham was when new glassware was used to prepare media and used for the biofilm assay (4.3.3.14). However, there was no consistency in this result and it was decided that the assay would be conducted by the lab in Pittsburgh to test our constructs for complementation.

Preliminary results from the complementation assay conducted in Pittsburgh revealed that from all the *cpn60* homologues tested only the *M. tuberculosis cpn60.1* could complement for loss of the *M. smegmatis cpn60.1*. The *M. tuberculosis cpn60.1* gene has previously been shown to play a role in granuloma formation in mice and guinea pig (Hu *et al*., 2008), and this is interesting as it further strengthens our hypothesis that this duplicated copy of the *cpn60* gene has evolved to take on a specialised function. However, whether there is a link between the biofilm formation (*M. smegmatis*) and granuloma formation (*M. tuberculosis*) is still unknown. Combined with the fact that the *cpn60.1* genes are similar to each other and fall in the same clade in the phylogenetic tree, the specialization in function of the Cpn60.1 proteins imply that this must be due to certain structural similarities between the two proteins. While *E. coli groEL* and *cpn60.3* appeared to show partial complementation, *cpn60.2* showed no complementation (figure 4.29). The ability of these *cpn60* homologues to only partially complement the loss of the *cpn60.1* gene may be due to some similarities in structure of the protein that contributes to its function. However, to investigate this further, experiments involving the construction of chimeric proteins by swapping regions between Cpn60.3 or *E. coli* GroEL with Cpn60.1 and testing for complementation would need to be done. This would enable the identification of the regions of the protein that are important in biofilm formation. Another issue to consider is that the expression

of the GroEL or Cpn60.3 protein could not be conclusively demonstrated and the lack of complete complementation could be due to poor expression of these proteins. The result that the Cpn60.2 protein did not complement for loss of Cpn60.1 was not entirely surprising, as they obviously play very different roles in the cell. Cpn60.2 is the housekeeping chaperonin and if it could have aided in assembly of the FASII complex (by interaction with KasA), it would have compensated for the loss of Cpn60.1 in the knock-out strain by forming mature biofilms.

Since the biofilm assay proved difficult to replicate in our lab, we began to look for alternative assays. A simple phenotypic assay using colony morphology has previously been shown to distinguish strains (Dahl *et al*., 2005; Kumar and Chatterji, 2008). We repeated this assay using different solid media and found a difference between the colony morphologies of the wild type *M. smegmatis* and knock-out strain *M. smegmatis* Δ*cpn60.1* when grown on Middlebrook 7H11 agar (4.3.4.1). However, when all the strains were tested, the complemented strain *M. smegmatis* Δ*cpn60.1* with pMsGroEL1 did not appear to resemble the wild type *M. smegmatis* strain (4.3.4.1 and 4.3.4.2). Since the levels of expression of the Cpn60.1 protein may vary in strain 1 (Cpn60.1 expressed from a single copy of *cpn60.1* located on the chromosome) and strain 3 (Cpn60.1 expressed from multiple copies due to its location on the complementing plasmid), it was decided that the use of colony morphology as an assay to analyse complementation of Cpn60.1 would not be ideal.

The original study in Pittsburgh also monitored the attachment and maturation of the wild type and mutant strains on a PVC surface using *gfp* plasmids (Ojha *et al*., 2005).

Results revealed that *cpn60.1* was not essential for attachment but was necessary for maturation of the biofilm. Hence, the PVC-*gfp* assay could be used as another possible assay to monitor the differences between the strains. Other studies have used the crystal violet staining assay in which crystal violet is used to stain the biofilm, and the bound dye is then solubilized using DMSO and quantified by measuring the absorbance at OD 590nm (Mathew *et al*., 2006). We did attempt to perform this assay in the lab, however, it proved extremely difficult, since the biofilm did not attach to the surface, but was present as a pellicle at the air-liquid interface.

The consistency of the biofilm (intact in the wild type strain 1 and disintegrating in the mutant strain 2) was another factor to be considered, and the results from the assays conducted in our lab revealed that strain 2 also formed intact biofilms similar to strain 1. However, interestingly, upon harvesting the biofilms to measure the mass, strain 5 (containing *cpn60.2*) always formed slimy biofilms. To investigate the reason for this slimy nature of the biofilm, lipids were extracted from all the different biofilm cultures. Results showed that there was no observable difference in the lipid profiles between the strains, suggesting that the lipids did not play a role in altering the texture of the biofilm of strain 5. Another possibility for the slimy texture could be due to changes in the capsular material. Intriguingly, as mentioned in chapter 1, it has recently been shown that Cpn60.2 is present in the capsular material in *M. tuberculosis* where it plays a role in macrophage attachment (Hickey *et al*., 2009). To confirm whether over expression of the Cpn60.2 protein in strain 5 causes changes in the capsular material which in turn contributes to the slimy texture of the biofilm, further experiments analysing the

capsular material from the biofilms will need to be done. When the mycolic acid profiles of *M. smegmatis* planktonic and biofilm cells were analysed, an increase in the prevalence of short-chain mycolic acids (C_{56}-C_{68}) from biofilm cells was observed (Ojha *et al*., 2005). When the mycolic acid profiles of the *M. smegmatis* Δ*cpn60.1* were monitored, a substantially lower amount of short chain mycolic acids was observed. This finding suggests that *cpn60.1* either directly or indirectly by association with KasA, plays a role in the induction of short chain mycolic acids which in turn contributes to biofilm maturation (Ojha *et al*., 2005). Sliding motility and biofilm formation in *M. smegmatis* has been shown to be due to glycopeptidolipids (GPL) (Martínez *et al*., 1999; Recht *et al*., 2000). GPL's are a class of glycosylated peptidolipids present in the outer most layer of the cell envelope (Ortalo-Magne *et al.*, 1996). A recent study has shown that biofilm formation in *M. smegmatis* is affected by the serine/threonine protein kinase PknF (Gopalaswamy *et al*., 2008). This PknF over-expressing strain showed altered levels of GPLs (Gopalaswamy *et al*., 2008). However, a recent study showed that while the *M. tuberculosis* Cpn60.1 is a substrate of PknF, its *M. smegmatis* homologue is not (Canova *et al*., 2009), indicating that even though they share a high degree of similarity, there are certain sequence or structural differences that make them unique in some areas.

In conclusion, in this part of the study we originally set out to test the various homologues for their ability to complement the *cpn60.1* and restore biofilm maturation. However, we failed to replicate the biofilm assay despite a substantial effort in trying to do so. The Pittsburgh lab appears to be able to consistently reproduce the original result and would be the best alternative to analyse our test strains further. However, the fact that the results obtained from strain 4 (containing *E. coli groEL*) are not consistent

cannot be ignored. One possible solution to overcome that issue would involve obtaining a number of independent reproducible replicates along with quantitative data (biofilm masses) before the data can be considered legitimate. Our results so far are preliminary and much work needs to be done in future to characterise specific regions of the protein that are responsible for this biofilm maturation. Together with the available information on the roles of other interactions that contribute to biofilm formation, we hope to have a better understanding of the role Cpn60.1 plays in biofilm formation.

5.1 Introduction

The previous two chapters discussed the uniqueness of the presence of three *cpn60* genes in *M. smegmatis* and looked into the phylogeny of the *cpn60* genes. It was also demonstrated that Cpn60.2 is likely to be the housekeeping chaperonin, which together with Cpn10 appears to be regulated by the repressor protein HrcA. This was followed by determining any sequence or structural properties that made Cpn60.1 have a more specialised role of biofilm maturation.

This part of the study focuses on the characterization of the *M. smegmatis* chaperonins in the extensively studied *E. coli*. As mentioned earlier, chaperonins are double ringed complexes, enclosing a cavity within which protein folding takes place. Substrate specificity has been studied and the *E. coli* GroEL has been shown to fold a large range of substrates, 13 of which were found to be essential for cell growth (Houry *et al*., 1999; Kerner *et al*., 2005). Although it has previously been demonstrated that a number of homologues of GroEL are capable of replacing it in *E. coli,* reports show that not all the copies of multiple chaperonins are capable of functioning when expressed in *E. coli* (Hu *et al*., 2008; Kovács, 2001, Lund, 2009). This implies that while there must be a certain level of specificity in substrate recognition and folding, there must also be a large overlap of substrates between GroEL and its homologues. Moreover, as a single copy has been shown to be sufficient for normal cell growth, the existence of these multiple copies of chaperonin genes is fascinating. So what is the significance and role of these other copies? To help answer these questions we designed experiments to identify which *cpn60* homologues from *M. smegmatis* could replace GroEL function in *E. coli*. Since there are conflicting reports that suggest that even though the *M. tuberculosis*

Cpn60.2 can complement for loss in function of *E. coli* GroEL (Hu *et al*., 2008), but it only forms lower oligomers (Qamra *et al*., 2004; Qamra and Mande, 2004), the analysis of the oligomeric states of *M. smegmatis* chaperonins that complement *E. coli* GroEL is a topic of interest. Thus, complementing proteins from this study will be analysed for their oligomeric structures in the next chapter.

5.2 Experimental set-up

A commonly used system for complementation makes use of a modified strain of *E. coli* (MGM100), in which the chromosomal *groE* promoter has been replaced by the pBAD promoter (McLennan and Masters, 1998). Mediated by the AraC protein, this promoter is induced in the presence of arabinose and repressed in the presence of glucose. The first complementation system makes use of this *E. coli* MGM100 strain, by transforming cells with various homologues present on an IPTG inducible p*trc*99A plasmid and checking for complementation on media containing glucose (figure 5.1). Thus all strains including the negative control (MGM100 + p*trc*99A) grow in the presence of arabinose and only strains containing genes that complement for loss of GroES and GroEL such as the positive control (MGM100 + p*trc*ESL) grow in presence of glucose (figure 5.2). Complementation was monitored at a range of temperatures (30°C, 37°C and 42°C) to determine if there was a temperature dependency. While the main focus of the complementation assay was to monitor the growth of test strains in the absence of the chromosomal *groESL* genes, the presence of a GroEL functional marker served as an additional assay. The depletion of GroEL in *E. coli* 431 strain resulted in the appearance of an 85kDa methionine biosynthetic enzyme known as MetE (Chapman *et al*., 2006). Previous work conducted in the lab also identified an 85kDa

protein in the *E. coli* MGM100 strain, which was analysed and confirmed to be MetE by mass spectrometry (Tabish Ahmed, Pers. Comm). Normally, in a rich medium such as LB which contains methionine, the expression of genes involved in the methionine biosynthetic pathway is repressed. However, the presence of the MetE protein can be explained; as mentioned previously (chapter 1), as follows: MetK, an obligate substrate of the GroES-EL folding machinery is responsible for the catalysis of methionine to S-adenosyl methione (SAM). Hence, in GroEL limiting conditions, folding of MetK is hindered, leading to decreased levels of SAM. SAM together with another gene in the *metE* operon, *metJ*, is responsible for co-repression of MetE. Thus, unfolded MetK eventually leads to derepression of MetE which is clearly visible on an SDS PAGE gel.

The other complementation system is more stringent and tests for complementation activity at higher temperatures of 42°C. This latter system makes use of the *E. coli* SF103 strain which has a temperature sensitive (ts 44) mutation in *groEL* that will allow for normal cell growth at 37°C but not at 42°C. The reason for using the SF103 strain to test for complementation is that it serves as a more stringent test for complementation as its determines restoration of cell growth at higher temperatures of 42°C. The use of this strain as a stringent test for complementation was previously demonstrated in a similar study involving the analyses of the chaperonins of *Rhizobium leguminosarum* (Ivic *et al*., 1997). The use of the SF103 strain to test for complementation of the *M. smegmatis* chaperonins at higher temperatures was a topic of interest, since previous complementation studies of the *M. tuberculosis* chaperonins have shown that Cpn60.2 cannot complement loss of GroEL at 42°C (Hu *et al*., 2008).

To test the functionality of the *M. smegmatis* homologues an expression system that would be compatible with both complementation systems was chosen. A plasmid containing the IPTG inducible *trc* promoter was selected into which to clone the various combinations of genes. The chaperonin 60 genes were cloned downstream from the *E. coli* co-chaperonin *groES* or the *M. smegmatis cpn10* gene to test whether the Cpn60 proteins could work and complement (loss of GroEL) with both Cpn10 and GroES.

Since lack of complementing ability, if seen, could be due to lack of expression, the expression of the Cpn60 proteins was monitored using the p*trc* constructs in the MGM100 strain.

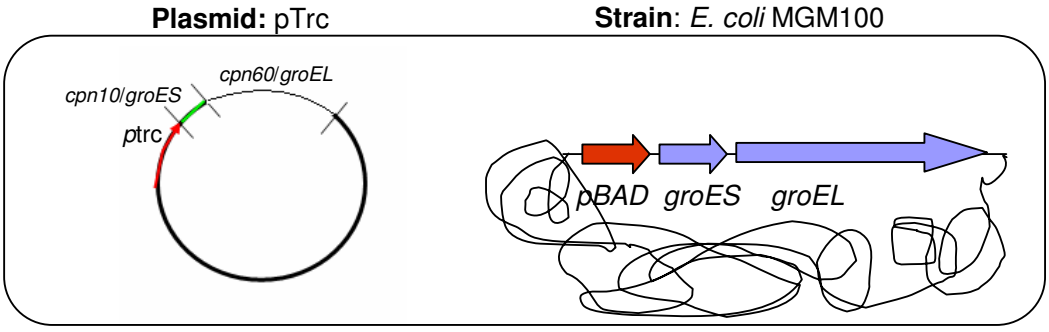


Figure 5. 1: **Schematic representation of experimental set-up for studying complementation using *E. coli* MGM100 strain**

E. coli MGM100 cell with plasmid p*trc* harbouring potentially complementing genes.

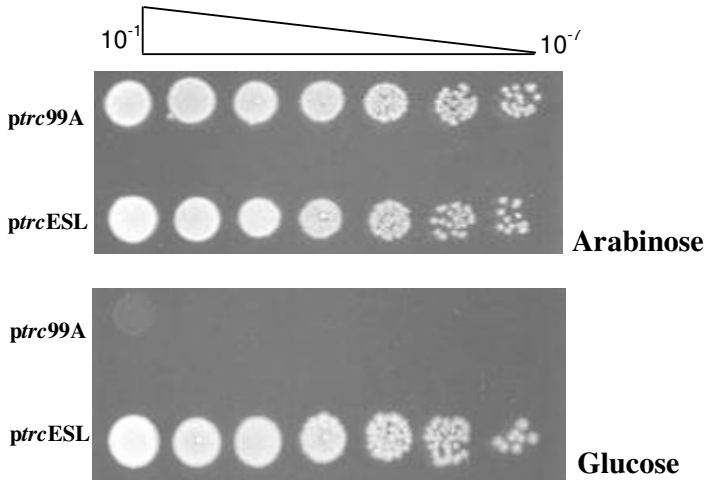


Figure 5. 2: **Typical result obtained from MGM100 complementation experiment**

Typical result showing growth of *E. coli* MGM100 with p*trc*99A and p*trc*ESL on media containing 0.2% arabinose (above) and complementation of only cells containing p*trc*ESL on 0.2% glucose (below).

5.3 Results

5.3.1 Construction of plasmids

The p*trc*99A plasmid into which the various combinations of genes were to be cloned was readily available in the lab. For cloning of the *M. smegmatis* genes, a master plasmid was first constructed into which various combinations of genes would be cloned. The p*trc*ES-Mtbcpn60.1 construct was available in the lab and the Mtb*cpn60.1* gene was excised using *Spe*I sites and the plasmid re-circularised. The resulting plasmid which contained the *trc* promoter upstream of the *E. coli groES* gene was named p*trc*ES-SpeI, and served as the master plasmid into which other *cpn60* gene would be cloned.

Due to the difficulties encountered in the cloning of *cpn60.1*, the construction of plasmids containing this gene will be discussed individually. After a great deal of optimisation, *cpn60.1* was PCR amplified from genomic DNA. The next step was to clone the gene into the holding vector pGEMTeasy to produce pGEMT-cpn60.1. It proved impossible to obtain plasmids with the *cpn60.1* gene in the correct orientation to be transcribed from the lac promoter, suggesting that expression of Cpn60 is deleterious to *E. coli* cells. In cases where the *cpn60.1* gene was successfully cloned into the holding vector, it was nearly always observed that the gene cloned in the opposite orientation to that of the *lac* promoter. In the few instances where the gene was cloned in the forward orientation, point mutations resulting in frame shifts were observed. This failure of directional cloning suggested that expression of this gene was not healthy for the cells. A great number of attempts at trying to sub-clone this gene into the p*trc*ES-SpeI plasmid downstream of the *E. coli* co-chaperonin *groES* also proved unsuccessful with potential positive clones revealing point mutations in the gene thus disrupting the reading frame of the gene. In the one attempt where the gene was successfully cloned, a 750bp insertion sequence element was found between the *trc* promoter and the start of the *groES* gene.

When combined, these results indicated that the difficulties in cloning the *cpn60.1* gene could be due to the toxicity of Cpn60.1 to *E. coli* cells. In order to test this hypothesis, the *cpn60.1* gene was cloned on its own into the pBAD24 plasmid where its expression could be tightly regulated. To do this, the pBAD24 plasmid was first digested with NcoI and the 5' ends filled in using Klenow polymerase. The *cpn60.1* gene was amplified and

blunt-end ligated into pBAD24 (figure 5.3). The cloning of this gene into the pBAD24 plasmid was straightforward.

The growth of cells containing this plasmid was then monitored in liquid and solid media. Results showed that in liquid media, there was a slight slowing in growth of cells in arabinose (gene expressed) when compared to cells in glucose (no expression). This was also observed on solid media, where cells containing the plasmid pBAD24-cpn60.1 showed slightly less growth on media containing arabinose, when compared to growth on media containing glucose. Protein expression of cells containing this plasmid was monitored and after induction with 0.2% arabinose, expression of Cpn60.1 steadily increased every hour till it peaked at 4 hours post induction (see figure 5.13). Taken together, these results suggest that while the *E. coli* cell can tolerate even high levels of expression of this gene, after a long growth phase (8 hours - overnight), the cells show a slight slowing in growth and the reason for this is unknown. From this result, it can be confirmed that the reason for the inability to clone the *cpn60.1* gene in the p*trc*99A plasmid is not due to a deleterious effect of the protein it encodes.

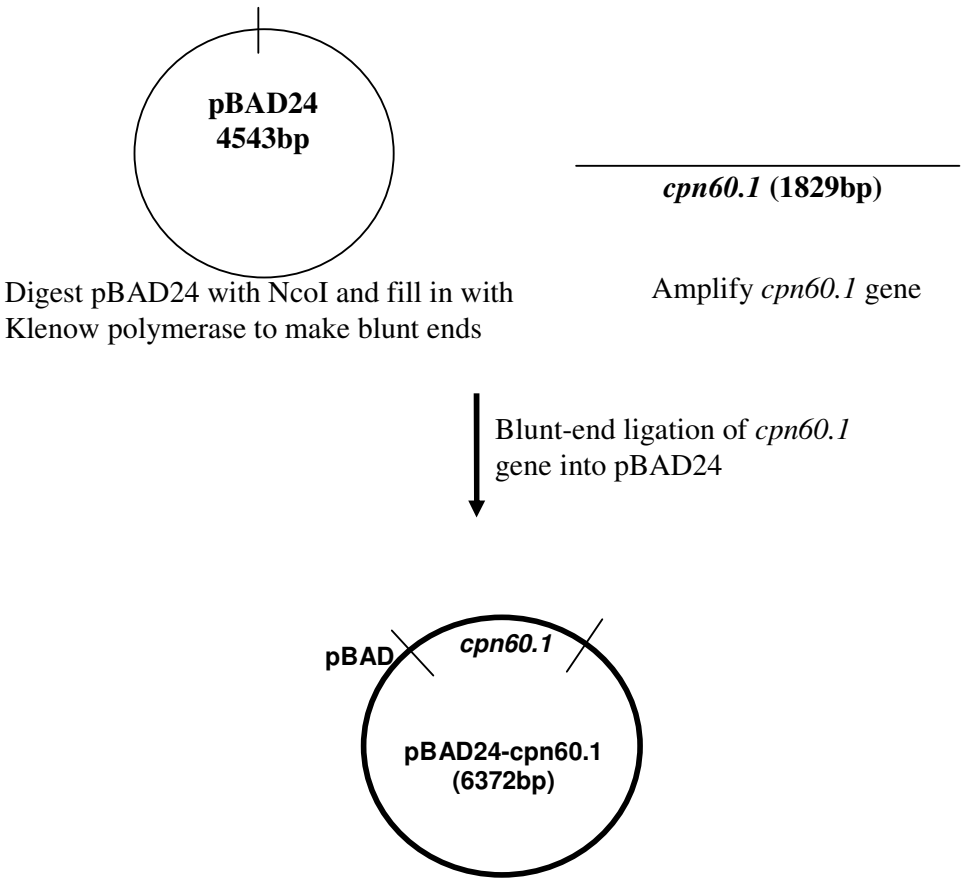


Figure 5. 3: **Schematic representation of construction of pBAD24-cpn60.1**

Having proved that the Cpn60.1 protein is not toxic to the cell, the cloning of this gene into another expression plasmid was attempted. The p*groE* plasmid which contains the *E. coli groE* promoter was selected. A p*groE* plasmid containing *E. coli groES* and *groEL* known as pOS2 was previously constructed in the lab. The *groEL* gene was excised from this plasmid using the sites BglII and HindIII. The *cpn60.1* gene was then cloned into the previously digested plasmid using the same restriction sites (figure 5.4). The cloning was successful and p*groE*-EScpn60.1 confirmed by restriction mapping and sequencing. While this proves that the *cpn60.1* gene can be cloned downstream of

the *groES* gene, the reasons for the failure to clone *cpn60.1* downstream of *groES* in p*trc*99A is not known. The complementation and protein expression results will be discussed later in this chapter.

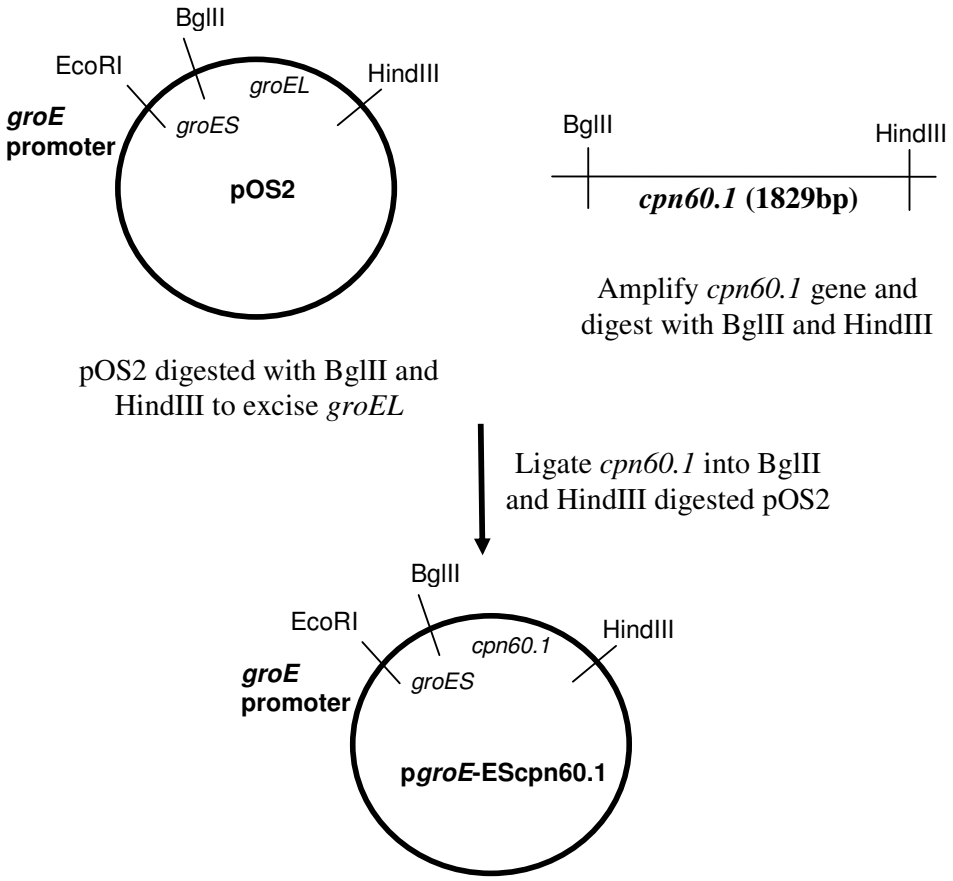


Figure 5. 4: **Schematic representation of construction of p*groE*-EScpn60.1**

The *cpn60.2* and *cpn60.3* genes were much simpler to amplify and clone into the pGEMTeasy holding vector to give pGEMT-cpn60.2 and pGEMT-cpn60.3. The genes were successfully sub-cloned into previously digested p*trc*ES-SpeI on SpeI fragments. The resulting plasmids were named p*trc*ES-cpn60.2 and p*trc*ES-cpn60.3 (figure 5.5).

The constructs were verified by restriction mapping and sequencing and found to be correct.

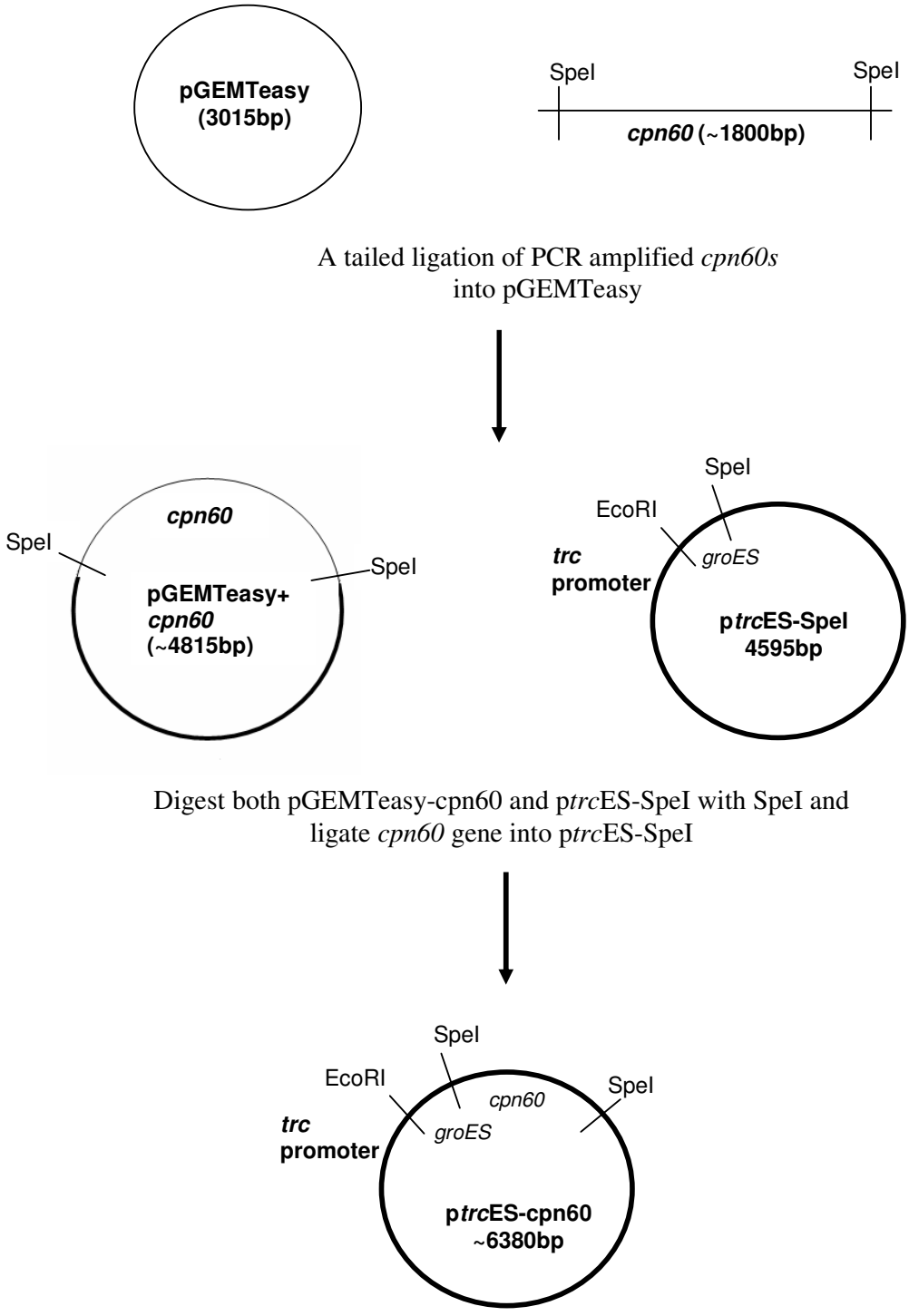


Figure 5. 5: Schematic representation of construction of p*trc*ES-cpn60.2 and p*trcES*-cpn60.3

To test if the Cpn60 proteins could complement when present with GroES or Cpn10, the *cpn60* genes were cloned downstream of *cpn10*. For the cloning of the genes downstream of the *M. smegmatis cpn10*, it was decided that the *cpn10* gene would be swapped with *groES* in all the constructs thereby keeping the downstream region the same so that any difference in complementation or expression observed would be solely due to the presence of the *M. smegmatis* co-chaperonin gene. The initial part of this *cpn10* cloning was done by an Undergraduate summer project student (Saira Kausar) under my supervision. The *cpn10* gene was amplified and cloned into pGEMTeasy to give pGEMT-cpn10. The *cpn10* gene was sub-cloned using *Eco*RI and *Hind*III sites into plasmid p*trc*ES-cpn60.2 which had previously been treated with the same set of restriction enzymes to excise the *groES* gene (figure 5.6). This construct was named p*trc*10-cpn60.2 and was confirmed by restriction mapping and sequencing.

Due to time constraints, the remaining work on the construction of plasmids containing *cpn10* was performed by me. The *E. coli groEL* gene was cloned downstream of the *cpn10* gene by excising *cpn60.2* from plasmid p*trc*10-cpn60.2 using *Hind*III sites. The resulting plasmid was named p*trc*10-groEL. As there was an *Eco*RI site within the *cpn60.1* gene and because the p*trc*ES-cpn60.1 could not be constructed, the *cpn60.1* gene was cloned downstream of the *cpn10* gene by replacing the *groEL* gene in p*trc*10-groEL using *Hind*III sites. The resulting plasmid was named p*trc*-10cpn60.1. The *cpn10* gene was cloned into plasmid p*trc*ES-cpn60.3 by replacing the *groES* gene using *Eco*RI and *Hind*III sites (figure 5.6). The resulting plasmid, p*trc*10-cpn60.3, was confirmed by restriction mapping and sequencing. All the constructs obtained from these cloning

experiments were then tested for complementation and protein expression as discussed below.

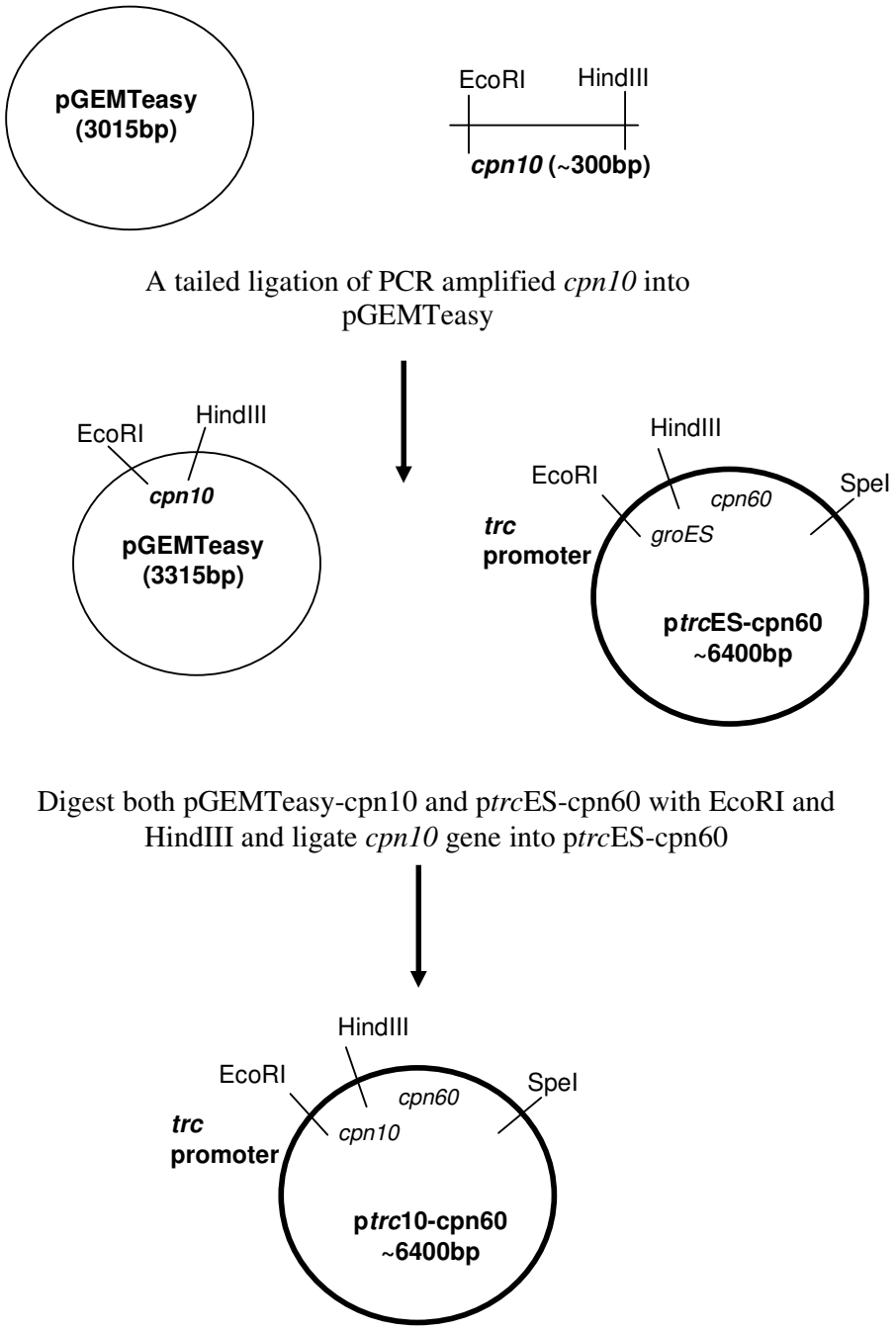


Figure 5. 6: Schematic representation of construction of p*trc*10-cpn60.2 and p*trc*10-cpn60.3

5.3.2 Complementation

5.3.2.1 Complementation in *E. coli* MGM100

As explained earlier, the MGM100 strain of *E. coli* has a tightly regulated pBAD promoter in place of the chromosomal *groE* promoter upstream of the *groES* and *groEL* genes. Hence in the presence of glucose the cells cannot grow unless a complementing gene is being expressed in the cell. To test which of the genes from *M. smegmatis* can complement for loss in function of *groEL*, MGM100 was transformed with each of the plasmids that were constructed previously (see 4.3.1). For the complementation experiments the positive control used was p*trc*ESL (p*trc*99A containing *E. coli groES* and *groEL*) and the negative control was p*trc*99A. The test plasmids were p*trc*10-cpn60.1, p*trc*ES-cpn60.2, p*trc*10-cpn60.2, p*trc*ES-cpn60.3, p*trc*10-cpn60.3 and p*trc*10-groEL. Cells containing p*groE*-EScpn60.1 were also checked for complementation separately.

All the transformed strains of MGM100 harbouring the p*trc* constructs were grown overnight in 5ml LB containing 0.2% arabinose and the complementation assay performed as described in Materials and Methods (section 2.6). While the plates kept at 37°C and 42°C were incubated overnight, the ones kept at 30°C were left for 20 hours and then analysed. Complementation was scored as complete (+++) if growth resembled that of cells containing positive control (p*trc*ESL), partial (++) if growth was slightly less than positive control, poor (+) if growth was only slightly better than negative control and negative (-) if no growth was observed similar to negative control cells (MGM100 + p*trc*). An example of plates showing complete or no complementation can be observed in figure 5.2 above. A summary of the results obtained from these

experiments is in table 5.1. These are the results obtained from three independent biological replicates which were highly reproducible. Images of complementation plates (arabinose and glucose) are also shown below.

Growth on arabinose: In summary, the growth of cells containing the *cpn60.*1, *cpn60.2* and *groEL* genes with either *E. coli groES* or *M. smegmatis cpn10* was normal at all three temperatures (30°C, 37°C, and 42°C) (figure 5.7). However, cells containing *cpn60.3* downstream of *groES* showed a poor growth phenotype even at permissive temperatures of 30°C and 37°C, and showed no growth at all at 42°C. The cells containing the same gene downstream of *cpn10* however did not show this growth defect. This phenotype will be discussed later in this chapter. In general it has been observed that the growth of cells on plates containing IPTG was very slightly less than that seen on plates not containing IPTG. This may be explained by the fact that when induced by IPTG, an over expression of the protein becomes harmful to cells and results in poor cell growth. Cells containing the p*groE*-EScpn60.1 plasmid showed normal growth on media containing arabinose at all three temperatures as well.

Growth on glucose: Only cells containing genes (on plasmids) that are able to complement for the loss of the chromosomal *groEL* are capable of growing on media containing glucose. The complementation results demonstrated that only cells containing p*trc*ES-cpn60.2, p*trc*10-cpn60.2 and p*trc*10-groEL were able to complement for loss in function of *E. coli groEL* at all three temperatures. Cells with constructs containing *M. smegmatis cpn10* gene appeared to show better growth when induced with 0.1mM IPTG compared to non-induced growth (figure 5.8).

The significance of these results will be discussed further in the chapter. In cases where no complementation was observed, cells will be analysed for expression of proteins.

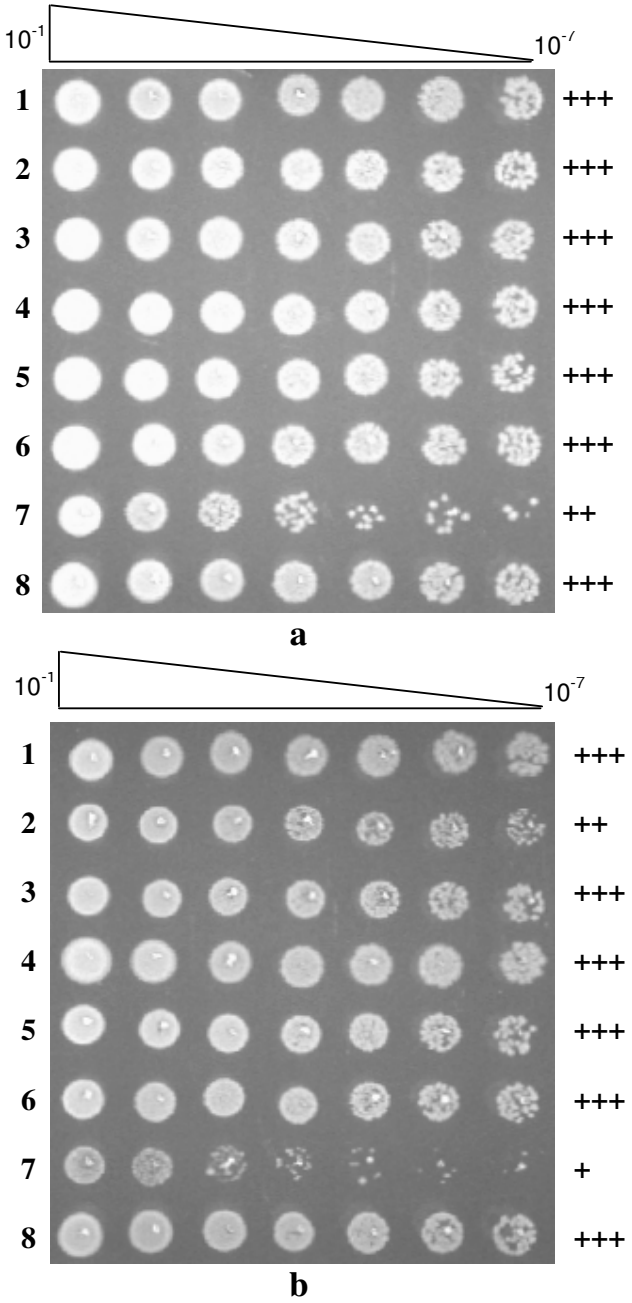


Figure 5. 7: **Growth of MGM100 strains at 30°C on media containing arabinose**

Growth of MGM100 at 30°C on media containing 0.2% arabinose (a) and 0.2% arabinose + 0.1mM IPTG (b) showing complete (+++) or partial (++ or +) growth.

E. coli MGM100 cells containing plasmids p*trc*99A (1), p*trc*ESL (2), p*trc*10-EL (3), p*trc*10-cpn60.1 (4), p*trc*ES-cpn60.2 (5), p*trc*10-cpn60.2 (6), p*trc*ES-cpn60.3 (7), and p*trc*10-cpn60.3 (8) were checked for growth on media containing 0.2% arabinose with or without 0.1mM IPTG.

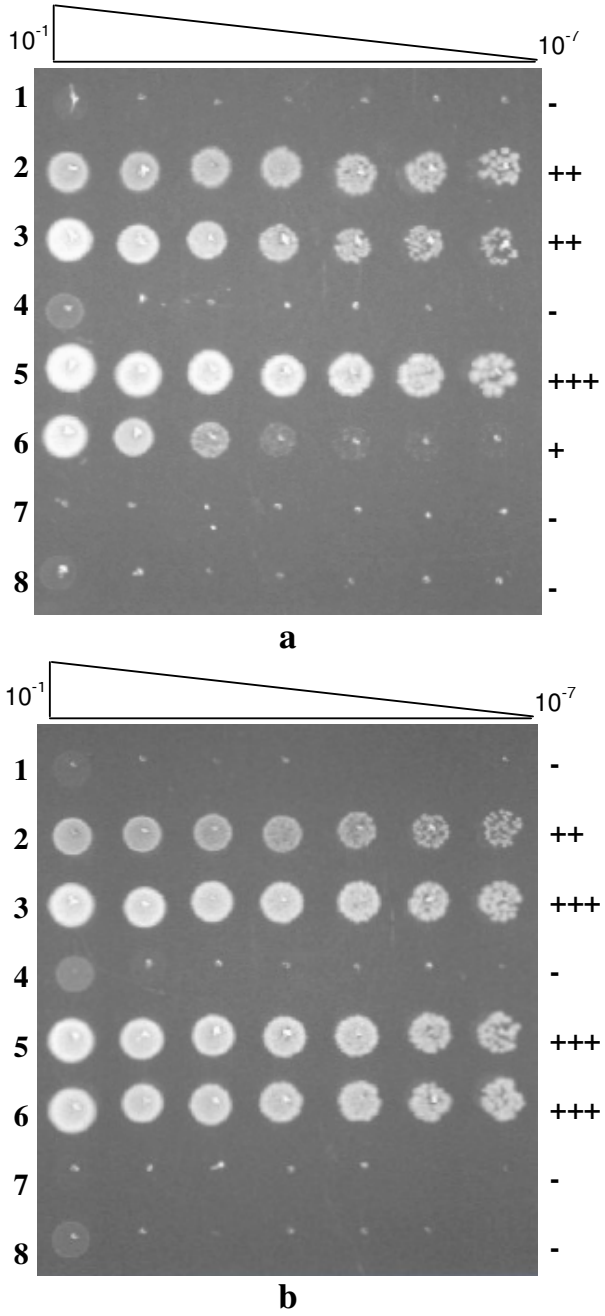


Figure 5. 8: Complementation of MGM100 strains at 30°C on media containing glucose

Complementation experiment conducted at 30°C on media containing 0.2% glucose (a) and 0.2% glucose + 0.1mM IPTG (b) showing complete (+++), partial (++ or +) or no (-) complementation.

E. coli MGM100 cells containing plasmids p*trc*99A (1), p*trc*ESL (2), p*trc*10-EL (3), p*trc*10-cpn60.1 (4), p*trc*ES-cpn60.2 (5), p*trc*10-cpn60.2 (6), p*trc*ES-cpn60.3 (7), and p*trc*10-cpn60.3 (8) were checked for complementation on media containing 0.2% glucose with or without 0.1mM IPTG.

Plasmid in *E. coli* MGM100	30°C				37°C				42°C			
	arabinose	arabinose + IPTG	glucose	glucose + IPTG	arabinose	arabinose + IPTG	glucose	glucose + IPTG	arabinose	arabinose + IPTG	glucose	glucose + IPTG
p*trc* 99A	+++	+++	-	-	+++	+++	-	-	+++	+++	-	-
p*trc*ESL	+++	++	++	+++	+++	++	+++	++	+++	++	++	++
p*trc*ES-cpn60.2	+++	+++	+++	++	+++	+++	+++	+++	+++	+++	+++	+++
p*trc*ES-cpn60.3	++	+	-	-	+	+	-	-	+	+	-	-
p*trc*10-*groEL*	+++	+++	++	+++	+++	+++	++	+++	+++	+++	-	+++
p*trc*10-cpn60.1	+++	+++	-	-	+++	+++	-	-	+++	+++	-	-
p*trc*10-cpn60.2	+++	+++	+	+++	+++	+++	+	+++	+++	+++	+	+++
p*trc*10-cpn60.3	+++	+++	-	-	+++	+++	-	-	+	-	-	-
p*groE*-EScpn60.1	+++	ND	-	ND	+++	ND	-	ND	+++	ND	-	ND

Table 5. 1: **Summary of complementation results using *E. coli* MGM100 strain**

Complementation of *E. coli* MGM100 by different combinations of Cpn60 and Cpn10 proteins showing complete complementation (+++), partial complementation (++ or +) or no complementation (-). Complementation of p*groE*-EScpn60.1 plasmid was not determined on media containing IPTG (ND).These experiments were conducted at least three times and the results found to be highly reproducible.

5.3.2.2 Complementation in *E. coli* SF103

SF103 is a strain of *E. coli* that has a mutation in its *groEL* gene which prevents it from growing at higher temperatures (42°C) but allows for normal growth at 37°C. The aim of using this temperature sensitive strain was to confirm the complementation results observed in the MGM100 strain, serve as a stringent test for complementation at 42°C (because of ts44 mutation in SF103 strain) and also to test genes that were cloned into a pBAD plasmid which could not be done using the MGM100 system. The complementation results obtained from MGM100 demonstrate complementation ability at a range of temperatures (30°C, 37°C or 42°C), while the complementation in the SF103 strain demonstrate the complementing ability at high temperatures (42°C) only. As with the complementation in MGM100, the positive control used with this set of experiments was p*trc*ESL and the negative control was p*trc*99A. The controls used for the pBAD24 set of constructs were pBAD-ESL (positive control) and pBAD24 (negative control). A summary of the results obtained is in table 5.2, where complete complementation is represented by +++, partial complementation by ++ or + and no growth by -. A summary of the tabulated results is presented below.

Growth at 37°C

p*trc* constructs: SF103 cells containing each of the p*trc* constructs was grown overnight in LB medium and the complementation experiment performed as described in materials and methods (chapter 2.6). From the results (figure 5.9 and table 5.2) it can be seen that while most of the strains containing the various plasmids showed normal growth, cells containing p*trc*ES-cpn60.3 once again were unhealthy and grew poorly

when compared to the other strains. This defect in growth is more prominent in media containing 0.1mM IPTG.

pBAD24 constructs: SF103 cells containing the pBAD24 constructs were grown overnight in LB medium and complementation experiments performed the next day. Cells containing the positive (pBAD-ESL) and negative (pBAD24) controls showed normal growth in media containing either 0.2% arabinose or 0.2% glucose (figure 5.11). Surprisingly, cells containing test plasmid pBAD24-cpn60.1 showed very poor growth in arabinose when compared to growth in glucose, which will be discussed further in this chapter (section 5.4).

Growth at 42°C

p*trc* constructs: Among all the strains with the various plasmids, only cells containing plasmids p*trc*10EL, p*trc*ES-cpn60.2 and p*trc*10-cpn60.2 showed complementation at 42°C. This complementation however was dependant on the presence of IPTG. In the absence of 0.1mM IPTG, only cells containing p*trc*ES-cpn60.2 showed good complementation. Only cells containing p*trc*10EL showed good complementation in presence of 0.1mM IPTG, while cells containing p*trc*10-cpn60.2 showed partial complementation (figure 5.10).

pBAD constructs: As predicted, no growth of cells containing pBAD24-cpn60.1 was observed, while positive control (pBAD-ESL) showed complementation when induced with 0.2% arabinose (figure 5.12).

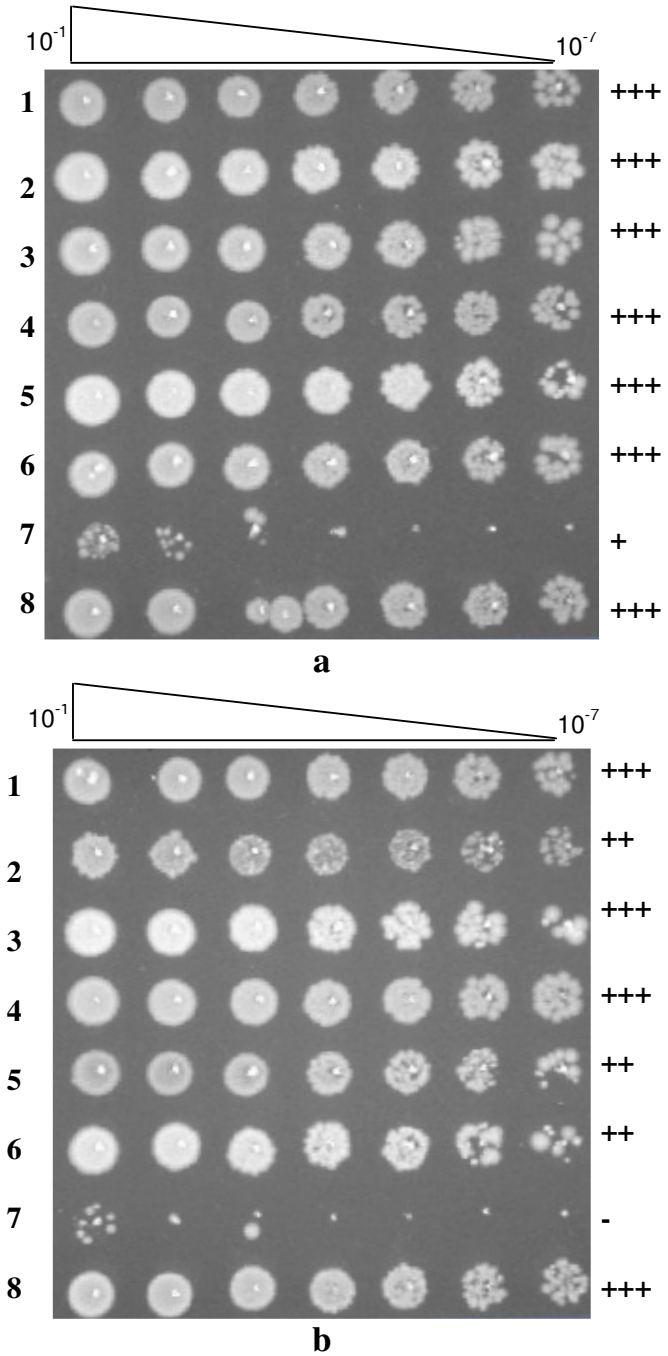


Figure 5. 9: **Growth of SF103 containing p*trc* constructs at 37°C**

Growth of SF103 at 37°C on media containing no IPTG (a) and 0.1mM IPTG (b) showing complete (+++), partial (++ or +) or no (-) growth.

E. coli SF103 cells containing plasmids p*trc*99A (1), p*trc*ESL (2), p*trc*10-EL (3), p*trc*10-cpn60.1 (4), p*trc*ES-cpn60.2 (5), p*trc*10-cpn60.2 (6), p*trc*ES-cpn60.3 (7), and p*trc*10-cpn60.3 (8) were checked for complementation at 37°C on media with or without 0.1mM IPTG.

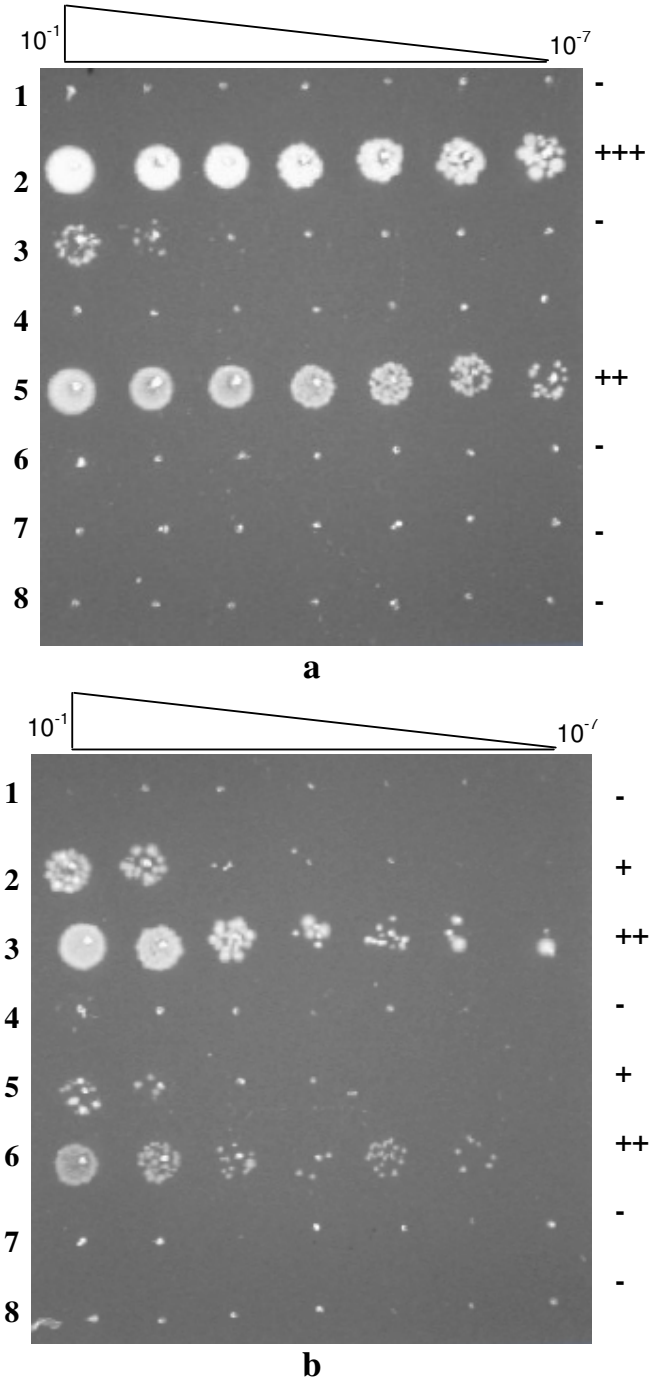


Figure 5. 10: **Complementation of SF103 containing p*trc* constructs at 42°C**

Complementation experiment conducted at 42°C on media containing no IPTG (a) and 0.1mM IPTG (b) showing complete (+++), partial (++ or +) or no (-) complementation.

E. coli SF103 cells containing plasmids p*trc*99A (1), p*trc*ESL (2), p*trc*10-EL (3), p*trc*10-cpn60.1 (4), p*trc*ES-cpn60.2 (5), p*trc*10-cpn60.2 (6), p*trc*ES-cpn60.3 (7), and p*trc*10-cpn60.3 (8) were checked for complementation at 42°C on media with or without 0.1mM IPTG.

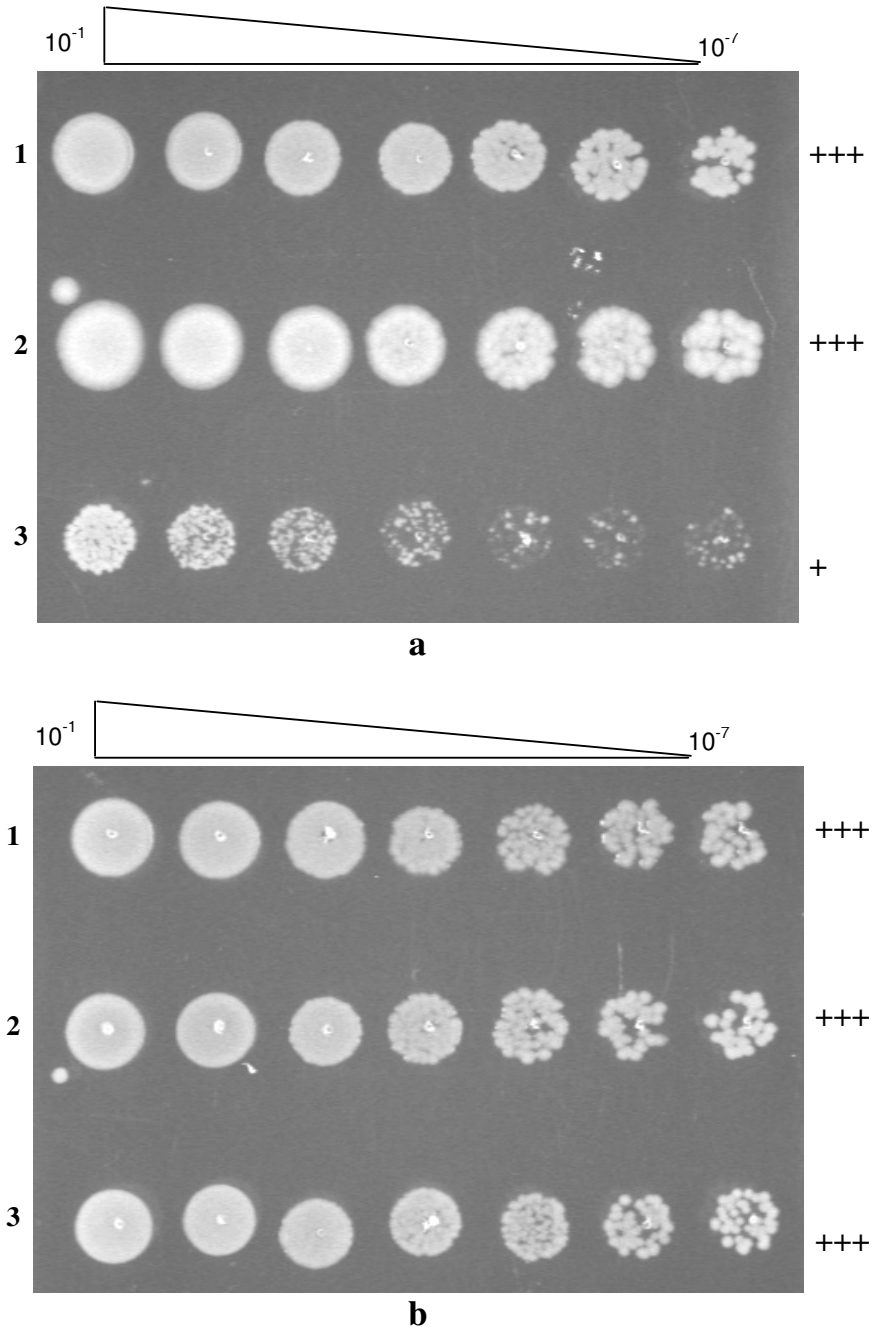


Figure 5. 11: **Growth of SF103 containing pBAD constructs at 37°C**

Growth of SF103 at 37°C on media containing 0.2% arabinose (a) and 0.2% glucose(b) showing complete (+++) or partial (++ or +) growth: *E. coli* SF103 cells containing plasmids pBAD24 (1), pBAD-ESL (2) and pBAD24-cpn60.1 (3)

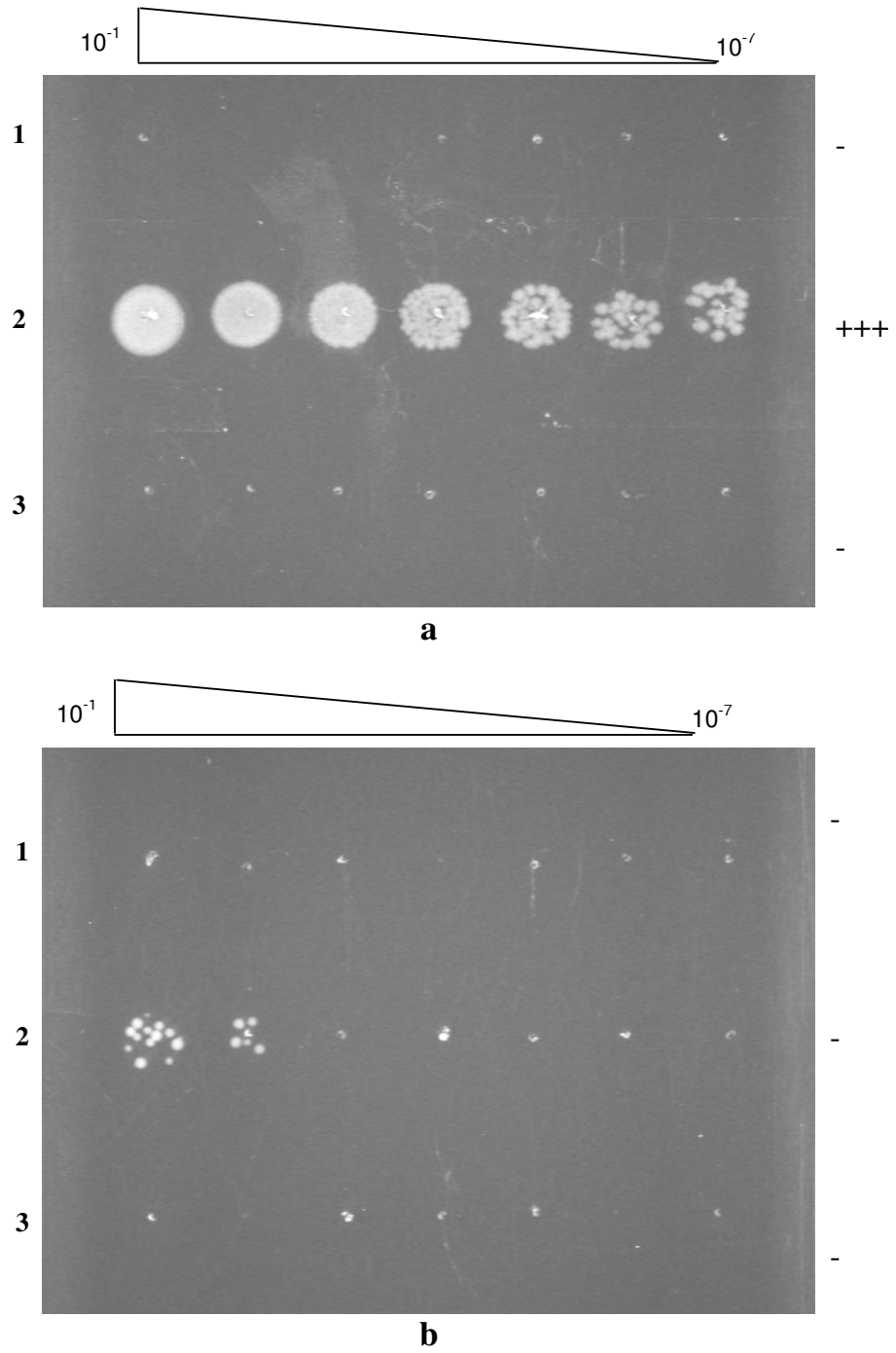


Figure 5. 12: **Complementation of SF103 containing pBAD constructs at 42°C**

Complementation experiment conducted at 42°C on media containing 0.2% arabinose (a) and 0.2% glucose (b) showing complete (+++) or no (-) complementation.

E. coli SF103 cells containing plasmids pBAD24 (1), pBAD-ESL (2) and pBAD24-cpn60.1 (3)

Plasmids in *E. coli* SF103	37°C		42°C	
	-IPTG	+ IPTG	-IPTG	+ IPTG
p*trc*99A	+++	+++	-	-
p*trc*ESL	+++	++	+++	+
p*trc*ES- SpeI	+++	+++	-	-
p*trc*ES-cpn60.2	+++	++	++	+
p*trc*ES-cpn60.3	+	-	-	-
p*trc*10-*groEL*	+++	+++	-	++
p*trc*10-cpn60.1	+++	+++	-	-
p*trc*10-cpn60.2	+++	++	-	++
p*trc*10-cpn60.3	+++	+++	-	-
	37°C		42°C	
	+ Arabinose	+ Glucose	+ Arabinose	+ Glucose
pBAD24	+++	+++	-	-
pBADESL	+++	+++	+++	-
pBAD24-cpn60.1	+	+++	-	-

Table 5. 2: **Summary of complementation results using *E. coli* SF103 strain**

Complementation of *E. coli* SF103 strain by different Cpn60 and Cpn10 proteins showing complete complementation (+++), partial complementation (++ or +) or no complementation (-). All the experiments were conducted at least three times and found to be highly reproducible.

5.3.3 Protein Expression

As lack of complementation could be due to either a failure of the protein to functionally replace GroEL or a lack of protein expression, we proceeded to analyse the expression of the various proteins. The expression of the various chaperonin genes present on the p*trc*99A plasmid was analysed, along with expression of Cpn60.1 from plasmids p*groE*-EScpn60.1 and pBAD24-cpn60.1. Total proteins were extracted and separated using an SDS-PAGE gel. In some cases, cell suspensions were sonicated and

centrifuged and then proteins extracted from soluble and insoluble fractions. In all cases, the amounts of sample loaded onto gels were adjusted so that they corresponded to an equal number of cells, as measured by optical density. The results obtained from expression studies of each of the chaperonin genes will be discussed individually and an interpretation of the observed results will follow in the Discussion section of this chapter.

Cpn60.1: Protein expression of the Cpn60.1 protein was analysed from cells containing pBAD24-cpn60.1, p*groE*-EScpn60.1 and p*trc*10-cpn60.1. Based on the sequence of the Cpn60.1 protein, the calculated molecular weight of a single subunit is 56.1kDa.

To determine if the Cpn60.1 protein was harmful to *E. coli* cells (as hypothesised from the difficulties encountered in cloning and expressing this gene in p*trc*99A), the *cpn60.1* gene was cloned into the tightly regulatable pBAD24 plasmid. Cultures of DH5α cells containing pBAD24-cpn60.1 were grown overnight and then sub-cultured during the day. The cultures were induced with 0.2% arabinose and samples were taken every hour to monitor protein expression (see chapter 2 for method). As can be seen in figure 5.13, expression of the Cpn60.1 protein is tightly regulated and there is no expression in the uninduced sample. In the induced samples, expression of the protein steadily increased every hour and peaked at four hours post induction. All the Cpn60.1 protein was present in the soluble cytosolic fraction. It is interesting to note that even though there is a large amount of protein being expressed, no complementation was observed in the SF103 strain containing this plasmid (section 5.3.2.2).

Expression of Cpn60.1 in MGM100 cells was analysed by extracting protein produced in plasmid p*groE*-EScpn60.1 and p*trc*10-cpn60.1. When cells containing p*groE*-EScpn60.1 were analysed no protein expression was observed (figure 5.14). In media containing arabinose (chromosomal GroESL present in cell), protein indicator MetE is absent in samples extracted from cells with p*groE*-EScpn60.1. However, when grown in glucose containing media, expression of MetE was observed indicating that there is no functional chaperonin protein in the cell (as explained in section 5.2). These results are in agreement with the complementation data which shows no complementing ability of cells containing this plasmid.

The expression of Cpn60.1 from *E. coli* MGM100 cells containing p*trc*10-cpn60.1 was analysed in different media (figure 5.15). No observable expression of Cpn60.1 can be detected in media containing 0.2% arabinose and as endogenous GroEL is expressed under these conditions, no MetE protein is observed either. However, in the presence of glucose, when no GroEL is present, and Cpn60.1 is not expressed, the 85kDa MetE protein is abundantly observed. When induced with IPTG, very low expression levels of Cpn60.1 can be observed (figure 5.15, lane 8) combined with a slight decrease in the levels of MetE as well. This may imply that the partial expression of Cpn60.1 is able to fold MetK, thus lowering MetE levels as well.

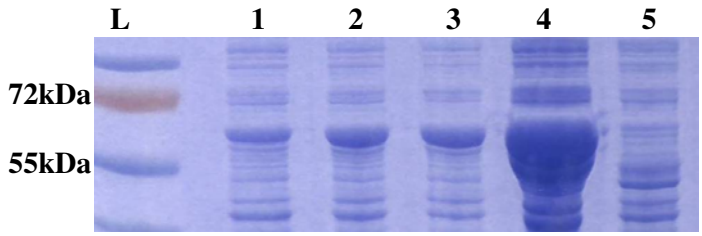


Figure 5. 13: Expression of Cpn60.1 from cells containing pBAD24-cpn60.1

12% SDS PAGE of soluble protein extracted from DH5α cells with plasmid pBAD24-cpn60.1 one hour after induction (1), two hours after induction (2), three hours post induction (3), four hours post induction (4) and uninduced sample (5).

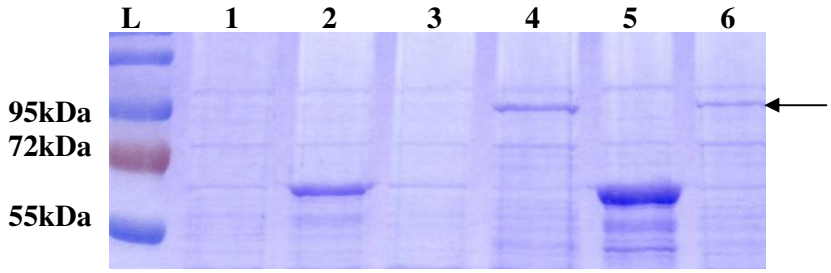


Figure 5. 14: Expression of Cpn60.1 from MGM100 cells containing p*groE*-EScpn60.1

(12% SDS PAGE gel)
L: molecular weight marker

1: MGM100, 0.2% arabinose

2: MGM100 + p*groE*-ESL, 0.2% arabinose

3: MGM100 + p*groE*-EScpn60.1, 0.2% arabinose

4: MGM100, 0.2% glucose

5: MGM100 + p*groE*-ESL, 0.2% glucose

6: MGM100 + p*groE*-EScpn60.1, 0.2% glucose

Presence of 85kDa MetE protein (arrow) can be seen in lanes 4 and 6.

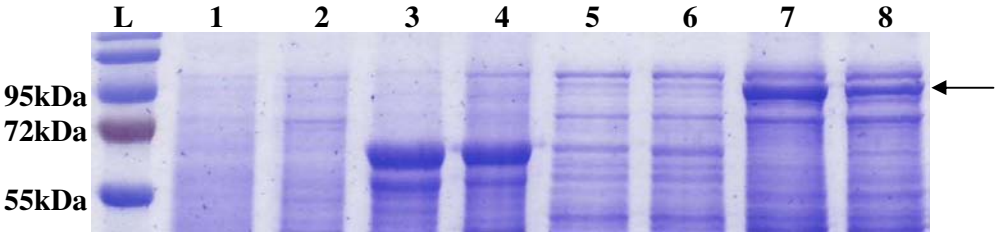


Figure 5. 15: Expression of Cpn60.1 from MGM100 cells containing p*trc*10-cpn60.1 (12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% arabinose, 0.1mM IPTG

2: MGM100 + p*trc*99A, 0.2% glucose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% arabinose, 0.1mM IPTG

4: MGM100 + p*trc*ESL, 0.2% glucose, 0.1mM IPTG

5: MGM100 + p*trc*10-cpn60.1, 0.2% arabinose

6: MGM100 + p*trc*10-cpn60.1, 0.2% arabinose, 0.1mM IPTG

7: MGM100 + p*trc*10-cpn60.1, 0.2% glucose

8: MGM100 + p*trc*10-cpn60.1, 0.2% glucose, 0.1mM IPTG

Presence of 85kDa MetE protein (arrow) can be seen in lanes 7 and 8.

Cpn60.2: The expression of the Cpn60.2 protein was analysed from *E. coli* MGM100 cells containing plasmid p*trc*ES-cpn60.2 and p*trc*10-cpn60.2. Interestingly, only cells containing these constructs showed the ability to replace functional GroEL in *E. coli*. As expected, results from these expression studies showed that Cpn60.2 is expressed irrespective of whether it is downstream of either the *E. coli groES* gene or the *M. smegmatis cpn10* gene (figures 5.16 and 5.17). The absence of the MetE protein in samples grown in 0.2% glucose further confirm that Cpn60.2 is able to completely replace *E. coli* GroEL.

Cpn60.3: As no complementation was observed previously, it was important to determine if any Cpn60.3 protein was being produced. Further more, cell growth was

very poor even when endogenous GroEL was present (in arabinose), possibly implying an interaction between Cpn60.3 and GroEL. *E. coli* MGM100 cells containing plasmid p*trc*ES-cpn60.3 and p*trc*10-cpn60.3 were analysed for expression of the Cpn60.3 protein.

When MGM100 cells with either p*trc*ES-cpn60.3 or p*trc*10-cpn60.3 were grown in the presence of 0.2% arabinose, no expression of protein was observed irrespective of whether the cells were induced with 0.1mM IPTG or not (figure 5.18 and 5.20). In the presence of medium containing only glucose, while no expression of Cpn60.3 protein was observed faint levels of MetE were observed. However, when grown in medium with glucose and induced with IPTG, very faint expression of Cpn60.3 is observed from cells containing p*trc*ES-cpn60.3 but not from cells with p*trc*10-cpn60.3 (figure 5.19 and 5.21). Importantly, even when Cpn60.3 is expressed at low levels (figure 5.19, lane 6), MetE levels decreased implying some level of functionality of Cpn60.3.

To check for the expression of Cpn60.3 protein in a background where the levels of endogenous GroEL was not controlled, but regulated by the cell itself, we proceeded to analyse the expression of Cpn60.3 from DH5α cells containing p*trc*ES-cpn60.3. Results showed that, when cultures were induced with 0.1mM IPTG, amounts of Cpn60.3 protein observed were less than those observed from uninduced cultures (figure 5.22). When the solubility of this protein was checked, it was found in inclusion bodies and could be solubilized with only 6M urea (figure 5.23). The significance of these results will be discussed shortly (section 5.4).

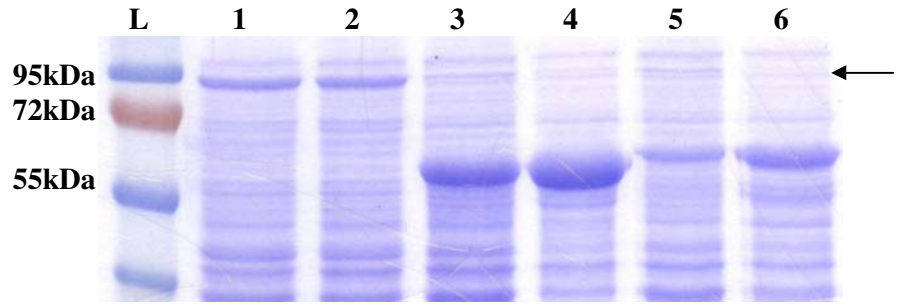


Figure 5. 16: Expression of Cpn60.2 from MGM100 cells containing p*trc*ES-cpn60.2

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% glucose

2: MGM100 + p*trc*99A, 0.2% glucose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% glucose

4: MGM100 + p*trc*ESL, 0.2% glucose, 0.1mM IPTG

5: MGM100 + p*trc*ES-cpn60.2, 0.2% glucose

6: MGM100 + p*trc*ES-cpn60.2, 0.2% glucose, 0.1mM IPTG

MetE protein (arrow) is present only in lanes 1 and 2

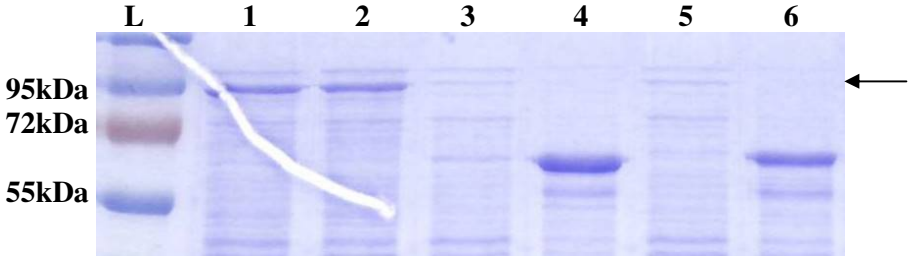


Figure 5. 17: Expression of Cpn60.2 from MGM100 cells containing p*trc*10-cpn60.2

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% glucose

2: MGM100 + p*trc*99A, 0.2% glucose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% glucose

4: MGM100 + p*trc*ESL, 0.2% glucose, 0.1mM IPTG

5: MGM100 + p*trc*10-cpn60.2, 0.2% glucose

6: MGM100 + p*trc*10-cpn60.2, 0.2% glucose, 0.1mM IPTG

MetE protein (arrow) is present only in lanes 1 and 2

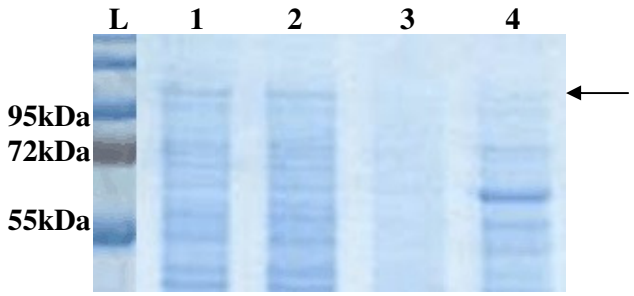


Figure 5. 18: Expression of Cpn60.3 from MGM100 cells containing p*trc*ES-cpn60.3

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*ES-cpn60.3, 0.2% arabinose

2: MGM100 + p*trc*ES-cpn60.3, 0.2% arabinose, 0.1mM IPTG

3: MGM100 + p*trc*ES-cpn60.3, 0.2% glucose

4: MGM100 + p*trc*ES-cpn60.3, 0.2% glucose, 0.1mM IPTG

MetE protein (arrow) is present only in lanes 1 and 2

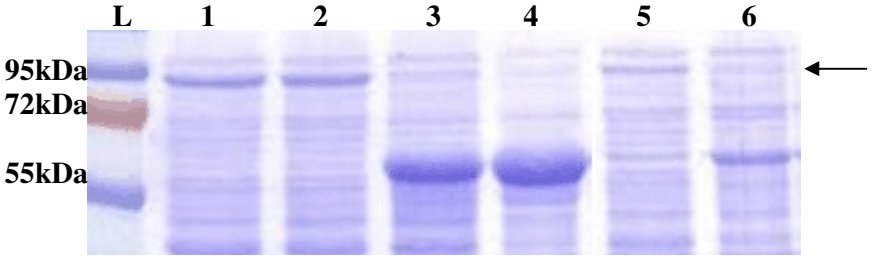


Figure 5. 19: Expression of Cpn60.3 from MGM100 cells containing p*trc*ES-cpn60.3

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% glucose

2: MGM100 + p*trc*99A, 0.2% glucose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% glucose

4: MGM100 + p*trc*ESL, 0.2% glucose, 0.1mM IPTG

5: MGM100 + p*trc*ES-cpn60.3, 0.2% glucose

6: MGM100 + p*trc*ES-cpn60.3, 0.2% glucose, 0.1mM IPTG

MetE protein (arrow) is present only in lanes 1, 2 and 5

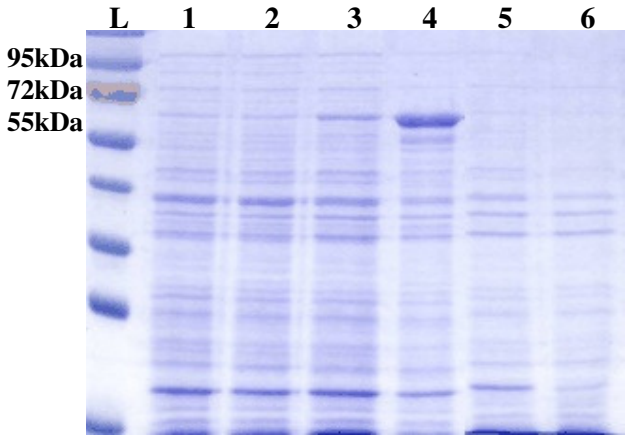


Figure 5. 20: **Expression of Cpn60.3 from MGM100 cells containing p*trc*10-cpn60.3 grown in 0.2% arabinose with or without 0.1mM IPTG**

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% arabinose

2: MGM100 + p*trc*99A, 0.2% arabinose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% arabinose

4: MGM100 + p*trc*ESL, 0.2% arabinose, 0.1mM IPTG

5: MGM100 + p*trc*10-cpn60.3, 0.2% arabinose

6: MGM100 + p*trc*10-cpn60.3, 0.2% arabinose, 0.1mM IPTG

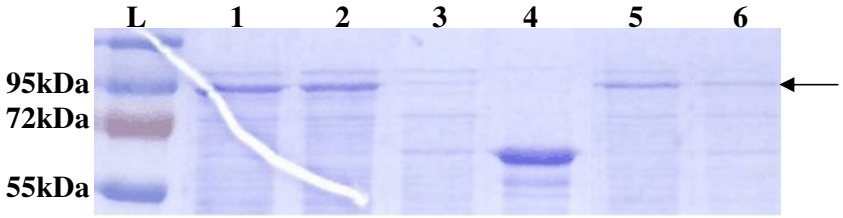


Figure 5. 21: **Expression of Cpn60.3 from MGM100 cells containing p*trc*10-cpn60.3 grown in 0.2% glucose with or without 0.1mM IPTG**

(12% SDS PAGE gel)

L: molecular weight marker

1: MGM100 + p*trc*99A, 0.2% glucose

2: MGM100 + p*trc*99A, 0.2% glucose, 0.1mM IPTG

3: MGM100 + p*trc*ESL, 0.2% glucose

4: MGM100 + p*trc*ESL, 0.2% glucose, 0.1mM IPTG

5: MGM100 + p*trc*10-cpn60.3, 0.2% glucose

6: MGM100 + p*trc*10-cpn60.3, 0.2% glucose, 0.1mM IPTG

MetE protein (arrow) is present only in lanes 1, 2, 5 and 6

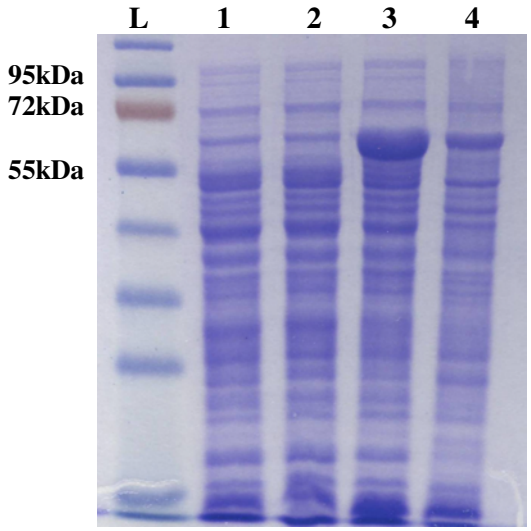


Figure 5. 22: Expression of Cpn60.3 from DH5α cells

(12% SDS PAGE gel)

L: molecular weight marker

1: DH5α + p*trc*99A

2: DH5α + p*trc*99A, 0.1mM IPTG

3: DH5α + p*trc*ES-cpn60.3

4: DH5α + p*trc*ES- cpn60.3, 0.1mM IPTG

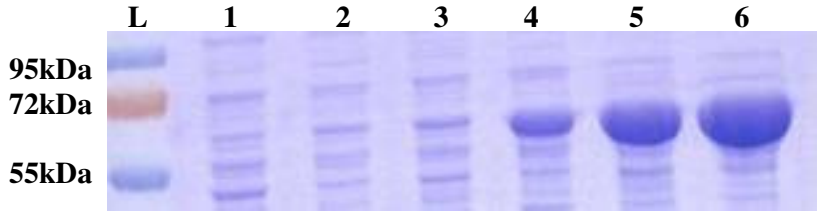


Figure 5. 23: Solubilisation of Cpn60.3 protein with Urea

(12% SDS PAGE gel)

L: molecular weight marker

1: DH5α + p*trc*ES-cpn60.3, 1M urea

2: DH5α + p*trc*ES-cpn60.3, 0.1mM IPTG, 1M urea

3: DH5α + p*trc*ES-cpn60.3, 2M urea

4: DH5α + p*trc*ES-cpn60.3, 0.1mM IPTG, 2M urea

5: DH5α + p*trc*ES-cpn60.3, 6M urea

6: DH5α + p*trc*ES-cpn60.3, 0.1mM IPTG, 6M urea

Strain (*E. coli*)	**Promoter**	**Co-chaperonin**	**Protein**	**Complementation**	**Expression**	**MetE**	**Solubility**
MGM100	*groE*	GroES	Cpn60.1	No	No	Yes	ND
MGM100	*trc*	Cpn10	Cpn60.1	No	No	Yes	ND
DH5α	pBAD	-	Cpn60.1	No	Yes	No	Soluble
MGM100	*trc*	GroES	Cpn60.2	Yes	Yes	No	Soluble
MGM100	*trc*	Cpn10	Cpn60.2	Yes	Yes	No	Soluble
MGM100	*trc*	GroES	Cpn60.3	No	Faint on induction	No	Soluble
MGM100	*trc*	Cpn10	Cpn60.3	No	Faint on induction	Yes	ND
DH5α	*trc*	GroES	Cpn60.3	ND	Yes	No	Inclusion bodies
MGM100	*trc*	GroES	GroEL	Yes	Yes	No	Soluble
MGM100	*trc*	Cpn10	GroEL	Yes	Yes	No	Soluble

Table 5. 3: Summary of results obtained from complementation and protein expression experiments

5.4 Discussion

A phylogenetic analysis of sequenced and not completely sequenced genomes reveals that in the Actinobacteria, when two copies of chaperonin 60 (*cpn60*) genes are present, one of them, *cpn60.1*, is always present in an operon with the co-chaperonin *cpn10*, while the second copy *cpn60.2* is always present on its own. In cases where only one copy of a cpn60 homologue is present, it has been found to be *cpn60.2* that is present (Goyal *et al*., 2006). This further confirms that the *cpn60.2* is essential while *cpn60.1* is not. The evolutionary rates of both genes are different, with *cpn60.1* evolving much faster than the essential copy (Goyal *et al*., 2006). The role the *cpn60.1* gene plays in biofilm maturation in *M. smegmatis* is indicative of functional divergence (Ojha *et al*., 2005). However, to date, the *cpn60.3* gene has not been knocked out and the role of Cpn60.3 remains unclear. While studies of expression levels under stress conditions could not reveal any particular function of this protein in the cell (chapter 3), we hoped to determine if this protein was able to function as a chaperonin in *E. coli*. Simultaneously, the expression and functionality of the other two chaperonins was also investigated. To this end, the three chaperonin genes were cloned into p*trc*99A or pBAD24 for co-expression with either the *E. coli* co-chaperonin or the *M. smegmatis* co-chaperonin.

While the cloning of the *cpn60.2* and *cpn60.3* genes was easy, the cloning of the *cpn60.1* gene proved extremely difficult. Initial cloning of *cpn60.1* into the pGEMTeasy vector suggested that expression of this protein may have been toxic to the cells, which is why the gene always recovered ligated in the opposite orientation to that of the upstream promoter. This hypothesis was further supported when no clones containing

the *cpn60.1* gene in the p*trc*ES plasmid were obtained. Once again, plasmids containing the *cpn60.1* gene were found to have point mutations and even insertion sequence elements. To check if the Cpn60.1 protein was indeed deleterious to *E. coli* cells, this gene was cloned into the tightly regulatable pBAD24 plasmid. Results obtained from this experiment suggested that while a slight slowing in growth was observed, the cells were still viable, even when the Cpn60.1 protein was expressed. This result proves that the Cpn60.1 protein is not deleterious to *E. coli*. However, while the *cpn60.1* gene could be cloned downstream of *groES* or *cpn10* in the p*groE* or p*trc*10 plasmids respectively, the reason for the inability to clone the gene in the p*trc*99A plasmid is still not known.

To determine if the presence of the co-chaperonin affected complementation or expression, two sets of plasmids containing either the *E. coli* co-chaperonin *groES* or the *M. smegmatis* co-chaperonin *cpn10* upstream of the *cpn60* genes were constructed.

In order to analyse whether the chaperonin proteins from *M. smegmatis* could complement for loss of function of the *E. coli* homologue, complementation studies were conducted in two strains *E. coli* MGM100 and *E. coli* SF103. While complementation in MGM100 can be tested at a range of temperatures, complementation in SF103 at 42°C is a more stringent test as previous studies from our lab have demonstrated (Ivic *et al*., 1997). The advantage however, of using the MGM100 system, is the presence of a functional GroEL marker, MetE (Chapman *et al*., 2006; Tabish Ahmed, personal commn.). On the other hand, genes cloned into the pBAD24 plasmid cannot be tested for complementation in MGM100, so SF103 provides a suitable yet stringent option.

When cells containing the p*trc*ES constructs were analysed, results obtained from this set of experiments showed that complementation was observed only in cells expressing Cpn60.2. Due to the leakiness of the *trc* promoter on the plasmid, growth was observed even in the absence of 0.1mM IPTG. This complementing ability was even observed at higher temperatures of 42°C. This is especially significant as at higher temperatures, only proteins capable of exhibiting complete chaperoning activity can allow for normal cell growth (Fayet *et al*., 1989). The lack of complementation of Cpn60.1 and Cpn60.3 from cells containing p*groE*-EScpn60.1 and p*trc*ES-cpn60.3 respectively could be either due to a failure of the proteins to functionally replace GroEL or because of no expression.

As expected, Cpn60.2 was expressed at levels comparable to that of the positive control GroEL. Furthermore, no MetE expression was observed. When expression of Cpn60.1 and Cpn60.3 from cells containing p*groE*-EScpn60.1 or p*trc*ES-cpn60.3 respectively, was monitored, low levels of Cpn60.1 and Cpn60.3 reduced levels of MetE. This strongly suggests that Cpn60.1 and Cpn60.3 can partially complement loss of GroEL. As it is known that GroEL folds 13 essential proteins in the cell, one hypothesis is that Cpn60.1 and Cpn60.3 can fold some of these essential proteins (MetK being one of them), that accounts for an observable decrease in MetE levels, however, due to the inability to fold the remaining essential proteins, restoration of growth and complementation is not observed.

When cells containing the p*trc*10 constructs were analysed, complementation by cells with p*trc*10-cpn60.2 could be seen only when induced by 0.1mM IPTG. Cells containing the other homologues, *cpn60.1* and *cpn60.3* failed to show any complementation. This complementation and functionality of the Cpn60.2 protein was further confirmed when protein expression was analysed and the MetE protein did not appear when Cpn60.2 was being expressed. Not surprisingly, no expression or very poor levels of expression of Cpn60.1 and Cpn60.3 respectively, was observed. The presence of MetE further confirms the lack of complementation of strains containing *cpn60.1* and *cpn60.*3. Interestingly, MGM100 cells containing *groEL* with *cpn10* appear to complement even at 42°C. Analysis of the expression of GroEL from these cells revealed that GroEL was expressed only when induced with 0.1mM IPTG.

These results shows that the *M. smegmatis* homologue is capable of folding an obligate GroEL substrate (MetK) that can be visualised. Recent proteomic studies investigating the substrates of *E. coli* GroEL have shown that 85 of the 350 interacting substrates are GroEL dependant (Kerner *et al*., 2005). Furthermore, 13 of these have been shown to be essential for cell growth. Hence, any complementing protein would have to exhibit complete chaperoning activity in folding these 13 essential proteins. This would indicate that substrate recognition and chaperoning activity between *E. coli* and *M. smegmatis* is conserved.

Two significant findings from the above results are, firstly, results from the expression studies reveal that the expression of the co-chaperonin Cpn10 appears to be tightly regulated and can be observed only when induced. The reason for this tight regulation

of only Cpn10 is unknown. Secondly, the *E. coli* GroEL has been shown to function with Cpn10. We know from previous work in our lab that although presence of *groES* is not essential in folding a few substrates, maximum chaperoning activity especially at higher temperatures can only be achieved by a fully functioning chaperonin and co-chaperonin (Erbse *et al*., 1999). Hence, this finding that *E. coli* GroEL can work together with *M. smegmatis* Cpn10 is significant.

MGM100 cells containing the p*trc*ES-cpn60.3 plasmid showed very poor growth on media containing arabinose, and this growth was further reduced when the gene was induced with 0.1mM IPTG. The reasons for this growth defect even though endogenous GroEL is being produced could be due to the fact that Cpn60.3 is interacting with available GroEL and forming mixed complexes as inclusion bodies, thereby inactivating GroEL. In a normal *E. coli* cell, expression of GroEL and most other chaperones is driven by their own promoters in response to levels of unfolded protein within the cell (Parsell and Sauer, 1989). However, in MGM100, since levels of GroEL are being regulated (by arabinose in media) and are not controlled by the cell, a lack of GroEL is probably resulting in the growth defect observed.

In order to test the hypothesis that in a strain where levels of GroEL were not artificially regulated (by pBAD promoter), the cell would respond to lower levels of GroEL (due to it's possible interaction with Cpn60.3) by up-regulating expression of GroEL, we transformed DH5α cells with the p*trc*ES-cpn60.3 plasmid and analysed the protein expression. Results obtained showed more protein (60kDa in size) in the uninduced samples when compared to the induced samples. This could be due to the fact that when

more Cpn60.3 protein is present in the cell, it interacts with existing GroEL resulting in non-functional complexes. This finding is not new as it has previously been shown that over-expression of GroEL results in the cell reaching a saturation point which results in cell death (Lund, PA, unpublished). This can be observed in our complementation results, where cells containing GroEL and GroES show poor growth when induced when compared with cells that were not induced. When the solubility of the visualised protein was tested, it was found to be insoluble and present in inclusion bodies. It is hypothesised that when the Cpn60.3 protein is expressed, it forms non-functional complexes with GroEL which are found as inclusion bodies within the cell. It is also hypothesised that this titrating effect of GroEL by Cpn60.3 to form non-functional complexes, is responsible for the poor growth of MGM100 cells containing *cpn60.3* on media containing arabinose. Due to time restrictions, these hypotheses could not be tested. However, possible further experiments could include co-immunoprecipitation studies followed by mass spectrometry or N-terminal sequencing to determine if the visualised protein (inclusion bodies) contains a mixture of GroEL and Cpn60.3.

The question of toxicity of Cpn60.1 is once again addressed due to results obtained in SF103 complementation. SF103 cells containing pBAD24-cpn60.1 show normal growth at the permissive temperature (37°C) when not induced (0.2% glucose). However, growth appears to reduce drastically, when cells are induced by 0.2% arabinose. This result suggests that when expressed, the Cpn60.1 protein is deleterious to SF103 cells.

In conclusion, in this part of the study we have demonstrated that Cpn60.2 is indeed the housekeeping chaperonin and can function extremely well in *E. coli* at higher

temperatures as well. This gives us a better understanding about substrate specificity being conserved between the two organisms. However, the mechanism by which this folding takes place needs to be addressed and the first step in determining that would be to compare the oligomeric states of the two proteins. Cpn60.1 has no chaperoning activity in *E. coli* and this could be due to low levels of expression or functional redundancy in *E. coli*. The reason for its toxicity to SF103 cells however is unknown. Cpn60.3 is similar in that it is also not able to completely functionally replace GroEL, and appears to interfere with normal cell growth when expressed with GroEL. As discussed, the hypothesis for this poor growth phenotype could be due to the interaction of the two proteins resulting in non-functional complexes.

Structural studies of the homologues from *M. tuberculosis* show that they always exist as lower oligomers and Cpn60.2 has even been crystallised as a dimer (Qamra *et al*., 2004; Qamra and Mande, 2004). Because of this and the fact that Cpn60.2, like its *M. tuberculosis* homologue can complement for loss of GroEL in *E. coli*, it is of interest to determine if the *M. smegmatis* chaperonins also form lower oligomers. For this purpose, the proteins would have to be purified and these results are presented in the next chapter.

6.1 Introduction

The structure of a chaperonin protein and the way each subunit interacts with the other to form a large functional species has always been a topic of interest, not only because of the precise mechanism by which the chaperonins function but also because of their specificity for certain substrates which make them essential for cell growth. The chaperonin family of proteins have been extensively studied with the *E. coli* tetradecamer GroEL and co-chaperonin GroES being the best characterised.

Protein folding within a cell is complicated by the fact that there are a large number of other macromolecules present (molecular crowding) causing proteins to aggregate in some instances (Ellis, 2001). To prevent this protein aggregation, which could have deleterious effects on the cell, the folding mechanism offered by the chaperonin proteins helps a number of proteins fold correctly. The chaperonin proteins have been shown to fold a large range of substrates some of which are essential (Houry *et al*., 1999; Kerner *et al*., 2005). This protein folding takes places in an ATP-dependant manner within an enclosed central cavity (Braig *et al*., 1994). This cavity is formed by homopolymeric subunits stacked in two rings and topped by a lid-like co-chaperonin (Xu *et al*., 1997).

The mechanism by which protein folding occurs has already been explained in detail in chapter 1. Briefly, protein folding takes place through a complex set of steps starting with binding of the unfolded protein at the hydrophobic apical end of one ring (*cis* ring) followed by the binding of seven ATP molecules to the *cis* ring. This results in allosteric changes in the ring (intra-ring positive co-operativity) and binding of co-chaperonin takes place which displaces the bound protein. ATP hydrolysis in the *cis*

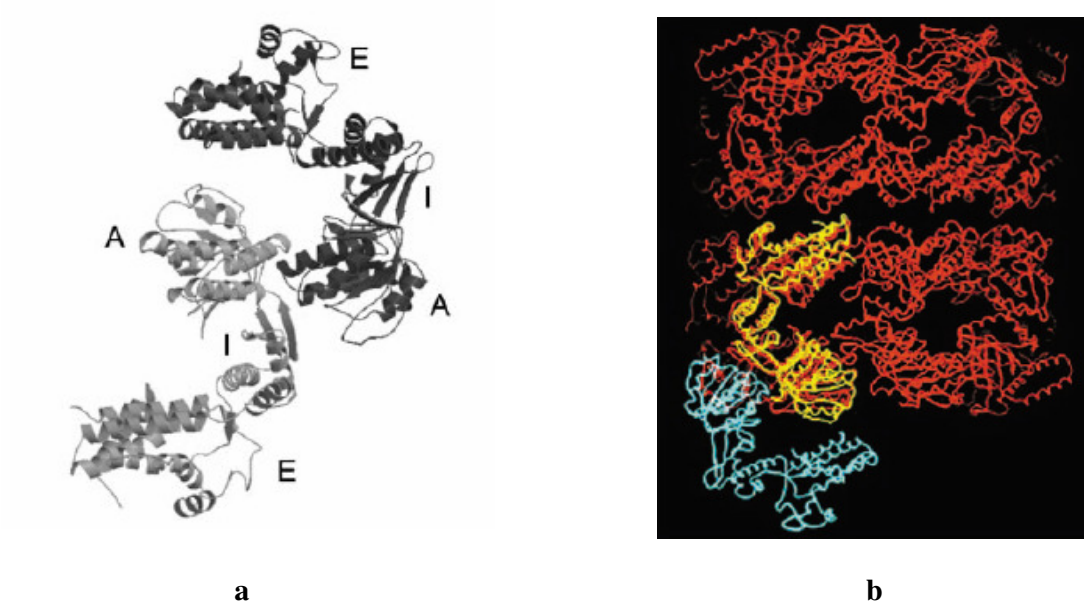
ring is followed by the binding of seven ATP molecules to the *trans* ring results and due to another set of allosteric changes (inter-ring negative co-operativity) the co-chaperonin disassociates from the *cis* ring. The binding of 7 ATP molecules to the *trans* ring is followed by GroES and substrate binding. The partially or completely folded protein is also released from the cavity.

This complex set of events illustrates the importance of the double ring structure of the chaperonin proteins in protein folding. However, there are certain exceptions where even in the absence of a functional double ring structure, protein folding still takes place. Previous work has shown that mutations in the equatorial region of the protein result in the formation of inactive single rings (Weissman *et al*., 1995). Although these single rings are capable of folding the protein, due to no inter-ring connectivity release of co-chaperonin is inhibited (Weissman *et al*., 1996, Xu *et al*., 1997). However, reports from work conducted in our lab have shown that chimeric single rings are sufficient for folding proteins that do not require *groES* as a co-factor (Erbse *et al*., 1999). Another study looking at single ring chimeras of *E. coli* GroEL and Mitochondrial Hsp60 has also shown them to be effective in folding proteins (Neilsen and Cowan, 1998). A screen of GroEL mutants further identified single rings that could function effectively (Sun *et al*., 2003). These reports suggest that single rings can function as well as the double ring, which raises the question, what is the adaptive value of the double ring? The answer is that even though the single rings are capable of folding substrates, their yield is much less when compared to that of the double rings. Furthermore, GroEL has been shown to be essential at all temperatures (Fayet *et al*., 1989), so the role of the double rings could be justified under conditions of stress. Thus it has been established

that even though single rings are capable of folding protein, the double rings structure is more efficient in protein folding.

Another set of proteins that deviate from the general rule of the necessity of double ring structures for functionality are the mycobacterial chaperonins. The *M. tuberculosis* Cpn60.1 and Cpn60.2 proteins have been shown to exist as lower oligomers under a range of different conditions aimed to mimic the physiology of the cell (Qamra *et al*., 2004). Furthermore, the Cpn60.2 protein was crystallised as a dimer (figure 6.1) (Qamra and Mande 2004). While the arrangements of the two subunits in the crystal structure of Cpn60.2 is vastly different from that of GroEL, the α carbon trace of each subunit of Cpn60.2 is very similar to that of GroEL (figure 6.2). However, genetic evidence from studies conducted in our lab show that the *M. tuberculosis* Cpn60.2 is capable of functioning in place of GroEL (Hu *et al*., 2008). More importantly, results obtained from this study (chapter 5) indicate that cells containing the *M. smegmatis* Cpn60.2 protein are capable of growth even at high temperatures of 42°C. As we know, efficient complementing ability cannot be achieved without the interaction of the chaperonin 60 protein with the heptameric co-chaperonin GroES. Thus, in order to functionally replace GroEL, Cpn60.2 must be interacting with GroES at some stage during the reaction cycle, strongly suggesting that Cpn60.2 must be forming large oligomers to facilitate this interaction. Thus, determining the oligomeric states of the chaperonins from *M. smegmatis* is of interest. While we hypothesise that the *M. smegmatis* chaperonins would probably resemble the oligomeric states of *M. tuberculosis* chaperonins, our convincing genetic evidence (chapter 5) is contradictory.

For the purpose of determining the oligomeric states of chaperonin proteins from *M. smegmatis*, we first purified the proteins and then analysed the proteins on native gels and also analysed the oligomers using Analytical Ultra Centrifugation (AUC). Because the Cpn60.3 protein was found in inclusion bodies, it was not purified.



a **b**

Figure 6. 1a: Ribbon diagram of crystal structure of the dimeric *M. tuberculosis* Cpn60.2 protein indicating the apical (A), intermediate (I) and equatorial (E) domains (taken from Qamra and Mande, 2004).

Figure 6. 1b: Superimposed carbon trace of *M. tuberculosis* Cpn60.2 and *E. coli* GroEL subunit α-carbon trace of the *M. tuberculosis* Cpn60.2 dimer, with a subunit superimposed onto a subunit of *E. coli* GroEL (taken from Qamra and Mande, 2004).

6.2 Results

6.2.1 Purification of Cpn60.1

The purification of the Cpn60.1 protein from *M. smegmatis* was made simpler by the presence of the naturally occurring C-terminal histidine tail (see chapter 3, figure 3.1), which made it possible to purify the protein using Ni- affinity chromatography. The Cpn60.1 gene was cloned into the pBAD24 plasmid as explained in section 5.3.1. *E. coli* DH5α cells containing the pBAD24-cpn60.1 were induced with 0.2% arabinose for 4 hours and the Cpn60.1 protein expressed in large quantities. After the cells were sonicated, the supernatant fraction was loaded onto a Nickel NTA column (GE healthcare), and bound using binding buffer (50mM NaCl, 50mM Tris pH 7.9). The unbound proteins were washed from the column using buffer (150mM NaCl, 50mM Tris pH 7.9 and 10mM Imidazole). The Cpn60.1 protein was eluted using buffer containing 150mM NaCl, 50mM Tris pH 7.9 and 200mM Imidazole. The various fractions collected were run on an SDS PAGE gel (figure 6.3 and 6.4). The trace of the absorbance (blue) and conductivity (red) can be seen in figure 6.2, where the initial peak is indicative of the unbound proteins (figure 6.3 lanes 1-5) and the small peak at the end, of the Cpn60.1 protein being eluted (figure 6.4 lanes 1-3).

The eluted Cpn60.1 obtained after purification using the Ni-NTA column was then subjected to a further round of purification using a hydrophobic purification column.

Figure 6. 2: **Trace of Ni-affinity chromatography of Cpn60.1**

Trace of Ni-affinity chromatography of Cpn60.1 showing the UV absorbance (blue line) and conductivity (red line). The various fractions collected during the series of washes on the column are indicated by red circle (unbound protein fraction), yellow circle (wash fractions) and blue circle (eluted fraction).

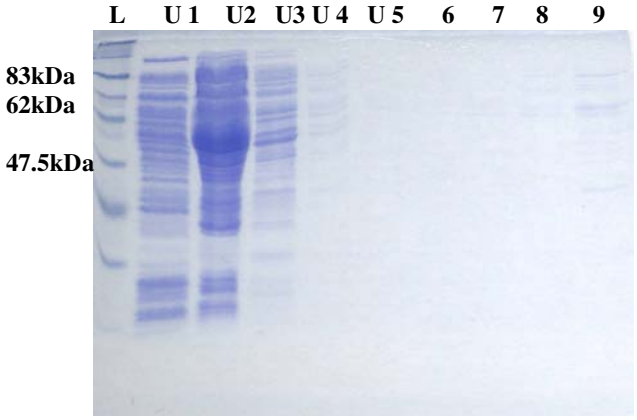


Figure 6. 3: SDS PAGE gel of fractions from the purification of Cpn60.1 using Ni-NTA column

Lanes U1-U5: Unbound protein eluants (samples collected during initial washes on column, red circle in figure 6.2), Lanes 6-9: Wash fractions (samples collected during washing stage, yellow circle on figure 6.2).

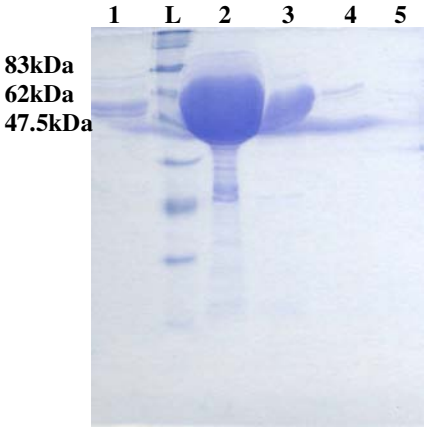


Figure 6. 4: SDS PAGE gel of fractions from the purification of Cpn60.1 using Ni-NTA column

Lanes 1-5: Eluted protein washes 1-5 (contain target protein, samples collected after addition of Imidazole to column, corresponds to blue circle in figure 6.2).

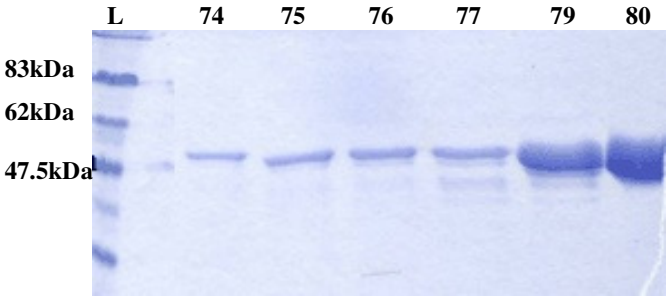


Figure 6. 5: SDS PAGE gel of fractions obtained from Cpn60.1 purification from hydrophobic column

Lanes 1-6: Eluted fractions

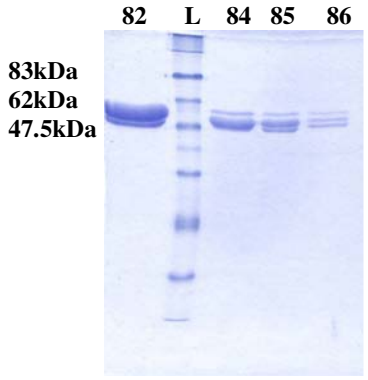


Figure 6. 6: SDS PAGE gel of fractions obtained from Cpn60.1 purification from hydrophobic column

Lanes 1-4: Eluted fractions

During the purification of Cpn60.1 using the Ni-NTA column, most of the target protein was eluted in fractions 2 and 3 as seen in figure 6.4. These fractions were pooled and the protein was then purified further using a hydrophobic column. The various fractions were then run on an SDS PAGE gel and the pure protein visualized (figures 6.5 and 6.6). Pure Cpn60.1 protein was observed in elution fractions 79, 80 and 82 (figures 6.5 and 6.6). Since the protein was in a high salt buffer $(NH_4)_2SO_4$, it was dialysed into a buffer containing a lower concentration of salt and stored at -80°C for further analysis.

6.2.2 Purification of Cpn60.2

As demonstrated in chapter 5, the Cpn60.2 protein is able to complement for the loss in function of its *E. coli* homologue, and can act either with its cognate Cpn10 or with *E. coli* GroES which is known to be heptameric. This implies that the protein must have a

structure similar to that of GroEL (higher oligomer). Although studies show that the *M. tuberculosis* Cpn60.2 protein crystallizes as a dimer (Qamra *et al.*, 2004), experiments conducted in our lab show that it is able to complement for GroEL (Hu *et al*., 2008). Since a similar phenotype is seen with the *M. smegmatis* Cpn60.2 proteins, we hypothesize that under similar conditions of purification (compared to the *M. tuberculosis*), the Cpn60.2 protein may also form lower oligomers

In order to test the above hypothesis, MGM100 cells containing the p*trc*ES-60.2 plasmid were grown in media containing glucose and induced with IPTG, after which total protein was extracted and the soluble fraction then used for purification.

6.2.2.1 DEAE sepharose ion exchange chromatography

The filtered soluble fraction obtained from MGM100 cells containing p*trc*ES-60.2, was applied to the DEAE sepharose column (see chapter 2 for method). The absorbance and conductivity were monitored through the chromatography (figure 6.7)

In the trace in figure 6.6, the first peak represents the unbound proteins (figure 6.8 lanes 5-8), while the next smaller peak (figure 6.8 within blue circle) is representative of fractions containing target protein (figure 6.9 lanes 1-4).

The eluted fractions (figure 6.9, lanes 1-4) containing the target protein were pooled and loaded on the hydrophobic column for further purification.

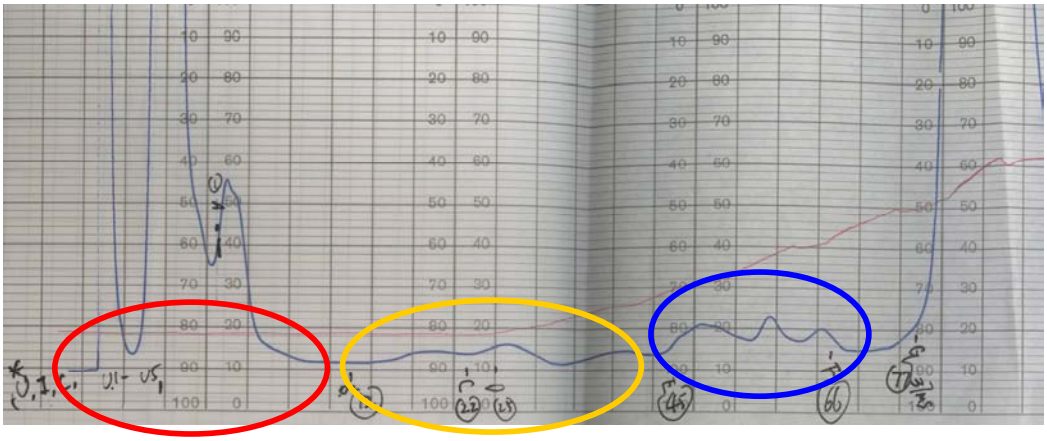


Figure 6. 7: DEAE sepharose ion exchange chromatography of Cpn60.2

Trace of UV absorbance at 280nm (blue line) and conductivity (red line). The various fractions collected during the series of washes on the column are indicated by red circle (unbound protein fraction), yellow circle (wash fractions) and blue circle (eluted fraction).

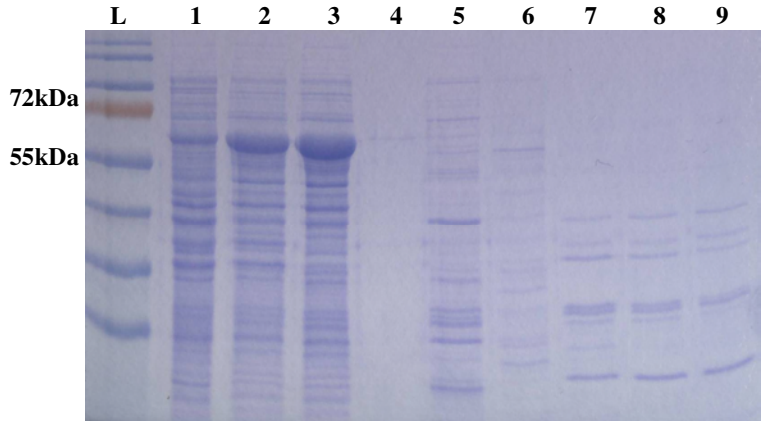


Figure 6. 8: SDS PAGE of fractions of proteins collected during purification of Cpn60.2 from DEAE ion exchange column

Lane 1: Uninduced sample MGM100 + p*trc*ES-60.2, Lane 2: IPTG induced sample MGM100 + p*trc*ES-60.2, Lane 3: Soluble fraction of induced sample MGM100 + p*trc*ES-60.2, Lanes 4-9: Unbound protein fractions (corresponding to red circle in figure 6.7).

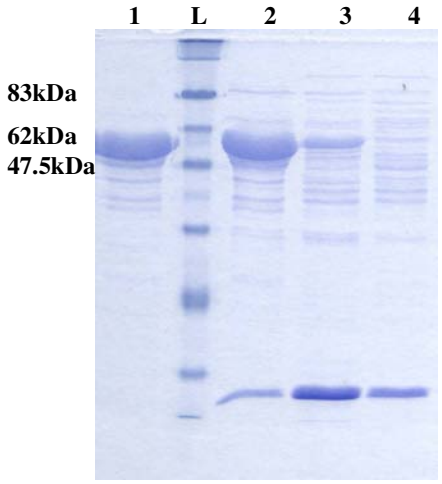


Figure 6. 9: SDS PAGE of fractions of proteins collected during purification of Cpn60.2 from DEAE ion exchange column

Lanes 1-4: Eluted sample fractions containing target protein (corresponding to blue circle in figure 6.7)

6.2.2.2 Hydrophobic column

The pooled protein from the above purification process was then loaded onto a hydrophobic column and the column washed using a gradient of high salt buffer (0.5M $(NH_4)_2SO_4$) represented by the high red line in the trace (figure 6.10) and going down to a low salt buffer (no $(NH_4)_2SO_4$, only phosphate buffer). The UV absorbance of the unbound proteins is represented by the initial blue peak on this trace in figure 6.10, while the second absorbance peak represents the eluted target protein. The various eluant fractions containing the Cpn60.2 protein were visualized on an SDS PAGE gel (figure 6.11, lanes 1-6) and pooled before being further purified using a size exclusion column.

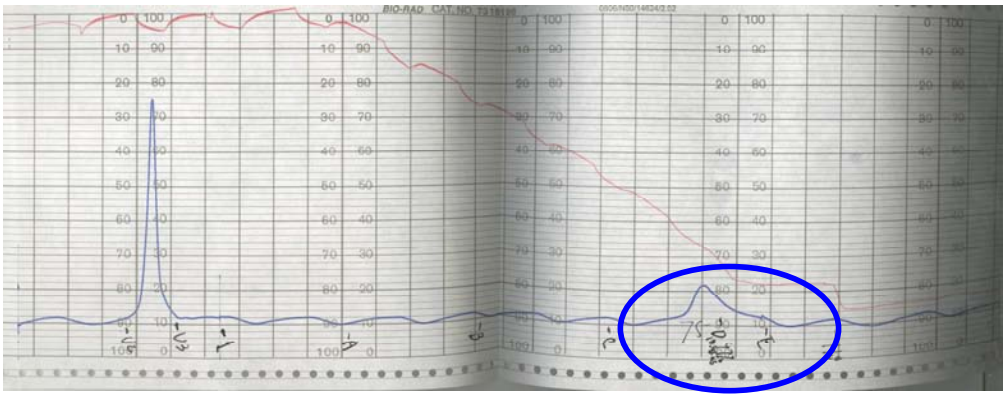


Figure 6. 10: Trace of hydrophobic column purification of Cpn60.2

showing the UV absorbance (blue) and conductivity (red) with the target protein being eluted (blue circle).

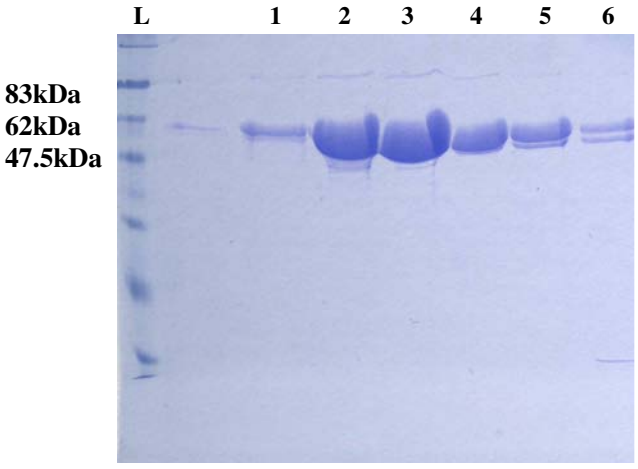


Figure 6. 11: SDS PAGE gel of eluted samples of Cpn60.2 after purification on hydrophobic column

Lanes 1-6: Eluted fraction samples containing the Cpn60.2 protein (correspond to blue circle in figure 6.10).

6.2.2.3 Size Exclusion purification

The pooled samples of Cpn60.2 protein obtained after elution from the hydrophobic column were loaded onto a size exclusion column as the last purification step. The trace in figure 6.12 represents the pure Cpn60.2 protein that was eluted as a single peak (one type of species present) and this pure protein can be seen on the SDS gel image (figure 6.13).

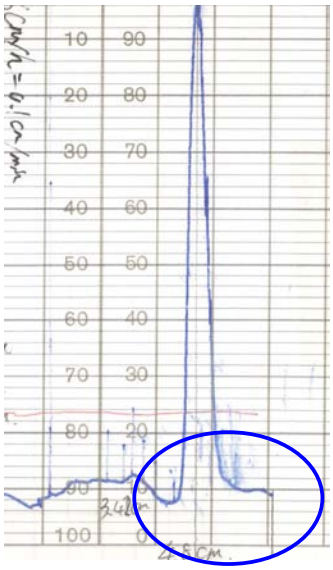


Figure 6. 12: Trace of size exclusion column purification of Cpn60.2 showing the UV absorbance (blue) and conductivity (red) with the target protein being eluted (blue circle).

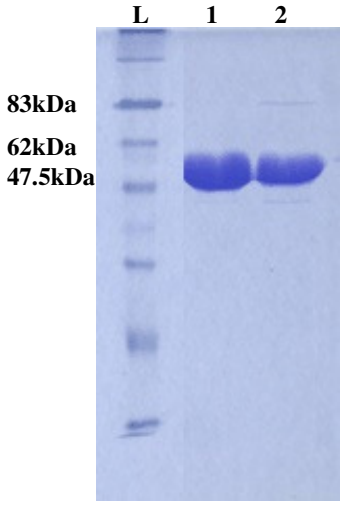


Figure 6. 13: **SDS PAGE gel of eluted samples of Cpn60.2 after purification on size exclusion column**

Lanes 1 and 2: Cpn60.2 protein (corresponding to blue circle in figure 6.12).

6.2.3 Native gels

A preliminary method of determining the oligomeric state of a protein is to extract protein complexes (native samples) and run the samples on a non-denaturing polyacrylamide gel. This method has previously been used by our lab and other groups to demonstrate the size of the un-denatured protein complex (Jones *et al*., 1998; Qamra *et al*., 2004; Liu *et al*., 2009).

Our initial findings suggest that the *M. smegmatis* Cpn60.1 and Cpn60.2 proteins form lower oligomers similar to their homologues in *M. tuberculosis*. From the native gel (figure 6.14), it can be observed that like the *M. tuberculosis* Cpn60.2, the *M. smegmatis* Cpn60.1 and Cpn60.2 form lower oligomers (possibly monomers or dimers). The calculated molecular weight of Cpn60.1 and Cpn60.2 are 56.1kDa and 56.4kDa respectively. This slight difference in size is apparent on the native gel where Cpn60.1

appears slightly smaller than Cpn60.2 (figure 6.14, lanes 5, 6 and 7). What is most clear though is that they do not form higher oligomers (tetradecamer) like the *E. coli* GroEL protein (figure 6.14, lane 2 and 3).

However, while this result shows the oligomeric state of the extracted protein, it does not give an indication of the oligomeric state of the protein within the cell, where it has been shown to function with the co-chaperonin GroES and complement.

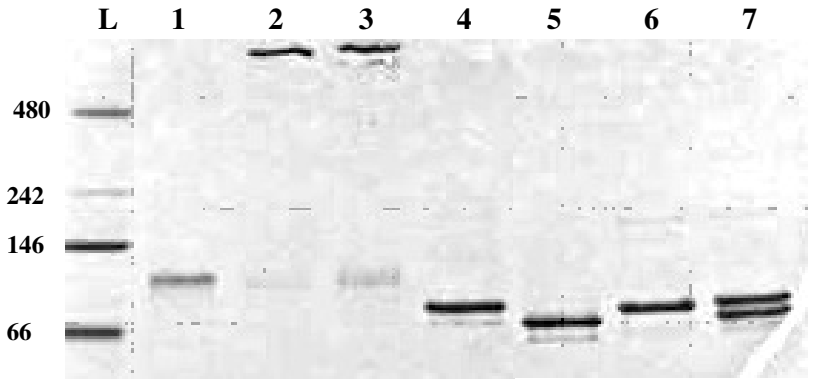


Figure 6. 14: **Native gel of various purified protein complexes**

(4-12% gradient native gel)

Lane 1: Purified GroES, Lane 2: Purified GroEL, Lane 3: Purified GroEL with urea

Lane 4: Purified Cpn60.2 *M. tuberculosis*, Lane 5: Purified Cpn60.1 *M. smegmatis*

Lane 6: Purified Cpn60.2 *M. smegmatis*, Lane 7: Purified Cpn60.1 and Cpn60.2 *M. smegmatis*

Analytical Ultra Centrifugation (AUC) has been used as a technique for determining the oligomeric state and shape of a protein complex for many years. It was originally described as a suitable technique by Svedberg in 1940 (Svedberg and Pedersen 1940). Briefly, AUC involves the application of a centrifugal force while simultaneously

observing the redistribution of the macromolecule in real-time. The advantage of this technique over various others is that the protein is not modified in any way as it does not require a tag or modification. Other techniques which can be used to study protein-protein interactions such as immuno-precipitation, cross-linking and fluorescence tagging involve labeling or interaction with a surface or matrix (Phizicky and Fields, 1995). But the major advantage of this technique is that the interaction and assembly of the various subunits can be monitored in solution which is similar to physiological conditions (Howlett *et al*., 2006). There are two main types of AUC. In the first type, a high centrifugal force is employed and the time-course of the sedimentation process is analysed (sedimentation velocity) and protein complexes are separated according to mass, density and shape. In the second type a lower centrifugal force is applied and the equilibrium between the diffused state and the sedimented state of the complex is measured (sedimentation equilibrium) (Scott and Schuck, 2005). The latter type is especially useful in studying dynamic interactions where the different forms of the complex are in equilibrium with each other.

The next set of experiments was conducted by a Dr. MingQi Fan in the lab and was to determine the oligomeric state of the proteins using this more accurate method of AUC. Sedimentation co-efficients were estimated using data from the AUC velocity run, using the SEDFIT programme, and molecular masses were estimated from the AUC equilibrium run (table 6.1). As can be seen in from the results, both proteins were monomeric under the conditions used (figure 6.16). The molecular masses of a single subunit obtained from these analyses are in the correct range as the calculated masses, however from these results it appears that Cpn60.1 is slightly larger than Cpn60.2. The

key point to note here is that under the same buffer conditions the *E. coli* GroEL protein appears to sediment as a teteradecamer as expected.

When the buffer conditions were altered and a higher concentration of salt and nucleotide were added (1M $(NH_4)_2SO_4$, 50mM Tris,50mM KCl, 7 or 10mM $MgCl_2$, pH 7.4, 1 or 0.5 mM ATP), it was observed that Cpn60.2 appeared to form a higher order oligomer (figure 6.17).

Protein	**Sedimentation coefficient (4°C)**	**MW (KDa)**	**Oligomer**
M. smegmatis Cpn60.1	1.6 S	58.8	Monomer
M. smegmatis Cpn60.2	1.6 S	58.6	Monomer
E. coli GroEL	12.8 S	820	Tetradecamer

Table 6. 1: **AUC results of purified *M. smegmatis* Cpn60.1 and Cpn60.2**

Results of AUC (velocity and equilibrium run) with purified *M. smegmatis* Cpn60.1 and Cpn60.2 proteins and *E. coli* GroEL (data from Dr. MingQi Fan).

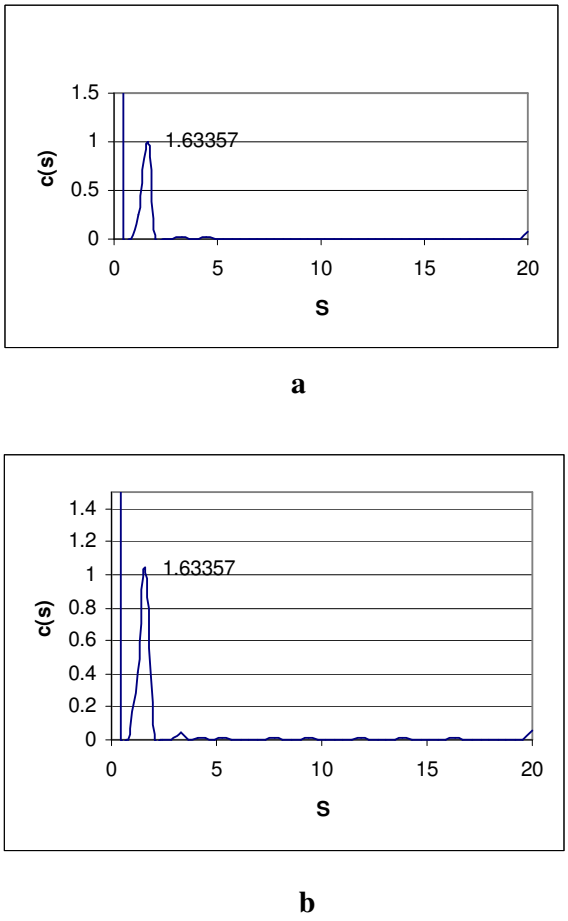


Figure 6. 15: Result from AUC (velocity run) using purified *M. smegmatis* Cpn60.1 and Cpn60.2

Result from AUC (velocity run) using purified *M. smegmatis* Cpn60.1 (a) and Cpn60.2 (b) in buffer containing 50mM Tris, 50mM KCl and 10mM MgCl2, pH 7.5 (figures from Dr. MingQi Fan).

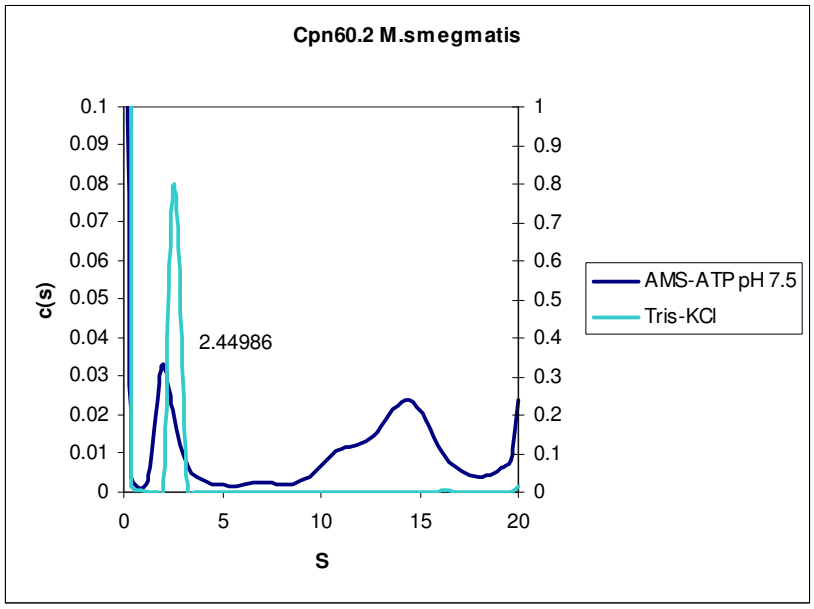


Figure 6. 16: Result from AUC (velocity run) using purified *M. smegmatis* Cpn60.2 in a high salt buffer

Result from AUC (velocity run) using purified *M. smegmatis* Cpn60.2 in buffer containing higher concentration of salt $(NH_4)_2SO_4$ and ATP (figure from Dr. MingQi Fan).

6.3 Discussion

In the previous chapter, we investigated the expression and complementing ability of the *M. smegmatis* chaperonins in *E. coli*. Our results from those experiments suggest that only the Cpn60.2 protein is able to complement for the loss of GroEL, while Cpn60.1 and Cpn60.3 do not appear to complement. This is similar to the results obtained from studies conducted using the *M. tuberculosis* homologues (Hu *et al*., 2008).

Since other studies have shown that there is a divergence of function between the various chaperonin proteins (Ojha *et al*., 2005; Hu *et al*., 2008), and that Cpn60.2 is the

housekeeping chaperonin, while Cpn60.1 appears to have taken on a different role in the cell, we decided to investigate whether this functional divergence was due to any structural differences between the proteins. For this purpose, we decided to purify the chaperonins from *M. smegmatis* and characterise them based on their oligomeric states.

As mentioned in the previous chapter, Cpn60.3 formed inclusion bodies in *E. coli* and was thus not purified. The purification of Cpn60.1 expressed from *E. coli* cells containing pBAD24-cpn60.1 was fairly simple, due to the presence of the histidine tail at the C-terminal end of the protein. Purification of Cpn60.2 from *E. coli* containing p*trc*ES-cpn60.2 was also performed and relatively pure protein was obtained at the end.

The initial analysis of the chaperonins was done using native PAGE which has previously been shown to be a relatively crude but fast method of determining the oligomeric states of a protein (Liu *et al.*, 2009). In the case of denaturing Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS PAGE), proteins are separated according to their electrophoretic mobility and not according to their charge (due to SDS binding). However with non-denaturing PAGE (native PAGE), the electrophoretic mobility of the protein is dependant on its size, shape and intrinsic charge.

Our results showed that both Cpn60.1 and Cpn60.2 migrated at relatively similar speeds and appeared to form smaller oligomers possibly monomers. However, as can be seen in lanes 2 and 3 in figure 6.14, the purified *E. coli* GroEL forms a tetradecamer.

However, to accurately determine the oligomeric state of these proteins, further work was conducted in the lab which made use of Analytical Ultra Centrifugation using the velocity and equilibrium methods. Even with this more accurate method the proteins were shown to be forming monomers.

As reiterated previously in the introduction of this chapter, the oligomeric states of the chaperonin play an important role in the protein folding reaction (Braig *et al*., 1994; Weissman *et al*., 1995; Xu *et al*., 1997). However, there are some exceptions where smaller oligomers such as single rings are found to function, albeit with less efficiency (Erbse *et al*., 1999; Sun *et al*., 2003).

However, contradictory structural studies from another group have shown that the *M. tuberculosis* Cpn60.2 exists as a lower oligomer and was crystallised as a dimer (Qamra *et al*., 2004; Qamra and Mande 2004). This implies that the conditions being used to test the oligomeric states of the chaperonins probably do not mimic the physiological conditions within the cell.

This lack of oligomer formation of chaperonins has been observed previously and was proposed to be due to dissociation during the extraction or purification process as seen in studies involving *Pseudomonas* sp #43 (Tokunaga *et al*., 1997), mammalian mitochondrial cpn60 (Viitanen *et al*., 1998) and *Thermus thermophilus* HB8 (Amada *et al*., 1995). The presence of nucleotides and co-factors has been found to play a big role in the reconstitution of these chaperonin proteins (Ybarra and Horowitz, 1995). A recent study found similar findings where the chaperonin protein from the halophilic

bacterium, *Tetragenococcus halophilia* was found to form a monomer *in vitro* but still functioned *in vivo* (Tosukhowong, A *et al*., 2005).

The reconstitution of the *Tetragenococcus halophilia* chaperonin protein was done in the presence of nucleotide, Mg^{2+} and higher concentrations of salt (Tosukhowong, A *et al*., 2005). To determine if the *M. smegmatis* chaperonins also formed larger oligomers under conditions of high salt, the formation of oligomers was tested under a range a different buffer conditions which were aimed at mimicking the physiology of the cell. Unsurprisingly, under a few conditions which used a very high concentration of salt (1M $(NH_4)_2SO_4$, 50mM Tris,50mM KCl, 7 or 10mM $MgCl_2$, pH 7.4, 1 or 0.5 mM ATPmM), Cpn60.2 appeared to form larger oligomers, possibly single rings. These results strongly indicate that under the correct conditions, even the mycobacterial chaperonins would be able to form larger oligomers.

In conclusion, we have successfully shown that the Cpn60.1 and Cpn60.2 proteins can be purified. Preliminary studies indicate that they form lower oligomers possibly monomers. However, under conditions where concentrations of salt and nucleotide are increased, the Cpn60.2 protein appears to form larger oligomers. This study needs to be taken further and a wide range of conditions under which large oligomers is formed need to be tested.

7.1 Conclusion

The main focus of this study was to characterise the multiple chaperonins of *M. smegmatis* based on their functional properties in *E. coli* and *M. smegmatis*. The background to the project was presented in the first chapter and then the results from each of the focal points of the project were reported and discussed in the results chapters.

At the end of the project we aimed to achieve a better understanding of the roles of the multiple chaperonins of *M. smegmatis*, which would provide an insight into the reason for the existence of multiple copies of *cpn60* genes. Simultaneously, due to their high similarity with homologues in the more pathogenic *M. tuberculosis*, any novel functions could serve as future areas of investigation into potential mechanisms for persistence or even as potential drug targets.

To this end, the studies in *M. smegmatis* were started by comparing the sequences of the *cpn60* genes with those in other sequenced mycobacteria and looking at the phylogeny and upstream regions. These studies gave early indications that, similar to other mycobacterial *cpn60* genes, the sequences of the two copies of *cpn60* genes were similar and conserved within the mycobacteria. Furthermore, due to its location in the genome in relation to the co-chaperonin *cpn10*, one gene (annotated as *cpn60.2* in most cases) could be assigned to be the housekeeping essential chaperonin gene while the other copy located together with the co-chaperonin was found to have more specialised functions (Ojha *et al*., 2005; Hu *et al*., 2008).

The levels of expression of these *cpn60* genes were then investigated under various stress conditions and it was observed that with the exception of *cpn60.3* all genes were very highly induced upon heat shock. While the change in expression levels of the genes due to the other stresses was not as high as those observed post heat stress, an induction up to 10 fold was observed in some cases. The *cpn60.1* gene was found to be dispensable for normal growth (Ojha *et al*., 2005) and this taken with the other findings so far allow us to conclusively state that *cpn60.2* is indeed the housekeeping chaperonin gene. While no distinct role for *cpn60.3* could be established from the experiments conducted in this study, future experiments will include knocking out the gene and studying the phenotypic changes and effects on any of the other genes on a whole genome level.

Another important finding from this study was confirmation of the fact that *cpn10* and *cpn60.1* are co-transcribed. Although present in a single operon, their individual expression levels varied considerably. When the 5' ends of the transcripts were analysed, it was observed that two such 5' ends existed. One of the 5' ends was in the expected position to be a transcriptional start site (upstream of *cpn10*), while the second 5' end was located upstream of *cpn60.1*. To investigate the possible presence of a second transcriptional start site within this operon, the intergenic region was analysed for the presence of promoter elements. However, no such putative promoter element could be identified under the conditions tested. This implies that the presence of two mRNA transcripts could be due to post-transcriptional processing of the mRNA which has been previously been discussed (chapter 3, section 3.4).

Next, to answer the question as to whether this specialised function of biofilm maturation is limited only to *cpn60.1* or if the other copies of the *cpn60* genes could perform the same function, complementation studies were performed. However, it proved difficult to replicate the original experiment demonstrating that biofilm formation in *M. smegmatis* is dependent on *cpn60.1* in our lab, and after a number of optimisation attempts, it was decided that the experiments would be conducted in the lab in Pittsburgh where the experiments were initially performed. Preliminary results obtained from those experiments strongly suggest this function of biofilm maturation appears to be unique to *cpn60.1*, as *cpn60.2* or *cpn60.3* did not appear to complement the loss in function of *cpn60.1*. A more interesting finding was that the *M. tuberculosis cpn60.1* homologue could complement for the loss in function of the *M. smegmatis cpn60.1*. These findings suggest that amino-acids which are conserved between the two *cpn60.1* genes could contribute to biofilm maturation function. To help identify these, future studies using protein alignment, domain swaps and site directed mutagenesis will need to be done. The aim of the domain swap experiments would be to construct chimeric proteins containing regions of homology in order to identify which particular region of the protein is involved in biofilm maturation. Once a particular region has been identified, this could then be taken further by performing single base pair substitutions to narrow down the particular bases responsible for the function.

In addition, Cpn60.1 was found to be associated with KasA which is involved with mycolic acid biosynthesis. Thus to characterise which region is involved in this association specific mutagenesis of bases or regions found to be conserved between the *cpn60.1* genes could be done followed by analysis of KasA association with the Cpn60.1 mutants. Once the region responsible for the association with KasA has been

determined, Cpn60.1 could potentially be used as a drug target, because of its indirect ability to impede mycolic acid biosynthesis.

As *E. coli* GroEL has served as the paradigm for the study of chaperonins and is essential for *E. coli* growth, the function of the three chaperonins and co-chaperonin from *M. smegmatis* was investigated in *E. coli*. For this purpose, the genes were cloned and growth and protein expression studies of the recombinant strains were performed. The complementation studies revealed that only the Cpn60.2 protein could complement for loss in function of the *E. coli* GroEL, further confirming that this protein plays the role of the housekeeping chaperonin in *M. smegmatis*. This proves that there must be an overlap of substrates between the mycobacteria and *E. coli*, providing insights into substrates folded by the chaperonin machinery.

As the oligomeric structure of the chaperonins is thought to be fundamental to their function, it was surprising to find that both the *M. tuberculosis* chaperonins form lower order oligomers, possibly dimers (Qamra *et al.*, 2004; Qamra and Mande, 2004) . To address this important aspect of apparent lack of oligomerisation of the proteins, studies on the oligomeric states of the *M. smegmatis* chaperonins were conducted. Similar to the homologues from *M. tuberculosis*, the Cpn60.1 and Cpn60.2 from *M. smegmatis* were also found to form lower order oligomers under the conditions tested in this study. However, because of our compelling genetic evidence which showed that Cpn60.2 could complement the loss of *E. coli* GroEL, we were persuaded that the Cpn60.2 must be forming higher oligomers during some stage in the reaction cycle. To this end, experiments are currently being conducted which test the oligomeric states of the

proteins under a range of conditions aimed at mimicking the physiology of the cell. Preliminary results obtained from these experiments strongly suggest that the *M. tuberculosis* Cpn60.2 does form a higher oligomer (heptamer and tetradecamer) under conditions of high salt and nucleotide. This study can be extended to the *M. smegmatis* Cpn60.2, which is predicted also to form the higher oligomer.

In conclusion, this study has enhanced the limited knowledge of the functions of the chaperonins of *M. smegmatis*, in that the essential role of Cpn60.2 has been confirmed by demonstrating its specific chaperoning function. Cpn60.1 on the other hand has not conclusively been shown to play a role in chaperoning unfolded proteins, but its role in biofilm maturation appears to be unique among the homologues of *M. smegmatis*. The similarities between the Cpn60.1 from *M. tuberculosis* and *M. smegmatis* provide convincing evidence that this protein must be playing a role in persistence mechanisms, which need to be further investigated. From our current study, the exact role of the third homologue, Cpn60.3, has not been established, but the foundations for future experiments have been laid out.

References

Adam, A., Petit, J.F., Wietzerbin-Falszpan, J., Sinay, P., Thomas, D.W., and Lederer, E. (1969) Mass spectrometric identification of N-glycolymuramic acid, a constituent of Mycobacterium smegmatis walls. . *FEBS Lett* **4**: 87-92.

Adams, J.L., and McLean, R.J. (1999) Impact of rpoS deletion on Escherichia coli biofilms. *Appl Environ Microbiol* **65**: 4285-4287.

Alsteens, D., Verbelen, C., Dague, E., Raze, D., Baulard, A.R., and Dufrene, Y.F. (2008) Organisation of the *Mycobacterial* cell wall: a nanoscale view. *Pflugers Arch* **456**: 117-125.

Alvarez, E.T. (1885) Recherches sur le bacille de Lustgarten. *Arch Physiol Normal Pathol*: 303-321.

Amada, K., Yohda, M., Odaka, M., Endo, I., Ishii, N., Taguchi, H., and Yoshida, M. (1995) Molecular cloning, expression, and characterization of chaperonin-60 and chaperonin-10
from a thermophilic bacterium, Thermus thermophilus HB8. *J. Biochem.*, **118**: 347–354.

Amann, E. (1985) Plasmid vectors for the regulated, high level expression of eukaryotic genes in Escherichia coli. *Dev Biol Stand* **59**: 11-22.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**: 1309-1314.

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* **181**: 223-230.

Aravindhan, V., Christy, A.J., Roy, S., Ajitkumar, P., Narayanan, P.R., and Narayanan, S. (2009) *Mycobacterium tuberculosis* groE promoter controls the expression of the bicistronicgroESL1 operon and shows differential regulation under stress conditions. FEMS Microbiol Lett **292:** 42-49.

Atlas, R.M. (1997) *Principles of Microbiology.* New York: WCB McGraw-Hill.

Av-Gay, Y., and Everett, M. (2000) The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. *Trends Microbiol* **8**: 238-244.

Baird, P.N., Hall, L.M., and Coates, A.R. (1988) A major antigen from Mycobacterium tuberculosis which is homologous to the heat shock proteins groES from E. coli and the htpA gene product of Coxiella burneti. *Nucleic Acids Res* **16**: 9047.

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

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., Collins, D., de Lisle, G., and Jacobs, W.R., Jr. (1994) inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science* **263**: 227-230.

Barreiro, C., Nakunst, D., Huser, A.T., de Paz, H.D., Kalinowski, J. and Martín, J.F. (2009) Microarray studies reveal a 'differential response' to moderate or severe heat shock of the HrcA- and HspR-dependent systems in Corynebacterium glutamicum. *Microbiol* **155**: 359-372.

Barry, C.E., 3rd, Lee, R.E., Mdluli, K., Sampson, A.E., Schroeder, B.G., Slayden, R.A., and Yuan, Y. (1998) Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* **37**: 143-179.

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

Belanger, A.E., Besra, G.S., Ford, M.E., Mikusova, K., Belisle, J.T., Brennan, P.J., and Inamine, J.M. (1996) The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the anti*Mycobacterial* drug ethambutol. *Proc Natl Acad Sci U S A* **93**: 11919-11924.

Belisle, J.T., Pascopella, L., Inamine, J.M., Brennan, P.J., and Jacobs, W.R., Jr. (1991) Isolation and expression of a gene cluster responsible for biosynthesis of the glycopeptidolipid antigens of Mycobacterium avium. *J Bacteriol* **173**: 6991-6997.

Berthet, F.X., Lagranderie, M., Gounon, P., Laurent-Winter, C., Ensergueix, D., Chavarot, P., Thouron, F., Maranghi, E., Pelicic, V., Portnoi, D., Marchal, G., and Gicquel, B. (1998) Attenuation of virulence by disruption of the Mycobacterium tuberculosis erp gene. *Science* **282**: 759-762.

Bhatt, A., Kremer, L., Dai, A.Z., Sacchettini, J.C., and Jacobs, W.R., Jr. (2005) Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to *Mycobacterial* cell lysis. *J Bacteriol* **187**: 7596-7606.

Bhatt, A., Fujiwara, N., Bhatt, K., Gurcha, S.S., Kremer, L., Chen, B., Chan, J., Porcelli, S.A., Kobayashi, K., Besra, G.S., and Jacobs, W.R., Jr. (2007) Deletion of kasB in Mycobacterium tuberculosis causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci U S A* **104**: 5157-5162.

Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7: 1513-1523.

Boone, D.R., Castenholz, R.W. and Garrity, G.M. (2001) *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1. New York: Springer.

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

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 A. *Nature* **371**: 578-586.

Brennan, P.J., and Nikaido, H. (1995) The envelope of *Mycobacteria*. *Annu Rev Biochem* **64**: 29-63.

Brennan, P.J. (2003) Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **83**: 91-97.

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

Brown, J.M., McNeil, M.M. and Desmond, E.P. (1999) Nocardia, Rhodococcus, Gordona, Actinomadura, Streptomyces, and other actinomycetes of medical importance. In *Manual of clinical microbiology*. Murray, P.R., Baron, E.J., Pfaller, M.A., Tenover, F.C. and Yolken, R.H. (ed). Washington, DC: American Society for Microbiology.

Bucca, G., Ferina, G., Puglia, A.M. and Smith, C.P., (1995) The *dnaK* operon of *Streptomyces coelicolor* encodes a novel heat-chock protein which binds to the promoter region of the operon. *Mol Microbiol* **17**: 663-74.

Bucca, G., Brassington, A.M., Schonfeld, H.J. and Smith, C.P. (2000) The HspR regulon of
streptomyces coelicolor: a role for the DnaK chaperone as a transcriptional co-repressor *Mol. Microbiol.* **38**: 1093–1103.

Bucca, G., Brassington, A. M., Hotchkiss, G., Mersinias, V. and Smith,C. P. (2003). Negative feedback regulation of dnaK, clpB and lon expression by the DnaK chaperone machine in Streptomyces coelicolor, identified by transcriptome and in vivo DnaK-depletion analysis. *Mol Microbiol* **50**: 153–166.

Bukau, B. (1993) Regulation of the Escherichia coli heat-shock response. *Mol Microbiol* **9**: 671-680.

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

Burnett, B.P., Horwich, A.L. and Low, K.B. (1994) A Carboxy-terminal deletion impairs the assembly of GroEL and confers a pleiotropic phenotype in *Escherichia coli* K-12. *J Bacteriol* **176**: 6980-6985.

Canova, M.J., Kremer, L., and Molle, V. (2009) The Mycobacterium tuberculosis GroEL1 chaperone is a substrate of Ser/Thr protein kinases. *J Bacteriol* **191**: 2876-2883.

Cerdeno-Tarraga, A.M., Efstratiou, A., Dover, L.G., Holden, M.T., Pallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M.A., Rabbinowitsch, E., Rutherford, K.M., Thomson, N.R., Unwin, L., Whitehead, S., Barrell, B.G., and Parkhill, J. (2003) The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. *Nucleic Acids Res* **31**: 6516-6523.

Chapman, E., Farr, G.W., Usaite, R., Furtak, K., Fenton, W.A., Chaudhuri, T.K., Hondorp, E.R., Matthews, R.G., Wolf, S.G., Yates, J.R., Pypaert, M., and Horwich, A.L. (2006) Global aggregation of newly translated proteins in an Escherichia coli strain deficient of the chaperonin GroEL. *Proc Natl Acad Sci U S A* **103**: 15800-15805.

Chaudhuri, R.R., and Pallen, M.J. (2006) xBASE, a collection of online databases for bacterial comparative genomics. Nucleic Acids Res. 34: 335-337.

Chen, J.M., German, G.J., Alexander, D.C., Ren, H., Tan, T. and Liu, J. (2006) Roles of Lsr2 in Colony Morphology and Biofilm Formation of *Mycobacterium smegmatis*. *J Bacteriol* **188**: 633-641.

Chen, J.M., Ren, H., Shaw, J.E., Wang, Y.J., Li, M., Leung, A.S., Tran, V., Berbenetz, N.M., Kocincova, D., Yip, C.M., Reyrat, J.M., and Liu, J. (2008) Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic Acids Res* **36**: 2123-2135.

Coates, A.R., Shinnick, T.M., and Ellis, R.J. (1993) Chaperonin nomenclature. *Mol Microbiol* **8**: 787.

Coburn, G.A., and Mackie, G.A. (1999) Degradation of mRNA in Escherichia coli: an old problem with some new twists. *Prog Nucleic Acid Res Mol Biol* **62**: 55-108

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, 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.

Cook, G.M., Berney, M., Gebhard, S., Heinemann, M., Cox, R.A., Danilchanka, O., and Niederweis, M. (2009) Physiology of *Mycobacteria*. *Adv Microb Physiol* **55**: 81-182, 318-189.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995) Microbial biofilms. *Annu Rev Microbiol* **49**: 711-745.
Daffe, M., and Draper, P. (1998) The envelope layers of *Mycobacteria* with reference to their pathogenicity. *Adv Microb Physiol* **39**: 131-203.

Dahl, J.L., Arora, K., Boshoff, H.I., Whiteford, D.C., Pacheco, S.A., Walsh, O.J., Lau-Bonilla, D., Davis, W.B., and Garza, A.G. (2005) The relA homolog of Mycobacterium smegmatis affects cell appearance, viability, and gene expression. *J Bacteriol* **187**: 2439-2447.

Dalbow, D.G., and Young, R. (1975) Synthesis time of □-galactosidase in Escherichia coli B/r as a function of growth rate. *Biochem J* **150**: 13-20.

Davey, M.E. and O'Toole, G.A. (2000) Microbial Biofilms: from Ecology to Molecular Genetics. *Microbiol Mol Bio Rev* **64**: 847-867.

De Bruyn, J., Bosmans, R., Turneer, M., Weckx, M., Nyabenda, J., Van Vooren, J.P., Falmagne, P., Wiker, H.G., and Harboe, M. (1987) Purification, partial characterization, and identification of a skin-reactive protein antigen of Mycobacterium bovis BCG. *Infect Immun* **55**: 245-252.

Deol, P., Vohra, R., Saini, A.K., Singh, A., Chandra, H., Chopra, P., Das, T.K., Tyagi, A.K., and Singh, Y. (2005) Role of Mycobacterium tuberculosis Ser/Thr kinase PknF: implications in glucose transport and cell division. *J Bacteriol* **187**: 3415-3420.

Ehira, S., Teramoto, H., Inui, M. and Yukawa, H. (2009) Regulation of *Corynebacterium glutamicum* Heat Shock Response by the Extracytoplasmic-Function Sigma Factor SigH and Transcriptional Regulators HspR and HrcA. *J Bacteriol* **191**: 2964-2972.

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

Ellis, R.J. (1993) The general concept of molecular chaperones. *Philos Trans R Soc Lond B Biol Sci* **339**: 257-261.

Ellis, R.J. (1994) Molecular chaperones. Opening and closing the Anfinsen cage. *Curr Biol* **4**: 633-635.

Ellis, R.J. (1997) Molecular chaperones: avoiding the crowd. *Curr Biol* **7**: R531-533.

Ellis, R.J. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**: 114-119.

Ellis, R.J., and Minton, A.P. (2003) Cell biology: join the crowd. *Nature* **425**: 27-28.

Embley, T.M., and Stackebrandt, E. (1994) The molecular phylogeny and systematics of the actinomycetes. *Annu Rev Microbiol* **48**: 257-289.

Erbse, A., Yifrach, O., Jones, S. and Lund, P.A. (1999) Chaperone Activity of a Chimeric GroEL Protein That Can Exist in a Single or Double Ring Form. *J Biol Chem* **274**: 20351-20357.

Esko, J.D., Doering, T.L. and Raetz, C.R.H. (2009) Eubacteria and Archaea. In *Essentials of Glycobiology*. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (ed). La Jolla, California: The Consortium of Glycobiology Editors.

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.

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

Fenton, W.A., and Horwich, A.L. (2003) Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q Rev Biophys* **36**: 229-256.

Fernandes, N.D., Wu, Q.L., Kong, D., Puyang, X., Garg, S. and Husson, R.N. (1999) A Mycobacterial Extracytoplasmic Sigma Factor Involved in Survival following Heat Shock and Oxidative Stress. *J Bacteriol* **181**: 4266-4274.

Fink, A.L. (1999) Chaperone-mediated protein folding. *Physiol Rev* **79**: 425-449.

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

Fitzkee, N.C., Fleming, P.J., Gong, H., Panasik, N., Jr., Street, T.O., and Rose, G.D. (2005) Are proteins made from a limited parts list? *Trends Biochem Sci* **30**: 73-80.

Fossati, G., Izzo, G., Rizzi, E., Gancia, E., Modena, D., Moras, M.L., Niccolai, N., Giannozzi, E., Spiga, O., Bono, L., Marone, P., Leone, E., Mangili, F., Harding, S., Errington, N., Walters, C., Henderson, B., Roberts, M.M., Coates, A.R., Casetta, B., and Mascagni, P. (2003) Mycobacterium tuberculosis chaperonin 10 is secreted in the macrophage phagosome: is secretion due to dissociation and adoption of a partially helical structure at the membrane? *J Bacteriol* **185**: 4256-4267.

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

Gao, B., and Gupta, R.S. (2005) Conserved indels in protein sequences that are characteristic of the phylum *Actinobacteria*. *Int J Syst Evol Microbiol* **55**: 2401-2412.

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.

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.

Glaziou, P., Floyd, K., and Raviglione, M. (2009) Global Burden and Epidemiology of Tuberculosis. *Clin Chest Med* **30**: 621-636.

Glickman, M.S., Cox, J.S., and Jacobs, W.R., Jr. (2000) A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. *Mol Cell* **5**: 717-727.

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

Gopalaswamy, R., Narayanan, S., Jacobs, W.R., Jr., and Av-Gay, Y. (2008) Mycobacterium smegmatis biofilm formation and sliding motility are affected by the serine/threonine protein kinase PknF. *FEMS Microbiol Lett* **278**: 121-127.

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.

Gragerov, A.I., Martin, E.S., Krupenko, M.A., Kashlev, M.V., and Nikiforov, V.G. (1991) Protein aggregation and inclusion body formation in Escherichia coli rpoH mutant defective in heat shock protein induction. *FEBS Lett* **291**: 222-224.

Grandvalet, C., de Crecy-Lagard, V. and Mazodier, P. (1999). The ClpB ATPase of *Streptomyces albus* G belongs to the HspR heat shock regulon. *Mol Microbiol* **31**: 521-532.

Grossman, A.D., Erickson, 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.

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

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.

Gutsche, I., Holzinger, J., Rauh, N., Baumeister, W., and May, R.P. (2001) ATP-induced structural change of the thermosome is temperature-dependent. *J Struct Biol* **135**: 139-146.

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.

Haber, E., and Anfinsen, C.B. (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J Biol Chem* **237**: 1839-1844.

Hall-Stoodley, L., and Lappin-Scott, H. (1998) Biofilm formation by the rapidly growing *Mycobacterial* species Mycobacterium fortuitum. *FEMS Microbiol Lett* **168**: 77-84.

Hall-Stoodley, L., and Stoodley, P. (2005) Biofilm formation and dispersal and the transmission of human pathogens. *Trends Microbiol* **13**: 7-10.

Hall-Stoodley, L., Brun, O.S., Polshyna, G., and Barker, L.P. (2006) Mycobacterium marinum biofilm formation reveals cording morphology. *FEMS Microbiol Lett* **257**: 43-49.

Hall-Stoodley, L. and Stoodley, P. (2009) Evolving concepts in biofilm infections. *Cell Microbiol* **11**: 1034-1043.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* 166: 557-580.

Hartl, F.U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**: 571–579.

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

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

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.

Henderson, B., Allan, E., and Coates, A.R. (2006) Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. *Infect Immun* **74**: 3693-3706.

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

Herman, C., Thevenet, D., D'Ari, R., and Bouloc, P. (1995) Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. *Proc Natl Acad Sci U S A* **92**: 3516-3520.

Hett, E.C., and Rubin, E.J. (2008) Bacterial growth and cell division: a *Mycobacterial* perspective. *Microbiol Mol Biol Rev* **72**: 126-156, table of contents.

Hickey, T.B., Thorson, L.M., Speert, D.P., Daffe, M., and Stokes, R.W. (2009) Mycobacterium tuberculosis Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient bacterial association with macrophages. *Infect Immun* **77**: 3389-3401.

Holmes, R.K. (2000) Biology and molecular epidemiology of diphtheria toxin and the tox gene. *J Infect Dis* **181 Suppl 1**: S156-167.

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*.

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

Hu, Y., Butcher, P.D., Mangan, J.A., Rajandream, M.A., and Coates, A.R. (1999) Regulation of hmp gene transcription in Mycobacterium tuberculosis: effects of oxygen limitation and nitrosative and oxidative stress. *J Bacteriol* **181**: 3486-3493.

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.

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 A resolution. *Nature* **379**: 37-45.

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

Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T., and Hattori, M. (2004) The complete genomic sequence of Nocardia farcinica IFM 10152. *Proc Natl Acad Sci U S A* **101**: 14925-14930.

Ivic, A., Olden, D., Wallington, E.J., Lund, P.A. (1997) Deletion of *Escherichia coli groEL* is complemented by a *Rhizobium leguminosarum groEL* homologue at 37 degrees C but not at 43 degrees C. *Gene* **194**: 1-8.

Jacobs, W.R., Jr. (2000) Mycobacterium tuberculosis: a once genetically intractable organism. In *Molecular Genetics of Mycobacteria*. Hatfull, G.F. and Jacobs, W.R., Jr (ed). Washington DC ASM Press, pp. 1-16.

Jain, S., Kaushal, D., DasGupta, S. K., and Tyagi, S.K. (1997) Construction of shuttle vectors for genetic manipulation and molecular analysis of Mycobacteria. *Gene* **190**: 37-44.

Jain, V., Sujatha, S., Ojha, A.K., and Chatterji, D. (2005) Identification and characterization of rel promoter element of Mycobacterium tuberculosis. *Gene* **351**: 149-157.

Jones, S., Wallington, E.J., George, R. and Lund, P.A. (1998) An arginine residue (Arg101), which is conserved in many GroEL homologues, is required for interactions between the two heptameric rings. *J Mol Biol* **282**: 789-800.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B.J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A.C., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D.A., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V.F., Wiegrabe, I., and Tauch, A. (2003) The complete Corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* **104**: 5-25.

Kapatai, G., Large, A., Benesch, J.L., Robinson, C.V., Carrascosa, J.L., Valpuesta, J.M., Gowrinathan, P., and Lund, P.A. (2006) All three chaperonin genes in the archaeon Haloferax volcanii are individually dispensable. *Mol Microbiol* **61**: 1583-1597.

Karatan, E. and Watnick, P. (2009) Signals, Regulatory Networks, and Materials That Build and Break Bacterial Biofilms. *Microbiol Mol Bio Rev* **73**: 310-347.

Karplus, M., and Weaver, D.L. (1994) Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **3**: 650-668.

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.

Kaufman, B.A., Newman, S.M., Hallberg, R.L., Slaughter, C.A., Perlman, P.S., and Butow, R.A. (2000) In organello formaldehyde crosslinking of proteins to mtDNA: identification of bifunctional proteins. *Proc Natl Acad Sci U S A* **97**: 7772-7777.

Kaufman, B.A., Kolesar, J.E., Perlman, P.S., and Butow, R.A. (2003) A function for the mitochondrial chaperonin Hsp60 in the structure and transmission of mitochondrial DNA nucleoids in Saccharomyces cerevisiae. *J Cell Biol* **163**: 457-461.

Kawasaki, Y., Wada, C., and Yura, T. (1990) Roles of Escherichia coli heat shock proteins DnaK, DnaJ and GrpE in mini-F plasmid replication. *Mol Gen Genet* **220**: 277-282.

Kawata, Y., Kawagoe, M., Hongo, K., Miyazaki, T., Higurashi, T., Mizobata, T., and Nagai, J. (1999) Functional communications between the apical and equatorial domains of GroEL through the intermediate domain. *Biochemistry* **38**: 15731-15740.

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.

Kim, P.S., and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **51**: 459-489.

Kim, A., Ghosh, P., Aaron, M.A., Bibb, L.A., Jain, S. and Hatfull, G.F. (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol. Microbiol.* **50**: 463–473.

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

Kong, T.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 U S A* **90**: 2608-2612.

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

Kumar, M., and Chatterji, D. (2008) Cyclic di-GMP: a second messenger required for long-term survival, but not for biofilm formation, in Mycobacterium smegmatis. *Microbiology* **154**: 2942-2955.

Kusukawa, N., and Yura, T. (1988) Heat shock protein GroE of Escherichia coli: key protective roles against thermal stress. *Genes Dev* **2**: 874-882.

Laal, S., Sharma, Y.D., Prasad, H.K., Murtaza, A., Singh, S., Tangri, S., Misra, R.S. and Nath, I. (1991) Recombinant fusion protein identified by lepromatous sera mimics native Mycobacterium leprae in T-cell responses across the leprosy spectrum. *Proc Natl Acad Sci U S A* **88**: 1054-1058.

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.

Lechevalier, H.A., and Lechevalier, M.P. (1967) Biology of actinomycetes. *Annu Rev Microbiol* **21**: 71-100.

Lederer, E., Adam, A., Ciorbaru, R., Petit, J.F., and Wietzerbin, J. (1975) Cell walls of *Mycobacteria* and related organisms; chemistry and immunostimulant properties. *Mol Cell Biochem* **7**: 87-104.

Lee, R.E., Brennan, P.J., and Besra, G.S. (1996) Mycobacterium tuberculosis cell envelope. *Curr Top Microbiol Immunol* **215**: 1-27.

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

Leopold, P.E., Montal, M., and Onuchic, J.N. (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* **89**: 8721-8725.

Leuchtenberger, W., Huthmacher, K., and Drauz, K. (2005) Biotechnological production of amino acids and derivatives: current status and prospects. *Appl Microbiol Biotechnol* **69**: 1-8.

Levinthal, C. (1968) Are there pathways for protein folding? *Journal de Chemie Physique et de Physico-Chemie Biologique* **65**.

Lewis, K. (2007) Persister cells, dormancy and infectious disease. *Nat RevMicrobiol* **5**: 48-56.

Lewthwaite, J.C., Clarkin, C.E., Coates, A.R., Poole, S., Lawrence, R.A., Wheeler-Jones, C.P., Pitsillides, A.A., Singh, M., and Henderson, B. (2007) Highly homologous Mycobacterium tuberculosis chaperonin 60 proteins with differential CD14 dependencies stimulate cytokine production by human monocytes through cooperative activation of p38 and ERK1/2 mitogen-activated protein kinases. *Int Immunopharmacol* **7**: 230-240.

Lin, Z., and Rye, H.S. (2006) GroEL-mediated protein folding: making the impossible, possible. *Crit Rev Biochem Mol Biol* **41**: 211-239.

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

Liu, H., Kovács, E. and Lund, P.A. (2009) Characterisation of mutations in GroES that allow GroEL to function as a single ring. *FEMS Micro. Lett*. **583**: 2365-2371.

Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Luckner, S.R., Machutta, C.A., Tonge, P.J., and Kisker, C. (2009) Crystal structures of Mycobacterium tuberculosis KasA show mode of action within cell wall biosynthesis and its inhibition by thiolactomycin. *Structure* **17**: 1004-1013.

Lund, P.A. (2001) Microbial molecular chaperones. *Adv Microb Physiol* **44**: 93-140.

Lund, P.A., Large, A.T., and Kapatai, G. (2003) The chaperonins: perspectives from the Archaea. *Biochem Soc Trans* **31**: 681-685.

Lund, P.A., and Ellis, R.J. (2008) The chaperone function: meanings and myths. *Novartis Found Symp* **291**: 23-36; discussion 36-44, 137-140.

Lund, P.A. (2009) Multiple chaperonins in bacteria--why so many? *FEMS Microbiol Rev* **33**: 785-800.

Lundrigan, M.D., Arceneaux, J.E.L., Zhu, W., and Byers, B.R. (1997) Enhanced hydrogen peroxide sensitivity and altered stress protein expression in iron-starved *Mycobacterium smegmatis*. *Bio metals* **10**: 215-225.

Macario, A.J., Lange, M., Ahring, B.K., and De Macario, E.C. (1999) Stress genes and proteins in the archaea. *Microbiol Mol Biol Rev* **63**: 923-967, table of contents.

Mah, T.F. and O'Toole, G.A. (2001) Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol* **9**: 34-39.

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.

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.

Martínez, A., Torello, S. and Kolter, R. (1999) Sliding motility in Mycobacteria. *J Bacteriol* **181**: 7331-7338.

Mathew, R., Mukherjee, R., Balachandar, R., and Chatterji, D. (2006) Deletion of the rpoZ gene, encoding the omega subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in Mycobacterium smegmatis. *Microbiology* **152**: 1741-1750.

McLeod, M.P., Warren, R.L., Hsiao, W.W., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A.L., Wang, D., Dosanjh, M., Hara, H., Petrescu, A., Morin, R.D., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Smailus, D., Siddiqui, A.S., Marra, M.A., Jones, S.J., Holt, R., Brinkman, F.S., Miyauchi, K., Fukuda, M., Davies, J.E., Mohn, W.W., and Eltis, L.D. (2006) The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A.* **103**:15582-15587.

McLennan, N., and Masters, M. (1998) GroE is vital for cell-wall synthesis. *Nature* **392**: 139.

Mdluli, K., Slayden, R.A., Zhu, Y., Ramaswamy, S., Pan, X., Mead, D., Crane, D.D., Musser, J.M., and Barry, C.E., 3rd (1998) Inhibition of a Mycobacterium tuberculosis □-ketoacyl ACP synthase by isoniazid. *Science* **280**: 1607-1610.

Meghji, S., White, P.A., Nair, S.P., Reddi, K., Heron, K., Henderson, B., Zaliani, A., Fossati, G., Mascagni, P., Hunt, J.F., Roberts, M.M., and Coates, A.R. (1997) Mycobacterium tuberculosis chaperonin 10 stimulates bone resorption: a potential contributory factor in Pott's disease. *J Exp Med* **186**: 1241-1246.

Milano, A., Forti, F., Sala, C., Riccardi, G. and Ghisotti, D. (2001) Transcriptional Regulation of *furA* and *katG* upon Oxidative Stress in *Mycobacterium smegmatis*. *J Bacteriol* **183**: 6801-6806.

Minnikin, D.E. (1982) In *The Biology of the Mycobacteria*. Vol. 1. Ratledge, C.a.S., J.L. (ed). London: Academic, pp. 95-184.

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

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

Nielsen, K. L. and Cowan, N. J. (1998) A single ring is sufficient for productive chaperonin-mediated folding in vivo. *Molecular Cell* **2**: 93–94

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920-930.

Nolting, B., and Andert, K. (2000) Mechanism of protein folding. *Proteins* **41**: 288-298.

O'Brien, L. M., Gordon, S.V., Roberts, I.S. and Andrew, P.W. (1996) Response of *Mycobacterium smegmatis* to acid stress *FEMS Micro. Lett.* **139**: 11-17.

O'Toole, G., Kaplan, H.B. and Kolter, R. (2000) Biofilm formation as microbial development. *Annu Rev Microbiol* **54**: 49-79.

O'Toole, R., Smeulders, M.J., Blokpoel, M.C., Kay, E.J., Lougheed, K. and Williams, H.D. (2003) A Two-Component Regulator of Universal Stress Protein Expression and Adaptation to Oxygen Starvation in *Mycobacterium smegmatis*. *J Bacteriol* **185**: 1543-1554.

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.

Ojha, A., and Hatfull, G.F. (2007) The role of iron in Mycobacterium smegmatis biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonic growth. *Mol Microbiol* **66**: 468-483.

Ojha, A.K., Baughn, A.D., Sambandan, D., Hsu, T., Trivelli, X., Guerardel, Y., Alahari, A., Kremer, L., Jacobs, W.R., Jr., and Hatfull, G.F. (2008) Growth of Mycobacterium

tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* **69**: 164-174.

Ortalo-Magne, A., Lemassu, A., Laneelle, M.A., Bardou, F., Silve, G., Gounon, P., Marchal, G., and Daffe, M. (1996) Identification of the surface-exposed lipids on the cell envelopes of Mycobacterium tuberculosis and other mycobacterial species. *J Bacteriol* **178**: 456-461.

Papavinasasundaram, K.G., Anderson, C., Brooks, P.C., Thomas, N.A., Movahedzadeh, F., Jenner, P.J., Colston, M.J., and Davis, E.O. (2001) Slow induction of RecA by DNA damage in Mycobacterium tuberculosis. *Microbiology* 147: 3271-3279.

Papavinasasundaram, K.G., Chan, B., Chung, J.H., Colston, M.J., Davis, E.O., and Av-Gay, Y. (2005) Deletion of the Mycobacterium tuberculosis pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**: 5751-5760.

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.

Pavelka, M.S., Jr. (2007) Another brick in the wall. *Trends Microbiol* **15**: 147-149.

Plotkin, S.S., and Onuchic, J.N. (2002) Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys* **35**: 111-167.

Qamra, R., and Mande, S.C. (2004) Crystal structure of the 65-kilodalton 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.

Qamra, R., Mande, S.C., Coates, A.R., and Henderson, B. (2005) The unusual chaperonins of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **85**: 385-394.

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

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., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L., and Saibil, H.R. (2001) ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* **107**: 869-879.

Ranson, N.A., Clare, D.K., Farr, G.W., Houldershaw, D., Horwich, A.L., and Saibil, H.R. (2006) Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nat Struct Mol Biol* **13**: 147-152.

Recht, J., Martinez, A., Torello, S., and Kolter, R. (2000) Genetic analysis of sliding motility in Mycobacterium smegmatis. *J Bacteriol* **182**: 4348-4351.

Reyrat, J.M., and Kahn, D. (2001) Mycobacterium smegmatis: an absurd model for tuberculosis? *Trends Microbiol* **9**: 472-474.

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.

Roberts, R.C, Toochinda, C., Avedissian, M., Baldini, R.L., Gomes, S.L. and Shapiro, L. (1996) Indentification of a *Caulobacter cresentus*operon encoding *hrcA*, involved in negatively regulating heat-inducible transcription, and the chaperone gene *grpE*. *J Bacteriol* **178**: 1829-1841.

Rodriguez-Quinones, F., Maguire, M., Wallington, E.J., Gould, P.S., Yerko, V., Downie, J.A., and Lund, P.A. (2005) Two of the three groEL homologues in Rhizobium leguminosarum are dispensable for normal growth. *Arch Microbiol* **183**: 253-265.

Röse, L., Kaufmann, S.H.E., Daugelat, S. (1994) Involvement of Mycobacterium smegmatis undecaprenyl phosphokinase in biofilm and smegma formation. *Microbes and Infection* **6**: 965-971.

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.

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.

Saito, H., and Uchida, H. (1978) Organisation and expression of the dnaJ and dnaK genes of Escherichia coli K12. *Mol Gen Genet* **164**: 1-8.

Sakakibara, Y. (1988) The dnaK gene of Escherichia coli functions in initiation of chromosome replication. *J Bacteriol* **170**: 972-979.

Sala, C., Forti, F., Magnoni, F., and Ghisotti, D. (2008) The katG mRNA of Mycobacterium tuberculosis and Mycobacterium smegmatis is processed at its 5' end and is stabilised by both a polypurine sequence and translation initiation. *BMC Mol Biol* **9**: 33.

Sambrook.J, R.D.W. (2001) *Molecular Cloning A Laboratory Manual.* New York: Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Schleifer, K.H., and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36**: 407-477.

Schroder, H., Langer, T., Hartl, F.U., and Bukau, B. (1993) DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *Embo J* **12**: 4137-4144.

Scott, D.J.and Schuck, P. (2005) A brief introduction to the analytical ultracentrifugation of proteins for beginners. In *Analytical Ultracentrifugation*. Scott, D.J., Harding, S.E. and Rowe, A.J. (ed). Cambridge, UK: Royal Society of Chemistry, pp. 1-25

Segal, G., and Ron, E.Z. (1995) The groESL operon of Agrobacterium tumefaciens: evidence for heat shock-dependent mRNA cleavage. *J Bacteriol* **177**: 750-757.

Sell, S.M., Eisen, C., Ang, D., Zylicz, M., and Georgopoulos, C. (1990) Isolation and characterization of dnaJ null mutants of Escherichia coli. *J Bacteriol* **172**: 4827-4835.

Shinnick, T.M., Vodkin, M.H., and Williams, J.C. (1988) The Mycobacterium tuberculosis 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the Escherichia coli GroEL protein. *Infect Immun* **56**: 446-451.

Sizaire, V., Nackers, F., Comte, E., and Portaels, F. (2006) Mycobacterium ulcerans infection: control, diagnosis, and treatment. *Lancet Infect Dis* **6**: 288-296.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol* **4**: 1911-1919.

Sridharan, S., Wang, L., Brown, A.K., Dover, L.G., Kremer, L., Besra, G.S., and Sacchettini, J.C. (2007) X-ray crystal structure of Mycobacterium tuberculosis □-ketoacyl acyl carrier protein synthase II (mtKasB). *J Mol Biol* **366**: 469-480.

Stewart, G. R., Snewin, 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, P.S. and Costerton, J.W. (2001) Antibiotic resistance of bacteria in biofilms. *Lancet* **358**: 135-138.

Stewart, G. R., Wernisch, L., Stabler, R., Mangan, J. A., Hinds, J., Laing, K. G., Young, D. B. and Butcher, P. D. (2002). Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. *Microbiol* **148**: 3129–3138.

Stock, J.B., Ninfa, A.J., and Stock, A.M. (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* **53**: 450-490.

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. (1988) Escherichia coli heat shock gene mutants are defective in proteolysis. *Genes Dev* **2**: 1851-1858.

Sun, Z., Scott, D.J. and Lund, P.A. (2003) Isolation and Characterisation of Mutants of GroEL that are Fully Functional as Single Rings. *J Mol Biol* **332**: 715-728.

Svedberg, T and Pedersen, K.O. (1940) *The Ultracentrifuge*, Oxford University Press, London.

Swanson, S., Gokulan, K., and Sacchettini, J.C. (2009) KasA, another brick in the *Mycobacterial* cell wall. *Structure* **17**: 914-915.

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

Tang, Y.C., Chang, H.C., 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.

Tehver, R., and Thirumalai, D. (2008) Kinetic model for the coupling between allosteric transitions in GroEL and substrate protein folding and aggregation. *J Mol Biol* **377**: 1279-1295.

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

Tilly, K., and Yarmolinsky, M. (1989) Participation of Escherichia coli heat shock proteins DnaJ, DnaK, and GrpE in P1 plasmid replication. *J Bacteriol* **171**: 6025-6029.

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.

Todd, M.J., Lorimer, G.H., and Thirumalai, D. (1996) Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc Natl Acad Sci U S A* **93**: 4030-4035.

Tokunaga, M., Miyawaki, H., Shiraishi, Y. and Tokunaga, H. (1997) Purification and characterization of a GroEL homologue from the moderately eubacterial halophile Pseudomonas sp.#43. *FEMS Microbiol. Lett.*, **152**: 321–326.

Tormay, P., Coates, A.R., and Henderson, B. (2005) The intercellular signaling activity of the Mycobacterium tuberculosis chaperonin 60.1 protein resides in the equatorial domain. *J Biol Chem* **280**: 14272-14277.

Torok, Z., Horvath, I., Goloubinoff, P., Kovacs, E., Glatz, A., Balogh, G., and Vigh, L. (1997) Evidence for a lipochaperonin: association of active protein-folding GroESL oligomers with lipids can stabilize membranes under heat shock conditions. *Proc Natl Acad Sci U S A* **94**: 2192-2197.

Tortoli, E. (2006) The new *Mycobacteria*: an update. *FEMS Immunol Med Microbiol* **48**: 159-178.

Tosukhowong, A., Nakayama, J., Mizunoe, Y., Sugimoto, S., Fukuda, D. and Sonomoto, K. (2005) Reconstitution and Function of Tetragenococcus halophila Chaperonin 60 Tetradecamer. *J. Biosci. Bioeng.* **99**: 30-37.

Tran, S.L., Rao, M., Simmers, C., Gebhard, S., Olsson, K. and Cook, G.M. (2005) Mutants of Mycobacterium smegmatis unable to grow at acidic pH in the presence of the protonophore carbonyl cyanide m-chlorophenylhydrazone. *Microbiol* **151**: 665-672.

Unniraman, S., Chatterji, M., and Nagaraja, V. (2002) A hairpin near the 5' end stabilises the DNA gyrase mRNA in Mycobacterium smegmatis. *Nucleic Acids Res* **30**: 5376-5381.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and van Sinderen, D. (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**: 495-548.

Viitanen, P. V., Lorimer, G., Bergmeier, W., Weiss, C., Kessel, M., and Goloubinoff, P. (1998) Purification of mammalian mitochondrial chaperonin 60 through in vitro reconstitution of active oligomers. *Methods Enzymol.,* **290**: 203–217.

Wada, M., and Itikawa, H. (1984) Participation of Escherichia coli K-12 groE gene products in the synthesis of cellular DNA and RNA. *J Bacteriol* **157**: 694-696.

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G., and Pieters, J. (2004) Protein kinase G from pathogenic *Mycobacteria* promotes survival within macrophages. *Science* **304**: 1800-1804.

Wallington, E.J., and Lund, P.A. (1994) Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes. *Microbiology* **140 (Pt 1)**: 113-122.

Walter, S. (2002) Structure and function of the GroE chaperone. *Cell Mol Life Sci* **59**: 1589-1597.

Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* **70**: 697-701.

Weissman, J.S., Kashi, Y., Fenton, W.A., and Horwich, A.L. (1994) GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78**: 693-702.

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**: 577–588.

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.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995) Navigating the folding routes. *Science* **267**: 1619-1620.

World Health Organisation: WHO Report 2009: Global Tuberculosis Control—Epidemiology, Strategy, Financing Geneva: WHO; 2009.

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.

Ybarra, J. and Horowitz, P.M. (1995) Inactive GroEL Monomers Can Be Isolated and Reassembled to Functional Tetradecamers That Contain Few Bound Peptides. *J Biol Chem* **270**: 22962-22967.

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.

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

Zhang, C.C. (1996) Bacterial signalling involving eukaryotic-type protein kinases. *Mol Microbiol* **20**: 9-15.

Zhang, Y., Post-Martens, K., and Denkin, S. (2006) New drug candidates and therapeutic targets for tuberculosis therapy. *Drug Discov Today* **11**: 21-27.

Zhou, Y.N., Kusukawa, N., Erickson, J.W., Gross, C.A., and Yura, T. (1988) Isolation and characterization of Escherichia coli mutants that lack the heat shock sigma factor sigma 32. *J Bacteriol* **170**: 3640-3649.

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