FUNCTIONAL AND STRUCTURAL CHARACTERISATION OF MYCOBACTERIAL CHAPERONINS

By

MOHAMMAD TABISH AHMED

A thesis submitted to the

University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

The University of Birmingham

September 2010

Abstract

Proteins are involved in almost every biochemical process within every living organism. Amino acids form the building blocks required to make proteins, however, on their own they only constitute the primary structure. For the protein to be biologically active, the linear amino acid polypeptide structure needs to be folded into its tertiary structure, or native state. Molecular chaperones are proteins that assist in the folding or unfolding of other proteins. Although essential for cell survival, these chaperones do not form a part of the final structure of the secondary proteins and are not involved in any of its biological functions. They play an important role in preventing unfavourable hydrophobic interactions between unfolded proteins in the crowded intracellular environment. A sub class of chaperones, referred to as chaperonins, are 60kDa monomers that usually assemble into large tetradecameric double ringed structures. They work by encapsulating misfolded or unfolded protein into their cage like central cavity and allow it to fold without influence from the crowded intracellular environment.

Most bacteria only have one GroEL (*E. coli* chaperonin) homologue; however some bacteria have been shown to have 2 or more copies of the *groEL* gene, including most *Mycobacteria*. In *Mycobacterium tuberculosis* there are 2 distinct *groEL* homologues which encode the chaperonins Cpn60.1 and Cpn60.2, while *M. smegmatis* has 3 copies of *cpn60*. Phylogenetic analysis of the Mycobacterial chaperonins has shown that the genes for the duplicated chaperonins diverged a long time ago, such that one of the *groEL* homologues in *M. tuberculosis* is more closely related to one in *M. smegmatis* than it is to the other one in *M. tuberculosis*. This suggested that the two chaperonins in *M. tuberculosis* have evolved to

fulfil different functions within the cell. In both cases Cpn60.2 is predicted to be the main house-keeping chaperonin. Interestingly, while most chaperonins occur as stable complexes with a characteristic double ring structure of 14 subunits, the Mycobacterial chaperonins are very unstable and appear to form much smaller complexes. Given that the large structures formed by most chaperonins are considered essential to their mechanism of action, it is unclear why the oligomers are so much less stable in Mycobacteria.

In this study I present detailed functional and oligomeric analysis of the *M. tuberculosis* chaperonins. Using various biological techniques, including complementation assays, site directed mutagenesis, and domain swap experiments; I have been able to demonstrate that both chaperonins have evolved to fulfil alternate functions within the cell. I have shown that while Cpn60.2 is fully functional in *E. coli*, Cpn60.1 is not. However, *M. tuberculosis* Cpn60.1 was able to function in *M. smegmatis* and complement in an assay of biofilm formation for the loss of its endogenous chaperonin. Domain swap experiments between Cpn60.2 and *E. coli* GroEL have provided further evidence to support the hypothesis that they function in a similar manner. Using analytical ultracentrifugation on purified proteins, I have been able to show that Cpn60.2 and its functional chimera do form larger oligomeric complexes *in vitro* under certain conditions and in the presence of ATP.

Dedicated to Mum, Dad, and Sharique

Thank you for everything!

Acknowledgement

First and foremost, I would like to take the privilege to express my deep sense of gratitude to my supervisor, Dr. Peter Lund. Without his limitless patience, help, encouragement, and guidance up to the very last minute, this PhD would not have been possible. I would also like to thank every member of the T101 lab who have come and gone during the course of my PhD, for they made this experience so much more enjoyable. I would like to thank Han, Martin, Ash, Riddhi, Elsa, MingQi, and everyone from Ian Henderson's end of the lab for keeping the lab lively and welcoming. Special mentions go to Tara, Andy, Preethy, Matt and Amanda for their constant support (scientific and otherwise), friendship, and for all the extremely entertaining conversations over the last few years.

A lot of thanks go to all the other postgraduate students who were a part of this short adventure and all the undergraduate students who treated me like one of their own, rather than as a demonstrator. It just wouldn't be possible to list them all. I would also like to thank all the members of the BGSC for all the effort they put along with me to organise all the events within biosciences. This PhD just would not have been the same without them all.

Last but not least, my greatest appreciation goes to my family; to my mother, father, and brother for their support and love; to my aunt, for always treating me like her own son; and to all my cousins, nieces, and nephews for all the support and laughter they have provided over the years.

Index of contents

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3: Expression and complementation analysis of *M. tuberculosis* chaperonins in *E. coli*

Chapter 4: Construction and analysis of chimeric chaperonins

List of illustrations

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3: Expression and complementation analysis of *M. tuberculosis* chaperonins in *E. coli*

Chapter 4: Construction and analysis of chimeric chaperonins

Chapter 5: Analysis of a chaperonin chimera

Chapter 6: Further characterisation of Mycobacterial chaperonins

Section 6.1: Testing the ability of *M. tuberculosis* Cpn60.1 to function in *M. smegmatis*

Section 6.2: Testing the existence of double ring *M. tuberculosis* Cpn60.2 by attempting to make single ring mutants

Section 6.3: Testing the hypothesis that phosphorylation is needed for improved oligomerisation

List of tables

Chapter 1: Introduction

Chapter 3: Expression and complementation analysis of *M. tuberculosis* chaperonins in *E. coli*

Chapter 4: Construction and analysis of chimeric chaperonins

Chapter 6: Further characterisation of Mycobacterial chaperonins

List of abbreviations

AAA	ATPases Associated with diverse cellular Activities
APS	Ammonium persulphate
AUC	Analytical ultra centrifugation
BSA	Bovine serum albumin
CCT	Chaperonin-containing TCP-1
CFU	Colony forming units
Cpn	Chaperonin
DEAE	Diethylaminoethyl
DMSO	Dimethylsulphoxide
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
FAS II	Fatty Acid Synthase II
Hsp	Heat Shock Protein
IPTG	Isopropyl-β-D-thiogalactoside
LB	Luria-Bertani
MOPS	4-morpholinepropanesulfonic acid
MS	Mass spectrometry
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
RBS	Ribosome binding site
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate

SDW	Sterile distilled water
STPKs	Ser/Thr protein kinases
TAE	Tris-acetate EDTA
TEMED	N.N.N', N' – tetramethylethylenediamine
TBS	Tris buffer saline
TCP-1	T-complex polypeptide- 1
Tris	2-amino-2-(hydroxymethyl)-1, 3-propanediol
WT	Wild type

Chapter 1

Introduction on the chaperonins of Mycobacteria

1.1 Protein Folding

The genome sequence of an organism is the basic information required by an organism to live. This code is read, translated and acted upon by the cell to maintain itself. Proteins represent the final stages of information transfer from an organism's genome sequence. It takes just 20 amino acids to form all the known and unknown proteins. A single error in the mechanism of protein folding in an organism can have an adverse affect on its survival capabilities. This, along with the importance of proteins in the medical and biotechnological fields makes the study of protein folding vital, as the final structure of the protein determines its activity. Even though the mechanism of protein folding is a very active area of research, exactly how the linear amino acid sequences are folded into fully functional proteins is yet to be conclusively answered.

1.1.1 A brief historical perspective of protein folding

Proteins were first recognised in the 18^{th} century, but it wasn't till the 19^{th} century that the term protein came into use. Even then it took almost another century before scientists started to study the processes involved in protein folding. In 1959, W. Kauzmann discovered that protein folding is mediated by hydrophobic interactions (Kauzmann, 1959). This then led to the idea that proteins fold in a trial and error manner. However, in 1968 Levinthal suggested that if proteins were folding in a 'trial and error' method, then it would take longer than the age of the universe for a protein to reach its native state (Levinthal, 1968) . A thermodynamic hypothesis was then put forward which suggested that a protein would reach its native state when its Gibbs free energy was at its lowest. Anfinsen demonstrated that the free energy of a protein was determined by its inter-atomic interactions, which were in turn determined by the

specific amino acids of the protein, and thus the native state of the protein was already encoded in its primary sequence (Anfinsen, 1973)

1.1.2 The energy landscape theory

During the process of understanding the mechanisms involved in protein folding, various models were put forward. One such model which emerged from various statistical and experimental studies, of the intermediate steps involved in protein folding, was the energy landscape theory (Onuchic *et al.*, 2000, Wolynes *et al.*, 1995). According to this view the energy landscape of a folding protein can be represented by a rugged funnel that represents all the energy states available to the protein as it folds (Wolynes *et al.*, 1995). The width of the funnel represents the entropy while the depth represents the free energy difference between folded and unfolded states. The unfolded state of a protein with the highest free energy then folds by one of many routes down the energy funnel to its native state which is associated with lower free energy and as such, higher stability (Onuchic *et al*., 2000) (fig.1.1). The rugged nature of the landscape is present due to the various intermediate forms a protein can take during its folding process.

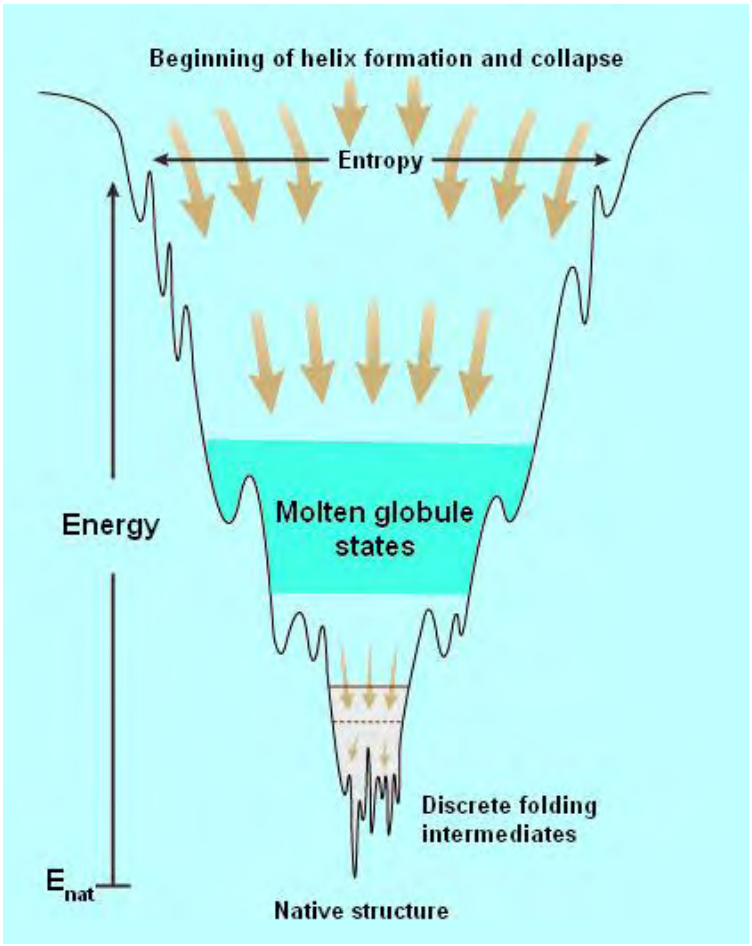


Fig 1.1: The energy landscape of protein folding. This represents the energy states that proteins must go through to reach their native states (from high free energy to low free energy). Image adapted from Wolynes at al., 1995.

1.2 Protein aggregation

The tertiary structure of a protein is one of the more stable states for it to be in. However, when proteins are synthesised within a cell, they are in their primary state with a lot of exposed hydrophobic residues. These hydrophobic residues can interact with each other and form aggregates to shield the residues from polar compounds like water, which is the main component of the cytoplasm. The internal nature of a cell is a lot more complex and crowded than *in vitro* (Ellis, 1997, Ellis, 2007). This crowded nature of the cell with macromolecules, such as other proteins, can promote undesired misfolding to occur. This mainly occurs because the possibility of unfolded proteins interacting with each other to form aggregates is increased (Young *et al.*, 2004, Ellis, 1997). Ultimately, these aggregates result in the loss of function of the proteins and can eventually lead to cell death (Ellis, 2007).

Gene product translation is also an aggregation prone period for the newly translated proteins. The exit channel of a ribosome has been shown to be quite long, at 100 Å (Nissen *et al.*, 2000). However it is also very narrow, at just 15 Å (Nissen *et al*., 2000, Ferbitz *et al.*, 2004). This allows basic helical formation within the ribosome, but does not allow the complete folding of a polypeptide chain into its native state. Also, to attain the native state the entire polypeptide sequence needs to have emerged from the ribosome (Hartl & Hayer-Hartl, 2002). As a result, the nascent polypeptide chain can end up being exposed for a fair length of time, as it can take up to a minute for synthesis to occur for larger chains. These exposed regions are prone to contact with other hydrophobic chains and can form aggregates.

To try and escape from the aggregating tendencies of the proteins, the cells make use of various systems for protein quality control. This quality control can occur in one of two ways; refolding, or degradation. This thesis will primarily be focusing on the bacterial protein quality control systems, but examples of eukaryotic systems will also be given where relevant.

1.3 Protein degradation

A very important aspect of the protein quality control process is protein degradation. In 1972, Alfred Goldberg first presented the evidence to show that selective degradation of proteins occurred in *E. coli* (Goldberg, 1972). Over the last few years our understanding of the importance and mechanisms of protein degradation has increased. It was the study of the ATP-dependent proteases that led to the discovery of one of the major chaperone families, Hsp100 (Schirmer *et al.*, 1996). This family of chaperones have been shown to interact with unfolded or misfolded proteins to direct them towards degradation (ref section 1.4.3.6). Degradation of proteins within a cell is an important way of regulating many important cellular processes. Another important aspect of proteolysis is to degrade any non functional proteins within the cell. This prevents potentially lethal protein aggregation from occurring. Proteolysis has also been shown to increase in response to stresses such as heat shock (Morimoto *et al.*, 1997).

Cellular protein degradation is predominantly via energy dependent proteases (Gottesman, 1996), which include the Lon protease, the ClpAP protease, and the ClpXP protease, and the 26S proteasome (the major cytosolic protease in eukaryotes) to name a few. One of the main

questions that were originally put forward was how these proteases recognise appropriate substrates to degrade. It has been shown that there are various methods by which this is achieved by different proteases. The Clp (Hsp100) proteases have been shown to both interact directly with their substrates and to also make use of adaptor proteins which bind to the substrate first and then bind to the protease to present them with the substrate (Striebel *et al.*, 2009b, Gottesman, 1996). The 26S proteasome on the other hand only degrades proteins that have been modified by ubiquitin ligases and other ubiquitin conjugating enzymes to include polyubiquitin chains (Pickart, 1997). This allows the proteases to interact with a broad range of unfolded or misfolded proteins to degrade (Inobe & Matouschek, 2008). An illustration showing some of the routes taken by unfolded or misfolded proteins to reach the proteases can be seen in figure 1.2.

The structural and functional properties of these proteases share a lot of similarities with molecular chaperones. Using ClpA and ClpX as examples, both have been shown to have chaperone activities (Singh *et al.*, 2000, Weber-Ban *et al.*, 1999). Structurally the ClpAP and ClpXP proteases consist of two heptameric rings of ClpP flanked on one or both side with hexameric rings of ClpA or ClpX creating a barrel shaped structure (Beuron *et al.*, 1998). This barrel structure contains a series of chambers connected via axial channels (figure 1.3). Functionally, the degradation site of substrates is present within the ClpP chamber. So the substrate is first recognised by either ClpA or ClpX, then it is unfolded by them in an ATP dependent manner. The resulting unfolded substrate is then translocated to the ClpP chamber where it is degraded. This mechanism is generally conserved in all three domains of life (figure 1.4). Further information on the Hsp100 family of chaperones can be found in section 1.4.3.6.

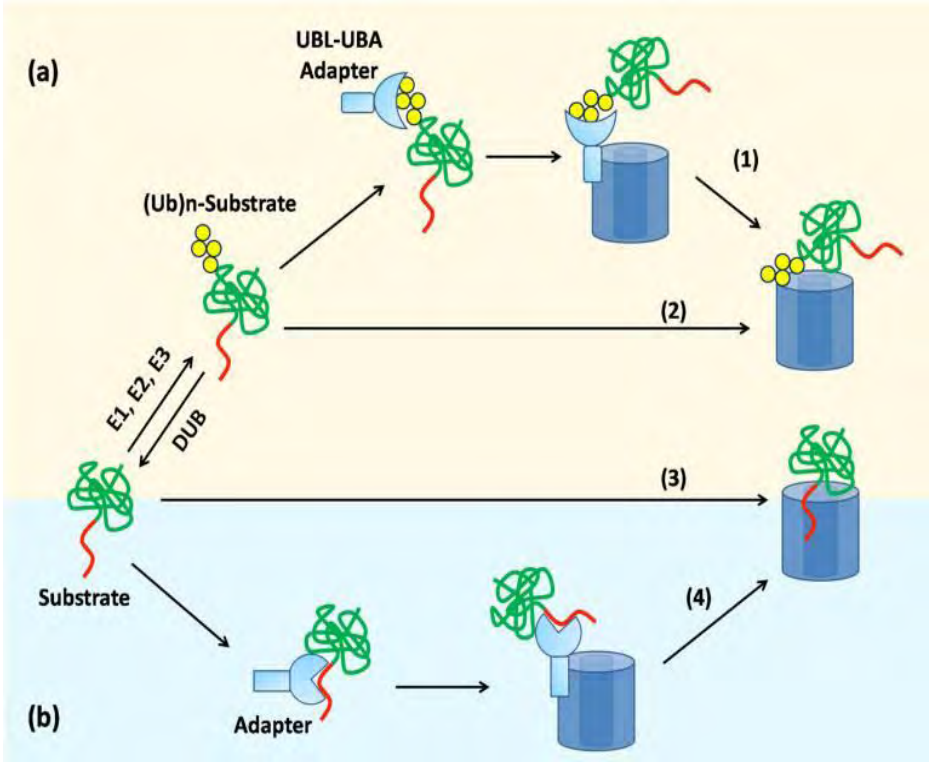


Fig 1.2: Pathways regulating the transfer of substrate to proteases in (a) eukaryotes and (b) bacteria. (1) Proteasome binding to an adaptor protein that is bound to a substrate protein. (2) Proteasome recognising and binding to the polyubiquitin chain of the modified substrate. (3) Proteasome binding to exposed polypeptide sequence of the substrate. (4) Bacterial proteases recognise substrate via adaptor proteins. Image taken from Inobe & Matouschek, 2008.

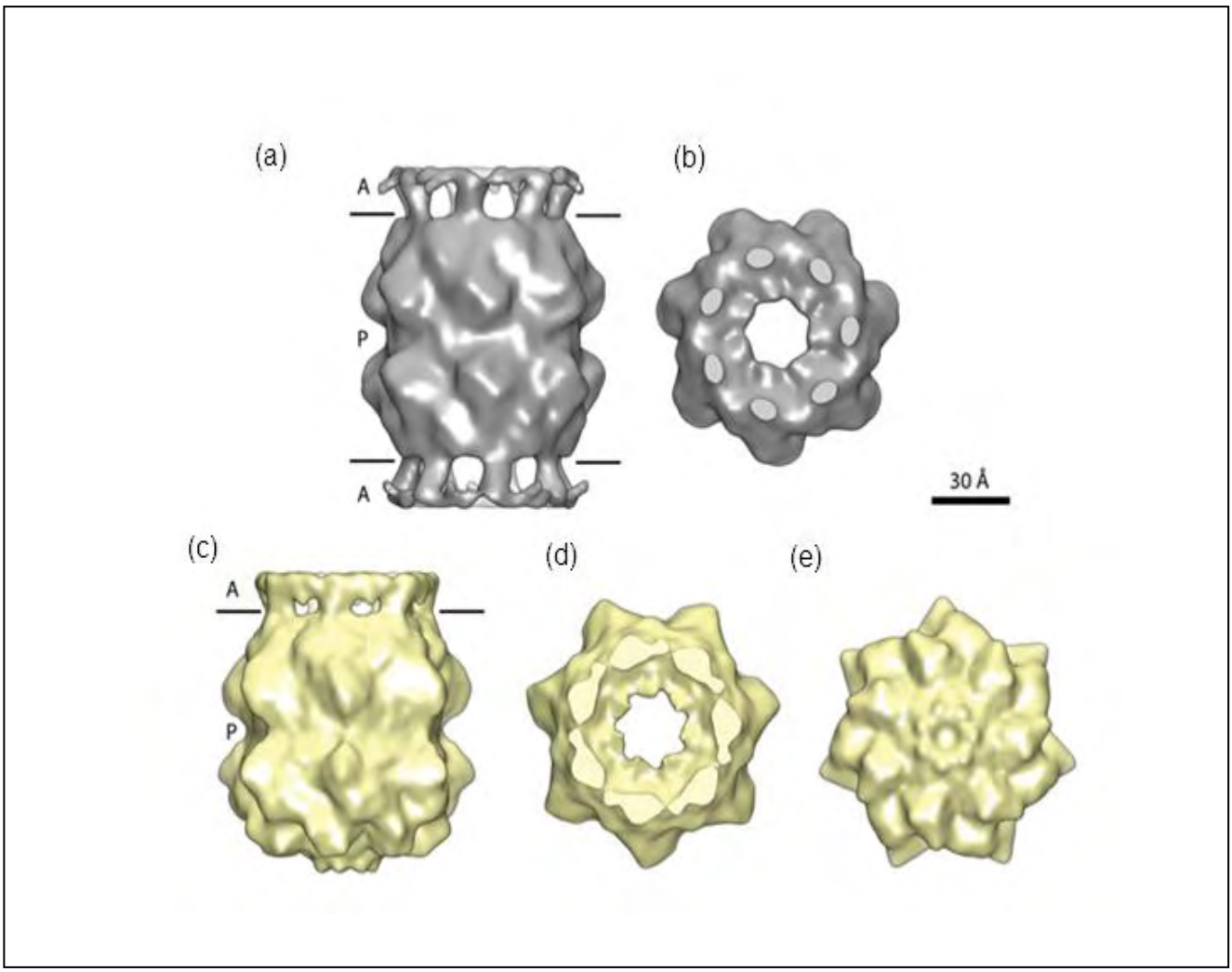


Fig 1.3: Cryo-EM reconstructions of ClpP attached to 1 or 2 ClpA. Two side views, of a 1:1 and a 2:1 complex are shown. Reconstructed 2:1 complex in side view (a), and axial view of 2:1 complex (b). Reconstructed 1:1 complex side view (c), and both axial views (d and e). Image taken from Effantin *et al.*, 2010.

As protein folding in *Mycobacterium tuberculosis* is the subject of this thesis, it is interesting to note that in *M. tuberculosis* and other actinomycetes, horizontal gene transfer has resulted in the acquisition of the 20S proteasome. Structurally it is composed of two rings of catalytic β subunits sandwiched by rings of α subunits making a barrel like structure (Hu *et al.*, 2006). The eukaryotic 26S proteasome consists of a 20S core flanked on one or both sides with a 19S ATPase cap. This cap is responsible for the recognition, unfolding and translocation of the ubquitinated substrate into the 20S chamber for degradation. In actinomycetes however, this function is performed by homologous ATPases like Mpa (*Mycobacterium* proteasome ATPase) or ARC (AAA ATPase forming Ring-shaped Complex) (Pearce *et al.*, 2008, Wolf *et al.*, 1998). While the eukaryotic protease system uses ubiquitin to recognise appropriate substrates for degradation, in *M. tuberculosis* it has been shown that Mpa recognises substrates that have small ubiquitin like proteins attached called Pup (prokaryotic ubiquitin-like protein) (Pearce *et al.*, 2008, Striebel *et al.*, 2009a, Burns *et al.*, 2009). An interesting point to note here is that although a large number of proteins (58 currently identified targets, including Cpn60.2 and Cpn10) in *M. tuberculosis* have pupylation sites, not all of them become proteasome substrates (Festa *et al.*, 2010). More in depth information about the Cpn60 family of chaperones is available in section 1.5. A cartoon representation of the protease degradation process can be seen in figure 1.4 below.

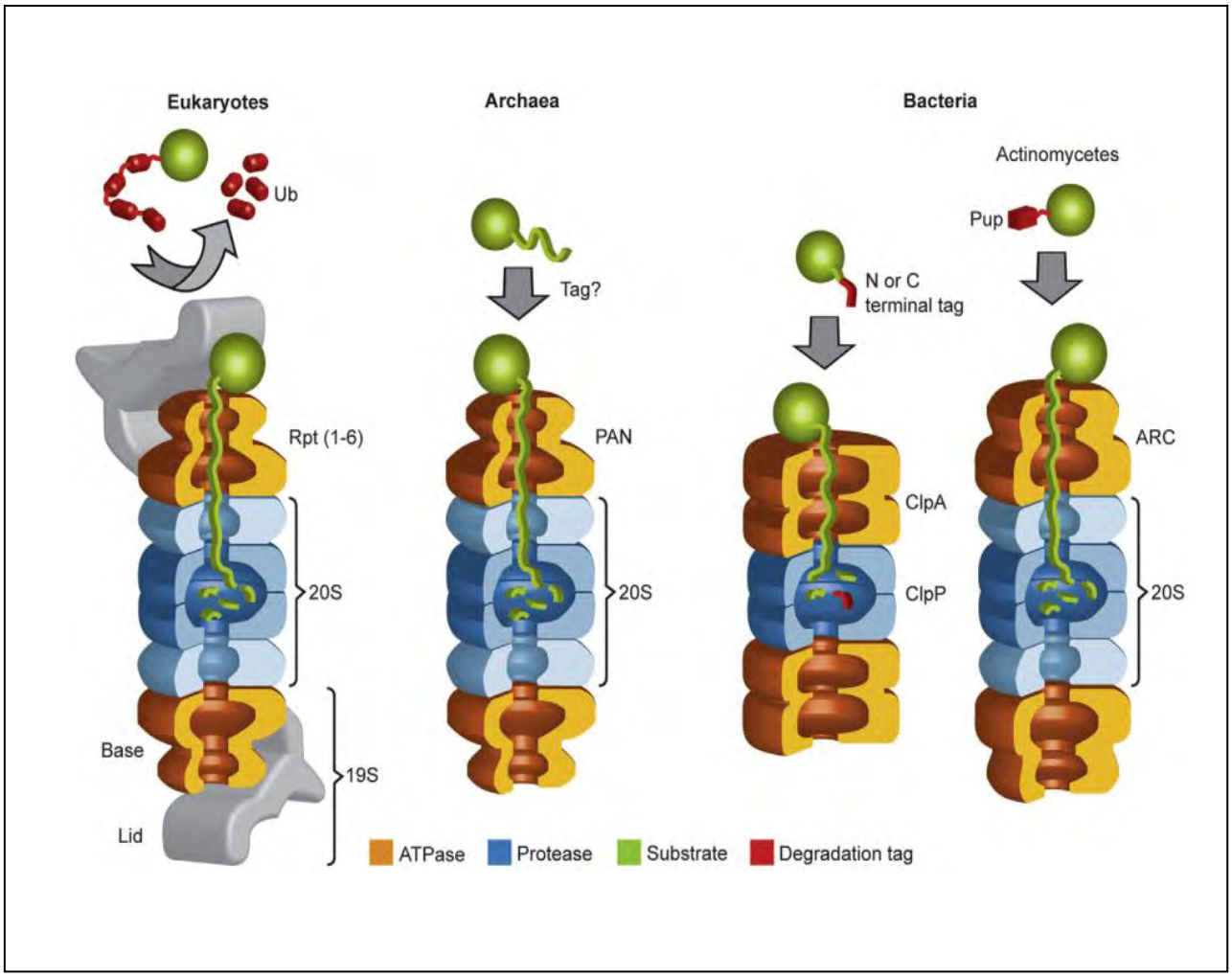


Fig 1.4: Representation of the mechanism by which protein degradation occurs in the 3 domains of life (Striebel *et al.*, 2009b). Protein quality control in cells makes use of both chaperones to fold/unfold proteins and proteases to degrade targeted unfolded proteins.

1.4 The molecular chaperones and cochaperones

Molecular chaperones are proteins that assist in the folding or unfolding of other proteins (fig 1.5) (Ellis, 1987). Although essential for cell survival, these chaperones do not form a part of the final structure of the secondary proteins and are not involved in any of its biological functions (Ellis & Hemmingsen, 1989, Ellis, 2005, Ellis, 2007, Hartl & Hayer-Hartl, 2002).

Molecular chaperones are able to recognise unfolded or partially folded proteins by a variety of mechanisms (Summers *et al.*, 2009, Li *et al.*, 2009, Gur & Sauer, 2008). Some of these chaperones use an ATP dependant mechanism to assist in the proper folding of the nascent chains, while others prevent aggregation by shielding the hydrophobic surfaces of the polypeptide from interacting with other proteins. It has also been shown that some forms of chaperones, e.g. Hsp100, have the ability to unfold proteins that have been misfolded (Ben-Zvi & Goloubinoff, 2001). In some cases the chaperones make use of another set of proteins referred to as the cochaperones to work cooperatively in the folding process. More detailed analysis of these cochaperones will be discussed in the context of their cognate chaperone.

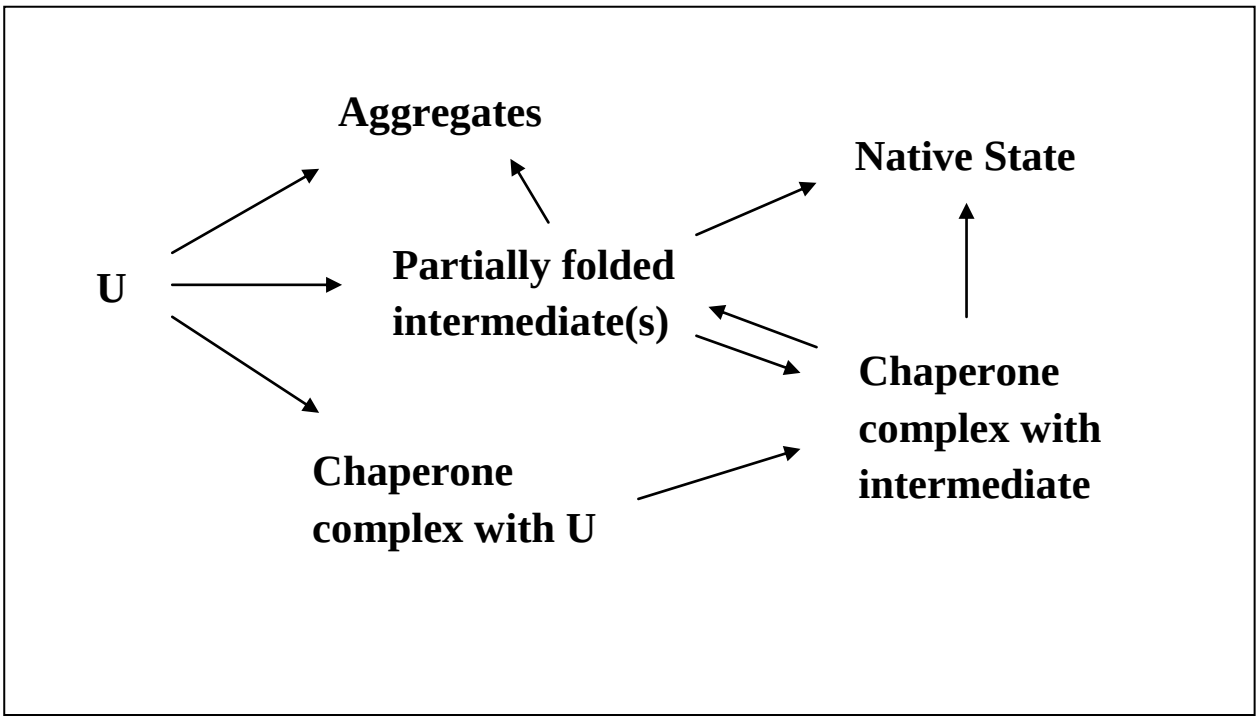


Fig 1.5: A general outline of the possible routes taken by nascent polypeptides (U). Unfolded proteins are prone to aggregation and as a result they, sometimes, require the assistance of molecular chaperones to reach their native state. Image adapted from Fink, 1999.

1.4.1 Classes of molecular chaperones

A range of different chaperones that function either independently or cooperatively to assist in the process of protein folding are present in the cytosol of prokaryotes and eukaryotes (table 1). The chaperones can be distinguished in various ways. One such way is based on their property to either be ATP dependent (e.g. Hsp60) or ATP independent (e.g. Trigger factor). The other method of distinguishing them depends on whether they just hold the proteins and prevent them from aggregating or if they actively help fold them *in vivo*, i.e. the 'holder' or 'folder' chaperones (Ying *et al.*, 2005). The most useful classification used to consider relationships between chaperones in different organisms is based on sequence similarity. Based on this classification the chaperones are grouped into families as shown in table 1.1 below.

1.4.2 Heat shock proteins (Hsp) and their regulation

Many molecular chaperones are also heat shock proteins (Gething & Sambrook, 1992). Subjecting the cells to an increase in temperatures just a few degrees above normal growth conditions causes the expression levels of the heat shock proteins to rise (Kregel, 2002). This increase in expression of HSPs is required by the cell to cope with the stress factors. When subjected to excess heat, the proteins within the cells start to denature and the HSPs then help fold or re-fold these proteins, some of which are required by the cell to increase its chances of survival. Other known factors that can induce heat shock proteins in cells include nalidixic acid, cadmium chloride, 4% ethanol, and puromycin (VanBogelen *et al.*, 1987). In E. coli it has also been shown that this response can also be attained by simply preventing the export of some periplasmic proteins (Ito *et al.*, 1986).

Table 1.1: Classes of molecular chaperone families. The examples shown are but a few of the known chaperones and are by no means the only example. Information obtained from Pete Lund, personal communication.

	Bacteria	**Mitochondria (m) / chloroplasts (c)**	**Archaea**	**Eukaryotic cytosol**	**Endoplasmic reticulum**	**Function**
Hsp100	ClpA ClpB ClpX	HSP78 (m) ClpC (c) ClpB (c)	Absent	HSP104; absent from cytosol of animals.	Absent	Disassembly of oligomers and aggregates.
Hsp90	HtpG	TRAP1 (m) HSP90 (c)	Absent	HSP90 HSP82 HSP83	GRP94	In eukaryotes: regulation of assembly of steroid receptors and signal transduction proteins; protein folding and degradation. In prokaryotes the function of HtpG is less clear, although there is evidence that it facilitates protein folding in stressed *E. coli.*
Hsp70	DnaK Hsc66	Grp75 (m)	HSP70 (absent from many archaea)	HSP70, HSC70, SSB, SSA	BiP	Prevention of aggregation of unfolded/newly translated/newly imported proteins; many other functions such as clathrin uncoating, regulation of heat shock response.
Chaperonin/ Hsp60/Cpn60	GroEL (also referred to as Cpn60)	HSP60 (m) Rubisco sub-unit binding protein (c)	Thermosome, TF55, CCT	CCT (also called TRiC)	Absent	Folding of newly translated and newly imported proteins. A vital part of the heat shock response.

Hsp40	DnaJ CbpA DjlA	Ydj1 (m) Cdj1 (c)	HSP40 (absent from many archaea)	HSP40 Sis1 Zuo1	Sec63	Binding of unfolded proteins, stimulation of ATPase activity of HSP70 (HSP70 cochaperone). Very heterologous group as many proteins possess "DnaJ-domain".
sHsps (small heat shock proteins)	IbpA IbpB	HSP22 (m) HSP21(c)	HSP16.5	HSP26 Alpha-crystallin	Absent	Stabilisation of unfolded proteins, prevention of protein aggregation.
Cochaperonin	GroES	HSP10	Absent	Absent	Absent	Essential for function of type I chaperonins.
Prefoldin	Absent	Absent	Prefoldin GimC	Prefoldin GimC	Absent	Prevents protein aggregation; functions with type II chaperonins but not essential.
Trigger factor	Tig	Absent (m); Present (c)	Absent	Absent	Absent	Assists folding of nascent ribosome-bound polypeptide.
Nascent polypeptide-associated complex	Absent	Absent	MTH177	NAC	Absent	Interacts with and probably assists folding of nascent ribosome-bound polypeptide.
ER chaperones	Absent	Absent	Absent	Absent	Calnexin Calreticulin ERp53 HSP47	Folding of ER glycosylated proteins. HSP47 is specific for collagen chains.

This can be achieved by using a null mutation of *secB*, as this prevents it from participating in the export of a subset of proteins into the periplasm or outer membrane results in the cell seeing these proteins as abnormal, which generates the signal for heat shock protein induction (Wild *et al.*, 1993). Under normal conditions heat shock proteins can make up to 5% of the total intracellular protein mass within a cell, rising up to 15% when the cell is put under stress (Qamra *et al.*, 2005).

In *E. coli* the increase in Hsps is regulated by the σ^{32} factor of RNA polymerase (Landick *et al.*, 1984, Cowing *et al.*, 1985, Diaz-Acosta *et al.*, 2006). The σ^{32} subunit of RNA polymerase is encoded by the *rpoH* gene. Under non-heat shock conditions FtsH mediated σ^{32} degradation ensures that the levels of σ^{32} within the cells do not rise high enough to cause a heat shock response (Herman *et al.*, 1995). Under heat shock conditions the levels of σ^{32} start to rise as the heat destabilizes a RNA structural element that overlaps the *rpoH* mRNA translation start point (Morita *et al.*, 1999). This causes the levels of gene expressions of the recognised promoters to increase as well (Diaz-Acosta *et al.*, 2006). In a lot of cases these recognised promoters are linked to the heat shock proteins. The *rpoH* gene was first discovered by Tetsuo Yamamori and Takashi Yura in 1982 and was referred to as the *hin* (heat shock induction) gene (Yamamori & Yura, 1982). They showed that strains of *E. coli* that had a mutation in the *rpoH* gene were unable to induce expressions of the GroE protein while the wild type cells did. It was later shown that a null mutation in the *rpoH* gene made the strain so sensitive to heat, that it would fail to grow over 20°C (Zhou *et al.*, 1988). This clearly showed the important role played by the heat shock proteins of *E. coli* not only in protecting the cell from increasing temperatures but also during normal growth at 37°C. Interestingly, psudo-revertants Δ*rpoH* cells that could grow at higher temperatures were also

obtained. These revertant cells were shown to be carrying a suppressor mutation, *suhA*, which would suppress all the nonsense and missense *rpoH* mutations. This mutation affected various factors within the cell, including the increase in the levels of σ^{32}, resulting in higher levels of heat shock mRNAs (Tobe *et al.*, 1987).

1.4.3 The chaperone pathways

A general model of some of the chaperone assisted protein folding mechanisms can be seen in figure 1.6. It can be seen that not all forms of protein require chaperones to fold, even if they may require them to stop aggregation. Most small proteins can rapidly fold without any assistance as soon as the entire sequence has been released by the ribosome. Chaperones such as the TF and Hsp70 are known to work by holding the nascent polypeptide chain in a state which allows them to fold once they have been released from the ribosome. Other chaperones such as the Hsp60 family provide chambers for the nascent chain to fold in and be kept away from cytosolic factors such as other folding proteins. It has also been shown that chaperones such as the TF bind directly to the ribosome exit site and interact with the newly formed polypeptide chains (Hoffmann *et al.*, 2006). Depending on the size of the polypeptide, these chaperones either allow folding to occur spontaneously, e.g. smaller chains, or they are then transferred to the next family of chaperones such as the Hsp70 family for further folding (Deuerling *et al.*, 1999, Thulasiraman *et al.*, 1999). There are however some polypeptide chains that are slow-folding and aggregation prone. These are folded within the cell with the help of the chaperonins (Hsp60 family) (Houry *et al.*, 1999, Saibil, 2008). Irrespective of the routes taken the different classes of chaperones are conserved in all domains of life. The importance of the molecular chaperones in cell survival makes it essential to understand their structural, biological, and physiological functions.

For this review, the practical and theoretical information of the Hsp60 family of chaperones will be discussed in detail after some of the other chaperone families, as it is the Hsp60 family of chaperones that is being studied in this project.

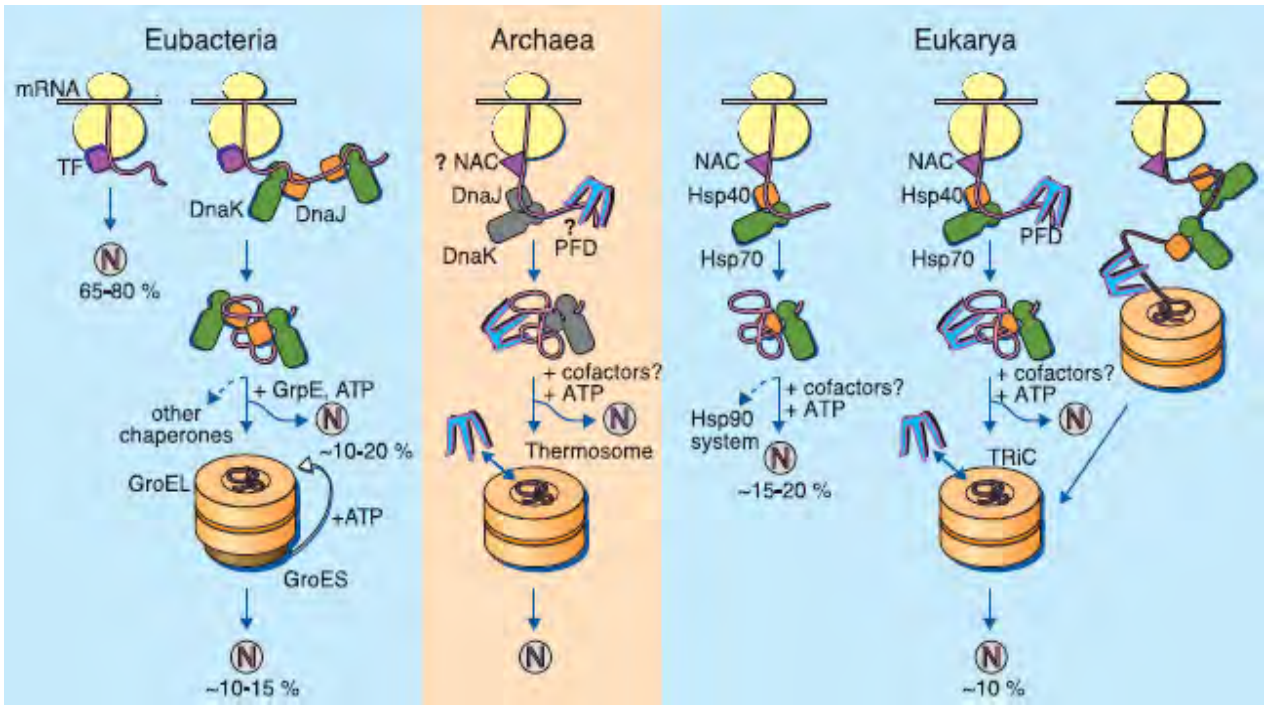


Fig 1.6: A general model for chaperone assisted protein folding within a cell. This figure shows just some of the routes a newly translated polypeptide chain can take to reach its native state. Image taken from Hartl & Hayer-Hartl, 2002.

1.4.3.1 The ribosome associated factor (trigger factor)

The trigger factor is a 48kDa protein encoded by the *tig* gene, which is highly conserved in the eubacteria. In 1998 it was shown to bind with the large subunit of the ribosome (Lill *et al.*, 1988). It is now known to bind to the L23 and L29 proteins near the exit channel of the ribosome where it interacts with nascent polypeptides (Blaha *et al.*, 2003, Kramer *et al.*, 2002, Lakshmipathy *et al.*, 2007). It has also been shown to be associated with up to 90% of all ribosomes in the cytosol (Patzelt *et al.*, 2002). The structure of the trigger factor can be seen in figure 1.7.

The trigger factor functions in an ATP independent manner and recognises nascent chains using hydrophobic interactions. It forms complexes with chains as small as 57 residues in length (Hesterkamp *et al.*, 1996). As has already been discussed, the next step in the process involves either allowing the protein to fold (if it is small enough to fold without further support) or if it is unable to do so, then it is passed on to the next set of chaperones like the HSP70 (DnaK) chaperone family (section 1.4.3.2). It has been shown that both TF and DnaK cooperate in chaperoning unfolded proteins. In *E. coli* this was achieved with the help of Δ*tig* mutants which resulted in DnaK transiently binding to 40% of nascent polypeptides instead of the usual 15% (Genevaux *et al.*, 2004, Deuerling *et al.*, 1999, Teter *et al.*, 1999). It is important to note that this deletion showed no defects in growth or protein folding in the mutant cells, however, a *tig dnaK* double mutant or *tig dnaK dnaJ* triple mutant exhibited synthetic lethality (Deuerling *et al.*, 1999, Teter *et al.*, 1999). In 2004 Genevaux et al also showed that both DnaK and TF could be deleted and the cells would still survive, but this was only achieved at a low temperature (20°C), showing the importance of these chaperones during normal growth conditions. However, research done on gram positive bacteria *B.*

subtilis has shown that the *tig dnaK* double mutant or even the *tig dnaK dnaJ*, triple mutant did not confer synthetic lethality and the cells were able to survive at temperatures up to 49°C. This may suggest the involvement of other chaperoning mechanisms involved in *B. subtilis* which are not present in *E. coli* (Reyes & Yoshikawa, 2002). Interestingly, recent research has suggested that the trigger factor may play a bigger role in protein folding and stabilisation. Other than its interaction with nascent polypeptide chains, it has been shown that the trigger factor may play a role in binding and stabilising native-like proteins to facilitate their assembly into larger protein complexes like the ribosome (Martinez-Hackert & Hendrickson, 2009).

In archaea and eukaryotes it is the nascent chain associated complex (NAC) or ribosome associated complex (RAC), that play a similar role, though they do not shown any signs of sequence homology with TF. These are the first proteins to come into contact with nascent polypeptide chains. NAC/RAC is a heterodimeric complex with two subunits α and β, of sizes 33kDa and 22kDa respectively (Wiedmann *et al.*, 1994). It shows a similar function to the TF in that it also binds to nascent chains and then releases them in an ATP independent manner when the chain has been released from the ribosome, but its direct role in protein folding is still not clear.

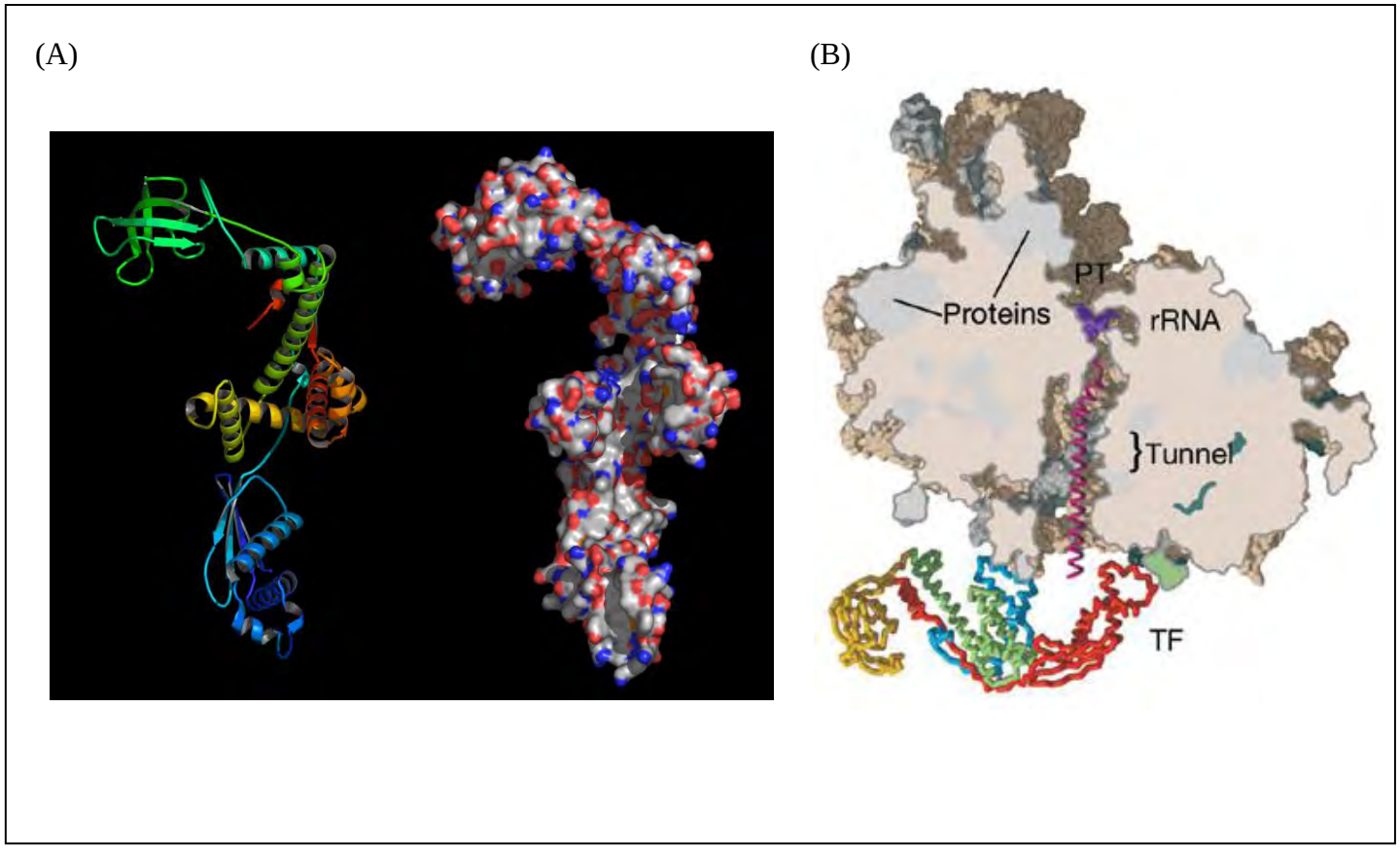


Fig 1.7: Structural and functional aspects of the trigger factor. (A) The 2.7 Å crystal structure of a 48kDa trigger factor. This is the best characterised ribosomal-associated chaperone. It assist in the folding mechanism of nascent polypeptide chains that are emerging from the ribosomal exit tunnel. In the space filled model, blue colour represents positively charged residues, while red represents negatively charged residues. (B) Trigger factor bound to a 50S ribosomal subunit. Image taken from Ferbitz *et al.*, 2004.

1.4.3.2 Hsp70 family of chaperones

The Hsp70 family of chaperones is very large and is present in both prokaryotes and eukaryotes. They are a 70kDa ATP dependent chaperone family and are known to be present not only in the cytosol but also in organelles such as the mitochondria, chloroplast and endoplasmic reticulum. Most eukaryotes have been shown to contain at least a dozen or more different HSP70 (DnaK) chaperones.

The N-terminus of Hsp70 is used to bind to and hydrolyse ATP while the C-terminal is responsible for peptide binding (Bukau & Horwich, 1998). A ribbon model of the structure of HSP70 can be seen in figure 1.8. Its cochaperones include DnaJ (HSP40) (refer to section 1.4.3.3), which has been shown to bind to the substrate before Hsp70, and GrpE which functions as a nucleotide exchange factor (Hartl & Hayer-Hartl, 2002, Fink, 1999) (figure 1.9). Together the cochaperones work by presenting the substrate to DnaK and also by modulating its ATPase activity.

Hsp70 binds to peptides by hydrophobic side chain interactions and hydrogen bonding (Frydman, 2001). Peptide binding occurs to the ATP bound state of DnaK when the "latch" (yellow in fig 1.9) is in an open conformation. This latch is the α helical segment of the C-terminal domain. The N-terminus of DnaJ then binds to DnaK and accelerates ATP hydrolysis by DnaK (Mayer *et al.*, 2000). This causes the latch to close and the peptide stays bound to the chaperone. GrpE then induces the release of ADP from the complex (Harrison *et al.*, 1997). This allows ATP to bind again to DnaK and causes it to release the peptide and

complete the reaction cycle. Nascent polypeptide chain can undergo repetitive cycles of folding to reach their native state (Szabo *et al.*, 1994).

The cellular concentration of DnaK exceeds that of the ribosomes, 50µM and 30µM respectively (Hartl, 1996). As discussed above, deleting the TF causes nascent chain polypeptide interactions with DnaK to rise from roughly 15% to about 40% (Deuerling *et al.*, 1999). It facilitates the folding of multi-domain proteins by cycles of binding and releasing (Teter *et al.*, 1999). There is however some evidence that suggests that Hsp70 may not interact with short lived partially folded intermediates (Strickland *et al.*, 1997).

1.4.3.3 Hsp40 family of chaperones

The Hsp40 family of chaperones consists of over 100 members, all of which have a highly conserved J domain of approximately 78 residues (Fink, 1999). To date, the best studied member of this family is the DnaJ protein which is a known cochaperone of DnaK. It is a 43kDa protein and is known to form a dimer in solution (Zylicz *et al.*, 1985, Ohki *et al.*, 1986). It is commonly suggested that Hsp40 is vital for efficient transfer of substrates to the HSP70 chaperone by stimulating its ATPase activity (Minami *et al.*, 1996) (Fig 1.9). The cochaperone function of DnaJ is known but there are some evidence to suggest that Hsp40 can function as chaperones on their own by binding to some nascent polypeptide chains (Hendershot *et al.*, 1996). This suggestion came about when it was shown that DnaJ binds to the denatured form of the protein firefly luciferase (Langer *et al.*, 1992, Szabo *et al.*, 1996). It was also shown that the Yeast homologue (Ydj1) can bind to denatured rhodanese (Cyr, 1995).

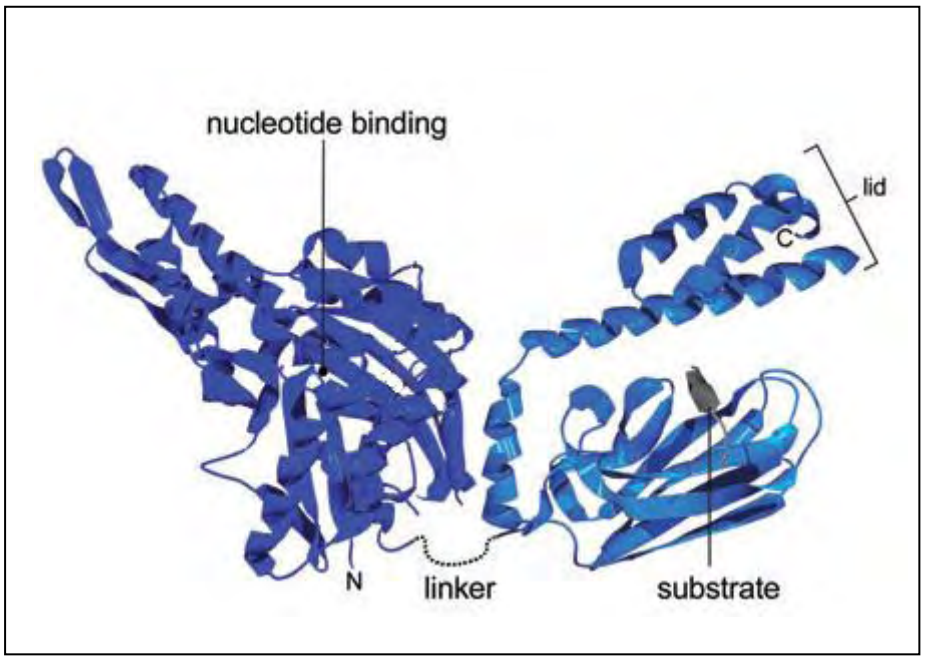


Fig 1.8: The ribbon model of HSP70 (DnaK) chaperone in *E. coli.* Image adapted from Genevaux *et al.*, 2007.

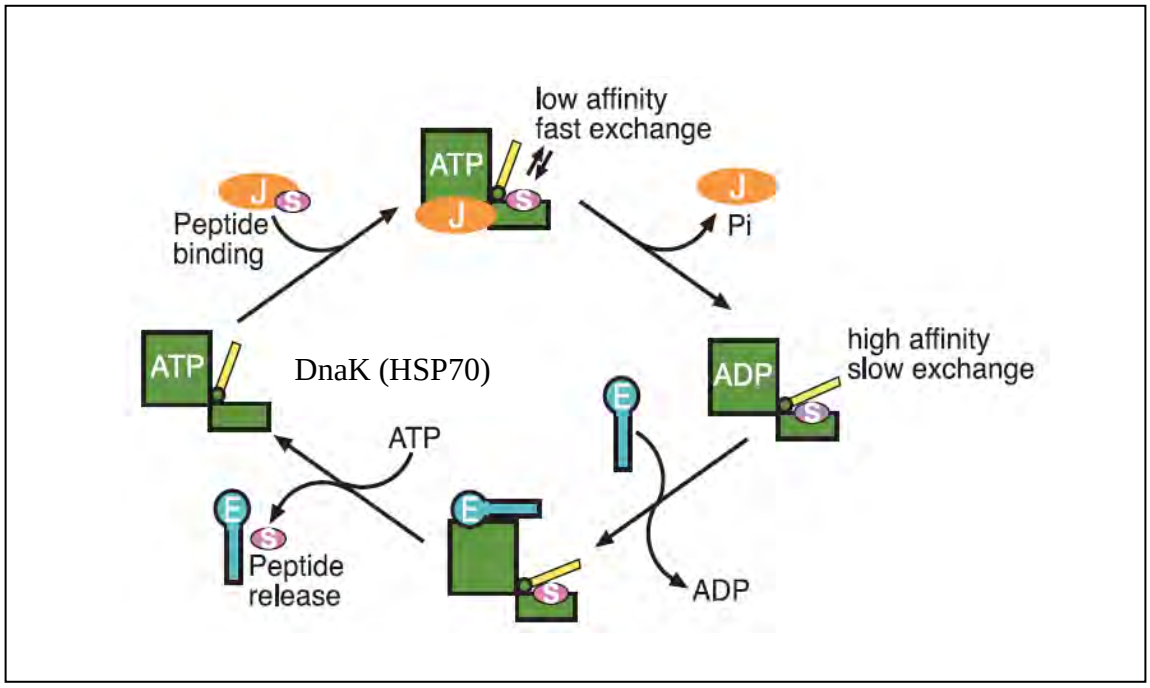


Fig 1.9: General mode of action of HSP70 chaperones. DnaJ (Hsp40), GrpE, and the peptide substrate are represented as J, E and S respectively. Taken from Hartl & Hayer-Hartl, 2002.

1.4.3.4 The small Hsps

One of the most divergent and widespread forms of chaperones, the small Hsps (or sHsps), are ubiquitous and their sizes vary from 16 to 42 kDa depending on the length of their N-terminal domain (Stromer *et al.*, 2003, de Jong *et al.*, 1998). However, they are also known to form large oligomeric structures with 9-30 subunits (Ehrnsperger *et al.*, 2000, Narberhaus, 2002). Structurally the stability of the small Hsp varies from one organism to another. The *M. tuberculosis* Hsp16.3 has a defined oligomeric structure (Chang *et al.*, 1996), while the *E. coli* IbpB seems to have a variable structure (Shearstone & Baneyx, 1999). The C-termini of small Hsps are known to be highly conserved and are shown to be homologous to the eye lens small heat shock protein α-crystallin (Ingolia & Craig, 1982).

Although extremely important in the maintenance of protein quality control within cells, the small Hsps are not as thoroughly understood as other chaperones (Nakamoto & Vigh, 2007). These small Hsps have been shown to function by preventing protein aggregation in an ATP independent manner. In vertebrates a vital role played by these chaperones is in the eye where they prevent cataracts by binding to denatured proteins. It has been shown that these chaperones are very efficient in that they can bind up to one substrate molecule per subunit and prevent them from aggregating (Haslbeck *et al.*, 1999, Ehrnsperger *et al.*, 2000, Lindner *et al.*, 2001) (Fig 1.10). They do not show any signs of substrate specificity and it has been proposed that they form complexes with denatured proteins when the cells are under stress so that they may act as a reservoir for the HSP70 chaperone machinery to refold these denatured polypeptides (Ehrnsperger *et al.*, 2000, Wang & Spector, 2001).

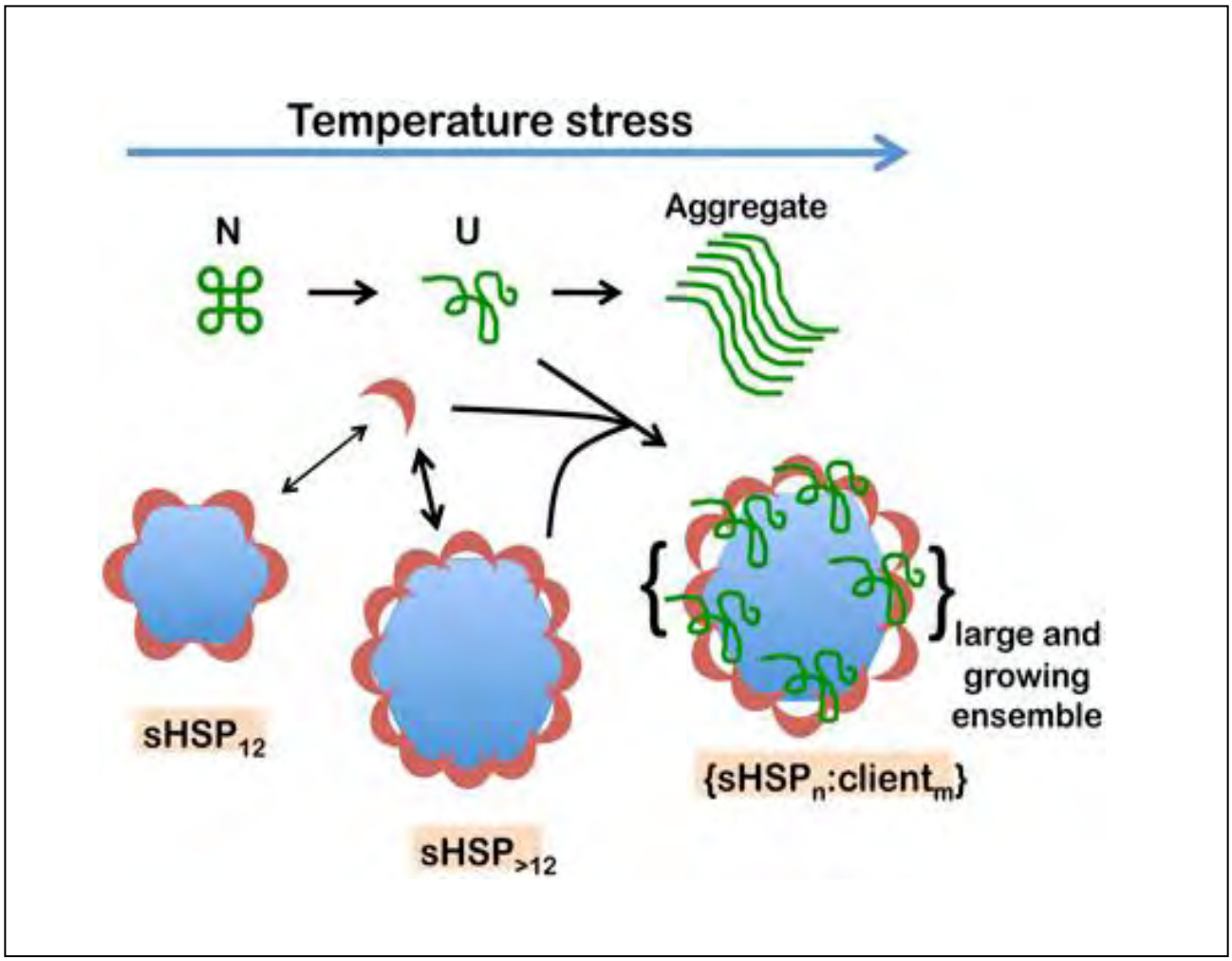


Fig 1.10: Mechanism of action of small Hsps. Heat stress causes native proteins (N) to become unstable. This mechanism highlights the ability of small Hsps to bind one unfolded substrate (U) per subunit and prevent aggregation (Eyles & Gierasch, 2010).

1.4.3.5 Hsp90 family of chaperones

One of the more abundant ATP-dependent forms of heat shock proteins, the Hsp90 molecular chaperone is 90kDa in size and found in bacteria (HtpG) and all eukaryotes (Hsp90) (fig 1.11). They have however not been found in archaea (Chen *et al.*, 2006). They are known to function as dimers for substrate binding. Some organelles have their own form of the HSP90 chaperone like Grp94 (ER form), while in *E. coli* the homologue is HtpG (high temperature protein G). As with all chaperones the Hsp90 family are also responsible in preventing protein aggregation, however, this is achieved by holding their substrates in a non active form until a signal is received to activate them. They also do not generally interact with nascent chains in vivo, but do so in vitro (Nathan *et al.*, 1997).

The prokaryotic homologue HtpG has been shown to be dispensable under stress condition unlike its eukaryotic homologue. The deletion of HtpG only affects cell growth at much higher temperatures although there is evidence that suggests HtpG does assist in the folding of protein in stressed *E. coli* cells (Thomas & Baneyx, 2000). In eukaryotes, Hsp70/Hsp40 chaperone machinery and the cochaperone Hop deliver substrates to Hsp90. This prevents the substrates from unfolding further or forming aggregates under stress conditions (Jakob *et al.*, 1995). The binding of ATP then allows the cochaperone p23 to bind to Hps90. ATP hydrolysis occurs due to the conformational changes in the Hsp90 dimer from p23 binding, resulting in the subsequent release of the substrate. Hsp90 does show signs of substrate specificity to signal transduction proteins (Pearl & Prodromou, 2006), such as protein kinases.

Structurally, HtpG consist of three domains and is found as a dimer when crystallised (fig 1.11). The N-terminal domain (NTD) contains the ATP binding site (lid), the carboxy terminal domain (CTD) is needed for homodimerization and the middle domain (MD) links the NTD and CTD together (Nemoto *et al.*, 1995). In its substrate free state, each of the 3 domains within each chain expose hydrophobic surfaces which has been suggested to act as substrate binding sites (Shiau *et al.*, 2006). Hsp90 interacts with its substrates when they are at a near native state and as such is believed to be involved in making much smaller conformational changes to its substrate proteins than that of Hsp70 or Hsp60 (Zhao *et al.*, 2005, Young *et al.*, 2001). Crystal structures of HtpG with bound nucleotide shows distinct conformational states and it is believed that these confirmations are a part of the reaction cycle (Shiau *et al.*, 2006, Krukenberg *et al.*, 2008, Harris *et al.*, 2004). A simple representation of the reaction cycle is as follows (fig 1.12):

- In the absence of ATP the chaperone is in an open state that has regions of hydrophobic residues from the MD (src loop in fig 1.11) and CTD (H21 and H21' in fig 1.11) are available for substrate interaction.
- The next step involves the loading of an appropriate substrate into the open cleft.
- ATP binding and hydrolysis at the active site lid then results in the binding of the substrate. Conformational changes then closes the cleft resulting in substrate protein remodelling. This change also hides the hydrophobic residues that were present on the surface.
- Once ATP is hydrolysed, the resulting ADP bound HtpG then releases the substrate and is ready for the next cycle.

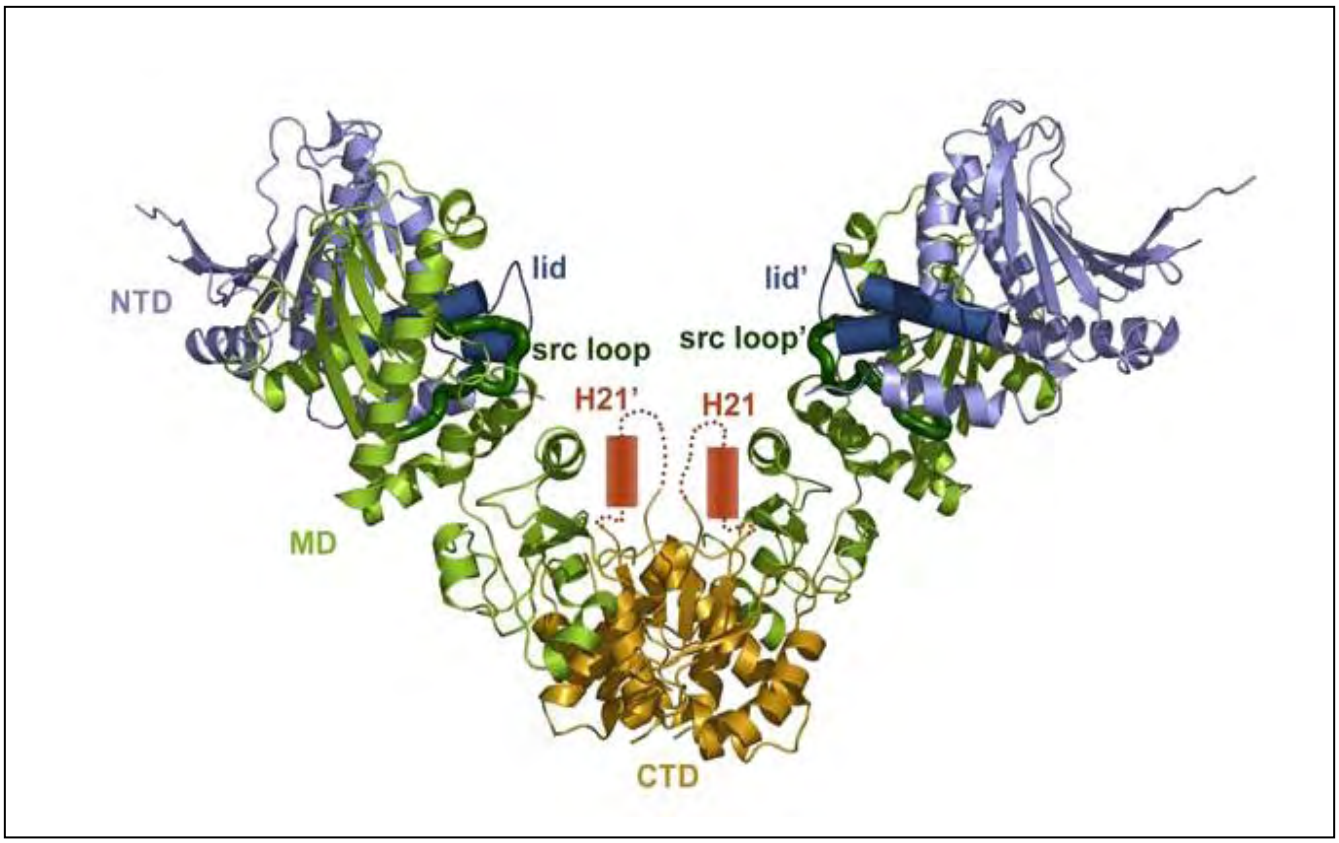


Fig 1.11: Structure of HtpG dimer (Prokaryotic Hsp90) (Shiau *et al.*, 2006). Ribbon view of the amino-terminal domain (NTD) (blue) contains the ‘‘active-site lid’’ region. The middle domain (MD) (green) contains the extended ‘‘src loop’’ (dark green tube), and the carboxy-terminal domain (CTD) is shown in gold. The exposed carboxy terminal helices (H21 and H21’) (red boxes) are also shown.

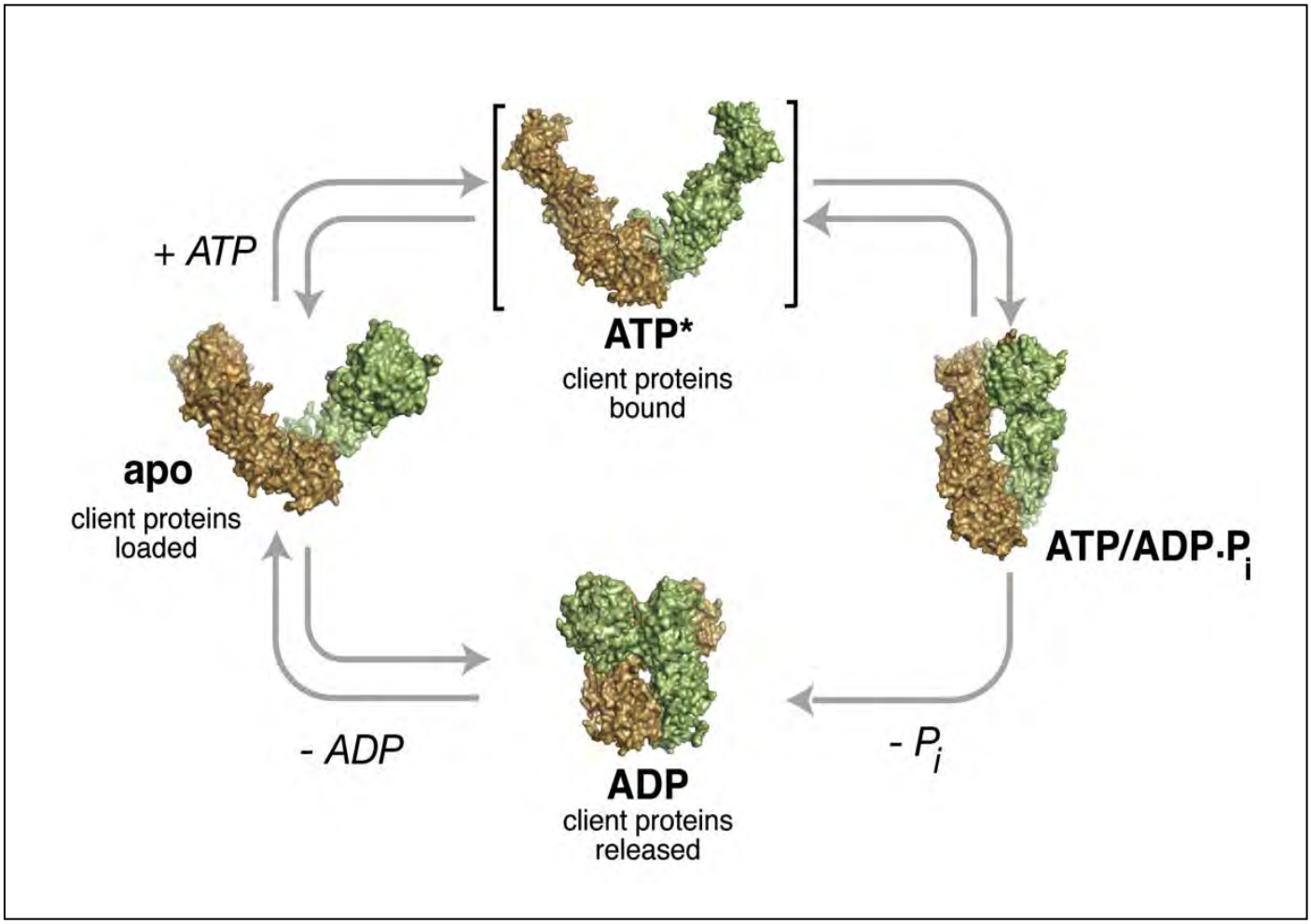


Fig 1.12: Prokaryotic Hsp90 reaction cycle. In the nucleotide free state (apo) the chaperone is in an "open" confirmation. The binding of substrate and ATP results in conformational changes that result in client protein remodelling. After the hydrolysis of ATP is complete the substrate is then released from the chaperone (Shiau *et al.*, 2006).

1.4.3.6 Hsp100 family of chaperones

The HSP100 (or Clp in prokaryotes) family of chaperones are ATP dependent and present in both eukaryotes and prokaryotes (Glover & Tkach, 2001). Hsp100 family members have been shown to form both heptameric (ClpP) and hexameric (ClpA) structures. If you recall from section 1.3, under conditions of extreme stress, when there are more proteins aggregating than the chaperone systems can prevent, Hsp100 acts on these aggregates and either resolubilize them or directs them towards proteolytic degradation (Schirmer *et al.*, 1996). The AAA (ATPases Associated with diverse cellular Activities) members ClpA, ClpB, ClpC, and ClpX are the best studied Hsp100 homologues in bacteria. They are known to have proteolytic properties. Hsp104 and ClpB unlike other chaperones do not promote the folding of non native proteins and they also do not prevent aggregation. They are responsible in resolubilizing stress-damaged proteins from aggregates (Hung & Masison, 2006, Mogk & Bukau, 2004), and require other chaperones like the Hsp70 system to act along side in a cooperative manner to fold them into their native state (Motohashi *et al.*, 1999, Glover & Lindquist, 1998). ClpA, ClpC and ClpX disaggregate proteins by degrading them along with the help of the protease ClpP (Maurizi & Xia, 2004, Kress *et al.*, 2007, Martin *et al.*, 2007). In effect, the Hsp100 family of chaperones functions in two different ways (fig 1.13).

This family is also sometimes referred to as the AAA+ super family due to the sequence and structures of the HSP100 resembling those of the ATPase associated with different cellular activity (AAA). They are subdivided into two groups based on their structural features and sequence similarities:

- Group 1: ClpA (Hsp104), ClpB, ClpC, ClpD, ClpE have two nucleotide binding domains
- Group 2: ClpX and ClpY have one nucleotide binding domain.

Both these groups can then be further subdivided based on specific signature motifs or by the length of the interdomain regions separating the two nucleotide binding domains (Schirmer *et al.*, 1996).

Structurally, all Hsp100 chaperones contain a conserved core, referred to as the AAA module, which consists of two nucleotide binding subdomains (NBD) (Maurizi & Xia, 2004). This area is the binding site for ATP and also contains the catalytic residues for hydrolysis (Lupas & Martin, 2002). The N-domain of the Hsp100 chaperones is not essential for all its members. Hsp78, which is a homologue of ClpB, does not have an N-domain but is active as a chaperone (Krzewska *et al.*, 2001). However, it is at this domain that the adapter protein ClpS associates on ClpA to alter its substrate specificity (Dougan *et al.*, 2002). It is the I-domain that is responsible for the biological activity of the Hsp100 chaperones. Deleting this region causes the proteins to lose the ability to hydrolyse ATP (Mogk *et al.*, 2003). A general representation of the mechanism by which some of the Hsp100 chaperones function is shown in figure 1.13.

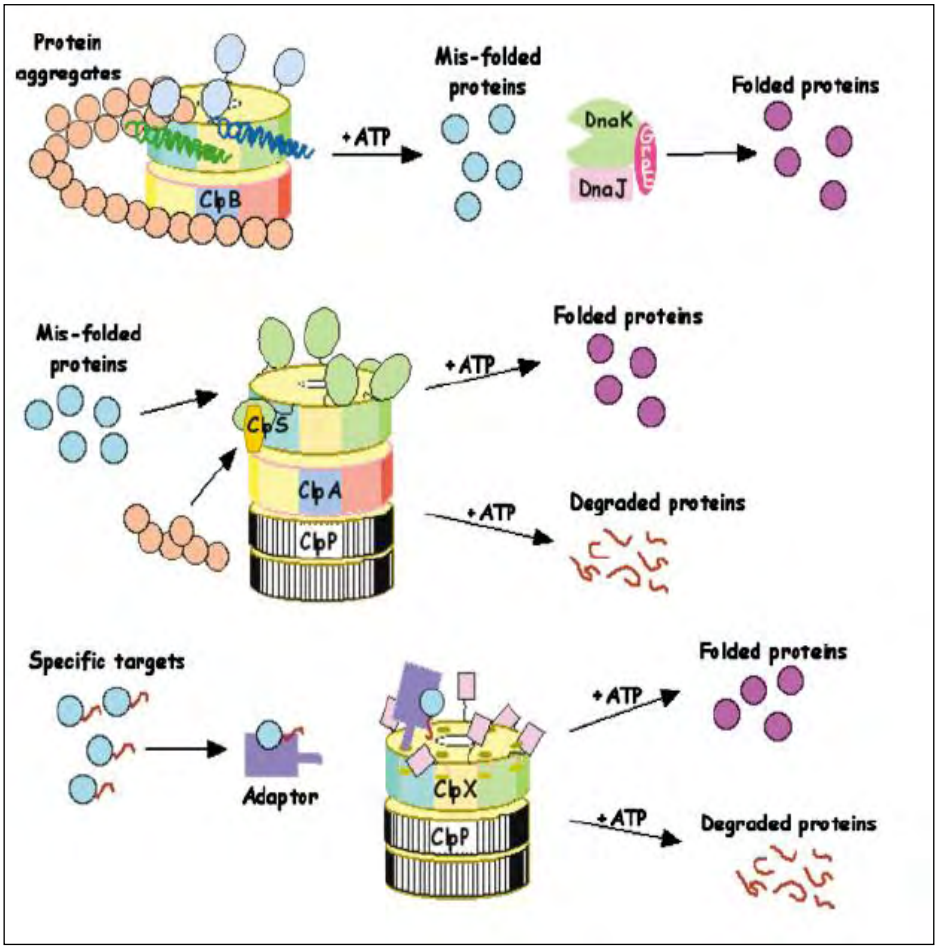


Fig 1.13: The different routes taken by different members of the Hsp100 family in protein quality control. ClpB functions by disaggregating damaged proteins and allows them to be refolded to their native state with the help of other chaperones like Hsp70. ClpA has the ability to recognise specific motifs to target substrate proteins but can also target substrate proteins without the motifs and degrade misfolded proteins with the help of ClpP. It can also make use of the ClpS adaptor protein to help disaggregate substrates. ClpX can only recognise substrates that have an adaptor protein attached and also have specific sequence motifs. ClpP are peptidases that break down misfolded polypeptides and prevent further aggregation. Image taken from Maurizi & Xia, 2004.

1.5 Hsp60 family of chaperones

The Hsp60 family of chaperones is one of the most abundant forms of ATP dependent heat shock proteins within a cell. The term 'chaperonin' (Cpn) was first used by Hemmingsen in 1988 to represent this subclass of molecular chaperones (Hemmingsen *et al.*, 1988). The chaperonins are a highly conserved class of chaperones and are found in all the domains of life. They are usually composed of two rings stacked together to form a larger oligomeric structure and enclose a central cavity in each ring. Each ring is made up of seven to nine 60kDa subunits. The central cavity is used by the chaperonins to enclose unfolded proteins and allow them to fold secluded from the risks of unwanted interactions from other folding proteins in the cell.

E. coli chaperonin GroEL and its cochaperonin GroES have been the most studied chaperones in the Hsp60 family. The details of their mechanism of action can be found in section 1.5.4.

1.5.1 Types of chaperonins

Chaperonins are classed into 2 sub groups based on sequence homologies (Horwich & Willison, 1993, Gutsche *et al.*, 1999) (Fig 1.14):

- Group 1 chaperones are members of the GroEL family and are found in eubacteria and eukaryotic organelles like the mitochondria. They are usually encoded by one gene and require the assistance of a cochaperonin like GroES.
- Group 2 chaperones are members of the TCP-1 ring complex (TriC) family and are found in archaea and the eukaryotic cytosol (Gutsche *et al.*, 1999). They are usually

encoded by more than one gene and do not require the assistance of any cochaperonins, though they may require other co-proteins to function, e.g. prefoldin.

The subunits of both groups of chaperonins consist of three domains (Fig 1.15):

- The apical domain (shown in purple): In group I chaperonins the binding sites for substrates and the cochaperonin GroES (cpn10) are located in the apical domain. In group II chaperonins the co-chaperoning function of GroES is replaced by a 25Å long flexible helical protrusion that extends from the tip of the apical domain and functions as a lid to cover the substrate when it is in the cavity (Klumpp *et al.*, 1997).

- The intermediate domain (shown in orange): This domain connects the equatorial and apical domains and is responsible for relaying a lot of the allosteric information to and from the other domains and between the rings. In group I chaperonins, this includes relaying the conformational changes generated by nucleotide binding from the equatorial to the apical domain to facilitate substrate binding and also binding of GroES to one end of GroEL transmits a signal to the other end that leads to GroES release.

- The equatorial domain (shown in green): This site contains a highly conserved nucleotide binding region and connects the two heptameric rings to each other. It is also responsible for most of the inter-subunit contacts within a ring. This is important as it not only controls the assembly of the ring, but also the cooperativity between the subunits.

For this review, the group 2 chaperonins will be dealt with before the group 1 chaperonins as it is GroEL and its homologues that are being studied in this report.

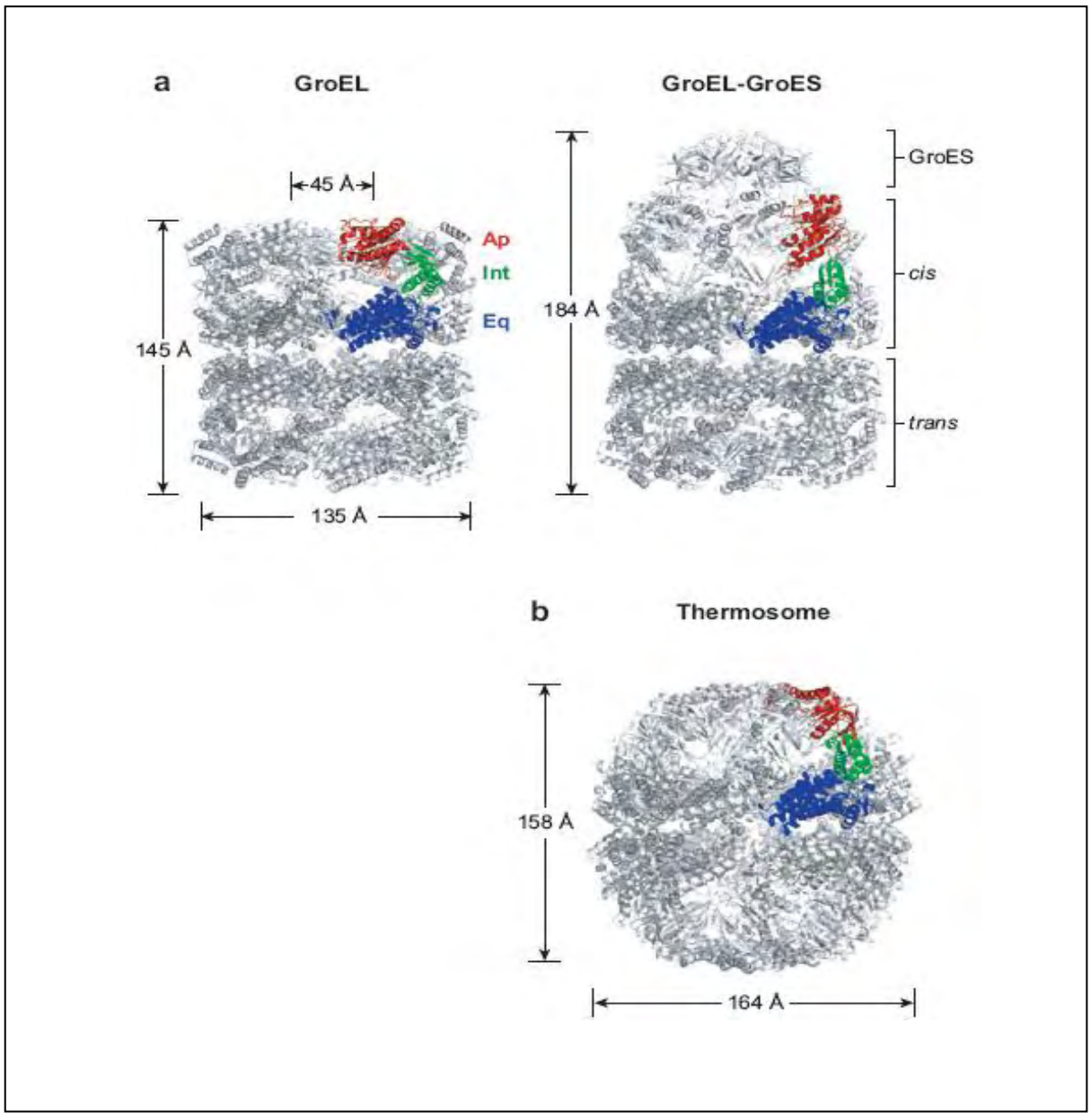


Fig 1.14: The structure of the 2 groups of chaperonins. A) Group I chaperonin, GroEL along with its cochaperonin GroES. B) Thermosome, a Group II chaperonin. Image taken from Horwich *et al.*, 2007. The Apical, intermediate and equatorial domains are coloured red, green and blue respectively.

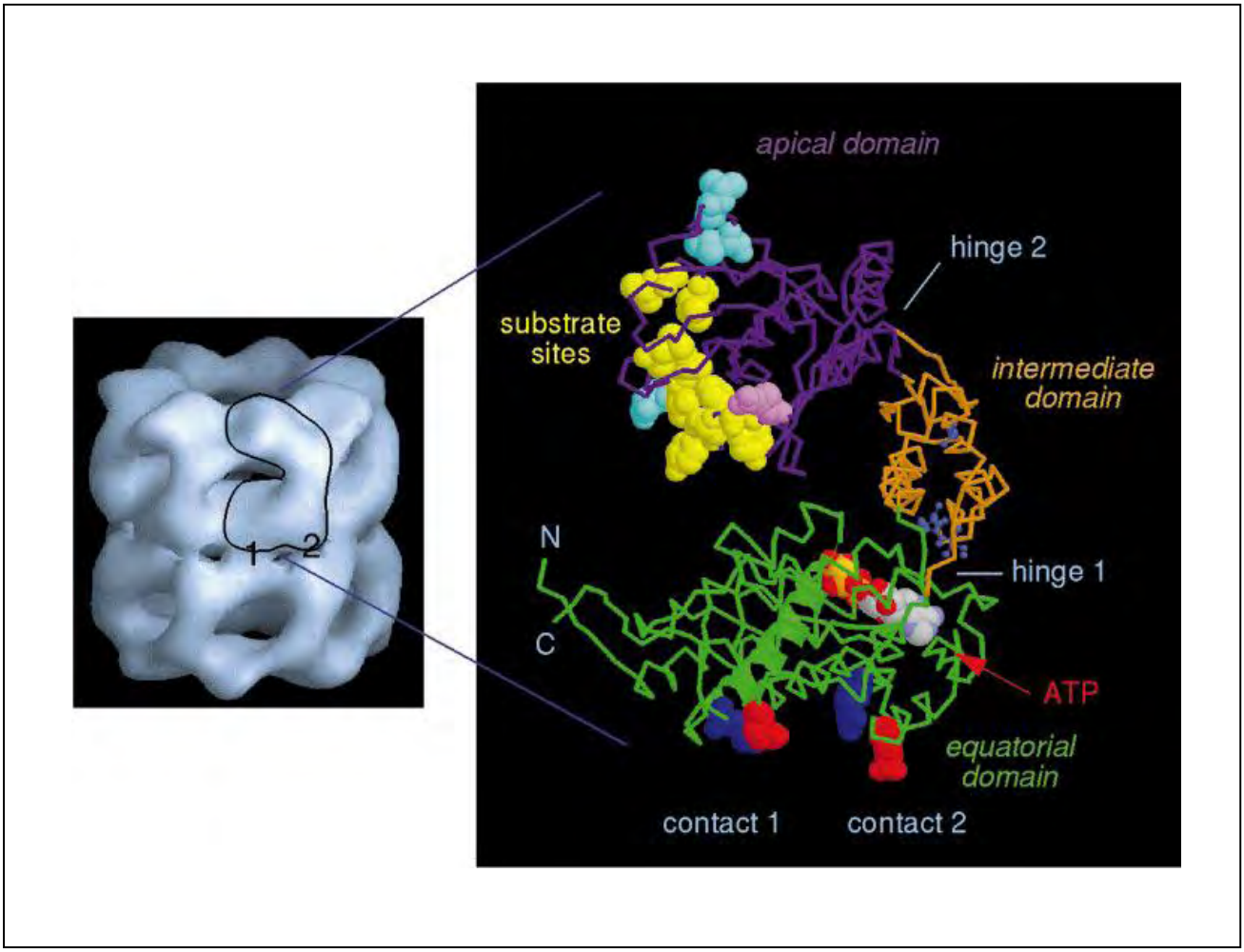


Fig 1.15: The illustration of a GroEL subunit. The three domains along with the substrate binding site, the ATP binding site, and the ring-ring contacts are also shown. Image taken from Ranson *et al.*, 1998.

1.5.2 Nomenclature

The term chaperonin is applied to all proteins that share homology with the *E. coli* GroEL. However, different studies use different terms for the same protein and as a result the nomenclature for these proteins is a little complicated. While the bacterial chaperonins are usually referred to as GroEL (mainly *E. coli*) or Cpn60, the mitochondrial homologues are usually called Hsp60. Similarly, the proteins that function along with the chaperonins called cochaperonins are referred to as GroES (in *E. coli*), Cpn10, or Hsp10. The eukaryotic group II chaperonin, which was originally called TCP-1, is now either referred to as TRiC (TCP containing ring complex) (Frydman *et al.*, 1992), or CCT (chaperonin containing TCP-1). The archaeal chaperonins are more commonly referred to as thermosomes (Trent *et al*., 1991).

For this thesis GroEL and GroES will be used exclusively to refer to the *E. coli* chaperonin and cochaperonin respectively. Cpn60 (Cpn60.1, Cpn60.2, or Cpn60.3 depending on the homologue) and Cpn10 will be used to refer to the Mycobacterial chaperonins. Thermosome will be used to refer to the archaeal chaperonin, and CCT will be used to refer to the eukaryotic chaperonin.

1.5.3 Group II chaperonins

As stated above, these chaperonins are found in the cytosol of eukaryotes and are also present in the archaea (Lewis *et al.*, 1992). These chaperonins are phylogenetically distinct from one another. The group II chaperonin in archaea, referred to as the thermosome, are made up of two ringed subunits each of which consist of 8-9 further subunits (Andra *et al.*, 1996, Knapp *et al.*, 1994) (fig 1.16). In eukaryotes it is the CCT (chaperonin containing TCP-1), that is the group II chaperone. They are made up of 8 different structurally related subunits

(Rommelaere *et al.*, 1993, Kubota *et al.*, 1995, Kubota *et al.*, 1994). Though the CCT and thermosome are both group II chaperonins, CCT is not induced by heat shock stress, whereas the thermosome is (Horwich *et al.*, 2007).

Group II chaperonins share a similar architecture with group I chaperonins in that the sub-units have the same 3 domain structure. The equatorial domain contains the inter-ring and inter-subunit interaction sites as well as the ATP binding site, the intermediate domain that connects the equatorial domain to the apical domain, and the apical domain which contains the binding site for client proteins. The main difference between the classes however lies in the absence of a GroES like cochaperonin to seal of the substrate protein within the cavity of the chaperonin. The group II chaperonins have instead been shown to contain a hydrophobic 25 Å long flexible motif that extends as a helical protrusion and acts as the lid to cap the central cavity (Klumpp *et al.*, 1997).

1.5.3.1 Archaeal thermosome

The architecture of group II archaeal thermosomes consists of two ringed structures stacked back to back (Schoehn *et al.*, 2000) (fig 1.16). The subunits possess an apical, intermediate and equatorial domain and are generally present as octameric rings but have also been seen as nonameric rings (Ditzel *et al.*, 1998, Trent *et al.*, 1991). The 8-9 subunits are encoded by up to three different genes. The thermosome from *Thermoplasma acidophilium*, for which a crystal structure is available, consists of two stacked eight-membered rings with alternating α and β subunits (Ditzel *et al.*, 1998). More recently the crystal structure of *Thermococcus* KS1 was shown to share the structural features of *T. acidophilium* (Iizuka *et al.*, 2005). One of the conformational differences between thermosomes and group I chaperonins, like GroEL, is

that while each subunit of GroEL interact with 2 subunits from the opposite ring, in thermosomes each subunit is aligned such that it makes contact with just one other subunit on the opposite ring.

Thermosomes do not have any cochaperones involved in their substrate reaction cycle, unlike group I chaperonins. Instead, their subunits contain approximately 30 residues that form a parallel β sheet that acts as the lid. Substrate folding takes place within the central cavity of the thermosome which is approximately 74% of the volume of the cis-GroEL-GroES cavity. It is known that the thermosomes are ATP dependent for their folding mechanism and also that substrate binding most probably occurs due to hydrophobic forces (Pereira *et al.*, 2010). Until recently all the crystal structures of archaeal thermosomes were of structures in the closed confirmation (sealed cavity). While they were informative, they did not provide much information about the conformational changes that occur during the reaction cycle. However recent structural studies of the thermosome of *Methanococcus maripaludis* has shed some light on these conformational changes by comparing the crystal structure of the closed state with that of the open state (Pereira *et al.*, 2010). It has been suggested that the closing mechanism is distinct from that of group I chaperonins in that all three domains of each subunit undergo large conformational changes between the open state and closed state following ATP hydrolysis, while in group I chaperonins only the apical and intermediate domains undergo changes and the equatorial domain is primarily stationary. These large conformational changes between all three domains results in the closure of the lid (Zhang *et al.*, 2010).

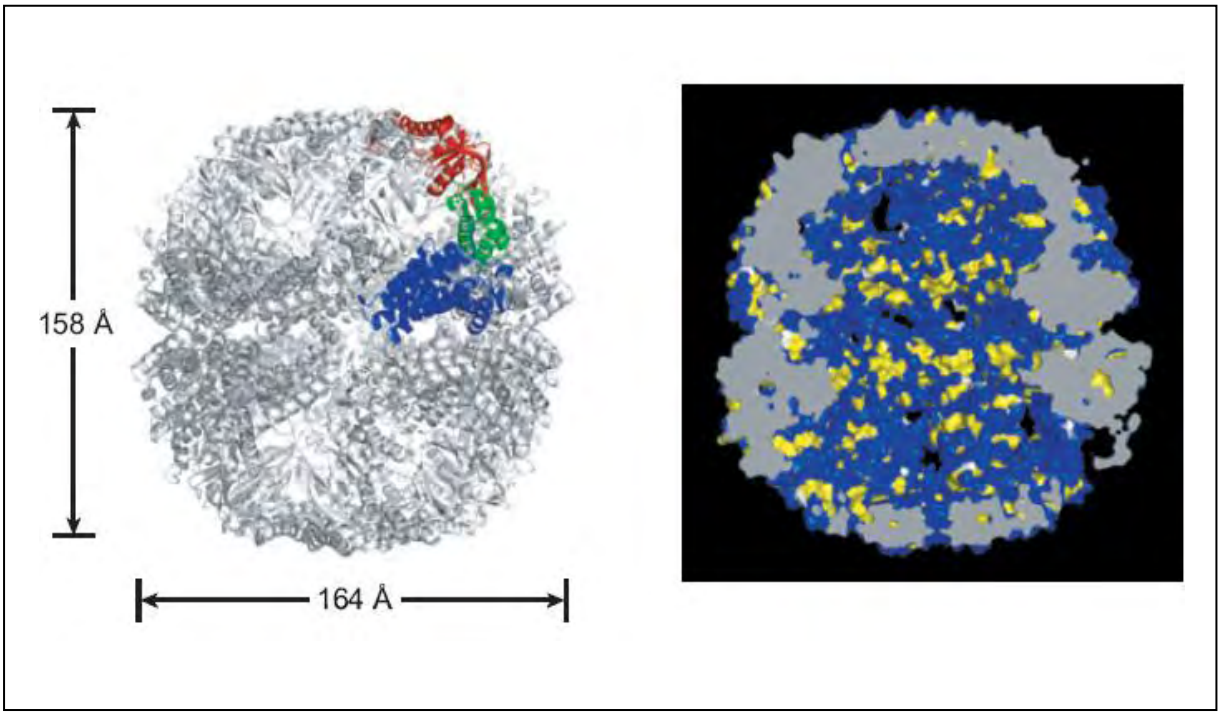


Fig 1.16: Octameric structure of the two thermosome rings from *Thermoplasma acidophilum*. The space-filling image on the right shows a cut-away view to show the properties of the cell wall in the central cavity. The blue colour represents hydrophilic residues and the yellow colour represents hydrophobic residues. Image taken from Horwich *et al.*, 2007. Original source of information Klumpp et al. 1997, Ditzel et al. 1998, Pappenberger et al. 2002.

The exact mechanism of the reaction cycle is still under investigation, but based on the reports so far it is possible to suggest that the thermosome interaction with ATP may be as follows:

1. ATP binding to one ring on the thermosome results in the opening of the cavity. The asymmetric nature of the rings results in the ATP affinity of the second ring to drop.
2. ATP hydrolysis on the first ring then results in large conformational changes that result in the substrate being released into the central cavity for the folding process while the lid closes the cavity.
3. At this stage, owing to negative cooperativity, the opposite ring can bind to ATP and maintain asymmetry. It has also been suggested that the dissociation of the product is required before ATP hydrolysis can occur on the second ring.

1.5.3.2 Eukaryotic CCT

CCT is a complex ATP dependent chaperonin as it possesses eight different but homologous subunits in each of its two rings which are referred to as α, β, γ, δ, ε, ζ, η, and θ (Kubota *et al.*, 1995, Kubota *et al.*, 1994). The entire structure is made up of two barrel shaped double ring cylinder which is approximately 160Å in height and 150Å in diameter (Llorca *et al.*, 1999) (Fig 1.17). Each of these eight subunits is encoded by eight specific genes. Although the presence of eight different subunits could result in a very large number of possible arrangements, the presence of preferential inter-subunit contacts ensures that the position of each subunit within each ring is highly specific in arrangement and is shared between all CCTs (Liou & Willison, 1997) (fig 1.17a). It has also been shown that these subunits are always in the same phase with the opposite ring (Martin-Benito *et al.*, 2007) (fig 1.17b). The

main difference between the eight subunits is present in the sequence of the apical domains which have specific interactions with unfolded proteins (Spiess *et al.*, 2006). Two of CCTs best studied specific substrates are the cytoskeletal proteins actin and tubulin. Both of these have been shown to interact specifically with select subunits from the CCT complex (Llorca *et al.*, 1999). As is the case with the thermosomes, it is the protrusions that are present on the apical domain that fold to cover the openings of the chaperonin and form the dome structure. EM analysis of asymmetrically arranged subunits revealed that the position of the subunits was directing a sequential mechanism of ATP binding and hydrolysis within the ring, unlike GroEL (Rivenzon-Segal *et al.*, 2005).

Although the exact mechanism of substrate folding by CCT is not yet fully understood, recent crystal structures have provided a lot of information about the nature of the residues on and in the CCT complex (Cong *et al.*, 2010). It has been shown that, like GroEL and thermosome, the inner surface of the closed chamber of CCT is also very hydrophilic but varies in charge distribution. The inner chamber of GroEL is a lot more negatively charged, while CCT has been shown to be predominantly positively charged. These differences in the charged properties of CCT have been suggested to be related to its ability to fold some substrates that cannot be folded by other chaperonins. These results implicate polar and electrostatic interactions in substrate binding; however, hydrophobic interactions have also been shown to occur (Rommelaere *et al.*, 1999, Yam *et al.*, 2008).

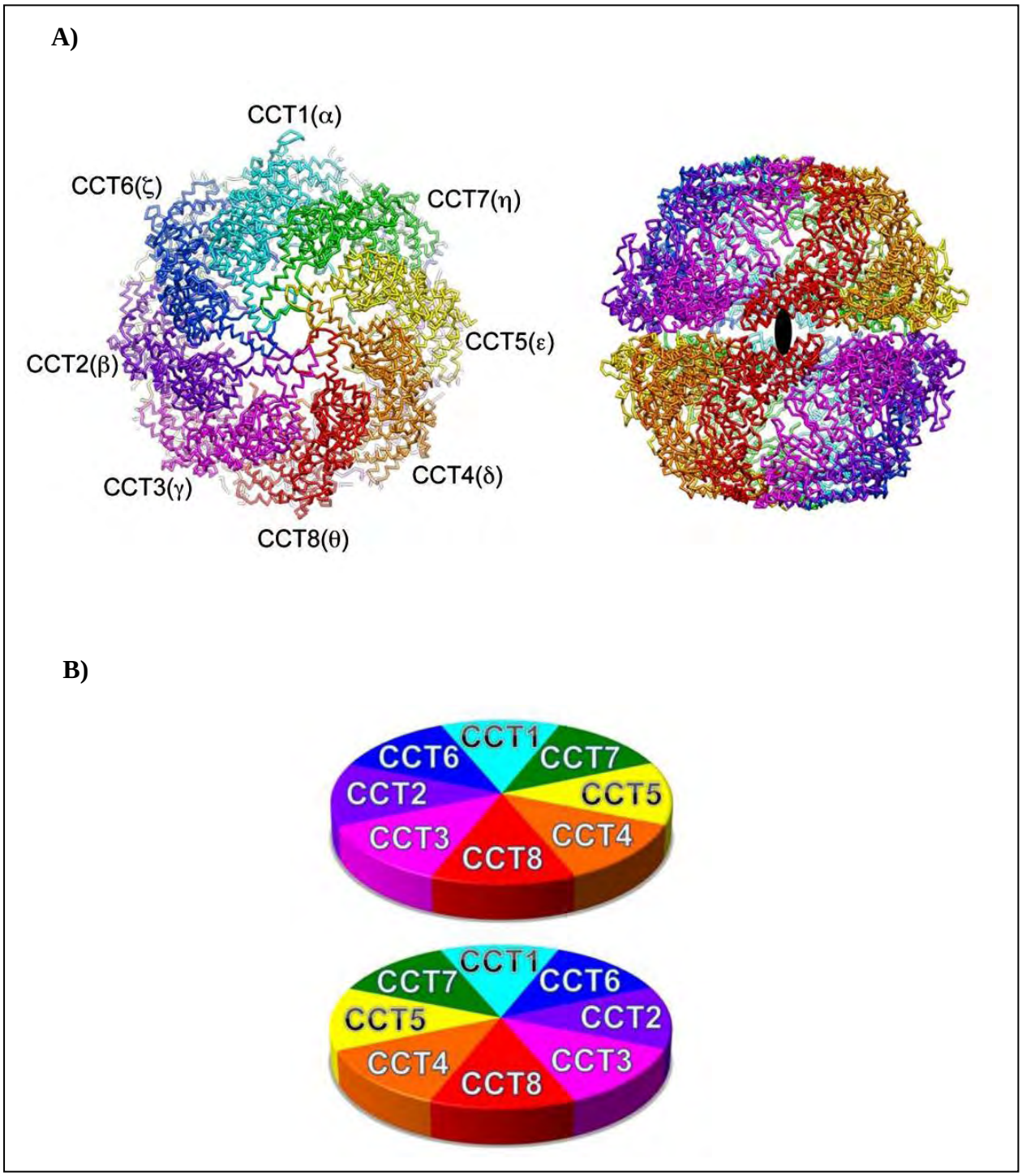


Fig 1.17: Structural properties of CCT. A) End on and side view of the double ring structure of CCT with 8 subunits per ring. B) Diagram illustrating the arrangement of the 16 subunits in the two rings. Taken from Cong *et al.*, 2010.

1.5.3.3 Prefoldin

It has been shown that, for some substrates, the group II chaperonins use a substrate delivery agent, prefoldin (also sometimes called GimC in archaea), which interacts with the substrates and then delivers it to the chaperonin (Geissler *et al.*, 1998). In this regard it can be regarded as a cochaperonin to the group II chaperones; however, its functional role is very different to that of classical cochaperonins such as GroES. Prefoldin from both archaea and eukaryotes are jellyfish shaped structures with six subunits. In most archaea the prefoldin complex is composed of two different subunits encoded by two different genes *pfdα* and *pfdβ* (Geissler *et al.*, 1998). At the tip of each of the subunits is a hydrophobic substrate binding site (Siegert *et al.*, 2000), although the mode of substrate binding seems to slightly differ between archaeal and eukaryotic prefoldin (Siegert *et al.*, 2000, Martin-Benito *et al.*, 2002, Zako *et al.*, 2006). Recent study on the prefoldin of a hyptherthermophilic archaea has suggested that it exhibits some refolding activity for lysozymes at lower than physiologically active temperatures (Zako *et al.*, 2010).

1.5.4 Group I chaperonins

As discussed above, the group I chaperonin family is found in eubacteria and eukaryotic organelles like the mitochondria. The *E. coli* GroEL-GroES system is one of the best studied examples of this group. It was first discovered by Georgopoulos when he was investigating mutants of *E. coli* that were incapable of allowing the growth of bacteriophage λ (Georgopoulos *et al.*, 1972, Georgopoulos *et al.*, 1973, Georgopoulos & Hohn, 1978, Tilly *et al.*, 1981). Experiments conducted with prokaryotic Rubisco and the Rubisco binding protein (Homologue of GroEL) showed that both the dimeric and hexadecameric forms of Rubisco could be synthesised in *E. coli* (Gatenby, 1984, Hemmingsen *et al.*, 1988). The chaperoning property of GroEL then became evident when it was shown to be responsible for the proper

folding of Rubisco in *E. coli* (Goloubinoff *et al.*, 1989). This chaperonin was also shown to be essential under all growth conditions (Fayet *et al.*, 1989).

1.5.4.1 GroEL and its cochaperonin GroES

The *E. coli* chaperonin and cochaperonin are generally referred to as GroEL and GroES. GroEL is a tetradecamer that is made up of two heptameric rings of 57kDa proteins stacked back to back to form two 7 subunit hollow cylinders. The entire structure is approximately 147 Å in length and 137 Å in diameter and the enclosed ring has a diameter of 47 Å (Braig *et al.*, 1994). GroES is the cochaperonin of GroEL. The GroES monomer is 10kDa in size. It forms a heptameric dome like structure with 7 identical subunits (figure 1.18). The cochaperonin can bind to either side of GroEL depending on the presence of ATP/ADP. GroES is largely composed of β sheets that form hydrophobic mobile loops which interact with specific residues on the apical domain (Bukau & Horwich, 1998, Sigler *et al.*, 1998, Horwich *et al.*, 2007). This interaction between GroEL and GroES is vital for the correct folding of its substrate proteins. When a substrate binds to the apical domain of one of the rings of GroEL, GroES also bind to the same ring (mechanism is discussed in section 1.5.4.2). The Binding of GroES to the apical domain not only displaces the bound substrate into the central cavity of GroEL but has also been shown to induce large conformational changes to GroEL as well. It has been shown that the binding of GroES to GroEL causes the apical domain to expand in an upward and outward motion. The results of this motion are still under debate. One viewpoint has shown that if a substrate is bound to more than one apical domain, then the expansion of the domain can subject the substrate with stretching forces that which could induce their unfolding (Coyle *et al.*, 1999, Lin *et al.*, 2008). While the other viewpoint has shown that the rigid body movements of the apical domain causes the hydrophobic residues of the apical domain to be hidden away from the substrate and as a

result the substrate protein is rapidly disassociated and pushed into the central cavity to fold (Horwich *et al.*, 2006, Rye *et al.*, 1997, Chaudhry *et al.*, 2003, Xu *et al.*, 1997). It is however agreed upon that GroES acts as the lid to seal off the GroEL cavity to allow a protein to fold without being influenced by the crowded nature of the cell. However, the mechanism involved in the folding of proteins within the cavity is also subjected to further debate.

1.5.4.2 GroEL and GroES function

The mechanism of action of GroEL and its cochaperonin is now understood in detail. As discussed above, GroES forms a homo-heptameric ring that caps one of the two GroEL rings forming an asymmetric complex (Horwich *et al.*, 2007, Bochkareva *et al.*, 1992, Burston *et al.*, 1995). This binding of GroES, displaces any bound substrate into the cavity of GroEL so the reaction cycle can continue. There are 3 states in which GroEL and GroES can be found during the reaction cycle (Ranson *et al.*, 1997):

1- The acceptor state (open state): In this state one ring of GroEL bind to the substrate (client protein) while the other ring remains bound to ADP and GroES.
2- The encapsulation state (closed state): In this state the substrate is released into the cavity to fold under hydrophilic conditions. This occurs when GroES and ATP bind to the same ring as the substrate.
3- The ejection state (release state): In this state the folded/semi folded protein along with GroES is released from GroEL. This occurs when the bound ATPs are hydrolysed and allows another 7 ATP molecule to bind to the second ring, followed by substrate and GroES binding.

During the process of trying to understand the mechanism behind the functions of the GroEL chaperonins, three main models were put forward:

1. The Anfinsen Cage model
2. The iterative annealing model
3. The assisted folding model

The Anfinsen cage model

According to this model, the polypeptide in the central cavity of the chaperonins follows the energy landscape guided by its own primary structure without any influence from other factors such as molecular crowding or interactions with the cavity itself (Ellis, 1994, Horwich *et al.*, 2009)

The iterative annealing model

This model suggests that GroEL allows proteins to undergo repeated attempts at folding into their native states by unfolding the substrates that are stuck in a non-native kinetic trap. This had been proposed to increase the speed at which proteins can be folded to their native states (Todd *et al.*, 1996).

The assisted folding model

This model suggests that GroEL, while providing a secluded environment for the protein to fold, also functions by accelerating the folding process of the substrates. This is achieved with the help of the interactions between the hydrophobic surfaces of the substrate and the hydrophilic surface of the GroEL internal wall. This model was proved for some polypeptides by Tang et al in 2006 (Radford, 2006, Tang *et al.*, 2006).

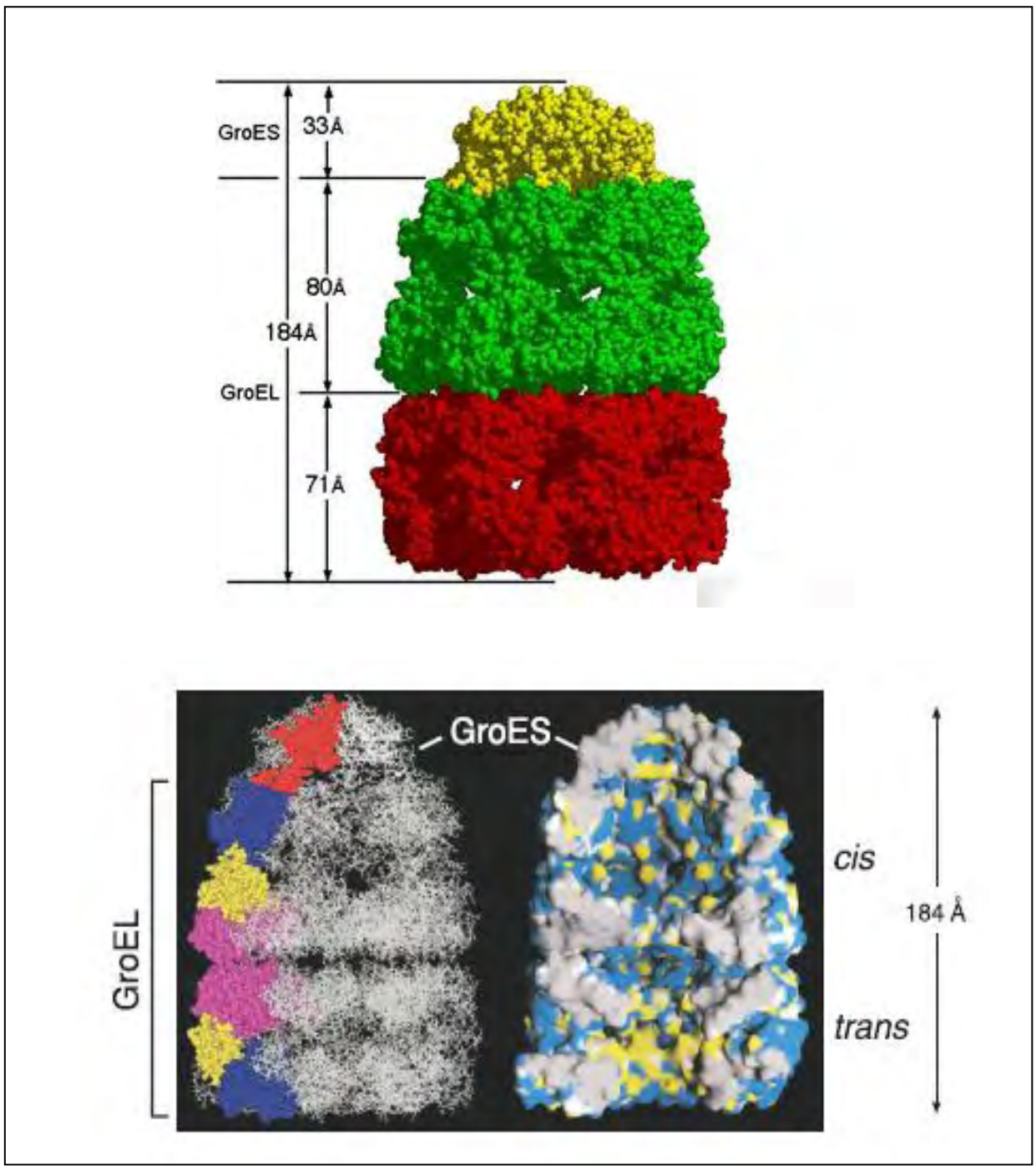


Fig 1.18: The structures of GroEL and GroES. The space-filling image shows a cut-away view to show the properties of the cell wall in the central cavity. The blue colour represents hydrophilic residues and the yellow colour represents hydrophobic residues. Images taken from Hartl & Hayer-Hartl, 2002.

Irrespective of the mechanisms used by GroEL in the folding of substrate proteins, in all cases a central cavity is required during the reaction cycle. This cavity provides a secluded environment for the substrate proteins to fold in within being influenced by the crowded cellular environment. The mechanism of the reaction can be summarised in the following steps (Tyagi *et al.*, 2009):

1- 7 ATP molecules bind to the equatorial domain of one of the uncapped GroEL rings (*cis* ring). This results in a small degree of elevation and counter clockwise twist of the apical domain, in effect, mobilising it for substrate binding. The binding of ATP to the *cis* ring also results in a large drop of affinity to ATP of the *trans* ring (opposite ring) due to negative cooperativity.

2- A substrate protein then binds to the apical domain of the same uncapped GroEL ring (*cis* ring) due to hydrophobic interactions between the rim of the central cavity and the polypeptide chain. This is soon followed by the binding of the cochaperonin GroES to same ring.

3- GroES binding to the *cis* ring causes a rigid body movement in the apical domain and causes the chamber to expand from 85,000$Å^3$ to 175,000$Å^3$. This movement results in the hydrophobic residues of the apical domains to move away from the cavity. As a result, the bound protein is displaced into the cavity and is encapsulated by GroES.

4- The next step involves the folding of the protein within the chamber and the hydrolysis of ATP to ADP in approximately 8-10 seconds. This hydrolysis causes the affinity of the *cis* ring of GroEL for GroES to drop. This hydrolysis also increases the affinity of the *trans* ring for ATP.

5- The binding of ATP molecules to the opposite ring causes the folded/semi folded protein and GroES to dissociate from the *cis* ring (Keskin *et al.*, 2002) (Fig 1.19).

6- If the released peptide is not in its native state it may interact with the same or another GroEL complex or even to a different chaperone system to fold.

This cycle can continue more than once to completely fold the polypeptide chain into its native form. By using this method, GroEL and the cochaperone GroES allow partially folded polypeptides to fold in a secluded environment free from the crowded internal environment of the cell. This would suggest that GroEL functions primarily as an Anfinsen cage (Horwich *et al.*, 2009, Apetri & Horwich, 2008, Horst *et al.*, 2007). However, since GroEL is also involved in assisting misfolded proteins to reach their native state (Chakraborty *et al.*, 2010, Lin *et al.*, 2008, Gupta *et al.*, 2010), it would suggest that GroEL is not just a passive Anfinsen cage but also helps unfold misfolded proteins before they can be folded back into its native state (Tehver & Thirumalai, 2008, Thirumalai & Lorimer, 2001). These studies along with others also favour the assisted folding model (Tang *et al.*, 2006, Radford, 2006, Marchenkov & Semisotnov, 2009). It could be argued that each of the models is substrate dependent, and that GroEL makes use of all three models depending on the substrate it needs to fold (Jewett & Shea, 2010). However, it is agreed upon that the folding of substrate proteins occurs in a central cavity that excludes the substrate from the crowded cellular environment.

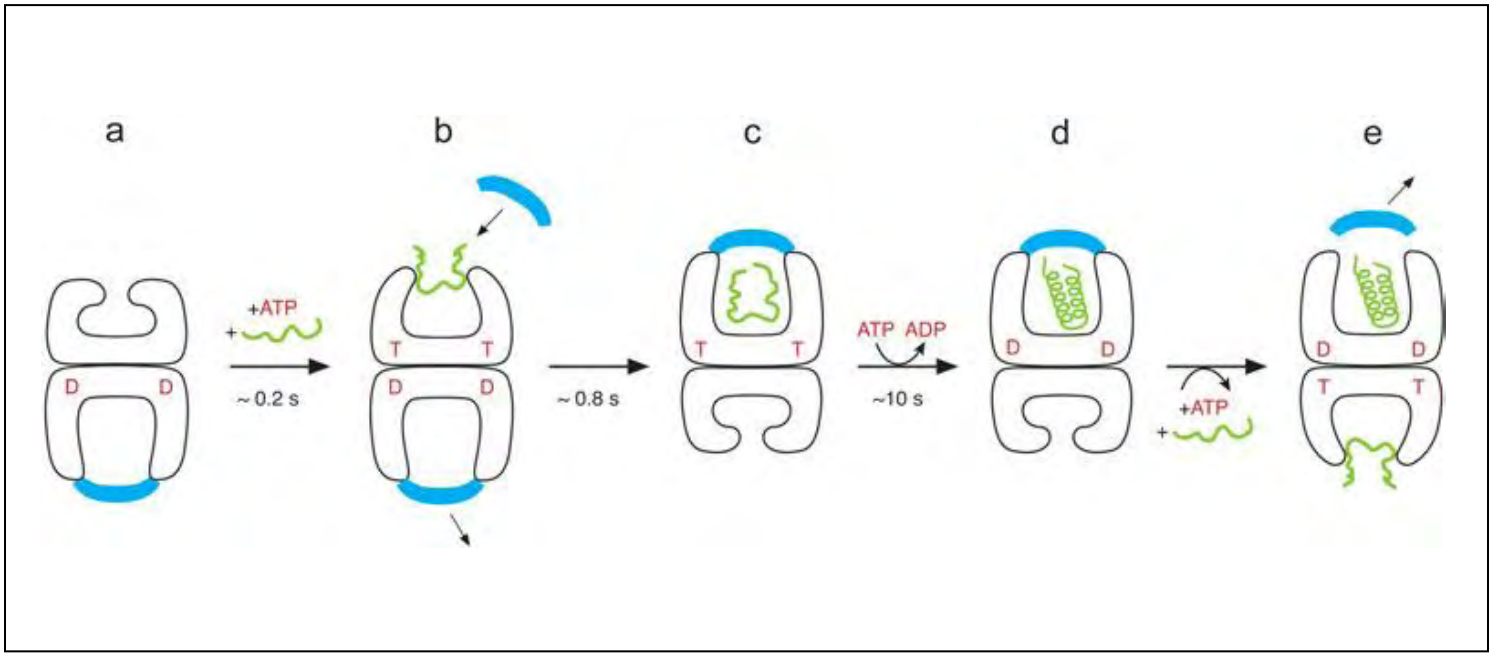


Fig 1.19: The protein folding mechanism of the GroEL-GroES complex (Horwich & Fenton, 2009). ATP binds to the apical domain of one of the rings of GroEL (a). It is then followed by substrate (green) and GroES (blue) (b). The substrate is then displaced by the binding of cochaperonin GroES and can fold within the central cavity without being influenced by other proteins (c). ATP hydrolysis at the original ring (d) allows ATP and substrate binding to occur on the 2nd ring (e).

1.5.4.3 GroEL substrates (client proteins)

GroEL allows the folding of a wide range of substrates in *E. coli*. It is the only chaperone system that is vital for cell survival under all growth conditions (Fayet *et al.*, 1989, Horwich *et al.*, 1993). Recent research conducted on *E. coli* has shed some light over the substrates that require the use of GroEL to fold and others that interact with it, but do not depend on it to reach their folded forms. It was shown using a combination of biochemical analysis and quantitative proteomics that around 250 proteins interact with GroEL, but only about 85 of those proteins were completely dependent on GroEL, and only 13 substrates out of those were vital for cell survival (Kerner *et al.*, 2005). This is still under active research and may not be accurate. The sizes of these substrates also vary greatly. Although the GroEL cavity can encapsulate substrates that are 60 kDa in size, larger proteins have also been shown to make use of the chaperonin folding mechanism by cycling on and off the GroEL ring in *trans* with GroES (Chaudhuri *et al.*, 2001). Experimental data has also shown that the loss of GroEL function in *E. coli* results in the accumulation of newly translated proteins (e.g. MetE) that were not present in the cell while GroEL was functional (discussed in further detail in section 3.1). It has also however been suggested that GroEL is responsible for the proper folding of a majority of the newly translated polypeptide and not just the limited number as suggested earlier (Chapman *et al.*, 2006). This suggestion came about when it was observed that in a temperature sensitive *E. coli* strain which affected the chaperonin GroEL, a biochemically isolated inclusion body of the bacterial cytoplasm contained a vast number of newly synthesised proteins (Chapman *et al.*, 2006).

1.5.5 The chaperonins of Mycobacteria

This thesis primarily focuses on the chaperonins of *Mycobacterium tuberculosis* with some work also being done on the chaperonins of *Mycobacterium smegmatis*. As such, this section highlights some of the reasons why the chaperonins of Mycobacteria are so interesting. Whenever possible, examples from both *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* have been used. This section also briefly looks into the pathogenic properties of *Mycobacterium tuberculosis*.

1.5.5.1 Multiple chaperonins of Mycobacteria

As discussed above, most bacteria only have one copy of the GroEL homologue; however it was shown using genome sequence analysis on 669 bacterial genomes that nearly 30% have 2 or more copies of the *groEL* gene (Lund, 2009). One of these bacteria is *Mycobacterium tuberculosis*, which is a major human pathogen. In 1987 Young *et al.*, showed the presence of a 65kDa antigenic protein which was initially referred to as "common antigen" (Young, 1987). This protein was later identified as Cpn60.2 along with the discovery of a second homologous gene which was termed *cpn60.1* (encoding chaperonin 60.1) (fig 1.20) (Kong *et al.*, 1993). *M. smegmatis* on the other hand has 3 copies of Cpn60 (figure 1.20). In both cases the cochaperonin *cpn10* gene is always in the same operon as *cpn60.1*, while *cpn60.2/cpn60.3* are present elsewhere in the genome. Even in Actinobacteria that only carry one *cpn60* gene the *cpn10* gene is always found separate in the genome (Maiwald *et al.*, 2003).

Recent work has suggested that *M. tuberculosis* chaperonins exist in lower oligomeric forms, and have weak ATPase activity along with partial activity with regards to substrate folding (Qamra *et al.*, 2004). Since the oligomeric state of a chaperonin in vital for its function, it is still unclear as to why they would exist as lower oligomers. Why these bacterial cells have evolved to contain two or more different forms of Cpn60 is also not fully understood, especially since GroEL can bind to a large variety of proteins without showing any signs of substrate specificity. One explanation could be that each copy is specialised for specific functions, especially those involved in cell survival in stressful environments. It was shown that the GroEL chaperonin could be forced to evolve to be more substrate optimised and be able to fold those substrates better than the WT chaperonin (Wang *et al.*, 2002). This optimisation however, came with the loss of the chaperonins ability to fully support cell growth. So specialised mycobacterial chaperonins could have arisen due to natural evolution of duplicated genetic copies, resulting in homologues that would be specialised in the folding of a different set of substrates. What is an interesting point to note is that in knock out experiments done to try and understand the function of these chaperonins, it was only possible to generate *cpn60.1* knockouts, as deleting either *cpn60.2* or *cpn10* resulted in cell death (Hu *et al.*, 2008). This experiment has posed further questions with regards to *cpn10* and *cpn60.1* being in the same operon since it would be logical to assume that the operon would consist of the house-keeping chaperonin and its cochaperone. Phylogenetic analysis has shown that the two Cpn60 proteins which are found in all Mycobacteria fall into two distinct phylogenetic groups, showing that the gene duplication which produced them was very ancient (Rao & Lund, 2010). Such is the difference, that Cpn60.1 from one Mycobacterial species is more similar to the Cpn60.1 from another Mycobacterial species, than it is to Cpn60.2 from own species (fig 1.20c). As discussed, this raises the possibility of the multiple Mycobacterial chaperonins being specialised for specific functions. Other

experiments have shown that both chaperonins as well as the cochaperonin are very potent immunomodulatory response factors (Henderson *et al.*, 2006).

Mycobacterium smegmatis is known to possess three copies of the GroEL homologue. Cpn60.1, referred to as GroEL1 in *M. Smegmatis* (Ojha *et al.*, 2005), has been shown to be required by the cells to synthesise mycolates during biofilm formation. The mycolates are long-chain fatty acids which are components of the mycobacterial cell wall. The loss of Cpn60.1 in *M. smegmatis* results in the cells losing the ability to form mature biofilms during growth, while the remaining growth pattern and the cell number remains unaffected (Ojha *et al.*, 2005). This seems to be related to loss of the ability of the cells to synthesise certain mycolic acids by the type II fatty acid synthase. It has been shown that Cpn60.1 interacts and associates with KasA, which is a key component of the FASII complex, and as such is directly responsible for the chaperoning of the FASII complex. It is still, however, unknown what parts of Cpn60.1 are responsible in this mechanism.

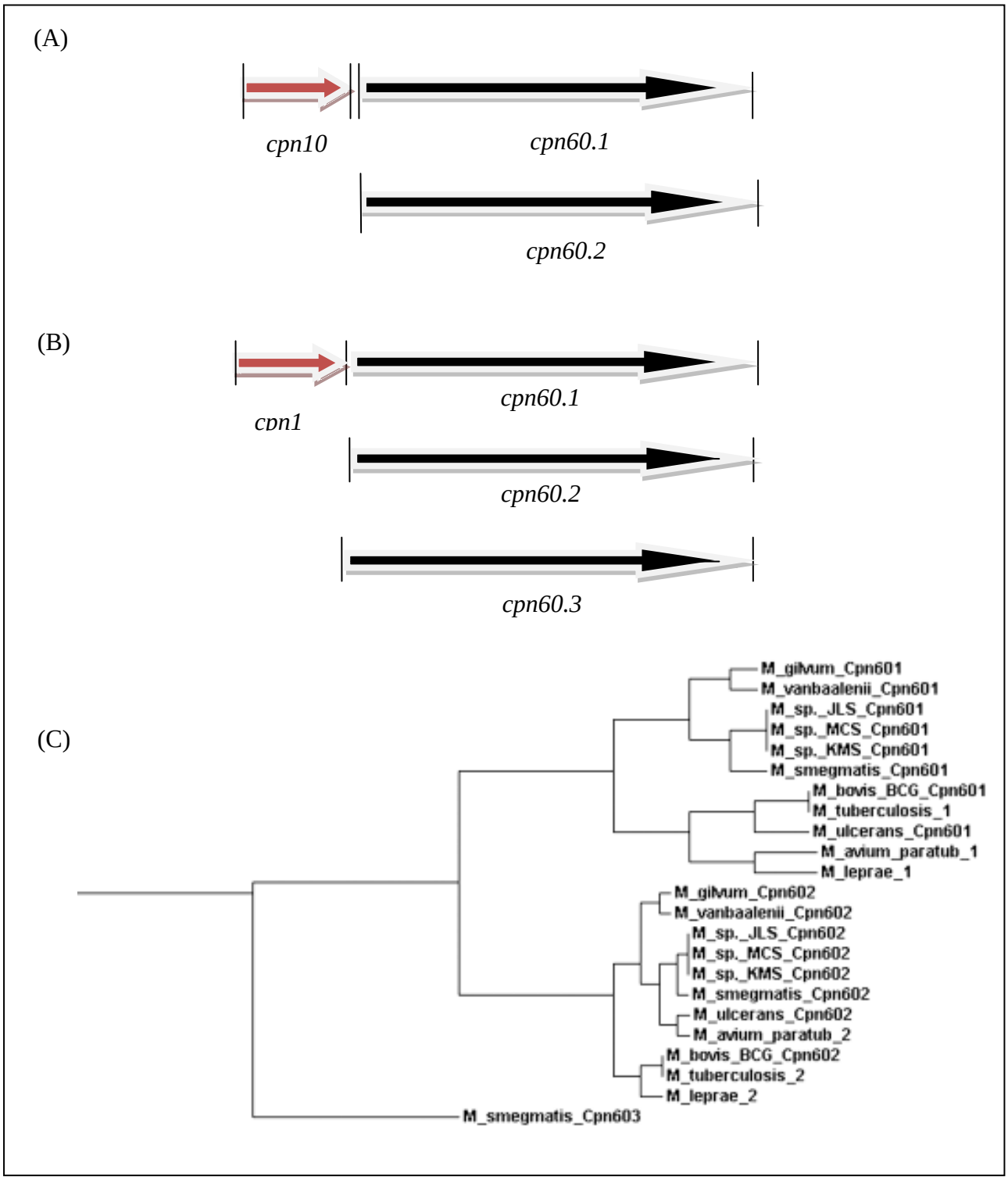


Fig 1.20: The Cpn60 homologues in *M. tuberculosis* (A) and *M. Smegmatis* (B). *Mycobacterium tuberculosis* has 2 Cpn60 homologues while *Mycobacterium smegmatis* has 3. In each case, the cochaperonin Cpn10 is in the same operon with Cpn60.1. (C) Phylogenetic distribution between Cpn60.1, Cpn60.2, and Cpn60.3 from various Mycobacteria (taken from Rao & Lund, 2010).

1.5.5.2 Pathogenic properties of *Mycobacterium tuberculosis* and the role of its chaperonins/cochaperonins as immunomodulatory response factors

Mycobacterium tuberculosis is a major human pathogen. It is spread through the air in droplets caused by coughing or sneezing by infected individuals. According to the world health organisation, approximately a third of the world population carries *M. tuberculosis*, with 10 million people being infected each year. Although only 5-10% of the newly infected individual will develop the disease, *M. tuberculosis* still causes approximately 1.5 million deaths a year. During its life cycle, the *M. tuberculosis* bacterium causes the formation of a cellular structure called the granuloma. This granuloma acts as a nest for the bacteria and is vital for the further spreading of the organism and in the causation of the disease. Research has shown that chaperonins of *Mycobacterium tuberculosis* are involved in more than just protein folding. Both the Cpn60s and the Cpn10 of *M. tuberculosis* are very potent immunomodulatory response factors and behave as antigens and cytokines (Qamra *et al.*, 2005). Immune responses caused by Cpn10 include T-cell proliferation as well as the production of antibodies (Young & Garbe, 1991). It has been shown that Cpn60.2 is present on the bacterial cell surface to facilitate efficient bacterial association with macrophages (Hickey *et al.*, 2009), and the Cpn60 proteins have been shown to be an immunodominant target for T-cell response in both humans and mice (Young, 1990). Cpn60.2 is one of the more potent immunogen in *M. tuberculosis*. During antibiotic treatment of TB, the concentration of antibodies specific to Cpn60.2 rises and persists in the cell for almost 18 months after the treatment (Sethna *et al.*, 1998). Cpn60.2 has also been used as a part of a DNA vaccine against *M. tuberculosis* (Lowrie, 2006). It may also play a role in autoimmunity, as there is an increased presence of Cpn60.2 specific antibodies for autoimmune diseases such as systemic sclerosis and rheumatoid arthritis (van Eden *et al.*,

1998). Cpn60.1 however has been shown to be a more potent cytokine than its homologue Cpn60.2 (Lewthwaite *et al.,* 2001). All the chaperonins of *M. tuberculosis* can also function as intracellular signals for a range of cells, such as myeloid cells (Maguire *et al.*, 2002). Research conducted in 1998 put forward the idea that the molecular chaperones behaved like virulence promoting factors as well as chaperoning the folding of proteins (Lewthwaite *et al.*, 1998).

Recent work has confirmed that the *cpn60.1* gene of *M. tuberculosis* is in fact not required by the cell for its survival at all even though it is in the same operon with *cpn10*, but rather it is in fact a virulence promoting factor. While *cpn60.2* and *cpn10* deletion resulted in cell death, the deletion of *cpn60.1* had no affect on the growth of the cells. It was only possible to delete the chromosomal *cpn60.2* and *cpn10* genes in the presence of a plasmid with the cloned chaperonin genes present. Various experiments including growing the cells under stressful conditions and growth in macrophages also showed that in all cases the deletion of *cpn60.1* had little to no affect on the cells survival capabilities (Hu *et al.*, 2008). One of the main phenotype of *cpn60.1* deletion was observed when *M. tuberculosis* strains lacking the *cpn60.1* gene was intravenously injected into immunocompetent mice. The results showed that while there was an initial difference in cell growth between the mutatnt and the wild type, with the mutant having fewer CFU at various weekly intervals, by 12 weeks there is was no significant difference between the two. It was however noticed that the mutant cells had lost the function to produce granulomatous inflammation in these mice and as such failed to cause classic tuberculosis (Hu *et al.*, 2008). This may in part explain the mechanisms by which *M. tuberculosis* causes TB and also makes the study of these chaperones a vital task.

1.5.5.3 Structural features of Mycobacterial chaperonins

Another factor which makes the Mycobacterial GroEL homologues interesting is their structure. The general structure of the GroEL proteins consists of a large and stable complex with a characteristic "double ring" structure of 14 subunits. Size exclusive chromatography and native PAGE has shown that the *M. tuberculosis* chaperonin may in fact have a natural tendency to exist as lower oligomers (Qamra *et al.*, 2004), and crystallise as a dimer (Qamra & Mande, 2004). Experiments using vectors to replace the *E. coli* GroEL chaperones showed that the complementation and expression levels of cpn60.1 are very weak, while the cpn60.2 is better in this regard (Liu H., 2007, PhD thesis, University of Birmingham). Since GroEL function is dependent on its structural stability, these results suggested that the *M. tuberculosis* Cpn60.2 must be able to form higher oligomeric structures to be able to complement for the loss of native GroEL. Another point that supports this idea is that Cpn10 has been shown to form heptameric structures (Roberts *et al.*, 2003). So there is a possibility that the presence of Cpn10 may provide a scaffold onto which the chaperonin subunits could form a higher oligomeric structure *in vivo*. Results obtained when Cpn60.2 was subjected to crowding agents to mimic *in vivo* environment conditions did show presence of higher oligomeric structures; however, it also suggested that these higher structures were unstable (Liu H., 2007, PhD thesis, University of Birmingham). Qamra and Mande also noted that the purified *M. tuberculosis* chaperonins did not show any signs of ATPase dependent activity (Qamra & Mande, 2004). Again this is unusual since the hydrolysis of ATP by the chaperonin GroEL is vital in the substrate folding reaction cycle and it has been shown that both Cpn60.1 and Cpn60.2 can function as chaperonins in an ATP dependent manner (Das Gupta *et al.*, 2008). Since the structure and ATP dependence of the GroEL proteins are essential for its mechanism of action, it is unclear why the Mycobacterial oligomers are much less stable and don't show signs of any ATPase activity *in vitro*.

1.6 Aims of the project

The main aim of this project is to better understand the properties of the multiple chaperonins of *Mycobacterium tuberculosis*. These properties include the functional and oligomeric properties of Cpn60.1 and Cpn60.2 and chimeras between them and *E. coli* GroEL. A list of the topics that are addressed in this project are as follows:

- Functional analysis of Mycobacterial chaperonins in *E. coli*. This was tested using protein expression and complementation assays.
- Functional analysis of the cochaperonin of *M. tuberculosis*. Since functional cochaperonins are a vital part of the chaperoning cycle, the ability of Cpn10 to function in *E. coli* was tested along with the chaperonins using complementation assays.
- Further characterisation of the proteins of *M. tuberculosis* by site directed mutagenesis and domain swaps experiments to see if the ability to oligomerise can be restored in *E. coli* (native gels and AUC), and the effect this has on the functional ability of the chaperonins (complementation assays).
- Testing the ATPase activity of purified functional chaperonins compared with GroEL, as the functional cycle of GroEL is dependent on its ATPase activity.
- Characterising the ability of *M. tuberculosis* Cpn60.1 to function in *M. smegmatis*. Since Cpn60.1 and Cpn60.2 in Mycobacteria may have evolved different functions, and the Cpn60.1 from one Mycobacteria is closely related to another Cpn60.1 from another Mycobacteria, the ability of Cpn60.1 from *M. tuberculosis* to complement for loss of Cpn60.1 in *M. smegmatis* was tested.

Chapter 2

Materials and Methods

Materials and Methods:

2.1 Bacterial strains

For this project the following bacterial strains were primarily used:

STRAIN	Genotype feature	Reference
E. coli DH5α	*F- gyr A96 (Nal^r) rec A1 relA1 endA1 thi-1 hsdR17 (r_k^- m_k^-) glnV44 deoR Δ (lacZYA-argF) U169 [φ80dΔ(lacZ) M15]*	Gibco- BRL
E. coli MGM100	MG1655 *kan^r nptII-pBAD-groESL*	McLennan *et al*., 1998
E. coli TG1	*supE hsdΔ5 thiΔ(lac-proAB)* F$^-$ *[traD36 proAB$^+$ lac1^q lacZΔ 15]*	Gibson, 1984
E. coli BL21 (DE3)	F– *ompT hsdS*B(rB– mB–) *gal dcm* (DE3)	Novagen
E. coli AI90	TG1 Δ*groEL::nptII*	Erbse *et al.*, 1999
E. coli XL1-Red	*endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutT* Tn*10* (Tetr)*a*	Agilent (formerly Stratagene)
Mycobacterium smegmatis MC^2 155	WT	Ojha *et al*., 2005

2.2 Plasmids

The pTrc expression plasmids used in this project were constructed by Dr. Yanmin Hu (St George's Hospital, London), and derived from the expression plasmid ptrc99A (fig 2.1). This plasmid has an amp^R selective marker, and contains a multiple cloning site downstream from an IPTG inducible *Ptrc* promoter.

Plasmid	Features/Genes Expressed	Reference
pTrc99A	An expression plasmid driven by the *trc* promoter (-35 region of the *trp* promoter together with the -10 region of the *lac* promoter), inducible by IPTG.	Amann *et al.*, 1988
pTrc-*groES-groEL*	A derivative of the pTrc plasmid containing *E.coli groEL* and *groES* genes under the control of the *Ptrc* promoter.	Lund, Unpublished
pTrc-*groES-cpn60.1*	A derivative of the pTrc plasmid containing *E.coli groES* gene along with *M. tuberculosis cpn60.1* gene under the control of the *Ptrc* promoter.	Lund, Unpublished
pTrc-*groES-cpn60.2*	A derivative of the pTrc plasmid containing *E.coli groES* gene along with *M. tuberculosis cpn60.2* gene under the control of the *Ptrc* promoter.	Lund, Unpublished
pTrc-*cpn10-cpn60.1*	A derivative of the pTrc plasmid containing *M. tuberculosis cpn10* gene and the *cpn60.1* gene under the control of the *Ptrc* promoter.	Lund, Unpublished
pTrc-*cpn10-cpn60.2*	A derivative of the pTrc plasmid containing *M. tuberculosis cpn10* gene and the *cpn60.2* gene under the control of the *Ptrc* promoter.	Lund, Unpublished
pTrc-*cpn60.2*	A derivative of the pTrc plasmid expressing *M. tuberculosis* Cpn60.2	This study
pBAD24	Expression plasmid driven by arabinose inducible P_{BAD} promoter	Guzman *et al.*, 1995
pRARE	Plasmid encoding tRNAs for codons that rarely occur in *E. coli*	Novagen
pET23a+	Expression plasmid driven by the T7 promoter	Novagen
pET-*groES-cpn60.1*	Derivative of pET23a+ expressing *E. coli* GroES and *M. tuberculosis* Cpn60.1	This study

pET-*cpn10-cpn60.1*	Derivative of pET23a+ expressing *M. tuberculosis* Cpn10 and Cpn60.1	This study
pSMGroEL1-H12	Expression plasmid containing *M. smegmatis* Cpn10 and Cpn60.1	Ojha *et al.*, 2005
pMs10-*mtbcpn60.1*	pSMGroEL1-H12 derivative expressing *M. smegmatis* Cpn10 and *M. tuberculosis* Cpn60.1	This study

2.2.1 Plasmid maps:

Plasmid map of pTrc99a

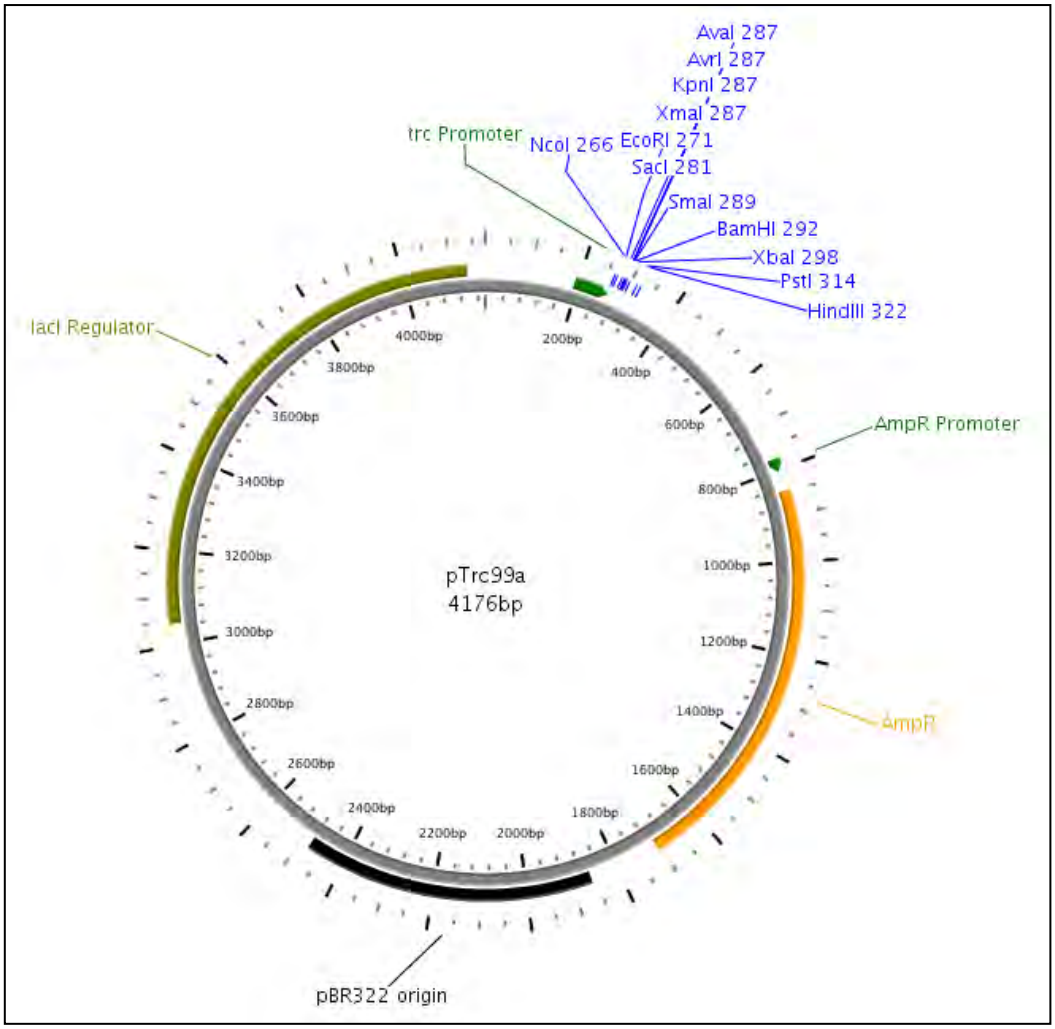


Fig 2.1: Plasmid map of pTrc99a. Image obtained using Plasmapper (Dong *et al.*, 2004)

Schematic representation of chaperonin Plasmid

Only a schematic representation of pTrc-GroES-GroEL has been shown. The remaining chaperonin and cochaperonin are interchangeable with GroEL and GroES. All the pTrc chaperonin plasmids have the chaperonins cloned into the MCS of pTrc99a. The remaining plasmids have a schematic representation presented in their respective results sections.

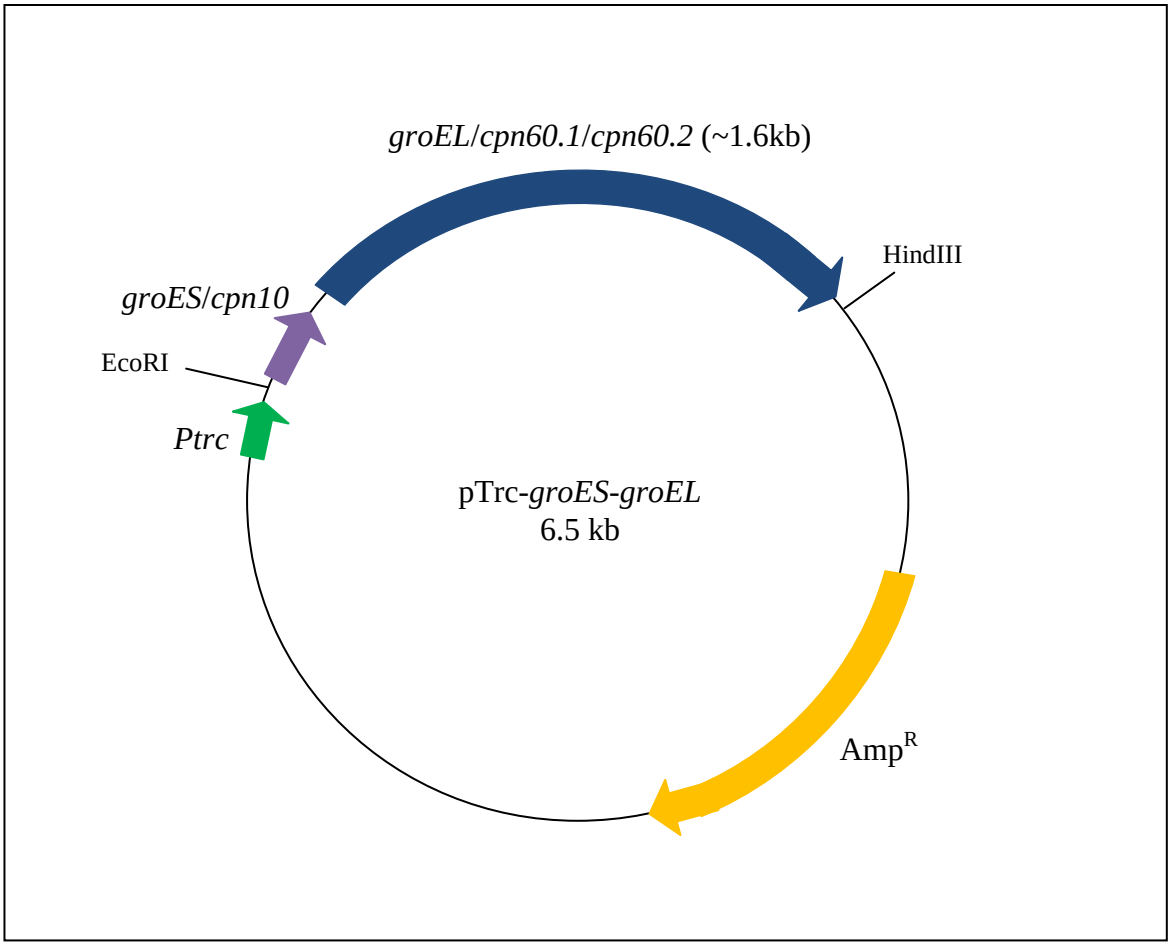


Fig 2.2: Schematic representation of pTrc-GroES-GroEL.

2.3 Primers: Below is a list of all the main primers that were used in this project. Primers that were only used once to correct mutations during the various experiments (such as in chapter 4) have not been listed.

2.3.1 Primers for sequencing

Name of Primer	Binding site on gene	Sequence (5′ to 3′)	Length (in bases)
Ptrc seq For	Within pTrc99a, upstream from both cochaperonin	GGA TAA TGT TTT TTG CGC	18
EL sequencing Prim1	*groEL* 470-488	CCG TAG GTA AAC TGA TCG C	19
60.2 sequencing Prim1	*cpn60.2* 438-454	CGC AGC GAT TTC GGC GG	17
60.1 sequencing Prim1	*cpn60.1* 472-489	GAC CTG GTT GGC GAA GCG	18
60.2 For1	126 bp upstream of *cpn60.2* (in *groES*)	TCC GAA AGC GAC ATT CTG G	19
60.2 SR Seq1	*cpn60.2* 331-349	AAC CCG CTC GGT CTC AAA C	19
60.2 SR Seq2	*cpn60.2* 1049-1068	AGA TCG AGA ACA GCG ACT CC	20
GroEL-SR Seq	*groEL* 1009-1030	GGT GAA GAA GCT GCA ATC CAG G	22
AI90 GroEL For	*groEL* 15-41	CGT AAA ATT CGG TAA CGA CGC TCG TGT	27
AI90 GroEL Rev	*groEL* 597-571	GTA GCC ACG GTC GAA CTG CAT ACC TTC	27
pET-60.1 Seq	Within pET23a+, 245 bp upstream from *groES*, 278 bp upstream from *cpn10*	GTC CTC AAC GAC AGG AGC ACG	21
pET-phos Rev	*cpn60.2* 602-585	TAC CCC GAG ATG TAG CCC	18

2.3.2 Primers for site directed mutagenesis

Name of Primer	Binding site on gene	Sequence (5' to 3')	Length (in bases)
ATG-cpn10 For	21 bp upstream of *cpn10* to 22 bp into *cpn10*	AAA TAG TGG AGG GCT CCA ATC ATG GCG AAG GTG AAC ATC AAG C	43
ATG-cpn10 Rev	22 bp in *cpn10* to 21 bp upstream of *cpn10*	GCT TGA TGT TCA CCT TCG CCA TGA TTG GAG CCC TCC ACT ATT T	43
Mut 60.21 For	*cpn60.2* 1332-1359	GGA GGC CCC GCT GGG GCA GAT CGC CTT C	28
Mut 60.21 Rev	*cpn60.2* 1359-1332	GAA GGC GAT CTG CCC CAG CGG GGC CTC C	28
Mut 60.22 For	*cpn60.2* 1362-1384	CTC CGG GCT GGC GCC GGG CGT GG	23
Mut 60.22 Rev	*cpn60.2* 1384-1362	CCA CGC CCG GCG CCA GCC CGG AG	23
Mut 60.23 For	*cpn60.2* 1368-1390	GCT GGA GCC GGC CGT GGT GGC CG	23
Mut 60.23 Rev	*cpn60.2* 1390-1368	CGG CCA CCA CGG CCG GCT CCA GC	23
Mut 60.24 For	*cpn60.2* 1371-1393	GGA GCC GGG CGC GGT GGC CGA GA	23
Mut 60.24 Rev	*cpn60.2* 1393-1371	TCT CGG CCA CCG CGC CCG GCT CC	23
Mut EL.1 For	*groEL* 1342-1367	GAA GCT CCG CTG GAA CAG ATC GTA TT	26
Mut EL.1 Rev	*groEL* 1367-1342	AAT ACG ATC TGT TCC AGC GGA GCT TC	26
Mut EL.2 For	*groEL* 1371-1393	CTG CGG CGA AGC ACC GTC TGT TG	23
Mut EL.2 Rev	*groEL* 1393-1371	CAA CAG ACG GTG CTT CGC CGC AG	23
Mut EL.3 For	*groEL* 1375-1401	GGC GAA GAA CCG GCT GTT GTT GCT AAC	27

Mut EL.3 Rev	*groEL* 1401-1375	GTT AGC AAC AAC AGC CGG TTC TTC GCC	27
Mut EL.4 For	*groEL* 1378-1404	GAA GAA CCG TCT GCT GTT GCT AAC ACC	27
Mut EL.4 Rev	*groEL* 1404-1378	GGT GTT AGC AAC AGC AGA CGG TTC TTC	27

Highlighted regions represent specific changes to the original sequences that were used for site directed mutagenesis.

2.3.3 Primers for domain swaps

Name of Primer	Binding site on gene	Sequence (5' to 3')	Length (in bases)
60.1 flanking fwd 1	*cpn60.1* 1-25	ATG AGC AAG CTG ATC GAA TAC GAC G	25
2F-60.1-60.2	*cpn60.1* 427-447	ACC AAG GAG CAG ATC GCG CAG GTG GCG ACG GTG	33
3F-60.1-60.2	*cpn60.1* 1207-1227	CGC AAT GCC AAG GCC GCG GTC GAG GAG GGC ATC	33
4R-60.1-60.2	*cpn60.1* 429-409	TGC GGT GGC CGC GAT GCC GGT CTT GCC GGA CAC	33
5R-60.1-60.2	*cpn60.1* 1209-1189	CTC CTC GAC GGC GGC CTT GGC GGC CGC GAC CGC	33
60.1 flanking rev 6	*cpn60.1* 1620-1600	TCA GTG CGC GTG CCC GTG GTG	21
60.2 flanking fwd 7	*cpn60.2* 1-25	ATG GCC AAG ACA ATT GCG TAC GAC G	25
8F-60.2-60.1	*cpn60.2* 427-447	GGC AAG ACC GGC ATC GCG GCC ACC GCA GCG ATT	33
9F-60.2-60.1	*cpn60.2* 1207-1227	GCG GCC GCC AAG GCC GCC GTC GAG GAG GGC ATC	33
10R-60.2-60.1	*cpn60.2* 429-409	CGC CAC CTG CGC AAT CTG CTC CTT GGT CTC GAC	33

11R-60.2-60.1	*cpn60.2* 1209-1189	CTC CTC GAC CGC GGC CTT GGC ATT GCG AAC CGC	33
60.2 flanking rev 12	*cpn60.2* 1623-1599	TCA GAA ATC CAT GCC ACC CAT GTC G	25
EL flanking fwd 13	*groEL* 1-25	ATG GCA GCT AAA GAC GTA AAA TTC G	25
14F-EL-60.1	*groEL* 430-450	GGC AAG ACC GGC ATT GCT CAG GTT GGT ACC ATC	33
15F-EL-60.1	*groEL* 1213-1233	GCG GCC GCC AAG GCT GCG GTA GAA GAA GGC GTG	33
16R-EL-60.1	*groEL* 432-412	CGC CAC CTG CGC AAT CGC TTT AGA GTC AGA GCA	33
17R-EL-60.1	*groEL* 1215-1195	CTC CTC GAC CGC AGC ACG GGT CGC GTG CAG GGC	33
EL flanking rev 18	*groEL* 1647-1623	TTA CAT CAT GCC GCC CAT GCC ACC C	25
19F-60.1-EL	*cpn60.1* 427-447	GAC TCT AAA GCG ATT GCG CAG GTG GCG ACG GTG	33
20F-60.1-EL	*cpn60.1* 1207-1227	CAC GCG ACC CGT GCC GCG GTC GAG GAG GGC ATC	33
21R-60.1-EL	*cpn60.1* 429-409	ACC AAC CTG AGC GAT GCC GGT CTT GCC GGA CAC	33
22R-60.1-EL	*cpn60.1* 1209-1189	TTC TTC TAC CGC AGC CTT GGC GGC CGC GAC CGC	33
23F-60.2-EL	*cpn60.2* 427-447	GAC TCT AAA GCG ATT GCG GCC ACC GCA GCG ATT	33
24F-60.2-EL	*cpn60.2* 1207-1227	CAC GCG ACC CGT GCC GCC GTC GAG GAG GGC ATC	33
25R-60.2-EL	*cpn60.2* 429-409	ACC AAC CTG AGC AAT CTG CTC CTT GGT CTC GAC	33
26R-60.2-EL	*cpn60.2* 1209-1189	TTC TTC TAC CGC GGC CTT GGC ATT GCG AAC CGC	33
27F-EL-60.2	*groEL* 430-450	ACC AAG GAG CAG ATT GCT CAG GTT GGT ACC ATC	33
28F-EL-60.2	*groEL* 1213-1233	CGC ATT GCC AAG GCT GCG GTA GAA GAA GGC GTG	33

29R-EL-60.2	*groEL* 432-412	TGC GGT GGC CGC AAT CGC TTT AGA GTC AGA GCA	33
30R-EL-60.2	*groEL* 1215-1195	CTC CTC GAC GGC GGC ACG GGT CGC GTG CAG GGC	33
31F-60.2-EL pre	*cpn60.2* 571-591	CTG GAC GTG GTT ATG CGG TTC GAC AAG GGC TAC	33
32F-60.2-EL pre	*cpn60.2* 1126-1146	CTG GCA GGC GGC GTG ATC AAG GCC GGT GCC GCC	33
33R-60.2-EL pre	*cpn60.2* 573-553	GAA CTG CAT ACC CAT ACC CTC GGT GAG CTC GAG	33
34R-60.2-EL pre	*cpn60.2* 1128-1108	TTT GAT AAC TGC CAC CGC GAC ACC ACC GGC CAG	33
35F-EL-60.2 pre	*groEL* 571-591	CTC ACC GAG GGT GAA GGT ATG CAG TTA GAC CGT	33
36F-EL-60.2 pre	*groEL* 1126-1146	GGT GGT GTC GCG GTT GCA GTT ATC AAA GTG GGT	33
37R-EL-60.2 pre	*groEL* 573-553	CTT GTC GAA CCG TTC AAC CAC GTC CAG TTC GTC	33
38R-EL-60.2 pre	*groEL* 1128-1108	ACC GGC CTT GAT AAC GCC GCC TGC CAG TTT CGC	33
39F-EL-60.1 pre	*groEL* 571-591	TTC ACC GAG GGT GAA GGT ATG CAG TTC GAC CGT	33
40F-EL-60.1 pre	*groEL* 1126-1146	GGC GGG GTT GCT GTT GCA GTT ATC AAA GTG GGT	33
41R-EL-60.1 pre	*groEL* 573-553	CTT GTC GAA GCC TTC AAC CAC GTC CAG TTC GTC	33
42R-EL-60.1 pre	*groEL* 1128-1108	ACC CAC CTT GAT AAC GCC GCC TGC CAG TTT CGC	33
43F-60.1-EL pre	*cpn60.1* 571-591	CTG GAC GTG GTT ATC GGC TTC GAC AAG GGC TTC	33
44F-60.1-EL pre	*cpn60.1* 1126-1146	CTG GCA GGC GGC GTC ATC AAG GTG GGT GCC GCC	33
45R-60.1-EL pre	*cpn60.1* 573-553	GAA CTG CAT ACC GAT ACC CTC GGT GAA CTC CAA	33
46R-60.1-EL pre	*cpn60.1* 1128-1108	TTT GAT AAC TGC GAC AGC AAC CCC GCC GGC CAG	33

47F-60.2-60.1 pre	*cpn60.2* 571-591	TTC ACC GAG GGT ATG CGG TTC GAC AAG GGC TAC	33
48F-60.2-60.1 pre	*cpn60.2* 1126-1146	GGC GGG GTT GCT GTG ATC AAG GCC GGT GCC GCC	33
49R-60.2-60.1 pre	*cpn60.2* 573-553	CTT GTC GAA GCC CAT ACC CTC GGT GAG CTC GAG	33
50R-60.2-60.1 pre	*cpn60.2* 1128-1108	ACC CAC CTT GAT CAC CGC GAC ACC ACC GGC CAG	33
51F-60.1-60.2 pre	*cpn60.1* 571-591	CTC ACC GAG GGT ATC GGC TTC GAC AAG GGC TTC	33
52F-60.1-60.2 pre	*cpn60.1* 1126-1146	GGT GGT GTC GCG GTC ATC AAG GTG GGT GCC GCC	33
53R-60.1-60.2 pre	*cpn60.1* 573-553	CTT GTC GAA CCG GAT AAC CTC GGT GAA CTC CAA	33
54R-60.1-60.2 pre	*cpn60.1* 1128-1108	ACC GGC CTT GAT GAC AGC AAC CCC GCC GGC CAG	33

2.4 General microbial techniques

2.4.1- Media: All media was autoclaved before use.

Luria-Bertani (LB): 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% NaCl.

LB agar: LB with 1.5% agar.

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

Difco Middlebrook 7H10 agar (BD Bioscience): 19g was added to 900ml water and autoclaved. 100ml Middlebrook ADC enrichment and 0.05% Tween80was added before use.

Biofilm base media: 13.6g KH_2PO_4, 2g $(NH_4)_2SO_4$ adjust the pH to 7.2 using KOH pellets then added 0.5mg $FeSO_4.7H_2O$, and 5g Casamino acid.

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

2.4.2- Additives

Additive	Stock concentration	Working concentration
L-Arabinose	20% (w/v)	0.2% (w/v)
D-Glucose	20% (w/v)	0.2% (w/v)
Ampicillin	100mg/ml	100µg/ml
Kanamycin	50mg/ml	50µg/ml
Chloramphenicol	50mg/ml	50µg/ml
Hygromycin	150mg/ml	150µg/ml
IPTG	1 M	1 mM

2.4.3- Growth conditions

E. coli growth conditions

E. coli was grown at 37°C, unless indicated otherwise, in Luria-Bertani (LB) media in a shaking incubator (180rpm). For overnight cultures, single colonies were picked from a plate or a loop full of frozen culture and inoculated into 5ml of LB. To subculture, 1ml of the

overnight culture was mixed into 49ml of LB in a 250ml conical flask. The flasks were shaken in the shaking incubator at 180rpm. LB agar plates were used for solid media experiments and were incubated at different temperatures. Incubation time varied depending on the strain used. In all cases appropriate amounts of additives were added depending on the requirement of the strains.

M. smegmatis growth conditions

M. smegmatis was grown at 37°C, unless states otherwise, in 5ml Difco Middlebrook 7H9 broth with ADC enrichment and 0.05% Tween 80. The cultures were grown for 36 hours with appropriate additives and aeration at 200rpm in a shaking incubator (Infors AG CH-4103). Plate cultures were incubated for 48 hours at 37°C on Difco Middlebrook 7H10 media with ADC, 0.05% Tween80 and appropriate additives.

2.4.4- Glycerol stocks

Glycerol stocks of all strains were made by mixing 875µl of fresh overnight culture with 125µl of 80% glycerol

2.4.5- Bacteriophage P1 Transduction

Reagents required

0.5M $CaCl_2.2H_2O$ (calcium chloride)

0.9M $Na_3C_6H_5O_7$ (sodium citrate)

1M $MgSO_4.7H_2O$ (magnesium sulphate)

0.1M $MgSO_4$/ 5 mM $CaCl_2$

Sloppy agar – 0.6% LBA with 2.5 mM $CaCl_2$

Preparation of bacteriophage P1 lysate

1 ml of fresh overnight (donor strain) was mixed with a pre-existing stock of bacteriophage P1 lysate at a range of dilutions ($10^0 – 10^{-5}$). $CaCl_2$ was then added at a final concentration of 2.5 mM and the solution was incubated at 37°C for 30 minutes. This was then added to 3ml of sloppy agar (autoclaved 0.6% agar in LB) and poured onto labelled LB plates for overnight incubation at 37°C. The plates were examined after the incubation period (roughly 14-16 hours) and the plates where the plaques were about touching but where lysis had not cleared the whole plate was picked. The sloppy agar was scraped off into a Falcon tube. The pate was also washed with 2ml LB/2.5 mM $CaCl_2$, and the washate was also added to the Falcon tube. The tube was then centrifuged at 10,000 rpm for 10 minutes and the supernatant was filtered though a 0.2 micron filter. The resultant suspension was stored at 4°C as the bacteriophage P1 lysate.

The titre of the bacteriophage P1 lysate was checked by spotting 10-fold serial dilutions of the prepared lysate onto a lawn of the recipient strain. The plates were then incubated at 37°C. The following day, the number of individually visible plaques was counted to give an estimate of the titre of the bacteriophage P1 lysate. Absence of plaques with certain strains indicates that the strain is not a recipient for the phage.

Transduction of recipient strain

The recipient strain was grown overnight in LB media with 2.5 mM $CaCl_2$ without shaking and then 1ml was spun down at 13000rpm for 5 minutes. The resultant pellet was then resuspended in 1ml 0.1M $MgSO_4$/ 5mM $CaCl_2$ and incubated at 37°C for 15 minutes with gentle shaking. 100µl of this suspension was mixed with neat, 10x, and 100x dilutions of 100µl bacteriophage P1 and incubated for a further 20 minutes at 37°C. To these cultures

200µl of 0.9M $Na_3C_6H_5O_7$ was added and the resultant suspension was spun at 13000rpm for 5 minutes. The supernatant was discarded and the pellet was then resuspended in 100µl LB and incubated at 37°C for 1 hour. Various dilutions, including neat, were then plated out on selective media plates e.g. Ampicillin plates to test for Ampicillin resistant gene transfer. Control plates were also used and included, plates with the recipient strain on its own (not transduced), as well as the bacteriophage P1 lysate on its own. These plates were incubated at 37°C for 1-2 days. The resulting colonies were screened by colony PCR.

2.4.6- Plasmid purification (Mini-prep)

Two methods of miniprep plasmid extraction were used during the project to extract plasmid DNA from *E.coli*.

Alkaline lysis method

Reagents required:

- GTE (50 mM glucose, 25 mM TRIS pH8.0, 10 mM EDTA) – autoclaved before use
- 1% SDS/0.2M NaOH (from stock of 10% SDS and 10M NaOH)
- 3M Na/5M acetate
- TE (10 mM TRIS pH8.0, 1 mM EDTA) – autoclaved before use
- Isopropanol
- 70% ethanol (in TE)
- 100% ethanol

1-5mls of overnight culture was centrifuged at 13000rpm for 2 minutes and the supernatant was poured off. The culture was then spun again for 20-30 seconds and the remaining LB was removed using a Gilson pipette. This is an important step as the left over LB can carry components into the prep, which can inhibit restriction enzymes and ligases. The pellet was

then resuspended in 200µl GTE by pipetting or vortexing. 400µl 1%SDS/0.2M NaOH was added to lyse the cells (tube inverted several times). 300µl of 3M Na/5m acetate was added to the solution which was then inverted several times until a white precipitate was formed. The tube was then incubated for 5 minutes at room temperature and then spun at 13000rpm for a further 5 minutes. The resulting supernatant, which contains the DNA plasmid, was poured into a clean labelled Eppendorf and 600µl isopropanol was added (and mixed very thoroughly). This tube was then cooled at -20°C for 30 minutes before being spun down at 13000rpm for 10 minutes to form a pellet of nucleic acid at the bottom of the Eppendorf. The pellet was then washed twice with 500µl of 70% ethanol and then again with 500µl of 100% ethanol (ensuring the pellet was not disturbed). The pellet was then dried at 37°C for 5-10 minutes and resuspended in 50µl TE.

Kit method

Plasmid DNA was purified from *E.coli* cultures using the Qiagen Qiaprep spin miniprep kit as per the manufacturers recommended guidelines. 5ml of overnight culture was spun down at 4000rpm and 4°C for 15 minutes. The pellet was resuspended in 250µl of P1 buffer solution and transferred to a micro centrifuge tube. The cells were then subjected to alkaline lysis for 5 minutes by adding 250µl of P2 solution, which degrades bacterial cell walls and membranes. After this, the reaction was neutralised by the addition of 350µl of N3 buffer. The cell debris was then spun down by centrifugation at 13000 rpm for 5 minutes. The supernatants were then added to a Qiaprep spin column and spun at 13000 rpm for 1 minute. This results in the plasmid DNA binding to the column. The plasmid was then washed in ethanol and eluted in 50µl of MilliQ water.

2.4.7- Plasmid maxi prep

This method of plasmid prep was used on large volumes of culture to extract a large amount of plasmid. Plasmid maxi prep was done using the GenElute plasmid maxiprep kit by Sigma-Aldrich as per the manufacturer's guidelines. A total of 50ml of overnight culture was centrifuged at 5000rpm for 10 minutes obtain the cell pellet. The pellet was resuspended in 6ml resuspension solution provided with the kit and transferred into a Falcon tube. To this 6ml of lysis solution was added, mixed thoroughly, and allowed to clear for 5 minutes. Once the solution had cleared, 8ml of neutralization solution was added and the tube was inverted 4-6 times. This was then centrifuged at 15000g for 15 minutes. The resultant supernatant was then transferred to the maxi spin column in a 50ml collection tube and spun in a swing bucket rotor at 4000g for 2 minutes. The flow through was discarded, 15ml of washing solution was added to the column and the tubes were spun again at 4000g for 5 minutes. The column was then transferred to a clean collection tube, 5ml of elution buffer was added, and the tube was spun at 4000 g for a further 5 minutes. The flow through obtained via this method contained the plasmids from the overnight cultures and was stored at -20° C.

2.4.8- Restriction digests of plasmids

For the restriction digest of a plasmid, approximately 300ng of plasmid DNA was used. For all digests the restriction enzymes and buffers used were either from New England Biolabs or Fermentas (20,000 units/ml) and approximately 2000 units of restriction enzyme was used in each case for complete digestion. The following contents were used to make a total volume of 20µl:

- Milli Q water 5-15 µl
- NEBuffer (depending on the restriction enzyme) 2 µl

- 10X BSA (depending on the restriction enzyme) 2 µl
- Restriction enzyme 1 µl
- Plasmid DNA approximately 300ng

The solution was then incubated at 37° C for 1-2 hours.

2.4.9- Blunt ending with Klenow

During the course of this project several ligation procedures involved the use of blunt ended DNA fragments. In cases where the restriction digest had not yielded a blunt ended product, Klenow (from Invitrogen) was used. The stock concentration of Klenow was 0.5U/µl. The reaction was mixed in an Eppendorf as follows:

- 10X React 2 Buffer 3µl
- 0.5 mM each dNTP (2 mM total) 4µl
- DNA 0.5-1µg
- Klenow 1µl
- SDW Made up to 30µl

The Eppendorf was gently mixed with a pipette and briefly centrifuged to collect the contents at the bottom of the tube. The tube was then incubated at room temperature for 10-15 minutes. The reaction was then terminated by incubation at 70°C for 10 minutes, and concentrated using ethanol precipitation as described below.

2.4.10- Ethanol precipitation

DNA was taken in an Eppendorf tube. 0.1 volumes of un-buffered 3M NaAc and 2.5 volumes of isopropanol were added. The content was gently mixed and incubated at -20°C for 20 minutes. After incubation, the tube was centrifuged at 13,000rpm for 10 minutes and the supernatant was discarded. The resulting pellet was washed twice with 70% ethanol and

100% ethanol before being dried at 37°C for 10-15 minutes. The dry pellet was then resuspended in 50µl SDW.

2.4.11- Ligation

The vector backbone and the DNA insert were usually mixed in a 1:3 ratio and treated with 1µl QuickStick DNA ligase (Bioline). The reaction included 5µl of 4X ligase buffer and was made up to 20µl with SDW. The mixture is incubated at 4°C overnight in a water bath. 10µl of the reaction is then transformed into competent *E. coli* cells.

2.4.12- Agarose gel electrophoresis

DNA (including plasmids, and digests) were run on 1% Agarose gel with SYBR-safe DNA gel stain (Invitrogen) at 5µl per 100 ml agarose. To 20µl of DNA, 5µl of loading buffer (0.25% bromophenol blue, 0.25% Xylene cyanol FF, 15% Ficoll type 400 [Pharmacia], made up using 10ml sterile water) was added, and 10-15µl was finally loaded on the gel. Bioline Hyperladder I was used as the marker for these gels. The gel was then run in a tank of 1X TAE at 90V for 1 hour (for a 10 well gel) or at 110V for 1.5 hrs (for a 24 well gel).

2.4.13- PCR

Colony PCR and gene amplification

The PCR mix for screening and gene amplification is as follows:

- 10 NH_4 reaction buffer 5µl
- 50 mM $MgCl_2$ 1.5µl
- dNTPs (100 mM stock) 0.5µl
- Primers (forward and reverse) 1-2µl
- Taq DNA polymerase (Fermentas) 0.5µl

- Colony/culture 1µl
- SDW made up to 50µl

The samples were PCR-amplified in a Eppendorf vapo.protect thermocycler under the following conditions:

Segment	Cycles	Temperature	Time
1	1	94°C	4 minutes
2	30	94°C	45 seconds
		55°C	1 minute
		72°C	1-5 minutes
3	1	72°C	10 minutes

Alternatively, the Thermo Scientific ReadyMix master mix was also used (23µl with 0.5µl of each primer and 1µl culture). The cycling conditions were the same as that listed above.

Site directed mutagenesis

The QuickChange Site-Directed Mutagenesis Kit (Agilent, formerly Stratagene) was used as per the manufacturer's instructions. The PCR protocol used is as follows:

- 10X reaction buffer 5µl
- DNA 50ng of dsDNA template
- Primers 12.5pmole (for each primer)
- dNTP mix 0.8 mM (final concentration)
- SDW made up to 50µl
- *PfuTurbo*/*PfuUltra* DNA polymerase 1µl

The PCR cycling parameters used were as follows:

Segment	Cycles	Temperature	Time
1	1	94°C	30 seconds
2	30	94°C	45 seconds
		55°C	1 minute
		72°C	1 minute/Kb
3	1	72°C	10 minutes

Following PCR, the samples were put on ice for 2 minutes to cool the reaction to ≤ 37°C. 1µl of *Dpn* I was added and the reaction was incubated at 37°C for 1 hour to allow the enzyme to digest the parental DNA plasmid. The digested product was then transformed into DH5α and plated out onto LB plates (with appropriate additives).

PCR in the construction of chimeric proteins

In chapter 4, various chimeric proteins were constructed by using a modified PCR extension technique. The details of the process have been outlined in the experimental design of chapter 4. The DNA polymerase and buffers for these reactions were from Finnzymes. The lists of additives in the PCR mix and the cycles have been listed below:

- 5X Phusion HF buffer 10µl
- 10 mM dNTPs 1µl
- Primer A 0.5 mM
- Primer B 0.5 mM
- Template DNA 20ng
- DMSO 1.5µl
- Phusion DNA Polymerase 0.5-1µl
- SDW Made up to 50µl

The PCR cycling parameters used were as follows:

Segment	Cycles	Temperature	Time
1	1	94°C	30 seconds
2	25	94°C	30 seconds
		65°C	1 minute
		72°C	1 minute/Kb
3	1	72°C	10 minutes

2.4.14- PCR purification

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

2.4.15- Competent cell preparation

E. coli competent cell preparation

Two different methods were used to prepare competent cells in this project for *E. coli*. Both methods have been detailed below. For the cells used in chapter three, method 1 was used, while method 2 was used for the remaining chapters.

Method 1

Solution 1 (TfbI): 30 mM CH_3COOK (potassium acetate)

10 mM RbCl (Rubidium chloride)

10 mM $CaCl_2$ (Calcium chloride)

50 mM $MnCl_2$ (manganese chloride)

15% glycerol (v/v)

The solution was made up in distilled water and the pH was adjusted to 5.8 with 0.2M acetic acid. Solution was filter sterilised before use.

Solution 2 (TfbII): 10 mM MOPS

75 mM $CaCl_2$

10 mM RbCl

15% glycerol (v/v)

The solution was made up in distilled water and the pH was adjusted to 6.5 with 0.1M KOH (potassium hydroxide). Solution was filter sterilized before use.

A flask with 100ml of pre-warmed LB was inoculated with 1ml of overnight culture and grown in a shaking water bath at 37°C till an OD_{600} of 0.4 was obtained. The flask was then placed on ice for 5 minutes. The culture was then spun down as 50ml aliquots at 10,000 rpm for 10 minutes. The supernatant was discarded and the pellet was gently resuspended in 25ml of TfbI. This suspension was then incubated on ice for 10 minutes and then spun down at 6000 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 solution were made and stored at -80°C.

Method 2

A flask with 100ml of pre-warmed LB was inoculated with 1ml of overnight culture and grown in a shaking water bath at 37°C till an OD_{600} of 0.4 was obtained. The culture was then transferred to two 50ml Falcon tubes and placed on ice for 15 minutes. The cells were then recovered by centrifugation at 4100rpm for 10 minutes (at 4°C). The resulting supernatant was disposed of and each pellet was resuspended in 30ml ice cold 0.1M $CaCl_2$. The Falcon

tubes were placed on ice for a further 15 minutes before the cells were recovered again by centrifugation at 4100rpm for 10 minutes (at 4°C). The supernatant was discarded and the pellets were resuspended in 2ml ice cold 0.1M $CaCl_2$ and 15% glycerol. 100µl aliquots of the solution were made and stored at -80°C.

M. smegmatis competent cell preparation (electrocompetent)

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. For a 100ml starting volume of culture, 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.

2.4.16- Transformation

E. coli transformation

To 100µl of competent cells (thawed on ice) approximately 100ng of plasmid DNA was added and incubated for 30 minutes on ice. The cells were then subjected to heat shock by placing them in a 42°C water bath for 30 seconds, followed by incubation on ice for 2 minutes. To this, 800µl of liquid LB was added and incubated at 37°C for 1 hour. The resultant solution was then plated out onto LB plates at a dilution of 10^{-1} and 10^{0}. The plates were incubated at 37°C overnight.

M. smegmatis transformation (electroporation)

1µg salt free DNA was added to 400µL of electrocompetent 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 Complementation analysis

With *E. coli* (spot test)

MGM100 was transformed with the pTrc plasmids carrying the chaperonin and cochaperonin genes. A single colony of the transformed strain was inoculated into 5ml LB broth (with appropriate additives) and grown over night at 37°C. The following morning it was sub-cultured and grown in LB + glucose for 2 hours at 37°C, and then serially diluted (in ten-fold steps) to spot on to square petri dishes (Sterilin). 5µl of each dilution was spotted onto the plates with the final dilution factor being $10^{7.}$ The plates were then incubated at 37°C, 30°C, and 26°C.

With *M. smegmatis* (biofilm formation)

M. smegmatis mc^2155 competent cells were transformed with the appropriate pMS10 plasmids. A single colony from the resulting strain was inoculated into 3ml Difco Middlebrook 7H9 broth with ADC, 0.05% Tween80, and appropriate antibiotics. This was incubated at 37°C in a shaking incubator for 24-36 hours. 10ml of the biofilm media was aliquoted into a 60 mm X 15 mm Petri dish. To this media, 10µl of the culture was added and kept stationary at 30°C for 4-7 days.

2.6 Protein analysis

2.6.1- Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

Reagents required:

- 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 MilliQ water and 500µl of β-Mercaptoethanol was added.

- Electrophoresis buffer: A stock solution of 5X electrophoresis buffer (1L) was made with 15.1g of TRIS, 94g of glycine, and 5g of SDS made up to 1L in MilliQ water. When running the gels a solution of 1X electrophoresis buffer was used. This comprised 200µl 5x buffer added to 800µl MilliQ water.

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

- Destaining solutions: This comprised 300ml Methanol and 100ml glacial acetic acid made up to 1L in MilliQ water.

Gel preparation:

Separating Gel				**Stacking Gel**	
1 Gel	**2 Gels**			**1 Gel**	**2 Gels**
2.3ml	4.6ml	MilliQ		1.35ml	2.7ml
1.25ml	2.5ml	30% Acrylamide		250µl	500µl
1.35ml	2.7ml	80% Glycerol		335µl	670µl
50µl	100µl	(pH 8.8) Tris Buffer (pH6.8)		20µl	40µl
50µl	100µl	10% APS		20µl	40µl
50µl	100µl	10% SDS		20µl	40µl
3µl	6µl	TEMED		2µl	4µl

SDS-PAGE gel sample preparation and SDS analysis: 1ml of overnight culture at an OD_{600} of 0.5 was spun down at 13000 rpm for 5 minutes and the pellet was washed twice (by spinning and resuspending each time) with 1ml of 1X PBS. The resultant pellet was then resuspended in 150µl of SDS loading buffer and frozen at -20°C. Protein samples were heated in a 100°C water bath for 10 minutes and then spun down at 13000 rpm for 5 minutes before being loaded on the gel.

The samples were run in a BIO-RAD mini-gel system assembled as per the manufacturer's instructions. Before loading the samples, the empty wells were flushed with 1xSDS running buffer to remove any un-polymerised acrylamide that might have gotten stuck to the wells. 15µl of each sample was then loaded into each well and the tank was filled with 1xSDS running buffer. The samples were run at 150V for 70-80 minutes.

2.6.2- Native gel

Reagents required:

- Resuspension buffer: 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 1 mM 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-HCl (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 preparation:

Separating Gel			**Stacking Gel**	
1 Gel	**2 Gels**		**1 Gel**	**2 Gels**
1.4ml	2.8ml	MilliQ	2.15ml	4.3ml
1.25ml	2.5ml	30% Acrylamide	0.85ml	1.7ml
1ml	2ml	80% Glycerol	1.25ml	2.5ml
1.25ml	2.5ml	(pH 8.8) Tris Buffer (pH6.8)	0.625ml	1.25ml
50µl	100µl	10% APS	50µl	100µl
4µl	8µl	TEMED	5µl	10µl

Native gel sample preparation: 1ml of overnight culture was spun down at 13000 rpm for 5 minutes and the pellet resuspended in 120µl of resuspension buffer. The solution was then allowed to stand for 10 minutes at room temperature. It was then sonicated 3 times with 20

second bursts and allowed to stand at room temperature for a further 10 minutes. The solution was then spun at 13000 rpm for 10 minutes and the supernatant was transferred to a new Eppendorf tube with 30µl of sample buffer to obtain the native sample preparation. The gel was run in a tank filled with 1x native gel running buffer at 220V for a period of 4-5 hours.

2.6.3- Western blot

Reagents required: The following reagents were made prior to the blotting process:

- Transfer Buffer (1L): 2.21g (10 mM) of CAPS in 10% methanol made up to 1 L with MilliQ. The pH was adjusted to 11 by 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 MilliQ water. The solution was then autoclaved and stored at room temperature.

- 1X PBS: 100ml of 10X PBS was mixed with 900ml of MilliQ water and mixed thoroughly.

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

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

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

- Diluted primary antibody solution: A dilution of 1:1,000 of primary antibody in blocking solution was made and 0.05% sodium azide was added to prolong the life of the solution. This was then kept cool in a fridge until required.

- Diluted secondary antibody: A dilution of 1:50,000 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 filter paper were cut to a size slightly larger than the gels. A PVDF membrane was also cut into 4 pieces of similar size and one corner was marked. The membrane was then placed in a Petri dish containing absolute methanol for 5 minutes before being moved to a container filled with transfer buffer. The four Whatman papers and two sheets of fibre pads were also placed into the container. Once the electrophoresis gel had run, the protein bands were transferred onto the membrane. This was achieved by first placing a fibre pad onto the negative side of the western blot apparatus, followed by two sheets of Whatman paper. The gel was then placed on top of the paper. On top of this was placed the PVDF membrane (facing the negative side, but placed on the positive side), followed by the remaining filter papers and then finally the last fibre pad. The gel holder (western blot apparatus) was placed in its tank facing the positive electrode with a cooling unit and pre-chilled transfer buffer in the tank. Electrophoresis was then commenced at 200mA for 70 minutes. This caused the proteins to move from the cathode to the anode, and as a result they get trapped in the membrane.

The next part of the process involved antibody binding and detection. The membrane was put in a tray with the protein containing side facing up. The membrane was then rinsed in TPBS for 30 minutes and then incubated in a blocking solution, made with 5% milk powder in TPBS, overnight. The membrane was then incubated in the diluted primary antibody for 1-2 hrs and then washed 3 times for 15 minutes in TPBS. After this, the membrane was incubated in the secondary antibody for a further 1-2 hrs. It was then washed twice with TPBS (15 min) and once with TBS (5 min). It was then laid onto a foil and layered with ECL developing solution and incubated for 5 minutes. The excess liquid was removed and the membrane was wrapped in cling film to be exposed on an x-ray film for 7-15 seconds.

2.6.4- Protein Purification

The purification of the JNL protein involved the use of ion exchange and size exclusion chromatography. In both cases the Bio-Rad biologic LP machine was used along with a fraction collector. Both of these processes are discussed after the lysate preparation below. After each process the protein sample was concentrated using the Vivaspin 20 5,000 MWCO columns (Sartorius sedim) in appropriate buffer and stored at -80°C.

Cell lysate preparation:

A single colony of *E. coli* MGM100 containing the pTrc plasmid was inoculated into 5 ml of LB broth and incubated at 37°C with shaking overnight (180rpm). 1ml of the overnight culture was added to 200ml of fresh LB broth in 3 x 500ml conical flasks with 0.2% glucose 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 0.5 mM IPTG for 4 hours. The cells were harvested by centrifuging at 6000rpm for 10 minutes at 4°C. The

supernatant was decanted and the cell pellet resuspended in 6ml of buffer 1. For cell lysate preparation, the resuspended pellet was sonicated (3x 1min, cooling intervals of 10sec).

(a) DEAE sepharose ion exchange chromatography: DEAE sepharose (GE Healthcare) was used for this process. This has an ion capacity of 0.11-0.16 mmol (Cl-)/m, with a bed height of 15cm, and a max flow rate of 300-600cm/hour.

Buffers used:

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

Chromatography:

The DEAE sepharose column (GE Healthcare, 1cm x 20cm, 12ml) was connected to a 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 125ml of buffer 1. The sample was run with a flow rate of 1ml/min. Proteins were eluted by using a linear gradient with increasing ionic concentration from buffer 1 to buffer 2 (0% - 100%). The column was then washed with ~72ml buffer 2. The column was cleaned using 1M NaCl buffer. Protein samples were collected (approximately 10ml of fraction was collected in each tube) and checked on 12% SDS-PAGE. The purified products were concentrated using Vivaspin 5K cut concentrators (Vivascience).

(b) Size exclusion chromatography (gel filtration): Superdex 200 (GE Healthcare) was used for this process. This has a bed dimension of 10 x 300 mm, bed volume of 24ml, a max flow rate of 1ml/min, and a separation range of 10,000–600,000 globular protein. For this process 250µl (approximately 1mg) of sample was used.

Buffer used:

- Buffer 5: 50 Tris, 50 mM KCl, 10 mM $MgCl_2$ at pH 7.4

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 5 at a flow rate of 1ml/min (approximately 10-15ml of fraction was collected in each tube). The collected sample was analyzed on 12% SDS-PAGE and concentrated using Vivaspin 5K cut concentrators (Vivascience).

2.6.5- ATPase assay

The ATPase assay was performed with the help of the P_iColorLock gold ATPase kit (Innova Biosciences). The assay was performed on clear 96 well plates and the Abs was read at 620 nm. For this assay 1µM protein was needed in a 100µl reaction (each well). A stock of 350µl buffer and protein was made and 100µl was transferred to 3 wells (triplicates). Into each of these wells ATP was added to the final concentration of 1mM in each well, and the plate was incubated at room temperature for 30 minutes. During this time, the required volume (25µl per well) of gold mix is prepared as per the manufacturer's instructions (P_iColorLock gold + 1% accelerator). After the 30 minutes of incubation time are over, 25µl of gold mix is added to each well and left for 5 minutes. 10µl of stabiliser is then added into each well and the plate is incubated at room temperature for a further 30 minutes before the absorbance is read at 620nm.

2.6.6- Analytical ultracentrifugation (AUC)

For this research, only sedimentation velocity experiments (chapter 5) were performed with the following parameters:

- 20 hours running time
- 40000 g
- 4°C

In this project, Midi D-tube dialyzers (Novagen) were used to exchange the sample buffers. The sample is placed inside the tube and put into a beaker containing the buffer that is required (this must be at least 100 times the sample volume). This is then left overnight at 4°C. The following morning, the sample is concentrated using the Vivaspin columns discussed above.

2.6.7- Protein quantification

In all cases the protein concentration is checked either with the Bradford assay (Bio-Rad protein assay reagent).

The process involved the following steps:

- 20X dilution of protein sample was made in filtered water
- 20µl of this mix was added to 1ml Bradford reagent in a 1ml cuvette for 5min
- Absorbance is read at 595nm (Bradford reagent on its own is the blank)
- The absorbance was then converted to concentration using the formula:

  $[P] = (1.79 \times Abs595nm + 0.0064) \times 20 = mg/ml$

Chapter 3

Expression and complementation analysis of *M. tuberculosis* chaperonins in *E. coli*

Background

As discussed above in section 1.5, most bacteria only have one GroEL homologue; however some bacteria have been shown to have 2 or more copies of the *groEL* gene, including most Mycobacteria. In *Mycobacterium tuberculosis* there are 2 distinct *groEL* homologues which encode the chaperonins Cpn60.1 and Cpn60.2 (Kong *et al.*, 1993). *cpn10* and *cpn60.1* constitute an operon, while *cpn60.2* is arranged separately on the chromosome. In section 1.6 the role played by these chaperonins in the pathogenicity of *M. tuberculosis* was discussed. *In vitro* studies have suggested that these chaperonins exist as lower oligomers (Qamra *et al.*, 2004). This finding was surprising, as in all other cases where this has been investigated, chaperonins are nearly always present as two or (occasionally) one single heptameric ring (Lin *et al.*, 2009, Nielsen & Cowan, 1998). This cage like ring structure allows substrate proteins to fold in a secluded environment, away from interactions by other sources.

It is known that the main house-keeping chaperonin in both *M. tuberculosis* and *M. smegmatis* is Cpn60.2, while Cpn60.1 seems to have evolved to perform alternate functions. Our current understanding of the functions of Cpn60.1 implicates its role in biofilm maturation for *M. smegmatis* while the *M. tuberculosis* homologue has been shown to play a role in granuloma formation during tuberculosis (Hu *et al.*, 2008, Kim *et al.*, 2003, Ojha *et al.*, 2005). Recent published data reports that Cpn60.2 from *M. tuberculosis* is unable to complement for the loss of GroEL function in *E. coli* (Kumar *et al.*, 2009). However, similar experiments conducted in our lab provided a different set of results (Han Liu, PhD thesis 2006). Han Liu was able to show that Cpn60.2 from *M. tuberculosis* could complement in *E. coli* for the loss of endogenous GroEL.

In this chapter I have built on the observations made by Han Liu by first replicating his results, and then expanding it with different *E. coli* strains and conditions. Both the *M. tuberculosis* chaperonins were expressed and analysed for their ability to complement in *E. coli*. This was done along with the cochaperonins from both *E. coli* (GroES) and *M. tuberculosis* (Cpn10). Since Cpn60.2 is the main house-keeping chaperonin in *M. tuberculosis,* while Cpn60.1 is not, it was hypothesised that Cpn60.2 would be able to complement in *E. coli* while Cpn60.1 would not. It was predicted that Cpn60.2 would be able to function with both cochaperonins Cpn10 and GroES to rescue cell growth.

Experimental design

For the majority of these experiments the main *E. coli* strain used was MGM100. In this strain the *groE* promoter has been replaced with a P_{BAD} promoter from the arabinose operon (Guzman *et al.*, 1995). This enables the control of the expression of the *E. coli* endogenous GroEL by using glucose (to suppress expression) or arabinose (to increase expression). As a result, the ability of *M. tuberculosis* chaperonins to replace the endogenous *E. coli* chaperonin GroEL can be assessed on LB plates containing glucose or arabinose. During the process of analysing the complementation and expression of the chaperonins other *E. coli* strains were also used, namely, Tab21, which is a BL21 derivative (described in section 3.3) and a Δ*groEL* strain called AI90 (described in section 3.6).

To begin with I was provided with pTrc plasmids carrying various combinations of the chaperonin and cochaperonin homologues (*groES+groEL*, *groES+cpn60.1*, *groES+cpn60.2*, *cpn10+cpn60.1*, and *cpn10+cpn60.2*). These plasmids carry the *trc* promoter which contains

the -35 region of the *trp* promoter together with the -10 region of the *lac* promoter. This promoter is much more tightly regulated than either *lac* or *trp*, and can be over induced with IPTG (Amann *et al*., 1988). Since IPTG cannot be metabolised within the cell, its concentration remains constant, and this enables continuous over expression of the required chaperonins from the plasmids within the cells. The cloning process for these plasmids was done by a collaborator Dr. Yanmin Hu (Coates group, St. George's University of London). These plasmids were first verified by restriction digests and sequencing to ensure the required genes were present without any unwanted mutations. After verification was completed, the plasmids were transformed into the required strains (MGM100, Tab21, and AI90) for the complementation assays.

3.1 Induction of MetE in GroEL depleted cells

Before work was carried out on the complementation assays, it was decided that the observed induction of a secondary protein, when GroEL levels were reduced in *E. coli*, should be investigated. These experiments were carried out in the MGM100 strain described above. Work conducted in the lab had shown that there was an increase in the expression of a protein called MetE when the *groE* operon was repressed in MGM100 (Darren Olden, Master's thesis, University of Birmingham). In retrospect, this result was not surprising as it is known that the *metE* gene is repressed by MetF and MetK, both of which have been found to be obligate GroEL substrates (Kerner *et al.*, 2005), so the lack of GroEL in the cell results in the over-expression of MetE (Chapman *et al.*, 2006). If this observation of MetE induction was in fact being caused by the lack of functional chaperonin in *E. coli*, then it could prove to be a visual marker on SDS-PAGE gels. An experiment was conducted to confirm if the increased expression of MetE was occurring due to the absence of GroEL in the system.

MGM100 was grown overnight in LB with 0.2% arabinose and 100µg/ml kanamycin. The OD_{600} of the resulting culture was measured and a sample was transferred to a flask with 40ml of LB + 0.2% arabinose, to give a starting OD_{600} of 0.01. The cells were grown and the OD_{600} was taken every 20 minutes for the first hour, and a final reading was taken at 130 minutes. From this culture the volume required to get an OD_{600} of 0.1 was calculated and 40ml was transferred into two 50ml Falcon tubes. Both the tubes were then centrifuged. One of the pellets was resuspended in 40ml of LB + 0.2% arabinose, while the second pellet was first washed with 20ml of 0.5% glucose, and then resuspended in 40ml LB + 0.5% glucose. Both cultures were then grown at 37°C and 1ml samples were collected every 20 minutes in 1.5ml Eppendorfs. The samples were diluted with LB, such that the culture in each Eppendorf had an OD_{600} of 0.8. In all cases the growth rate of each culture was the same. Protein sample was then prepared from these cultures and run on a SDS-PAGE gel (Fig 3.1).

The results showed that even after 1 hour of incubation, no MetE over-induction was distinguishable. However, after about 100 minutes the levels of a 100kDa protein started to increase in the cells grown on glucose but not in the cells grown on arabinose. This band was then cut out from the gel and analysed by trypsin digestion and mass spectrometry (service provided by the Functional Genomics laboratory in the School of Biosciences). The result of the mass spectrometry confirmed that the protein was the 92kDa protein, MetE. This result correlated with earlier results in the lab, and also with work done by Chapman *et al.* on the aggregation of certain proteins in the absence of GroEL (Chapman *et al.*, 2006).

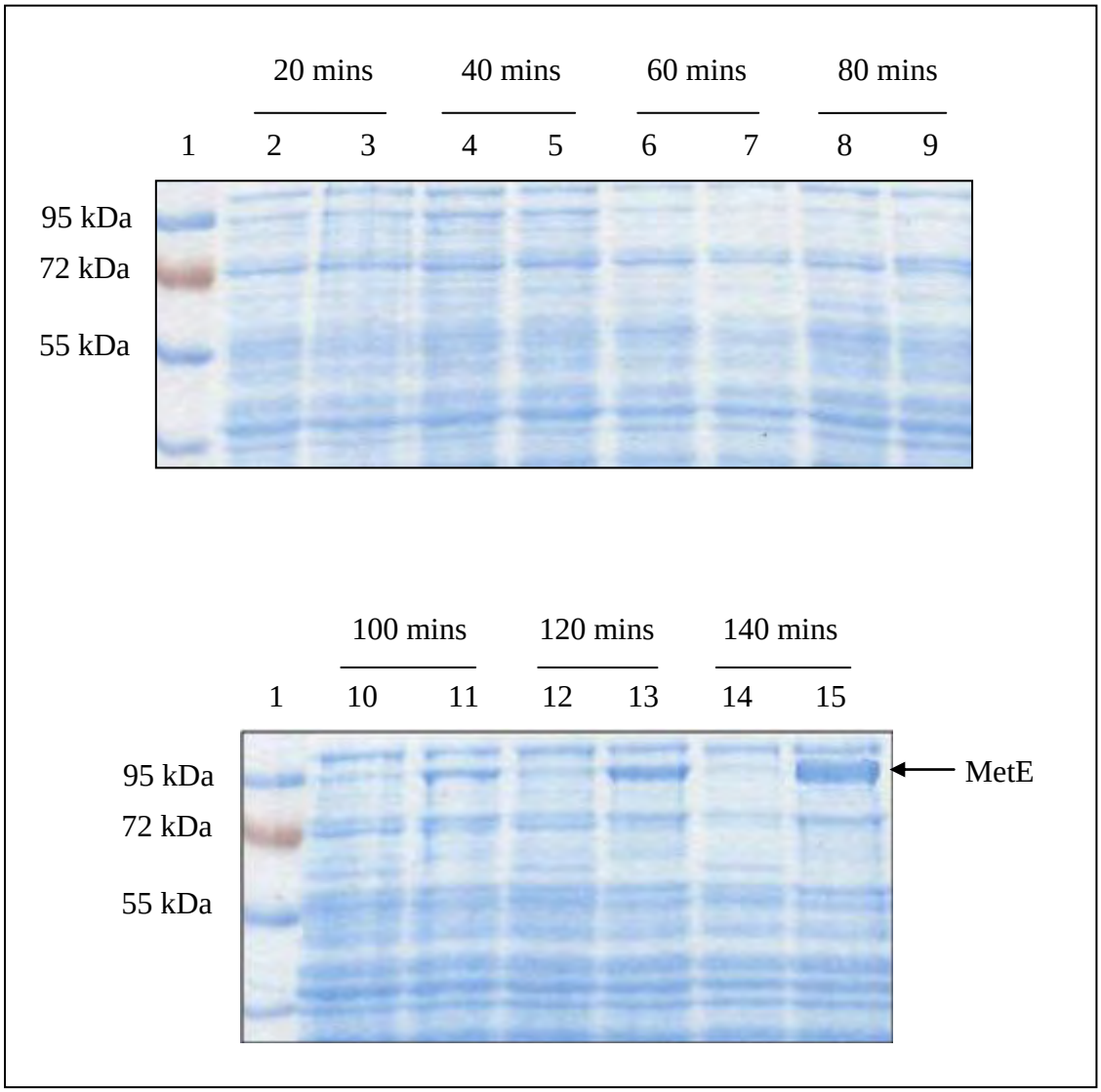


Fig 3.1: SDS-PAGE gel of protein extracts from MGM100 grown on arabinose or glucose at 20 minute intervals. Even numbered lanes are MGM100 in arabinose, while odd numbered lanes are MGM100 in glucose. Lane 1- Prestained protein marker (Fermentas).

Since a GroEL band wasn't distinguishable on the Coomassie stained gel in either arabinose or glucose, its presence was determined with western blotting (Fig 3.2). The results of the western blot confirmed the presence of GroEL in the samples from cultures grown with arabinose as well as very low levels in the samples from cells grown with glucose. This was expected, and also confirms that the P_{BAD} promoter can be used to express or repress GroEL in MGM100. This also demonstrated that even after 1.5 hours (approximately 4-5 generations), although the levels of GroEL in the glucose culture was extremely low, there was still enough to sustain some growth. This result was again not surprising as it has been shown by previous work conducted in the lab that a very small concentration of GroEL is required to keep the cells viable (Ivic *et al.*, 1997, Lorimer, 1996). It was however noted that after the initial 3 hour incubation period, cell growth in glucose drastically reduced when compared with those in arabinose. Cells grown on arabinose show a much more prominent GroEL band on the film. It was observed that the expression level of MetE started to rise at about 2 hours after incubation in the presence of glucose and not in the presence of arabinose. Taken together, these results confirmed the earlier findings that the reduction in the levels of GroEL within *E. coli* causes over-expression of the newly translated MetE protein. This result can also be used as a potential visual marker for a lack of functional chaperonin in a cell.

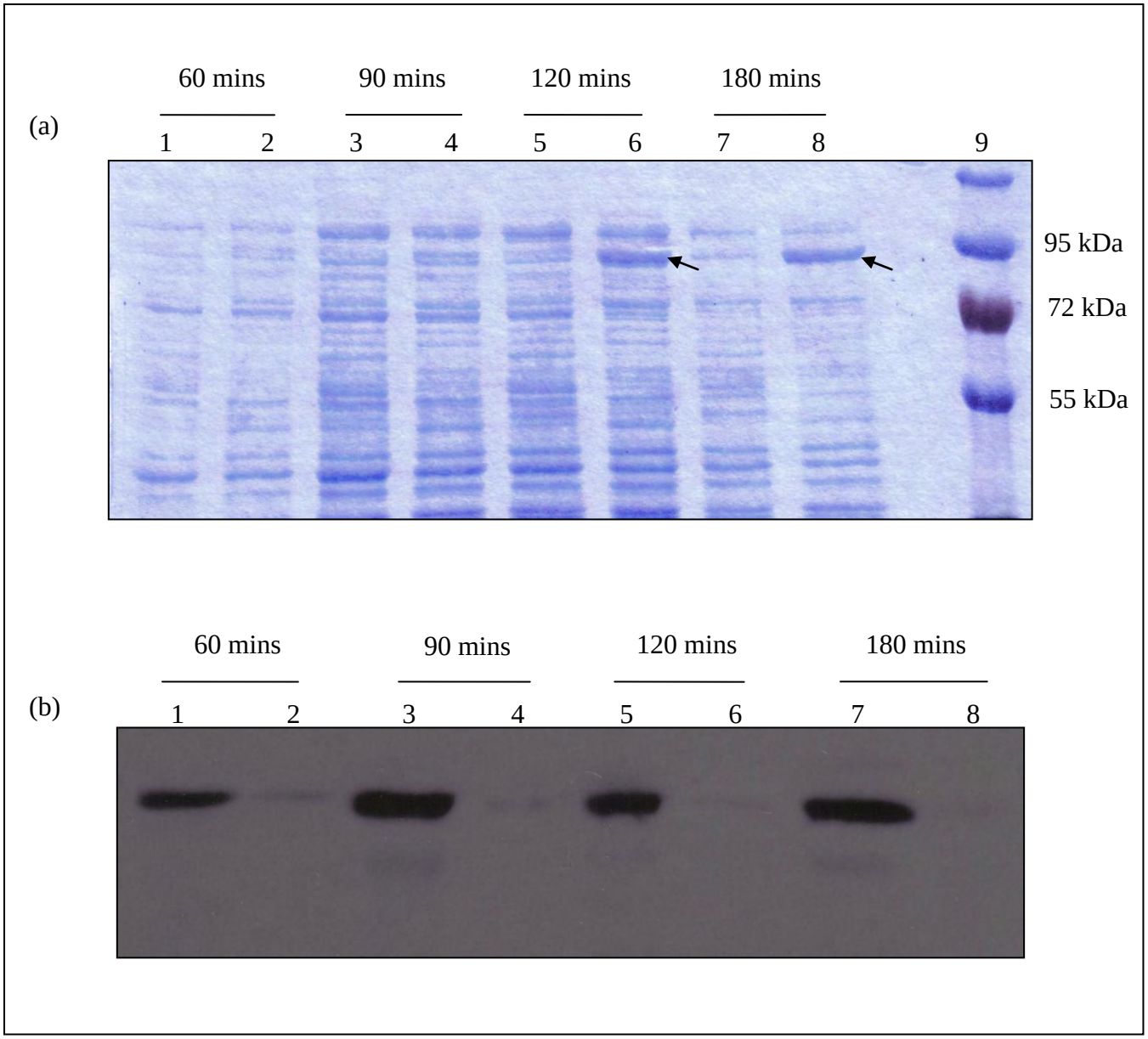


Fig 3.2: Western blot of GroEL from MGM100 grown in arabinose or glucose. (a) SDS-PAGE protein gel of MGM100 grown in arabinose or glucose (arrows pointing at MetE). (b) Western blot using GroEL specific antibody. In both cases odd numbered lanes are MGM100 in arabinose, while even numbered lanes are MGM100 in glucose. Lane 9- Prestained protein marker (Fermentas).

3.2 Expression and complementation of the *M. tuberculosis* GroEL homologues in *E. coli* MGM100

Research conducted on *Mycobacterium tuberculosis* has shown that Cpn60.2, and not Cpn60.1, is the house-keeping chaperonin of this organism (Hu *et al.*, 2008). Prior work conducted in the lab suggested that the Cpn60.2 chaperonin of *Mycobacterium tuberculosis* has the ability to complement for the loss of endogenous GroEL in *E. coli*. It was also noted that Cpn60.1 did not show any signs of complementation in such a system (Han Liu, PhD thesis 2006, University of Birmingham). However, the expression level of Cpn60.1 was also very low and could not be distinguished on SDS-PAGE gels. So it was possible that the lack of complementation by Cpn60.1 was caused by the lack of protein expression. In this section, the hypothesis that *M. tuberculosis* chaperonin 60.2 can function in *E. coli* MGM100 was put to the test. It was expected that the complementation data would confirm the original observations made by Han Liu that Cpn60.2 can function as a chaperonin in *E. coli* while Cpn60.1 cannot.

As discussed in the experimental design, the reason behind the use of the MGM100 strain was that in this strain the chromosomal *groE* promoter has been replaced with a P_{BAD} promoter from the arabinose operon. As a result, the expression of endogenous GroEL can be controlled with the use of either arabinose (to induce expression) or glucose (to suppress expression). This enables the assessment of the *M. tuberculosis* chaperonins to replace the endogenous *E. coli* GroEL and function as a chaperonin.

For this experiment, MGM100 was transformed with the pTrc plasmids carrying the chaperonin and cochaperonin genes. The resultant MGM100 + pTrc strains were grown overnight in arabinose, then sub-cultured and grown in LB + glucose the following morning for 2 hours, and then serially diluted (in ten-fold steps) to spot on to square petri dishes (Sterilin). 5µl of each dilution was spotted onto the plates with the final dilution factor being $10^{7.}$ The plates were then incubated at 37°C, 30°C, and 26°C. The remaining overnight culture was used to extract total cell proteins which were then run on SDS-PAGE gels. Four different plates were used for complementation under each temperature; LB + 0.2% arabinose, LB + 0.2% arabinose + 0.5 mM IPTG, LB + 0.2% glucose, and LB + 0.2% glucose + 0.5 mM IPTG. The pTrc plasmids in question are IPTG inducible.

The complementation results showed that in the presence of arabinose all the cultures grew very well (figure 3.3). This was expected, as the arabinose induces the P_{BAD} promoter and consequently the expression of the chromosomally encoded *groEL* and *groES* genes are sustained in all the cultures. In the presence of glucose, when P_{BAD} is tightly repressed and hence levels of GroEL and GroES fall as the cells divide, only the GroEL homologues that have the ability to complement for the loss of the endogenous GroEL would be able to sustain cell growth. In this case it was seen that Cpn60.2, expressed either with GroES or Cpn10 as the cochaperonin, was able to maintain cell growth at all the temperatures tested. It was also noted that Cpn60.2 with GroES or Cpn10 gave improved growth relative to the endogenous GroES-GroEL form in the presence of 0.5 mM IPTG.

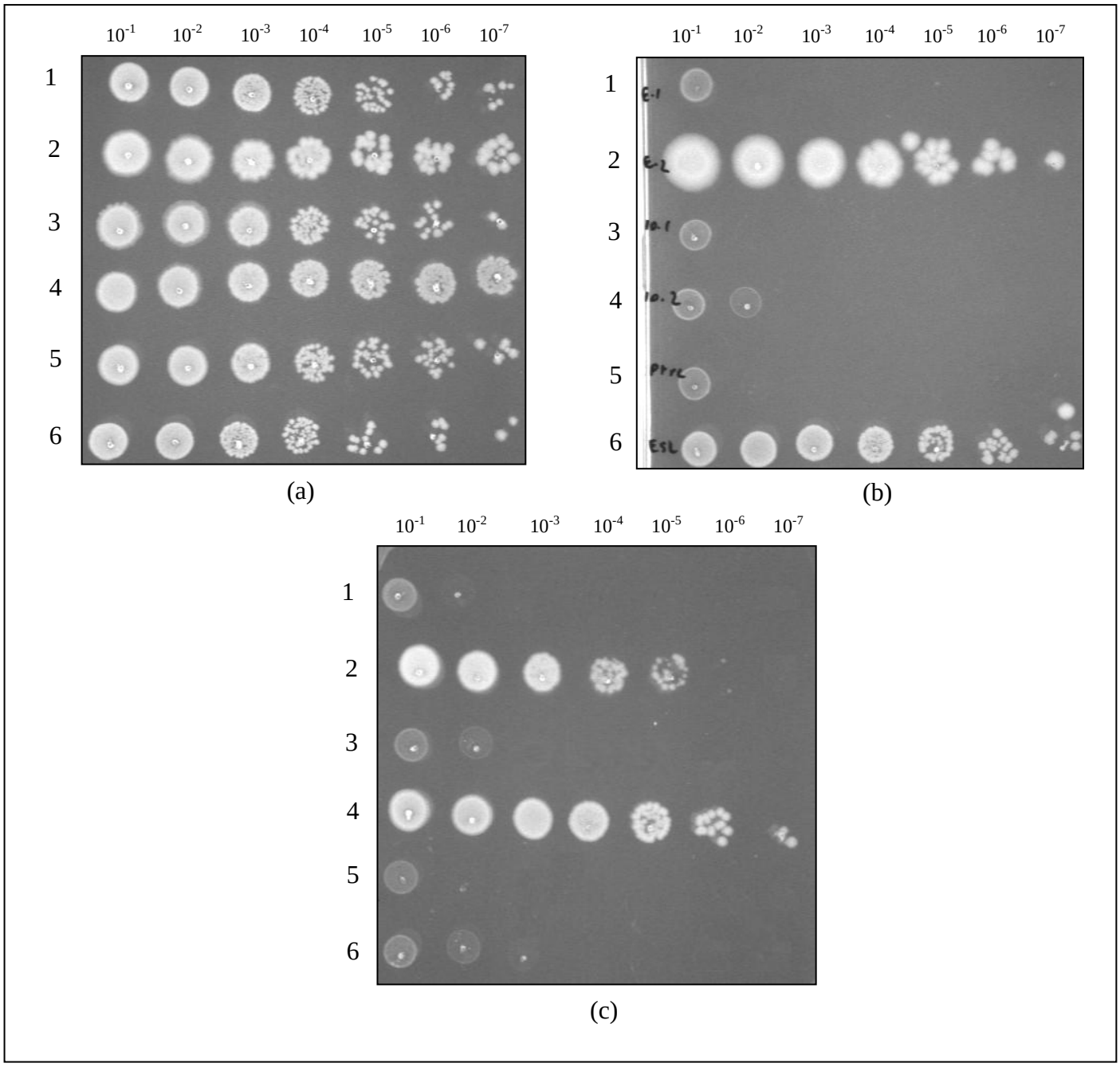


Fig 3.3: Mycobacterial Cpn60.2 can rescue growth of *E. coli* MGM100 under GroESL limiting conditions. (a) Plate grown at 37°C in the presence of arabinose. (b) Plate grown at 37°C in the presence of glucose without IPTG. (c) Plate grown at 37°C in the presence of glucose with 0.5 mM IPTG.

1- MGM100/ pTrc-GroES-Cpn60.1
2- MGM100/ pTrc-GroES-Cpn60.2
3- MGM100/ pTrc-Cpn10-Cpn60.1
4- MGM100/ pTrc-Cpn10-Cpn60.2
5- MGM100/ pTrc99a
6- MGM100/ pTrc-GroES-GroEL

When IPTG was not included Cpn60.2 only showed complementation when co-expressed with GroES and not with Cpn10. Whether this is due to the efficiency of translation, which may be dependent on the exact sequence upstream of Cpn60.2, is unclear. The data suggests that even without IPTG the pTrc plasmids had leaky expression and that the levels of chaperonins within the cells were high enough to complement under certain conditions.

However, the results also show that the addition of IPTG at a molarity of 0.5 mM was causing retarded growth of the strains carrying the pTrc-GroES-GroEL plasmid, suggesting that very high concentrations of chaperonin may be deleterious for the cells (as evidenced later – see Fig 3.4). The reason for this is unclear. These colonies were also usually smaller in size than those without IPTG. Only the images of plates grown at 37°C are shown as the results from the remaining temperatures were very similar. A summary of the results can be seen in table 3.1 below.

	30°C/26°C				37°C			
MGM100	**Arabinose – IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**	**Arabinose – IPTG**	**Arabinose + IPTG**	**Glucose - IPTG**	**Glucose + IPTG**
GroES + Cpn60.1	+++	+++	-	-	+++	+++	-	-
GroES + Cpn60.2	+++	++	+++	++	+++	++	+++	++
Cpn10 + Cpn60.1	+++	+++	-	-	+++	+++	-	-
Cpn10 + Cpn60.2	+++	+++	-	+++	+++	+++	-	+++
pTrc Vector	+++	+++	-	-	+++	+++	-	-
GroES + GroEL	+++	++	+++	+	+++	++	+++	-

Table 3.1: Summary of the complementation data in MGM100. This table summarises the complementation data of cells expressing different combination of chaperonins and cochaperonin from the pTrc plasmid at different temperatures, with or without 0.5 mM IPTG.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

+ Denotes 0.1 – 1 % survival - Partial complementation

- Denotes <0.1 % survival - No complementation

The results from these complementation plates showed that Cpn60.2 can complement in MGM100 under appropriate conditions. However, Cpn60.1 failed to complement, irrespective of which cochaperonin it was co-expressed with. There are two possible reasons for this result. Either Cpn60.1 was incapable of recognising and/or folding the required GroEL substrates or the expression level of Cpn60.1 was too low. Therefore, the levels of Cpn60.1 expression were estimated from SDS-PAGE gels of whole cell lysates. It was expected that the expression of all the chaperonins and cochaperonins from the pTrc plasmid would increase in the presence of IPTG, and that this would be distinguishable on a SDS-PAGE gel. A lower concentration of IPTG (0.1 mM) was used to reduce cell death in the over-night cultures. The results obtained from the SDS-PAGE gels showed that the expression of Cpn60.2 was distinguishable with both cochaperonins. The expression level of Cpn60.1 however, was much lower and could only be seen when co-expressed with GroES (fig 3.4a). MS confirmed that the band was Cpn60.1. As a result, it could be speculated that the level of expression of Cpn60.1 may not have been high enough to sustain cell growth. However a point to note here is that although the level of expression of Cpn60.1 was low, it was not nonexistent. Interestingly, it was also observed that MetE was over-expressed in cells expressing Cpn60.1 both with and without IPTG, but only in cells expressing Cpn60.2 when IPTG was not present (fig 3.4b). When Cpn60.2 is over-expressed with IPTG, the MetE band cannot be seen. This further supports the hypothesis that Cpn60.2 can function as a chaperonin in *E. coli* while Cpn60.1 cannot.

Based on the results obtained in this experiment it can be concluded that Cpn60.2 from *M. tuberculosis* is capable of functioning as a chaperonin in *E. coli*. Cpn60.1 was unable to complement for the loss of endogenous GroEL, but this could be attributed to its low expression levels. To try and remedy this lack of expression various experiments were

designed which will be discussed shortly. First, the observation that excessive chaperonins can result in cell death was further analysed by altering their expression levels.

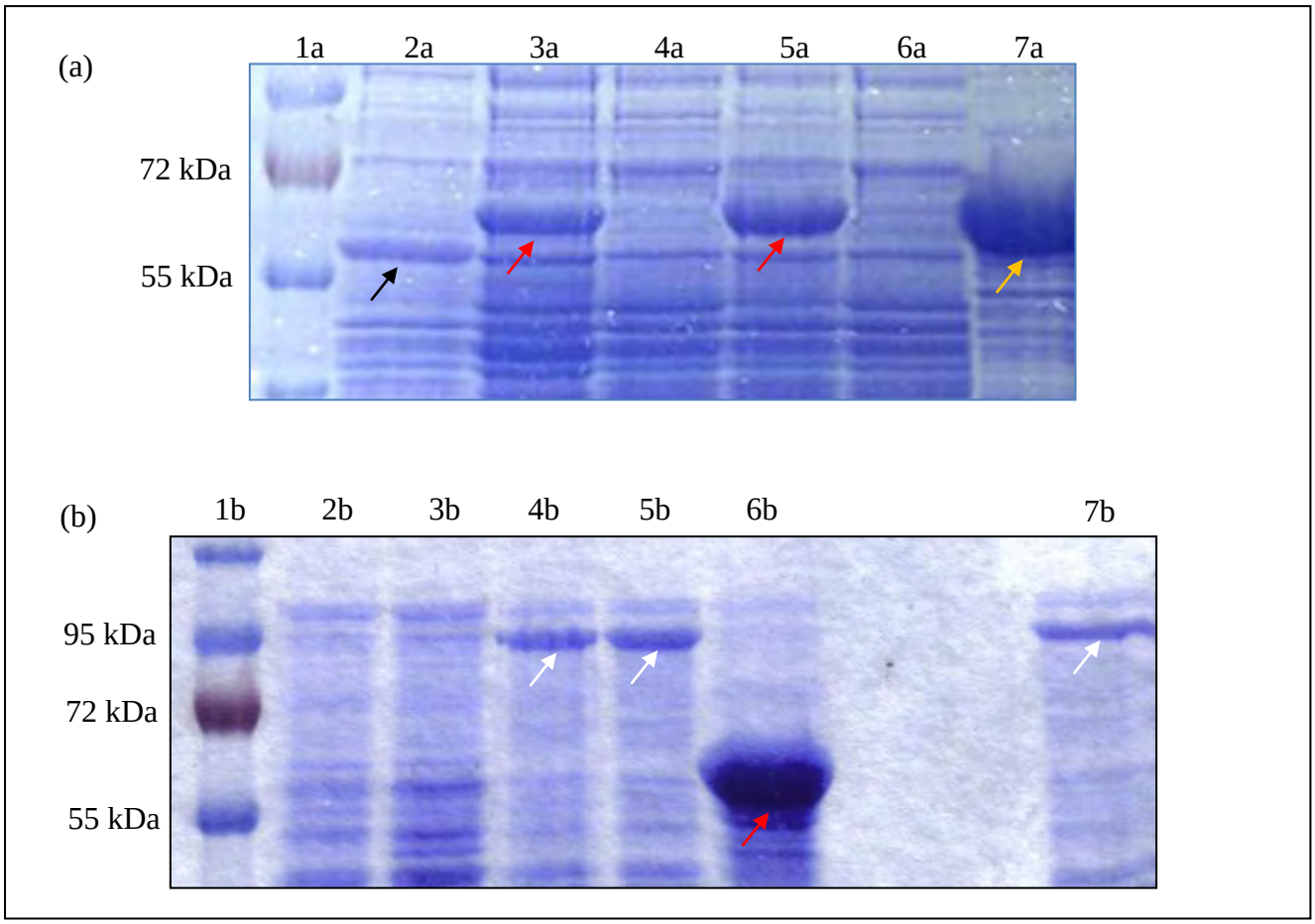


Fig 3.4: SDS-PAGE gels. (a) Gel showing over-expression of Cpn60.2, and some over-expression of Cpn60.1 in MGM100. The samples were taken from cultures grown in 0.2% arabinose with 0.1 mM IPTG. (b) MetE induction seen with non functional chaperonins. Samples for 2b and 3b were taken from cells grown in 0.2% arabinose, while the remaining samples were taken from cells grown in 0.2% glucose with 0.5 mM IPTG. Black arrow indicates Cpn60.1, red arrow indicates Cpn60.2, yellow arrow indicates GroEL, and white arrows indicate MetE.

Lane 1a and 1b- Prestained protein marker (Fermentas)

Lane 2a- MGM100/ pTrc-GroES-Cpn60.1
Lane 3a- MGM100/ pTrc-GroES-Cpn60.2
Lane 4a- MGM100/ pTrc-Cpn10-Cpn60.1
Lane 5a- MGM100/ pTrc-Cpn10-Cpn60.2
Lane 6a- MGM100/ pTrc99a
Lane 7a- MGM100/ pTrc-GroES-GroEL

Lane 2b- MGM100/ pTrc99a
Lane 3b- MGM100/ pTrc99a
Lane 4b- MGM100/ pTrc-GroES-Cpn60.1 (-IPTG)
Lane 5b- MGM100/ pTrc-GroES-Cpn60.1 (+IPTG)
Lane 6b- MGM100/ pTrc-GroES-Cpn60.2 (+IPTG)
Lane 7b- MGM100/ pTrc-GroES-Cpn60.2 (-IPTG)

3.2.1 The effect of changing IPTG concentration on complementation

In section 3.2 the ability of Mycobacterial chaperonins to complement for the loss of endogenous GroEL in *E. coli* MGM100 was tested. It was observed that the use of IPTG at a concentration of 0.5 mM resulted in a drastic reduction is the number of viable colonies with the strain carrying the pTrc-GroES-GroEL plasmid. It was hypothesised that excessively high levels of GroEL were deleterious for the cells. Since the plasmids in question are induces with IPTG, its concentration was altered before reattempting the complementation assay. It was decided that 3 different IPTG concentrations 0.5 mM, 0.25 mM, and 0.1 mM would be tested. It was predicted that the colonies grown on the lowest IPTG concentration would be able to complement better than those on higher IPTG concentrations. The complementation assay remained unaltered from that used in section 3.2.

The results that were obtained were as predicted. The plates showed a reduction in the number of viable colonies when the IPTG concentration was increased (Fig 3.5). This confirmed the original hypothesis that high levels of expression of GroEL, and to a lesser extent Cpn60.2, is lethal for the cells. As a result, in all further complementation assays that required IPTG induction, a concentration of 0.1 mM was used.

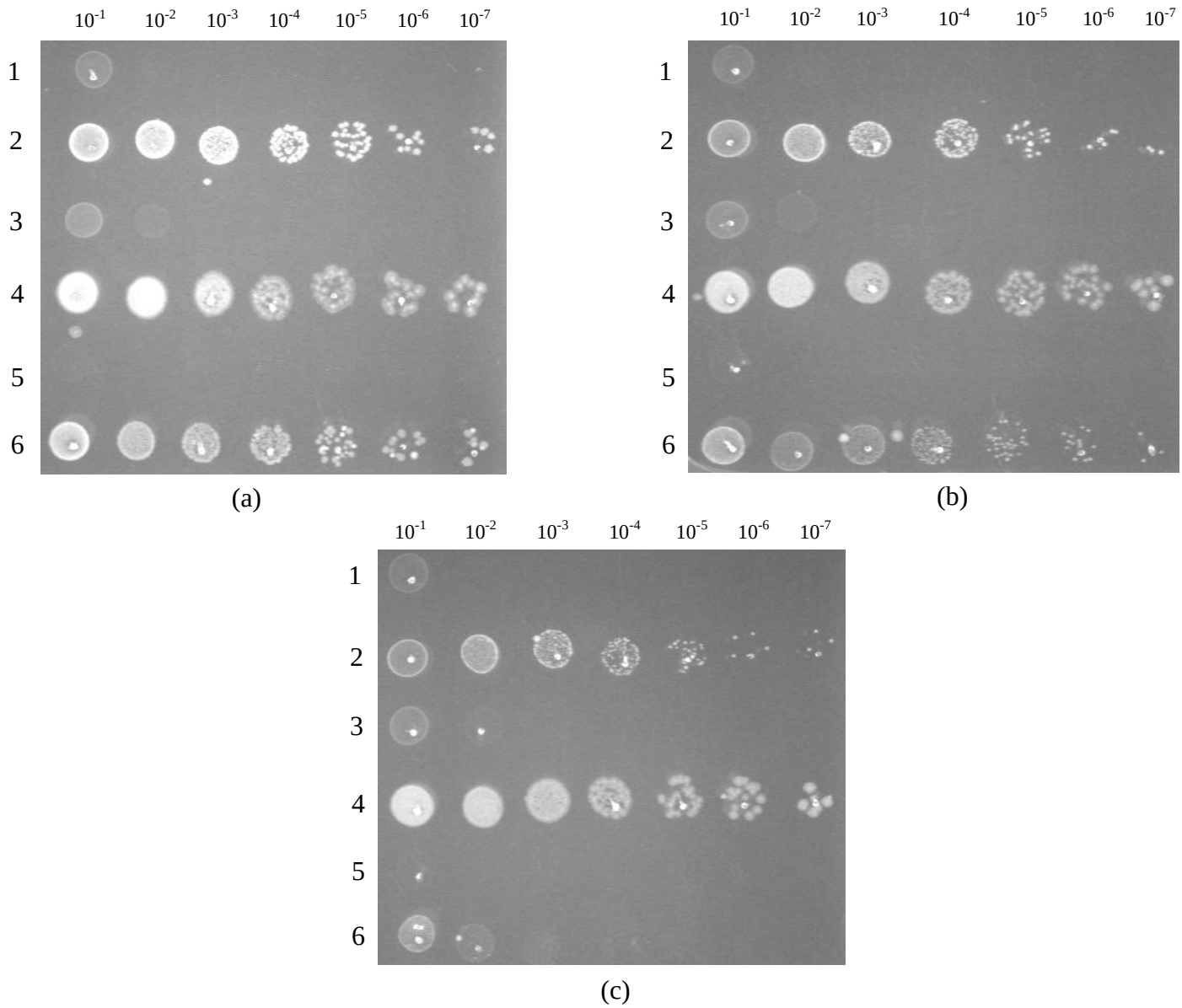


Fig 3.5: Complementation at varying concentrations of IPTG at 37°C. (a) Plate grown at 37°C with 0.2% glucose and 0.1 mM IPTG. (b) Plate grown at 37°C with 0.2% glucose and 0.25 mM IPTG. (c) Plate grown at 37°C with 0.2% glucose and 0.5 mM IPTG.

1- MGM100/ pTrc-GroES-Cpn60.1

2- MGM100/ pTrc-GroES-Cpn60.2

3- MGM100/ pTrc-Cpn10-Cpn60.1

4- MGM100/ pTrc-Cpn10-Cpn60.2

5- MGM100/ pTrc99a

6- MGM100/ pTrc-GroES-GroEL

3.2.2 Using pRARE plasmids with MGM100 to reduce codon bias

Codon bias is a phenomenon that involves different organisms preferring one of the several codons that encode the same given amino acid. The exact reason behind this is under constant debate. One theory is that it may occur due to natural selection, so as to optimise gene translation depending on the GC content of the DNA (Sharp *et al.*, 2005). Since *M. tuberculosis* has a higher GC content than *E. coli,* it was expected that *M. tuberculosis* would prefer a different combination of codons as compared with *E. coli*. This was confirmed by statistical analysis done on their codon usage (Pan *et al.*, 1998). To ensure that the Mycobacterial homologues were not being subjected to codon bias in *E. coli*, a second plasmid was transformed into MGM100 along with the pTrc plasmid combinations. This plasmid, pRARE, encodes tRNAs for codons that rarely occur in *E. coli*, and therefore enhances the expression levels of proteins otherwise limited by codon usage.

The transformation, complementation, and protein expression protocols remained primarily unaltered from section 3.2. The only difference in this complementation assay was the use of 0.1 mM IPTG to induce the expression of the chaperonins. The results showed that the use of pRARE had no prominent affect on the complementation data (Table 3.2). Images of the plates have not been shown as there was no significant difference between these results and the complementation results in section 3.2.1. As was previously seen, Cpn60.2 showed complementation with GroES in the presence and absence of IPTG, and with Cpn10 only in the presence of IPTG. Cpn60.1 did not show any signs of complementation again with either cochaperonin. The levels of protein expression also remained unaltered.

	30°C/26°C				37°C			
MGM100	**Arabinose - IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**	**Arabinose – IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**
GroES + Cpn60.1 + pRARE	+++	+++	-	-	+++	+++	-	-
GroES + Cpn60.2 + pRARE	+++	++	+++	+++	+++	+	+++	+++
Cpn10 + Cpn60.1 + pRARE	+++	+++	-	-	+++	+++	-	-
Cpn10 + Cpn60.2 + pRARE	+++	+++	-	+++	+++	+++	-	+++
pTrc vector + pRARE	+++	+++	-	-	+++	+++	-	-
GroES + GroEL + pRARE	+++	++	+++	+++	+++	+	+++	+++

Table 3.2: A summary of the complementation data in MGM100 with pRARE. This table summarises the complementation data of cells expressing different combination of chaperonins and cochaperonin from the pTrc plasmid along with pRARE at different temperatures, with or without 0.1 mM IPTG.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

+ Denotes 0.1 – 1 % survival - Partial complementation

- Denotes <0.1 % survival - No complementation

3.3 Complementation and expression of *M. tuberculosis* GroEL homologues in Tab21

Based on the observations made in section 3.2, it was speculated that the lack of viability of the Cpn60.1 strains in the complementation experiments could be due to the lower expression levels of the Cpn60.1 protein compared with GroEL and Cpn60.2. As a result a number of strategies were tested for their ability to improve the expression of the Cpn60.1 protein. One of these methods was to use the strain Tab21, a BL21 (DE3) derivative.

BL21 is derived from the *E. coli* B strain and is a very good protein expression strain. It has been modified to carry the T7 RNA polymerase gene under the control of an IPTG inducible *lac* promoter. This polymerase is extremely active and can synthesise RNA at a rate several times that of *E. coli* RNA polymerase. As a result this strain can be used to induce protein expression from plasmids that make use of the T7 transcription and translation mechanism (e.g. pET vectors).

Before BL21 could be used in the complementation assays it was necessary to be able to control the expression of endogenous GroEL and GroES. This was necessary so that the complementing ability of Cpn60.1 and Cpn60.2 could be tested without the interference of endogenous chaperonins. For this reason it was decided that the pBAD expression system used in MGM100 would also be used with BL21. To achieve this, the *groE* promoter in BL21 would need to be replaced with the P_{BAD} promoter. However, since GroEL is vital for cell survival, gene deletion was not an option. Instead, the *groESL* operon under control of the P_{BAD} promoter, which was available on the MGM100 chromosome, had to be inserted into the BL21 strains. This was achieved by using P1 bacteriophage transduction.

P1 phage was grown with MGM100 and then used to infect the BL21 strain as per the protocols described in section 2.4.5. This results in bacteriophage P1 injecting the donor bacterial DNA into the BL21 strain. The resultant recipient cell can then recombine the donor bacterial DNA into its own DNA. For the purpose of the complementation experiments, the *pBAD-ESL* operon had to replace the endogenous *groESL* operon. BL21 is not resistant to any antibiotics, but the *pBAD-ESL* operon has a Kanamycin resistance gene just downstream of the GroEL sequence which was also used in the screening process. So to test if the resultant BL21 strains had incorporated the *pBAD-ESL* operon into their genome, the cells were grown on LB plates with 0.2% arabinose and 50µg/ml kanamycin. Any colonies that were obtained were then re-streaked onto plates with 0.2% glucose and 50µg/ml kanamycin. BL21 cells that were able to grow in the presence of arabinose but not in the presence of glucose were then selected for the complementation assays.

This new BL21 pBAD-*groESL* strain was named Tab21. The original pTrc plasmids containing the Mycobacterial homologues were transformed into Tab21 using the protocol described in section 2.4.16. This was done to test the complementation and expression of the chaperonins from the pTrc plasmid in Tab21 before attempts were made using pET vectors (section 3.3.1). It was hypothesised that Cpn60.2 would be able to complement in Tab21 but Cpn60.1 would not. Since the same pTrc plasmids were being used, an increase in protein expression compared with those seen in MGM100 was not expected.

The results that were obtained showed that there was no variation in the complementation data at any of the temperatures tested (see Fig. 3.3). In the presence of arabinose all the strains grew as the P_{BAD} promoter was inducing the expression of endogenous GroES and GroEL. In the presence of glucose, the complementation results obtained were similar to that

of MGM100. The use of Tab21 did not have any affects on the overall complementation data and similar levels of protein expression were obtained. It was concluded that, as was the case with MGM100, the Cpn60.2 chaperonin complements for the loss of endogenous GroEL in Tab21, while Cpn60.1 does not. A summary of the complementation and expression results can be seen in table 3.2 and figure 3.6.

As discussed above, since Tab21 is a protein expression strain that makes use of the T7 RNA polymerase expression system, the next step in trying to achieve higher levels of protein expression of Cpn60.1 involved the transfer of the *groES-cpn60.1* operon and the *cpn10-cpn60.1* operon onto a pET plasmid (section 3.3.1).

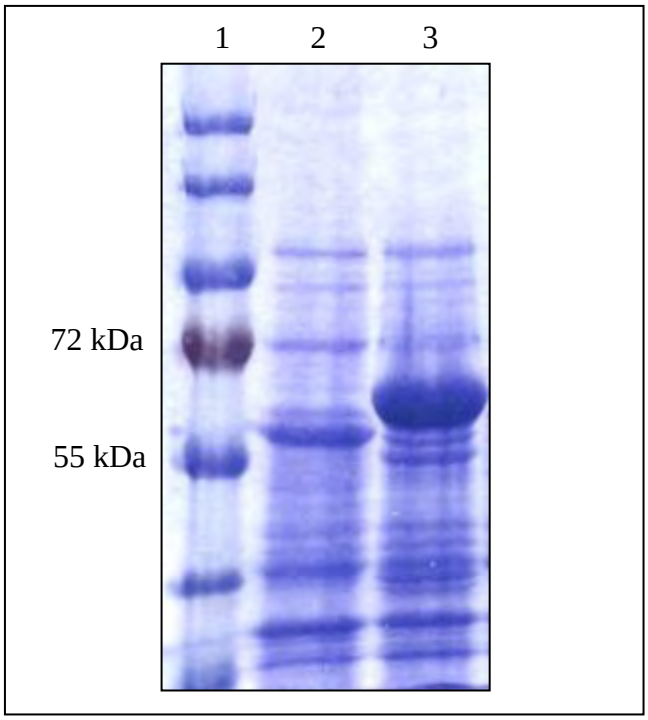


Fig 3.6: SDS-PAGE gel showing over-expression of Cpn60.1 and Cpn60.2 in Tab21. All the samples were taken from cultures after 5 hours incubation with 0.1 mM IPTG and 0.2% arabinose.

Lane 1- Prestained protein marker (Fermentas) Lane 3- Tab21/ pTrc-GroES-Cpn60.2

Lane 2- Tab21/ pTrc-GroES-Cpn60.1

	30°C/26°C				37°C			
Tab21	**Arabinose - IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**	**Arabinose - IPTG**	**Arabinose + IPTG**	**Glucose - IPTG**	**Glucose + IPTG**
GroES + Cpn60.1	+++	+++	-	-	+++	+++	-	-
GroES + Cpn60.2	+++	++	+++	+++	+++	++	+++	+++
Cpn10 + Cpn60.1	+++	+++	-	-	+++	+++	-	-
Cpn10 + Cpn60.2	+++	+++	-	+++	+++	+++	-	+++
pTrc Vector	+++	+++	-	-	+++	+++	-	-
GroES + GroEL	+++	++	+++	+++	+++	++	+++	+++

Table 3.3: Summary of the complementation data in Tab21. The table shows the growth of cells expressing different combination of chaperonins and cochaperonin from the pTrc plasmid in Tab21 at different temperatures, with or without 0.1 mM IPTG.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

+ Denotes 0.1 – 1 % survival - Partial complementation

- Denotes <0.1 % survival - No complementation

3.3.1 Using pET plasmid in Tab21

To try and obtain a higher level of protein expression for Cpn60.1, its gene was cloned into a pET vector to be used with Tab21. It was expected that this would allow the required gene products to be expressed in greater quantity as compared with the pTrc plasmid by making use of the T7 transcription and translation mechanism.

A pET vector with the ampicillin resistance gene was chosen, as Tab21 is kanamycin resistant. It was decided that pET23a+ would be used as the cloning vector as it had the required ampicillin resistance gene, and also had the appropriate restriction sites to allow a direct swap of genes from the pTrc plasmids. A schematic representation of the swaps is show in figure 3.7. As the level of protein expression of Cpn60.2 was high with the pTrc plasmid, it was decided that only the Cpn60.1 gene along with both the cochaperonins would be cloned into the pET vector. Once the required pET plasmid combinations were obtained and confirmed by sequencing, they were transformed into Tab21 as per the protocol described in section 2.4.16 and their expression levels were analysed.

Unfortunately expression of Cpn60.1 could not be distinguished on the gel (fig 3.8). Only the over-expression of GroEL was seen from the cells containing pET-GroES-GroEL. It was also observed that the strains carrying the pET-GroES-Cpn60.1 and pET-Cpn10-Cpn60.1 plasmids showed slower growth in both liquid and solid media compared with cells carrying the pTrc versions. An accurate complementation analysis could not be done as cell growth was extremely slow on solid media with glucose and arabinose. The reason for this observation is unclear. It is possible that the loss of the cells ability to regulate its endogenous

GroEL in both cases could have resulted in cell death. However this does not explain why cell growth was not affected by the pTrc plasmids as the cells are unable to regulate their expression either.

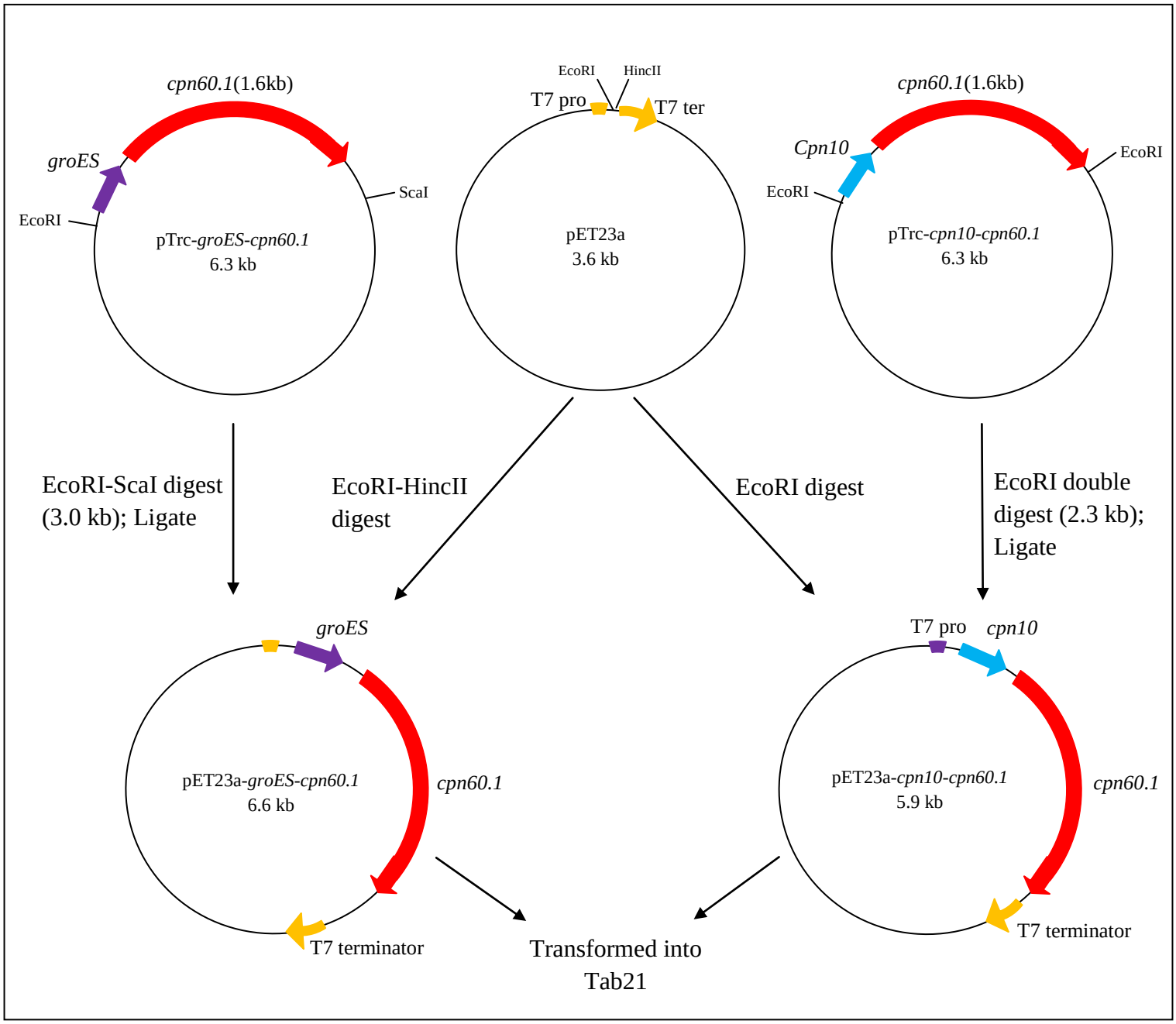


Fig 3.7: Schematic illustration of the steps involved in cloning GroES-Cpn60.1 and Cpn10-Cpn60.1 into pET23a+.

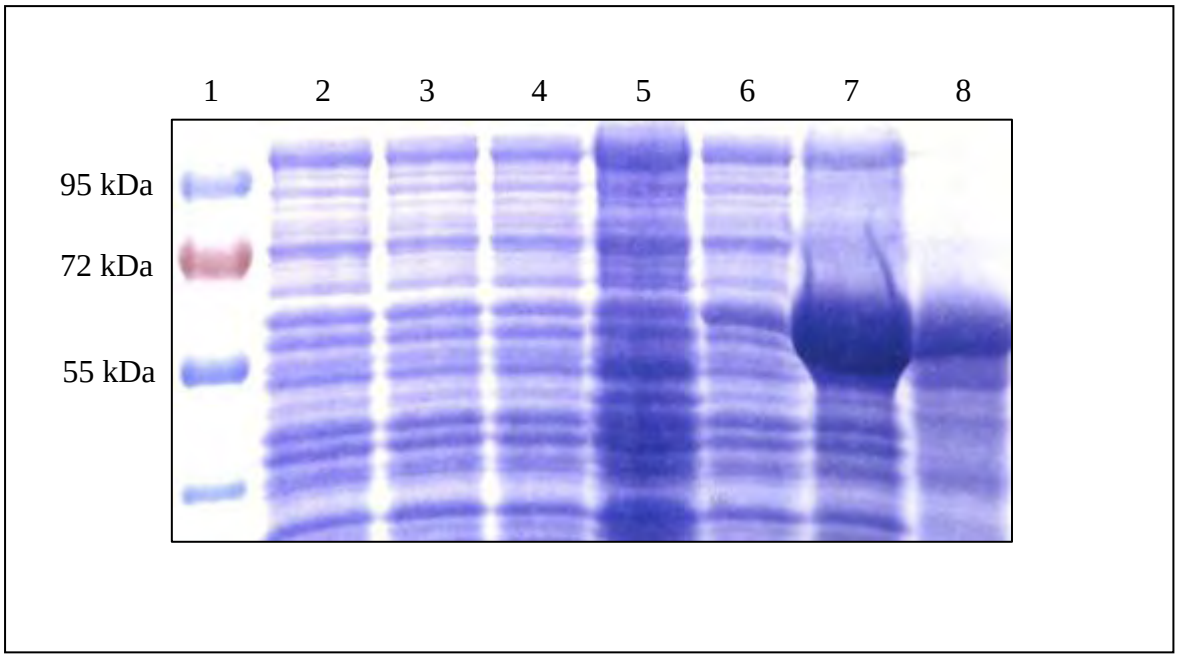


Fig 3.8: SDS-PAGE of protein extracts from cells containing the pET vectors. The method used for protein extraction was the same as that use for the pTrc plasmids in MGM100. In all cases protein samples were made from cells grown in 0.2% arabinose for 5 hours.

Lane 1- Prestained protein marker (Fermentas)

Lane 2- Tab21/ pET-GroES-Cpn60.1 (0.1 mM IPTG)

Lane 3- Tab21/ pET-GroES-Cpn60.1 (no IPTG)

Lane 4- Tab21/ pET-Cpn10-Cpn60.1 (0.1 mM IPTG)

Lane 5- Tab21/ pET-Cpn10-Cpn60.1 (no IPTG)

Lane 6- Tab21/ pET23a+ vector

Lane 7- Tab21/ pET GroES-GroEL (0.1 mM IPTG)

Lane 8- Tab21/ pET GroES-GroEL (no IPTG)

3.4 Cpn10 site directed mutagenesis

As was discussed in section 3.2, during the initial experiments to test protein expression, some expression of Cpn60.1 was also observed. However, its expression was only distinguishable when the chaperonin was expressed with GroES and not with Cpn10. Since over induction of Cpn60.1 using pET vectors was unsuccessful, an alternate approach was looked into.

SDS-PAGE gels showed the presence of both GroES and Cpn10 (figure 3.9 and 3.10). Sequence data showed that the start codon of Cpn10 was GTG (encoding valine), rather than the more widely used start codon of ATG (encoding methionine). It was hypothesised that perhaps this was the reason why Cpn60.1 was visible on Coomassie stained SDS-PAGE gels when co-expressed with GroES, but not with Cpn10. It may be that the starting codon of Cpn10, which is valine, does not allow the ribosome to form a strong enough initiation complex (Sussman *et al.*, 1996). Instead the complex may be transient and as a result the expression of Cpn60.1, which is just downstream from Cpn10, may be affected. To address this, the starting codon of Cpn10 was changed from a GTG (valine) to ATG (methionine) using site directed mutagenesis for both the pTrc-Cpn10-Cpn60.2 and pTrc-Cpn10-Cpn60.1 plasmids.

The primer design and PCR protocol used for this purpose has been described in the materials and methods section. The resulting PRC products were digested with DpnI before being transformed into MGM100. DpnI digestion is a necessary part of site directed mutagenesis as

it only digests the parental plasmid and not the PCR products. This drastically reduces the chances of transformed cells to carry the parental plasmid. DNA sequencing was done to confirm the presence of an ATG start codon for the *cpn10* gene and to ensure no other mutations were present.

After confirming the presence of the ATG forms of Cpn10-60.1 and Cpn10-60.2 (referred to as ATG-Cpn10-60.1 and ATG-Cpn10-60.2), the MGM100 strains expressing them were subjected to the same complementation and protein expression analysis as the original plasmids. The ATG-Cpn10-60.2 pTrc plasmid results showed a slight increase in the expression of Cpn10 but not in Cpn60.2. It was also noticed that there was a slight improvement in the complementing ability of ATG-Cpn10-Cpn60.2 in the presence of glucose and absence of IPTG as compared with the original Cpn10-60.2. Unfortunately no difference in the expression of Cpn60.1 was seen with ATG-Cpn10-60.1 (fig 3.11). The complementation data of ATG-Cpn10-Cpn60.1 also remained unchanged and no viable cells were seen on any of the complementation plates. A summary of the complementation data can be seen in table 3.4 below.

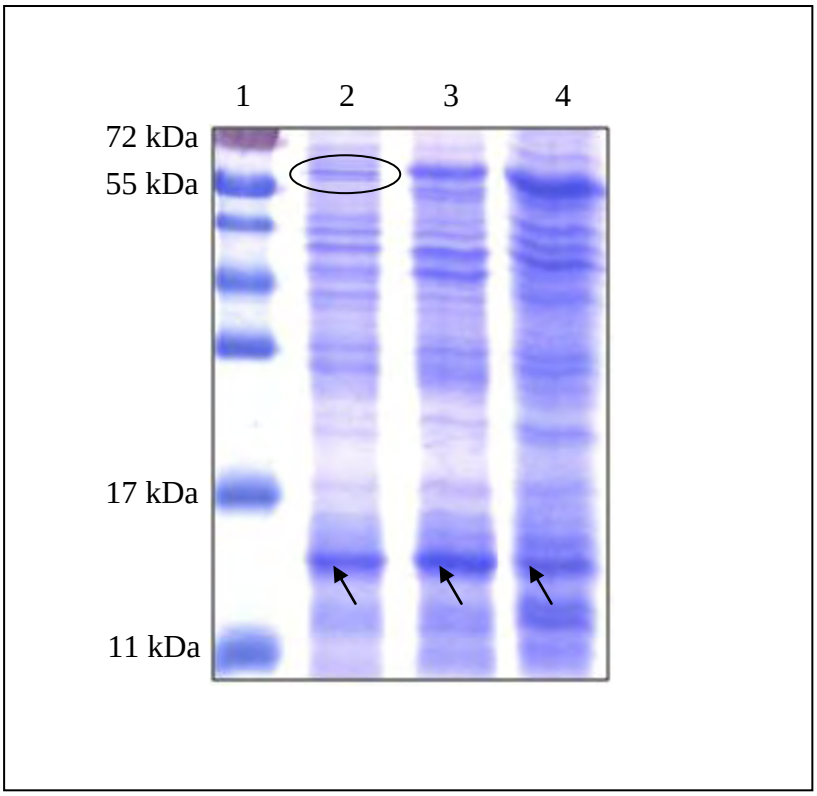


Fig 3.9: 15% SDS-PAGE gel of GroES in MGM100 with 0.2% arabinose and 0.1 mM IPTG using the pTrc plasmids. All protein samples were taken from 2ml of OD_{600} 1 culture. Arrows are pointing at the GroES protein band. Arrows are pointing at GroES while Cpn60.1 is circled.

Lane 1- Prestained protein marker (Fermentas)

Lane 2- MGM100/ pTrc-GroES-Cpn60.1

Lane 3- MGM100/ pTrc-GroES-Cpn60.2

Lane 4- MGM100/ pTrc-GroES-GroEL

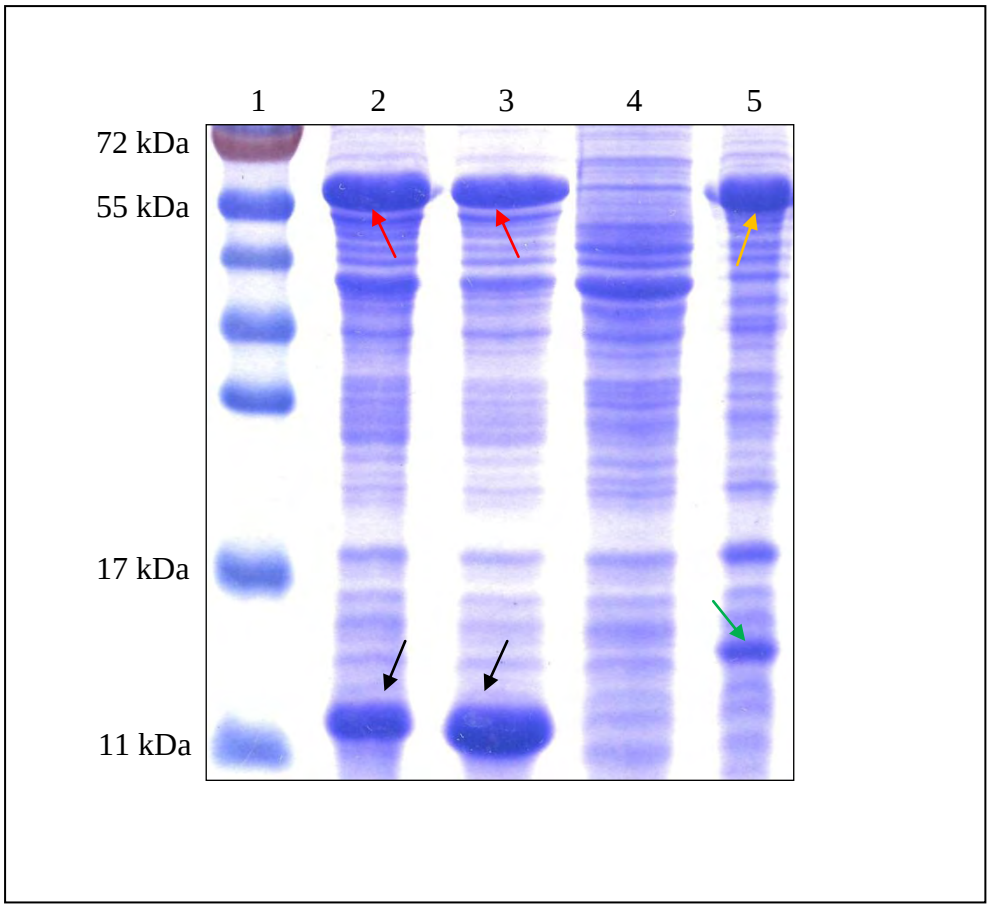


Fig 3.10: SDS-PAGE gel showing the slight increase in expression of the ATG from of Cpn10 from the plasmid ATG-Cpn10-60.2 as compared with the original Cpn10-60.2 plasmid. All the samples were taken from 2ml of OD_{600} 1 MGM100 culture with 0.2% arabinose and 0.1 mM IPTG. Black arrows are pointing at Cpn10 band, red arrows are pointing at Cpn60.2, the green arrow is pointing at GroES, and the yellow arrow is pointing at GroEL.

Lane 1- Prestained protein marker (Fermentas)

Lane 2- MGM100/ Original pTrc-Cpn10-60.2

Lane 3- MGM100/ pTrc-ATG-Cpn10-60.2

Lane 4- MGM100/ pTrc99a vector (negative control)

Lane 5- MGM100/ pTrc-GroES-GroEL

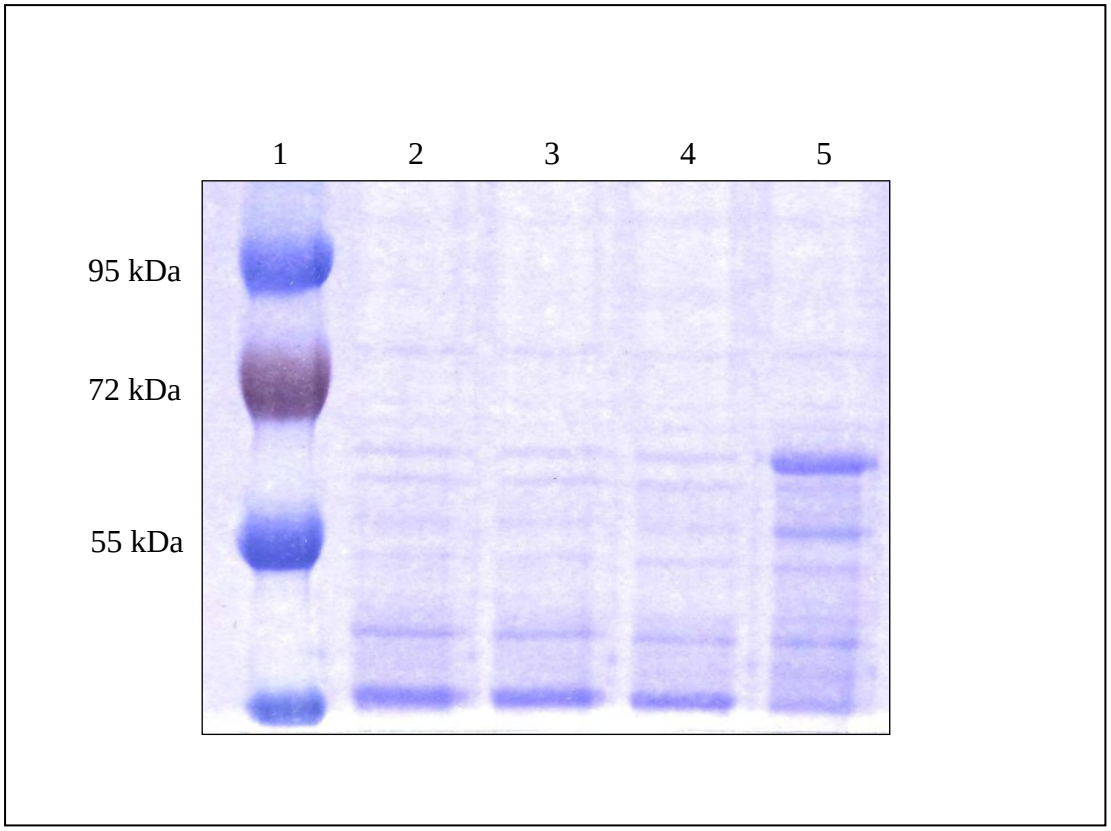


Fig 3.11: SDS-PAGE gel of ATG-Cpn10-60.1 showing no increase in expression in MGM100 with 0.2% arabinose and 0.1 mM IPTG. The protein samples were taken from 1ml of OD_{600} 0.5 cultures.

Lane 1- Prestained protein marker (Fermentas)

Lane 2- Original pTrc-Cpn10-60.1

Lane 3- pTrc-ATG-Cpn10-60.1

Lane 4- pTrc99a vector

Lane 5- pTrc-GroES-GroEL

	30°C/26°C				37°C			
MGM100	**Arabinose – IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**	**Arabinose - IPTG**	**Arabinose + IPTG**	**Glucose – IPTG**	**Glucose + IPTG**
Original Cpn10 + Cpn60.1	+++	+++	-	-	+++	+++	-	-
ATG Cpn10 + Cpn60.1	+++	+++	-	-	+++	+++	-	-
Original Cpn10 + Cpn60.2	+++	++	-	+++	+++	+++	-	+++
ATG Cpn10 + Cpn60.2	+++	++	+	+++	+++	+++	+	+++
pTrc Vector	+++	+++	-	-	+++	+++	-	-
GroES + GroEL	+++	+	+++	+	+++	+	+++	+

Table 3.4: A summary of the complementation data of ATG-Cpn10 when expressed with Cpn60.1 and Cpn60.2 in MGM100.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

\+ Denotes 0.1 – 1 % survival - Partial complementation

\- Denotes <0.1 % survival - No complementation

3.5 Random mutagenesis of Cpn60.1

It is already know that in *E. coli* a relatively small concentration of GroEL is needed to sustain cell growth at 37°C (Herendeen *et al.*, 1979, VanBogelen *et al.*, 1987, Lorimer, 1996). It was discussed in section 3.1 that even after 3 hours of being induced with arabinose, GroEL bands could not be distinguished on the SDS-PAGE gel without using western blots. Although attempts to improve the overall level of expression of Cpn60.1 in the last few experiments were unsuccessful, it can still be speculated that, if Cpn60.1 could function as a replacement for GroEL in *E. coli* then the level of expression seen in MGM100 with GroES should have been sufficient to do so. However, it is unknown how active Cpn60.1 is compared with GroEL, so it is possible that if its activity is much lower than GroEL then a higher concentration of Cpn60.1 would be needed to complement.

All the experiments done so far on Cpn60.1 have shown that it cannot complement in *E. coli* in the absence of endogenous GroEL, while Cpn60.2 is very capable of sustaining cell growth up to 37°C. Cpn60.1 and Cpn60.2 share 61% sequence identity and 76% sequence similarity (Fig 3.12). It was hypothesised that it may be possible to obtain a functional form of Cpn60.1 that could complement for the loss of endogenous GroEL in *E. coli* MGM100 with random mutagenesis. For this reason, the XL1-Red *E. coli* strain provided by Agilent technologies (formerly Stratagene) was used to mutagenise *cpn60.1*. This *E. coli* strain is deficient in three of the primary DNA repair pathways and according to its manufacturer, has a mutation rate ~5000 fold higher than that of wild-type. Unfortunately, after several attempts, no such mutants were found.

Fig 3.12: Comparison of the amino acid sequence of *M. tuberculosis* chaperonins. The asterisk indicates identical residues, a colon indicates a conserved substitution, and a full stop indicates a semi conserved substitution. The sequence was aligned with the help of ClustalW2 on the European Bioinformatics Institute website.

3.6 Complementation of *M. tuberculosis* and *M. smegmatis* Cpn60.2 in Δ*groEL* AI90

All the results so far have supported the understanding that Cpn60.2 is the main house-keeping chaperonin of Mycobacteria. In this chapter it has been shown that Cpn60.2 can function very well in *E. coli* MGM100 and Tab21. However, in all these experiments it was possible that the P_{BAD} promoter was a little leaky and that there was just enough GroEL present within the cell, when grown on glucose, that it was capable of functioning along with Cpn60.2 and sustaining cell growth. A lab colleague, Andrew Large, was able to show that MGM100 cells grown on glucose did have very small amounts of residual GroEL by heavily overloading SDS-PAGE gels and western blotting. So, to better test the hypothesis that Cpn60.2 can fully complement in *E. coli* in the absence of endogenous GroEL, the chromosomally encoded *groEL* gene would need to be removed from *E. coli*.

For this experiment the *ΔgroEL* strain, AI90, was used (Ivic *et al.*, 1997). In this strain the *groES* and *groEL* genes are present on the pACYC plasmid (pACYC-Ptrc-GroESL-SacB), as it is not possible to delete the *groEL* gene without providing a complementing copy. It has been shown that the *sacB* gene expression in gram negative bacteria like *E. coli* confers sucrose sensitivity (Gay *et al.*, 1985). When AI90 is grown on sucrose, it allows for the selection of cells that no longer posses the *sacB* gene. In a lot of these cases, it is achieved with the removal of the pACYC plasmid from the cells prior to the selection process. In such a case, the AI90 strain would only grow if pACYC was displaced with a plasmid carrying chaperonins capable of functioning in the complete absence of GroEL (plasmid shuffling). It was expected that both *M. tuberculosis* and *M. smegmatis* Cpn60.2 would be able to rescue cell growth in the complete absence of GroEL, as it had also been shown (T. Rao, PhD thesis) that Cpn60.2 from *M. smegmatis* could complement in MGM100.

AI90/pACYC-Ptrc-GroESL-SacB was transformed with the pTrc-GroES-Cpn60.2 constructs, the pTr99a vector, the pTrc-GroES-GroEL construct, and also pOFESL (constitutively expresses GroES and GroEL) using the protocol described in section 2.4.16. The cells were then grown over night in LB with 100µg/ml ampicillin, 50µg/ml chloramphenicol, and 0.1 mM IPTG. The following day the cultures were serially diluted and plated out on selectable plates with and without sucrose at 30°C and 37°C (fig 3.13).

The results showed that in the presence of sucrose and chloramphenicol none of the strains grew (fig 3.13a). This was expected, as the presence of sucrose causes the pACYC plasmid to be removed from the cell, and since this plasmid confers chloramphenicol resistance, its absence results in cell death. On plate (b) where there is no sucrose present, all the cells grew as the pACYC plasmid expressing the endogenous *groES* and *groEL* gene was still present. Finally on plate (c) in presence of sucrose and ampicillin, but in the absence of chloramphenicol, the positive controls along with the Cpn60.2 constructs were able to grow.

To confirm the absence of GroEL from the colonies that were present on the plates, colony PCR was done using primers that were specific to the N-terminal region of GroEL (fig 3.14). The PCR yields a ~570bp fragment if GroEL is present in the cells, as can been seen in lane 2 of figure 3.14, which is a colony PCR of a DH5α strain. The negative control for this PCR involved the use of a plasmid expressing *Methanococcus maripaludis* CCT, which is capable of complementing in *E. coli* (Andrew Large, personal communication). The results confirm that there was no GroEL present within the complemented strains. These results proved that both *M. tuberculosis* and *M. smegmatis* Cpn60.2 are capable of complementing in the complete absence of endogenous GroEL.

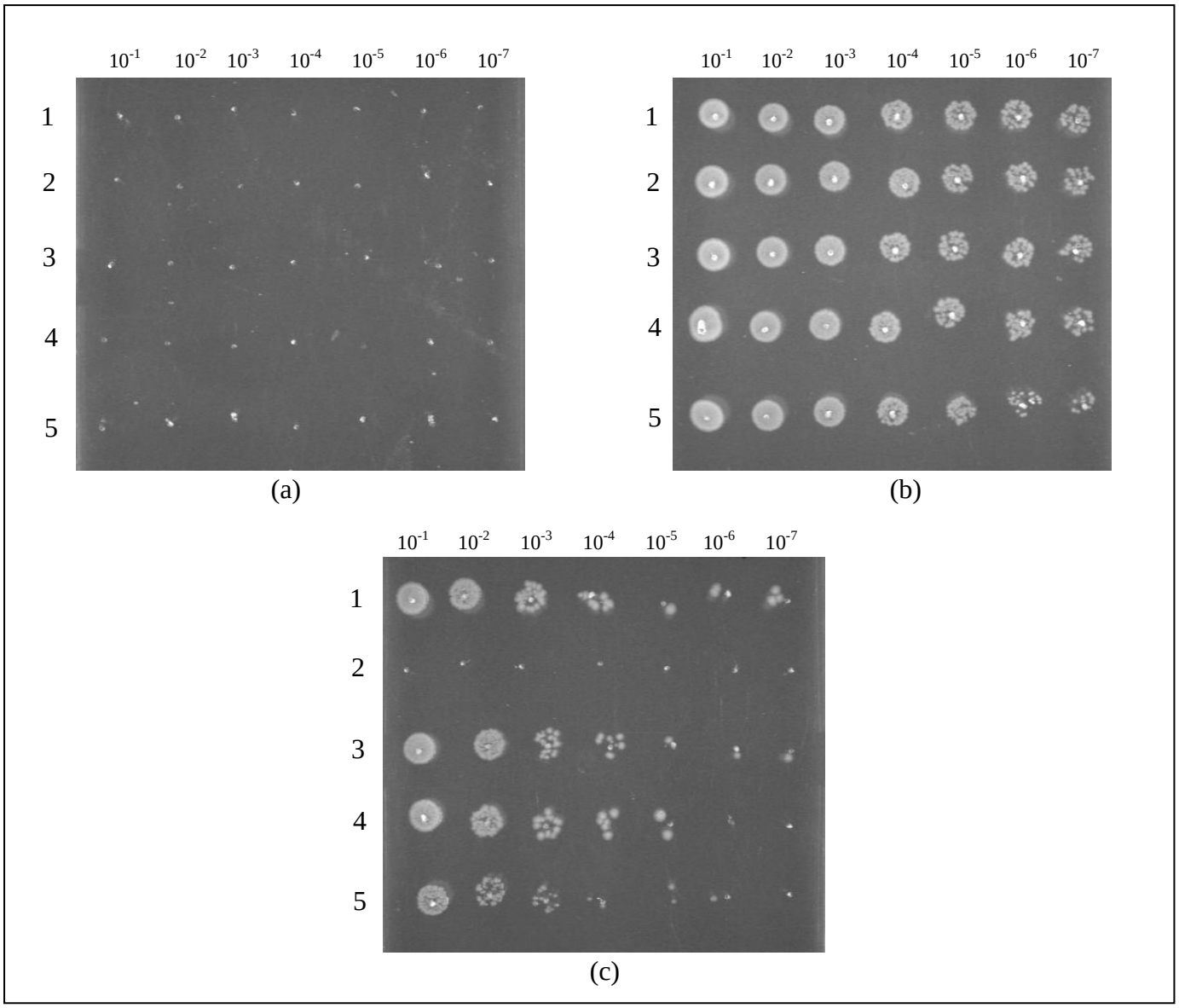


Fig 3.13: Images showing complementation of Cpn60.2 from *M. smegmatis* and *M. tuberculosis* in AI90 at 37 deg Celsius. A) Cells plated on LB with 10% sucrose, 100µg/ml ampicillin, 50µg/ml chloramphenicol, and 0.1 mM IPTG. B) Cells plated on LB with 100µg/ml ampicillin, 50µg/ml chloramphenicol, and 0.1 mM IPTG. C) Cells plated on LB with 10% sucrose, 100µg/ml ampicillin, and 0.1 mM IPTG.

1- AI90/ pOFESL

2- AI90/ pTrc99a vector

3- AI90/ pTrc-GroES-GroEL

4- AI90/ pTrc-GroEs-Cpn60.2 (*M. smegmatis*)

5- AI90/ pTrc-GroES-Cpn60.2 (*M. tuberculosis*)

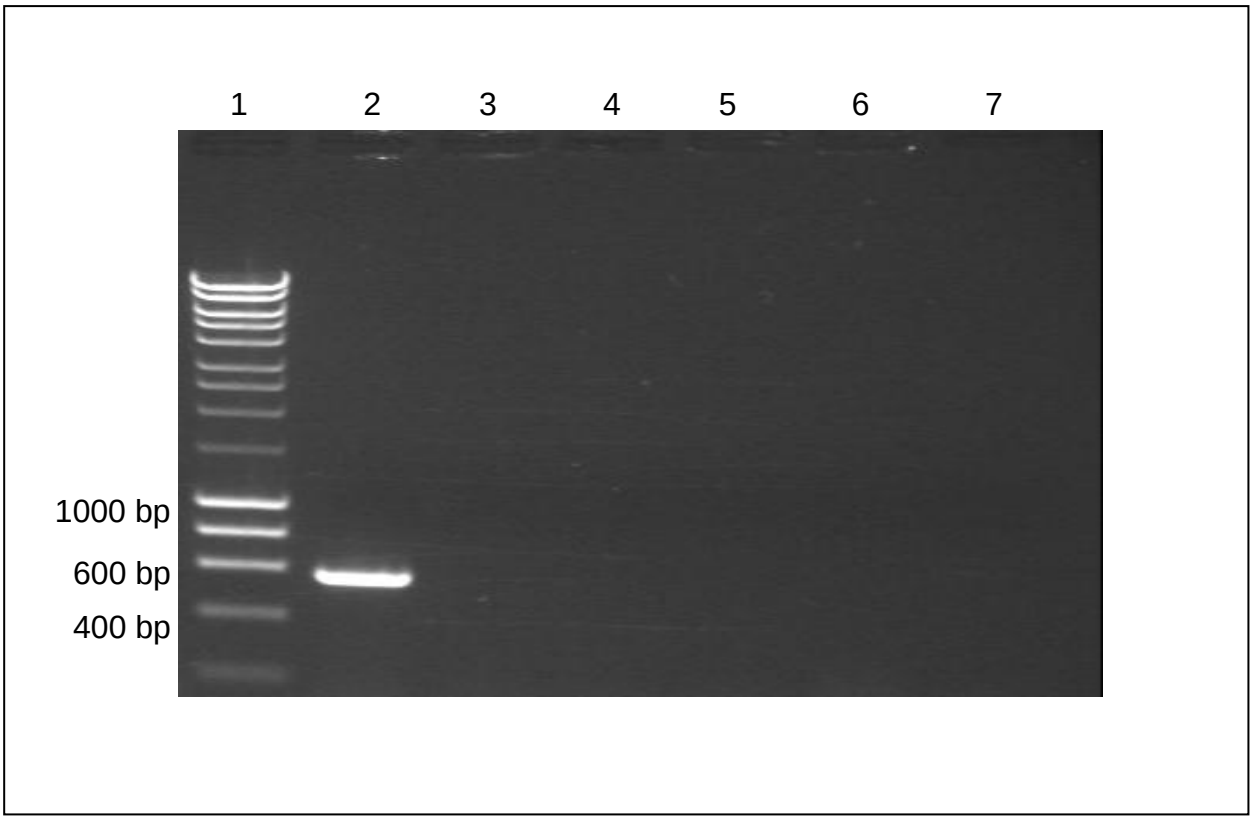


Fig 3.14: PCR using *groEL*-specific primers. The protocol used for colony PCR is described in section 2.4.13.

Lane 1- Kilobase ladder – Hyper ladder I (Bioline)

Lane 2- DH5α (positive control)

Lane 3- CCT (negative control)

Lane 4- Cpn60.2 (*M. smegmatis*)

Lane 5- Cpn60.2 (*M. tuberculosis*)

Lane 6- Cpn60.2 (*M. smegmatis)*

Lane 7- Cpn60.2 (*M. tuberculosis)*

3.7 Complementation by *M. tuberculosis* Cpn60.2 without a cochaperonin

Throughout this chapter it was assumed that Cpn60.2 functions along with GroES and Cpn10 to complement in *E. coli*. However, the ability of Cpn60.2 to function on its own has never been tested, even though it is known that in *M. tuberculosis* Cpn60.1 and Cpn10 constitute an operon while Cpn60.2 is arranged separately on the genome. It is unclear why the main house-keeping chaperonin is not present on the same operon as its cochaperonin. In this section, the ability of Cpn60.2 to complement in the absence of any cochaperonin was analysed. It was hypothesised that Cpn60.2 needs a cochaperonin to function.

For the purpose of this experiment the *groES* gene from the pTrc-GroES-Cpn60.2 plasmid was removed. Sequence analysis reviled that there was an NcoI site just upstream from the *groES* gene and an XbaI site just downstream. Restriction digest was done with both these enzymes as per the protocol described in section 2.4.8. The digested sample was then run on a 1% agarose gel and the backbone vector, excluding the GroES fragment, was gel extracted. This fragment was blunt ended with Klenow and then religated. The resultant pTrc-Cpn60.2 plasmid was transformed into MGM100 for the complementation and expression analysis. It was predicted that in the absence of any cochaperonin, Cpn60.2 would not be able to complement for the loss of endogenous GroEL in *E. coli*.

The results that were obtained showed that in the presence of arabinose the strain grew well, but in the presence of glucose only the strain carrying both the chaperonin and cochaperonin was able to complement. To ensure the lack of complementation was not caused by the lack of expression, SDS-PAGE gels were run. The gel showed that Cpn60.2 was being over-expressed without GroES. This confirmed the hypothesis that Cpn60.2 cannot function without a cochaperonin.

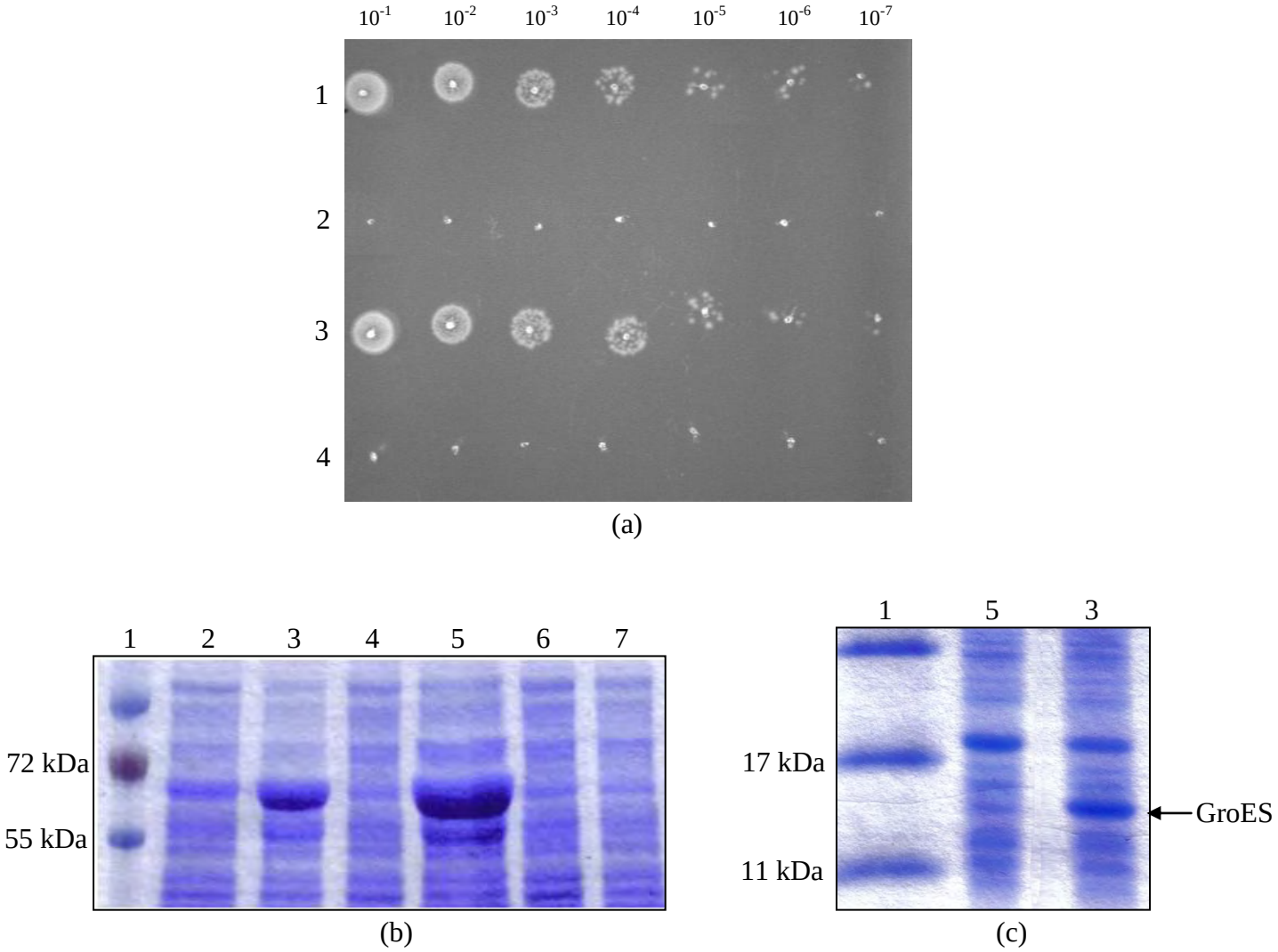


Fig 3.15: Complementation and expression of pTrc-60.2.

(a) Lack of complementation of pTrc-Cpn60.2 in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG. 1- MGM100 pTrc-GroES-Cpn60.2; 2- MGM100 pTrc-Cpn60.2; 3- MGM100 pTrc-GroES-GroEL; 4- MGM100 pTrc99a.

(b) & (c) SDS-PAGE showing over-expression of Cpn60.2, and lack of expression of GroES. Samples were taken from MGM100 cultures after 5 hours incubation in 0.2% arabinose. Every even numbered lane has no IPTG and every odd numbered lane has 0.1 mM IPTG. Lane 1- Prestained protein marker (Fermentas); Lane 2 & 3- pTrc-GroES-Cpn60.2; Lane 4 & 5- pTrc-Cpn60.2; Lane 6 & 7- pTrc99a.

Discussion

It has been estimated that approximately 10% of the cellular proteins in *E. coli* interact with GroEL under normal conditions (Houry *et al.*, 1999, Kerner *et al.*, 2005). MetF and MetK are two of those proteins. It is known that the *metE* gene is repressed by MetF and MetK both of which have been found to be obligate GroEL substrates (Kerner *et al.*, 2005), so the lack of GroEL in the cell would result in the over-expression of MetE (Chapman *et al.*, 2006). This observation was confirmed in earlier experiments conducted in the lab, where the levels of MetE started to increase during the reduction of endogenous GroEL. In the results obtained here, MetE was distinguishable on the gels when the endogenous GroEL was suppressed. This result is interesting as it proved to be a very useful marker for the reduction or loss of function of GroEL in the cells. The main reason why no MetE induction was observed on the majority of protein gels shown in this chapter was because all the samples were taken from cells grown in arabinose, which does not repress the expression of GroEL. As such, these samples had enough GroEL present within them to stop the aggregation of MetE. However, MetE induction could be seen from cells containing Cpn60.1 (both with GroES and Cpn10) when the protein samples were made from cultures grown on glucose (fig 3.4b).

When the multiple *groEL* genes were found in *Mycobacterium tuberculosis* it was unknown why such a system would be needed. It was initially thought that since *cpn60.1* gene was in the same operon as the cochaperonin *cpn10*, it would be the house-keeping chaperonin. However, it was found that this is not the case, and that Cpn60.1 is not required by *M. tuberculosis* for its survival (Hu *et al.*, 2008). Earlier work conducted in the lab suggested that when *M. tuberculosis* Cpn60.1 and Cpn60.2 were introduced into *E. coli* and the endogenous GroEL expression was suppressed, only the cells with Cpn60.2 were able to survive. The results that have been obtained from the experiments in this chapter agree with

this observation. The results show that only Cpn60.2 and not Cpn60.1 is able to complement for the loss of endogenous GroEL in *E. coli*. It also shows that the cochaperonin Cpn10 is capable of functioning together with Cpn60.2 in *E. coli*. This result agrees with the understanding that Cpn60.2 is the main house-keeping chaperonin in *Mycobacterium tuberculosis*, while Cpn60.1 seems to have been adapted to play a role in other cellular functions. It is already known that Cpn60.1 has very potent cytokine properties as compared with Cpn60.2 (Lewthwaite *et al.*, 2001). So it is very probable that, with regards to *M. tuberculosis*, the Cpn60.1 is in fact a very important factor in the causation of infection. Further experiments will need to be done to determine the reasons behind the differences in the overall function of these chaperonins. One of these experiments involves domain swaps between the *M. tuberculosis* chaperonins and *E. coli* chaperonins and has been discussed in chapter 4.

While there may be an issue of low levels of protein expression for Cpn60.1, the use of IPTG at higher concentrations showed that sustained high levels of chaperonins can be lethal for cell survival. The addition of IPTG at a concentration of 0.5 mM showed a loss in the number of viable cells with the combinations which otherwise would have complemented without IPTG. The use of the pBAD system while providing a good solution to suppress any background expression of endogenous GroEL, has resulted in the loss of MGM100 being able to control the expression of the homologues according to its requirements. As a result, it may have proved deleterious when expressed at very high levels. Since chaperonins like GroEL use hydrophobic residues to interact with substrates, it can be speculated that when GroEL is expressed at higher than needed levels they may interact with substrates other than their own, some of which could be vital for cell survival. In such a case if GroEL is unable to properly fold these proteins and instead just traps them, then over a period of time the lack of these essential proteins would lead to cell death. Reducing the concentration of IPTG, and in

doing so, reducing the over-expression of the chaperonins resulted in better cell growth and better complementation.

It is known that *E. coli* codon usage differs from Mycobacterial codon usage (Pan *et al.*, 1998, Sharp *et al.*, 2005). As a result, the GroEL homologues may not be expressing as well in *E. coli* as they would have been in *Mycobacterium tuberculosis*. To ensure no codon bias was involved in the expression of the chaperonins, the plasmid pRARE was used in section 3.2.2. This pRARE plasmid encodes tRNAs for codons that rarely occur in *E. coli*. The results that were obtained, however, showed no difference to the original complementation data. Under appropriate conditions Cpn60.2 complemented well for the loss of endogenous GroEL, while Cpn60.1 showed no signs of increased expression or any signs of complementation. This result was not surprising as it was shown that Cpn60.2 could complement in *E. coli* without the pRARE plasmid in section 3.2. So the lack of complementation by Cpn60.1 was assumed to be its inability to function in *E. coli* rather than its reduced expression. However, low expression as the cause of lack of complementation was not formally ruled out.

Results in Tab21 were the same as those in MGM100. The lack of complementation by Cpn60.1 could still be speculated to be due to low levels of protein expression within the cells. However, a key point to note here is that, while the expression of the Mycobacterial chaperonin 60.1 may have been too low to allow complementation, it was still higher than that of GroEL under normal cellular conditions. One method of trying to increase the protein expression further was by transferring the genes onto a pET vector. This vector has a T7 RNA polymerase promoter and will increase the expression levels of any gene located downstream when induced with T7 RNA polymerase. Since Tab21 has a T7 RNA polymerase gene, it should have, at least theoretically, resulted in higher levels of protein

expression. However, from the results obtained it was observed that the protein expression levels were lower than that achieved by the pTrc plasmids in MGM100. As has already been discussed, an accurate complementation analysis could not be done as cell growth was extremely slow on solid media with glucose and arabinose. A possible reason could be that in the presence of glucose, Cpn60.1 is not capable of complementing for the loss of GroEL and the loss of the cells ability to express its endogenous GroEL resulted in cell death. In the presence of arabinose however, the reason is unclear. As discussed in sections 3.2 and 3.3, it was only the Cpn10-60.1 combination that did not show any prominent increase in expression. Further work will need to be done to understand why it is that this particular combination of chaperonin/cochaperonin fails to express at higher levels and if possible to try and achieve levels of expression matching those seen in the other chaperonin/cochaperonin combinations.

In this chapter the cochaperonins were also analysed. The experiments tested the ability of GroES to function with the Mycobacterial homologues, while also testing the ability of Cpn10 to function in *E. coli*. It was shown that that GroES can function as a cochaperonin for Mycobacterial chaperonins and also that Cpn10 can function as a cochaperonin in *E. coli*. However, from sequencing data it was noted that the starting codon for Cpn10 was valine rather than methionine, which is less prominent as a start codon. It was hypothesised that the starting codon of Cpn10 may not allow the ribosome to form a strong enough initiation complex resulting in lower expression of Cpn60.1, which is just downstream from Cpn10 (Sussman *et al.*, 1996). However, this does not explain why Cpn60.2 when co-expressed with Cpn10 had much higher expression levels. Even so, it could be speculated that Cpn60.1 needs the Mycobacterial cochaperonin to function. In such a case, over-expression of Cpn60.1 in the presence of GroES may not yield a fully functional chaperonin. Since it was only possible

to distinguish over-expression of Cpn60.1 in the presence of GroES, attempts were made to try and increase the expression of Cpn60.1 along with Cpn10. The results showed that the complementation data with the ATG form of Cpn10 was unaltered. It also showed that changing the start codon of Cpn10 only slightly increased the expression of cochaperonin but had no affects on the expression levels of the chaperonins downstream from it.

Even attempts made at muteganising Cpn60.1 using XL1-Red did not yield any desirable products. There is however some issues with this experiment, including the fact that the pTrc-GroES-Cpn60.1 plasmid was only passaged once through the XL1-Red strain. Although it's a highly mutagenic strain, the difference between Cpn60.1 and Cpn60.2 may be large enough to require more than a few mutations, and as such may need to be mutagenised multiple times. In the end, no desired Cpn60.1 mutant could be found. Future work in this regard could also make use of other mutagenic chemicals such as hydroxylamine or error prone PCR.

The aim of this study was to characterize the functional properties of Cpn60.1 and Cpn60.2. From all the results obtained here it has become evident that even though Cpn60.1 can be expressed at a higher than basal levels of GroEL, it still fails to complement. Cpn60.2 on the other hand complements extremely well both with Cpn10 and GroES, suggesting it is the main house-keeping chaperonin in Mycobacteria. This was confirmed with the complementation tests done in the Δ*groEL* strain AI90. It proved that Cpn60.2 from both *M. tuberculosis* and *M. smegmatis* can function along with GroES in *E. coli* in the complete absence of GroEL.

Chapter 4

Construction and analysis of chimeric chaperonins

This laminated page contains the table listing all the chimeras that were attempted in this chapter. Please detach the page and use it as a reference while reading through the results, as only the names of the constructs will be used.

Name	Construct	GroEL amino acid seqs.	Cpn60.1 amino acid seqs.	Cpn60.2 amino acid seqs.
	Blue –Cpn60.2; Red - Cpn60.1; Green – GroEL			
AHC		144-403		1-143; 404-542
GBI		1-144; 405-548		145-504
JNL		192-375		1-191; 376-541
MKO		1-191; 376-541		192-375
DHF		144-403	1-143; 404-541	
GEI		1-144; 405-548	145-404	
PNR		192-375	1-191; 376-540	
MQO		1-191; 376-549	192-375	
AEC			144-403	1-143; 404-541
PKR			1-191; 376-549	192-375
DBF			1-143; 404-541	144-402
JQL			192-375	1-191; 376-541

The white, orange, and black bars above the chimeric bars represent the approximate location of the equatorial, intermediate, and apical domains respectively.

Background

In the last chapter I showed that Cpn60.2 from *M. tuberculosis* can function in *E. coli* in the absence of endogenous GroEL. It was also demonstrated that Cpn60.1 was unable to do the same. As a result, it was concluded that there was a possibility of the chaperonins being specialised to fulfil specific functions within *M. tuberculosis,* such that only Cpn60.2 could function as the main house-keeping chaperone. There are a number of references that support this idea in *Actinobacteria* (Hu *et al.*, 2008, Kim *et al.*, 2003, Ojha *et al.*, 2005, Qamra *et al.*, 2005, Cehovin *et al.*, 2010).

In most cases, chaperonins are large oligomeric complexes with a cage like structure. This allows substrate proteins to fold within the cavity without being subjected to interference by other complexes (Horwich *et al.*, 2006). Although it has already been shown that Cpn60.2 can complement in *E. coli*, one of the aspects of chaperonin function that we did not look at in the last chapter was structural stability, as in, the ability of the chaperonins to maintain higher oligomeric structures. It has been reported that that the Mycobacterial chaperonins have a tendency to exist as lower oligomers (Qamra *et al.*, 2004). However, since Cpn60.2 can function in *E. coli* in the absence of GroEL, it can be predicted that Cpn60.2 should be forming some higher oligomeric structure during its reaction cycle, even if it is very unstable. This was shown to be the case in some recent experiments done in our lab where Cpn60.2 was analysed using AUC in various buffer conditions. It was noted that under high salt conditions and in the presence of either ATP or ADP Cpn60.2 does form larger oligomeric structures and this was confirmed to be tetradecameric by MS (Elsa Zacco, University of Birmingham; Dr. J Freeke and Dr. J Benesch, University of Oxford). This was also supported by the fact that Cpn60.2 can function with both Cpn10 and GroES, both of which are known to oligomerise as heptamers (Roberts *et al.*, 2003, Hunt *et al.*, 1996). Another aspect that

could determine the complementation ability of chaperonins is substrate specificity. As both chaperonins have evolved at different rates (Hughes, 1993), and have been shown to perform different functions in *M. tuberculosis* (Hu et al., 2008), there is a possibility that they can recognise a different set of substrates.

This chapter will be focusing on two primary objectives. Firstly, to attempt to change the substrate specificity of Cpn60.1 such that it is then able to function in *E. coli*. Secondly, to see if it is possible to improve the unstable oligomeric nature of the *M. tuberculosis* chaperonins and to see how this affects their function in *E. coli*. To try and achieve these objectives, I decided to make chimeric proteins using domain swap techniques. The plan was to swap the apical domains of the three chaperonins between each other. In the process of doing so, I would also in effect be making chimeras where the equatorial and intermediate domains were being swapped. It was hypothesised that the apical domain of GroEL or Cpn60.2, if swapped with the apical domain of Cpn60.1 would allow the chimeric Cpn60.1 to recognise and fold GroEL substrates. It was also hypothesised that the domain swaps between GroEL and Cpn60.2 would improve the oligomeric stability of the Cpn60.2 chaperonin when the equatorial domain of GroEL is present in the chimera.

As such it was predicted that apical domain swap between the chaperonins would confer the ability to complement in cases where they normally wouldn't, and also that the swapped equatorial regions would result in higher stability in the oligomeric forms of the chaperonins.

Experimental design

To make the various chimeric proteins, a modified PCR extension technique was used (Warrens *et al.*, 1997)(Fig 4.1). The general protocol involved the use of specifically designed primers that had overlapping sequences, which allows the amplification of the different domains from the chaperonin sequence.

Using GroEL as an example, the equatorial domain is present at the two extremities of the chaperonin polypeptide and consists of amino acid residues 6 to 133 and 409 to 523. The intermediate domain that links the equatorial and apical domains of GroEL also spans two regions of the polypeptide and includes residues 134 to 190 and 377 to 408. The apical domain is present in the centre of the polypeptide chain and constitutes residues 191 to 376. The different sections would then be annealed together in the required order and amplified further. The details of primers, numbered 1 to 54, used in this chapter are listed in the material and methods section.

Using Cpn60.2 (referred to as ABC) and Cpn60.1 (referred to as DEF) as examples, the process has been outlined in the figure below (fig 4.1). The first PCR amplifies the three sections of the chaperonin while leaving 12bp overhangs that were designed to overlap the equivalent regions on the other chaperonin. Sections B and E constitutes the apical domain of Cpn60.2 and Cpn60.1 respectively, while sections A, C, and D, F constitute the equatorial and intermediate domains. For the 2nd PCR stage, the required sections from the 1st PCR stage were mixed and used as templates. This template was then amplified further with appropriate primers. In the 3rd and final PCR step the products from the 2nd PCR step and the final remaining section from the 1st PCR step were combined to complete the sequence and

introduce the apical domain of one chaperonin into the sequence of the other. The resulting chimera was then amplified using the original flanking primers and purified for the cloning process.

Since I already had plasmids with the original chaperonins and cochaperonins under the control of the IPTG inducible promoter, it was decided that rather than trying to clone the entire PCR product into a new plasmid, the chimeras would be swapped with the chaperonins using specific restriction enzymes. The restriction enzymes used were picked based on the protein sequences to find sites that would only be cut once at either side of the apical domain, but also ensured that the digested fragment contained the swapped domains (fig 4.2). For Cpn60.2 the restriction enzymes were BamHI and BsmI, for Cpn60.1 they were Bsu36I and BmgBI, and for GroEL they were BstBI and DraIII. The plasmids used for the cloning process were pTrc-GroES-Cpn60.1, pTrc-GroES-Cpn60.2, and pTrc-GroES-GroEL. Once cloned, the plasmids were sequenced to ensure there were no errors present in the genetic sequence and that the required swaps had been made.

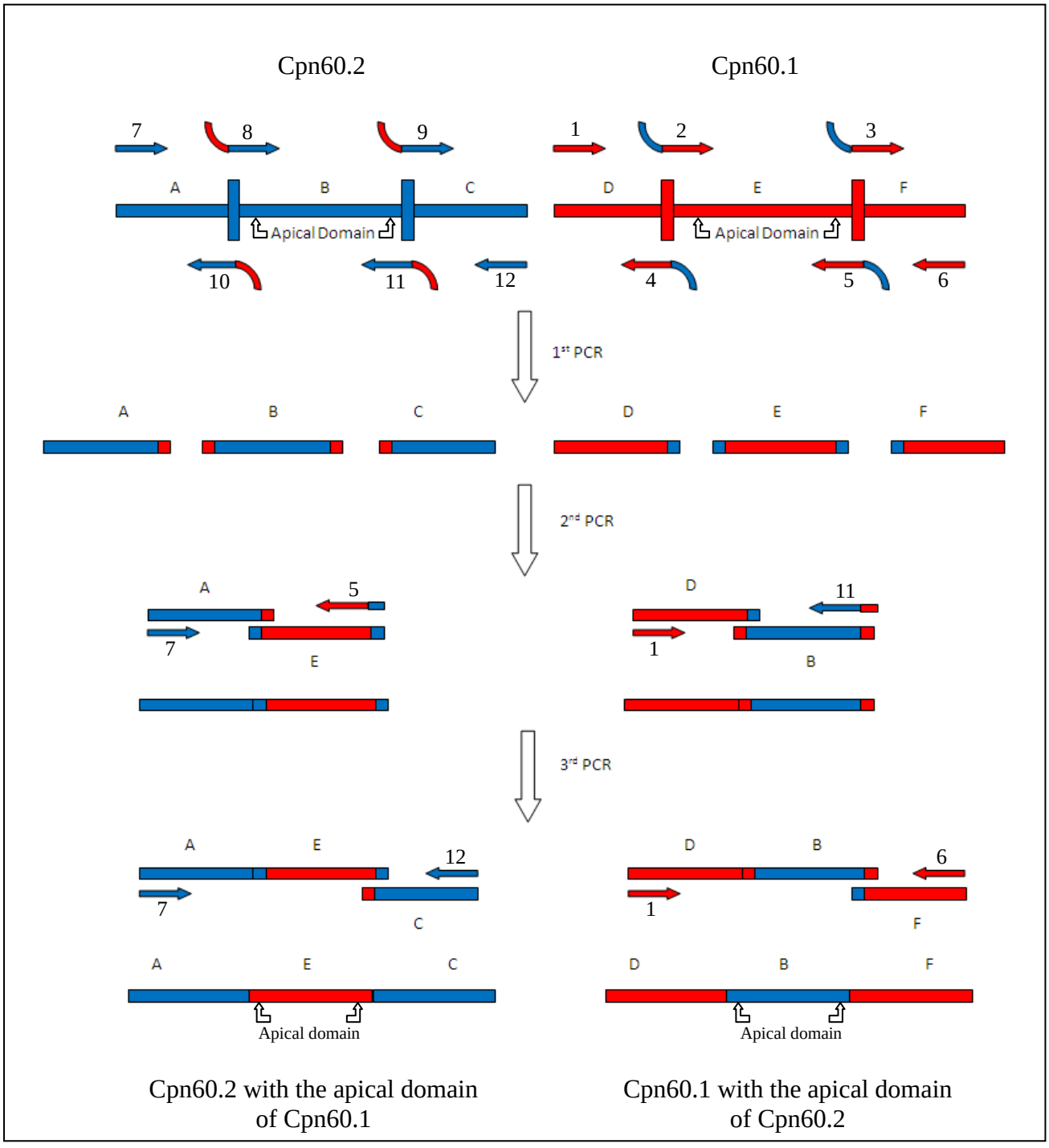


Fig 4.1: Schematic representation of the steps involved in constructing chimeric proteins using a modified PCR extension technique (not to scale). The red bar represents the sequence for Cpn60.1, while the blue bar represents Cpn60.2. The central segment of each bar is the location of the apical domain sequence, represented by B and E for Cpn60.2 and Cpn60.1 respectively. The segments on either side of the apical domain include both the intermediate domain and the equatorial domain.

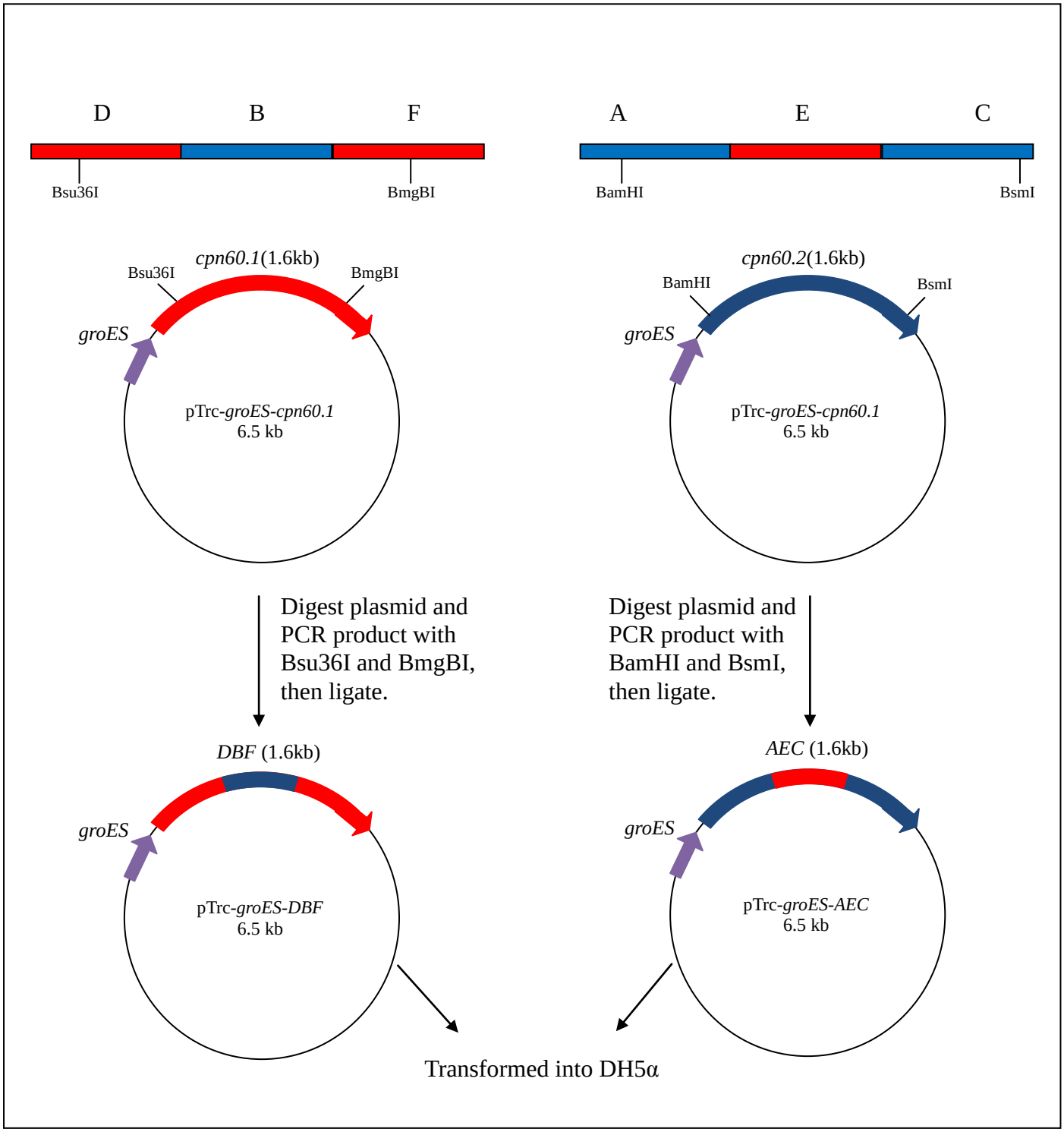


Fig 4.2: Schematic representation of the cloning process of the chaperonin chimeras. In all cases the chimeras were cloned into pTrc plasmids with GroES as the cochaperone. The restriction enzymes used for Cpn60.1 as mentioned were Bsu36I and BmgBI, while for Cpn60.2 they were BamHI and BsmI. For GroEL, the sites used were BstBI and DraIII. The backbone and insert were then ligated to obtain the required plasmids and transformed into DH5α.

The full list of all the chimeras that were attempted for this chapter have been shown below in table 4.1. In these experiments, two different groups of chimeras were attempted. Initially, I constructed chimeras where the apical domains and segments of the intermediate domains were swapped. These swaps were attempted as a result of similar work done on mitochondrial chaperonin by Nielsen and Cowan (Nielsen & Cowan, 1998). In these chimeras the junction points of the fusions were at I144 and A405 for GroEL and I143 and A403 for Cpn60.1 and Cpn60.2. For the purpose of this chapter, this group of chimeras will be referred to as imprecise chimeras (due to the imprecise nature of the apical domain swap). The complementation and expression results obtained from these chimeras can be seen in section 4.1. After the construction of these chimeras it was noted that the inclusion of partial segments of the intermediate domain, rather than the entire domain, was causing issues during the chaperonin cycle. In some cases this would result in cell death, while in others the complete loss of ability to function where they would not be expected to. To ensure that the results of these experiments were not being influenced by the inclusion of a partial segment of the intermediate domain, it was decided that a second group of chimeras, where only the apical domain was precisely swapped, would be made. The junction points for the fusions of the precise domain swaps were at E191 and V377 for GroEL and G190 and V376 for Cpn60.1 and Cpn60.2. For the remainder of this chapter, these chimeras will be referred to as precise chimeras (due to the precise nature of the apical domain swaps). The complementation and expression results of these chimeras can be seen in section 4.2.

Using these chimeric proteins, attempts were made to ascertain the roles played by the apical and equatorial domains in the functional and structural properties of chaperonins. The table below lists all the chimeras attempted, and a simple hypothesis associated with each one.

Name	Construct	GroEL amino acid seqs.	Cpn60.1 amino acid seqs.	Cpn60.2 amino acid seqs.	Expected to complement?	Expected to oligomerise?
	Blue –Cpn60.2; Red - Cpn60.1; Green – GroEL					
AHC		144-403		1-143; 404-542	Yes	No
GBI		1-144; 405-548		145-504	Yes	Yes
JNL		192-375		1-191; 376-541	Yes	No
MKO		1-191; 376-541		192-375	Yes	Yes
DHF		144-403	1-143; 404-541		Yes	No
GEI		1-144; 405-548	145-404		No	Yes
PNR		192-375	1-191; 376-540		Yes	No
MQO		1-191; 376-549	192-375		No	Yes
AEC			144-403	1-143; 404-541	No	No
JQL			192-375	1-191; 376-541	No	No
PKR			1-191; 376-549	192-375	Yes	No
DBF			1-143; 404-541	144-402	Yes	No

Table 4.1: The list of chimeric chaperonins made. The white, orange, and black bars above the chimeric bars represent the approximate location of the equatorial, intermediate, and apical domains, respectively. The table also includes some predictions on the functional and structural properties.

4.1 Imprecise domain swaps

As discussed above, the first group of domain swaps were imprecise. The segments that were swapped included not only the apical domain but also a large part of the intermediate domain. These chimeras were constructed so that the junctions between the different genes were in the same relative positions as those described in a study which used chimeras between GroEL and the mitochondrial protein Hsp60 (Nielsen & Cowan, 1998). Since the PCR protocol used in the construction of all the imprecise chimeras was the same, a representative example of this approach is shown in figure 4.3, using Cpn60.2 (ABC) and GroEL (GHI) as examples. A schematic representation of this process has been shown in figure 4.1.

The first PCR resulted in 3 segments being amplified from each chaperonin (fig 4.3a). The first band constituted the N-terminal equatorial domain sequence and a small section of the intermediate domain sequence. The second band included the remainder of the N-terminal intermediate domain sequence, the entire apical domain sequence, and a large section of the C-terminal intermediate domain sequence. The third band included the remaining sequence of the C-terminal intermediate and equatorial domains. The bands were then gel extracted using the protocol described in section 2.4 and used in the next reaction. For the second PCR the products of the first reaction were used as the templates, and then amplified further for the third and final reaction (fig 4.3b). The final reaction combines the product of the second PCR with the remaining C-terminal product of the first reaction and resulted in the completed chimera (fig4.3c). This was then cloned into the pTrc plasmids using restriction digests, as described above, and then transformed into DH5α using the protocol described in section 2.4.16. Every construct was then sequenced to ensure the swaps were made at the right points, and did not have any undesired mutations.

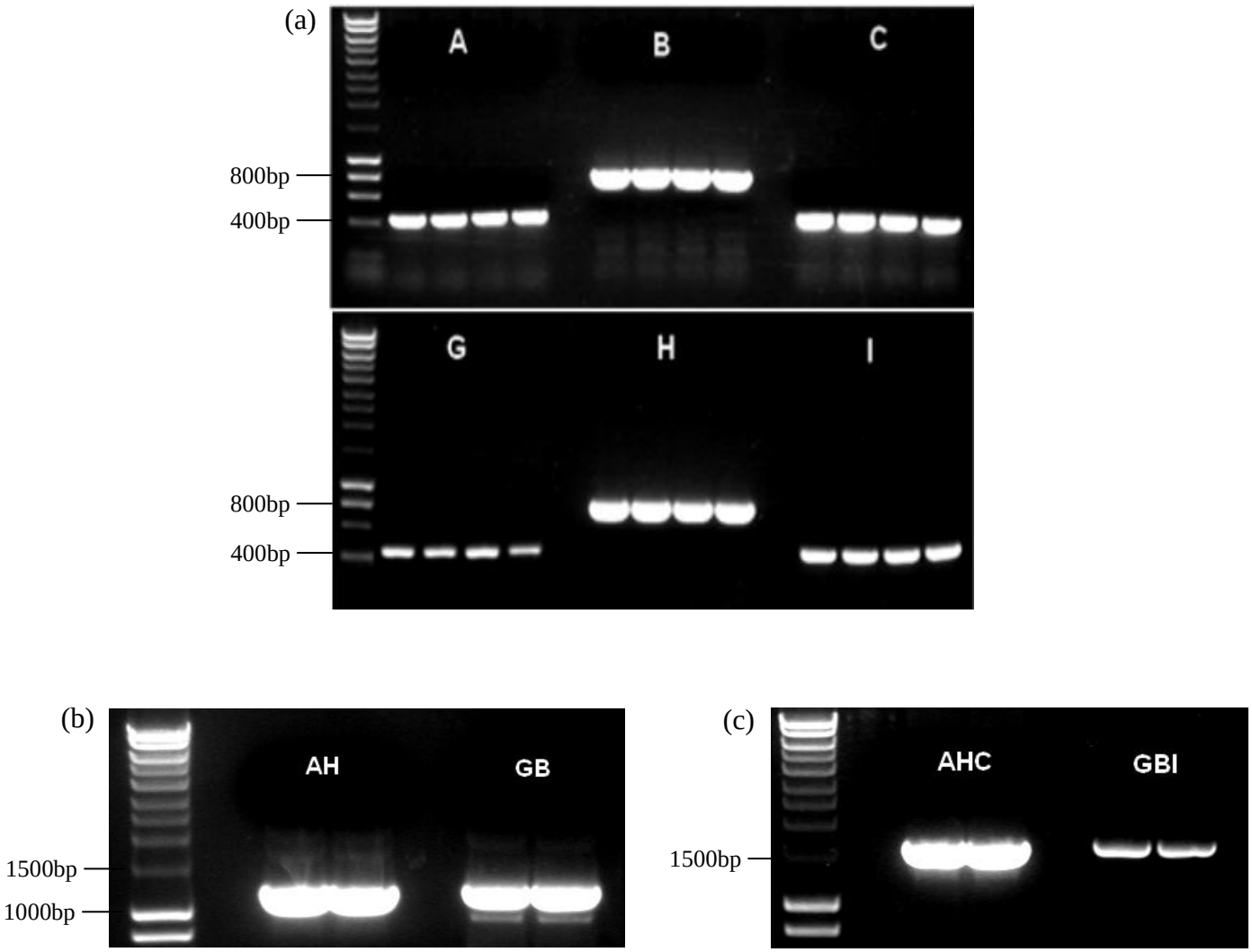


Fig 4.3: Agarose gel for each of the 3 PCRs for Cpn60.2 (ABC) and GroEL (GHI). (a) The three bands obtained from the first PCR for each chaperonin. (b) Single band obtained after the second PCR. (c) The third PCR yields the final combination of chimera.

4.1.1 Chimera AEC

This is a chimera where an imprecise domain swap has been made between Cpn60.2 and Cpn60.1. In this chimera the equatorial domain sequence and a partial segment of the intermediate domain sequence are from Cpn60.2 while the apical domain sequence and the remainder of the intermediate domain sequence come from Cpn60.1.

This chimera was cloned into pTrc99A, to give the plasmid pTrc-GroES-AEC, and was sequenced to confirm the junctions between cpn60.1 and cpn60.2, and that no other mutations were present. This pTrc plasmid was then transformed into MGM100 to test its ability to complement in *E. coli*. It was hypothesised that the introduction of the apical domain from Cpn60.1 into Cpn60.2 would result in this chimera failing to complement for the loss of endogenous GroEL because the apical domain from Cpn60.1 is unable to recognise the same substrates as Cpn60.2.

The complementation protocol was the same as that used in the last chapter. The resultant MGM100 + pTrc strain was grown overnight in arabinose, then sub-cultured and grown in 200 ml LB + 0.2% glucose the following morning for 2 hours, and then serially diluted (in ten-fold steps) to spot on to square petri dishes. 5µl of each dilution was spotted onto the plates with the final dilution factor being $10^{7.}$ The plates were then incubated at 37°C, 30°C, and 26°C. Proteins samples were prepared from the same cultures and run on SDS-PAGE gels to check expression.

The results obtained from the complementation plates were as expected. While the cells were viable in the presence of arabinose (endogenous GroEL is expressed), they were unable to rescue *E. coli* MGM100 growth in the presence of glucose (fig 4.4a). SDS-PAGE gels

confirmed that AEC was being expressed in the cells, so the loss of ability to function in *E. coli* was not due to lack of expression (fig 4.4b). This result supported the original hypothesis that the apical domain from Cpn60.1 would not be able to recognise the same substrates as Cpn60.2 and as such it would fail to rescue *E. coli* MGM100 growth. To try and confirm this hypothesis we made a chimera where the apical domain of Cpn60.1 was replaced with that of Cpn60.2. The results and discussion of this can be seen below in section 4.1.2 below.

4.1.2 Chimera DBF

In this chimera the equatorial domain sequence and a partial segment of the intermediate domain sequence is from Cpn60.1, while the apical domain sequence and the remainder of the intermediate domain sequence comes from Cpn60.2.

Attempts at cloning this product into the pTrc plasmid were unsuccessful. After the digestion and ligation process, the pTrc plasmids were transformed into DH5α. Of the 42 clones screened using colony PCRs, restriction digests, and plasmid sequencing, a larger number had truncated sequences where approximately 200-300bp were missing from the C-terminus. In cases where positive colony PCR results were obtained, sequencing results showed that frameshift mutations had occurred. All the attempts to correct the frameshift mutations by using site directed mutagenesis were unsuccessful. Attempts to better control the over expression of this chimera using the pBAD24 plasmid was also unsuccessful. The reason for this is unclear. To ensure this wasn't an issue in the protocol, the same process was used with the plasmid pTrc-GroES-Cpn60.1 (without any chimeric sequence) successfully. It can only be speculated that this chimera was lethal for the cells and as such only cells that contained non-functional variants were able to survive. Due to the results obtained here, the original hypothesis could neither be accepted, nor rejected.

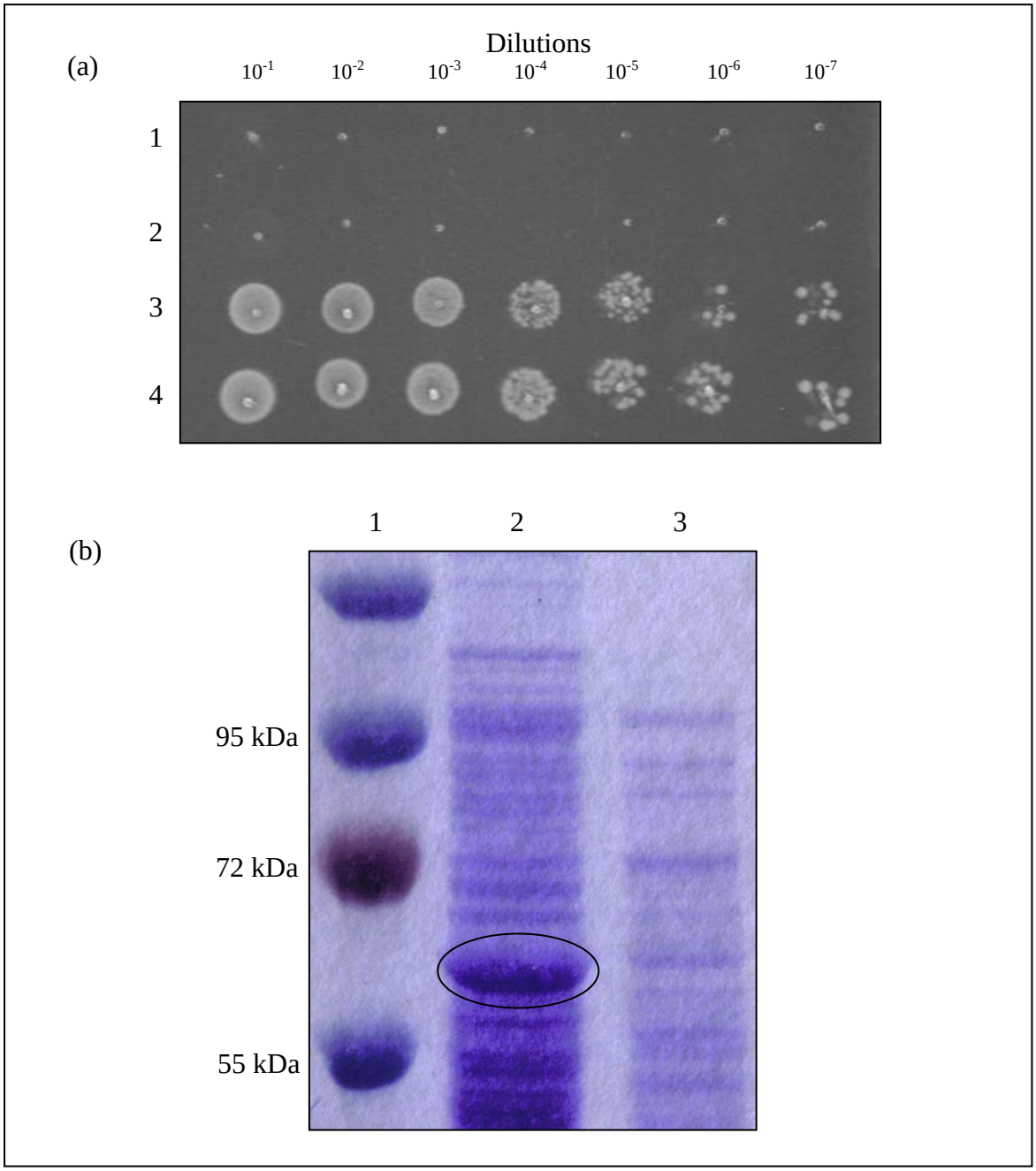


Fig 4.4: Complementation and expression of AEC.

(a) Lack of complementation of AEC in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG. 1- MGM100 pTrc-GroES-AEC, 2- MGM100 pTrc99a, 3- MGM100 pTrc-GroES-GroEL, 4- MGM100 pTrc-GroES-Cpn60.2.

(b) SDS-PAGE showing over expression of AEC (circled). Samples were taken from cultures after 5 hours incubation with 0.1 mM IPTG and 0.2% arabinose. Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc-GroES-AEC; Lane 3: MGM100 pTrc99a.

4.1.3 Chimera AHC

In this chimera the equatorial domain sequence and a partial segment of the intermediate domain sequence is from Cpn60.2, while the apical domain sequence and the remainder of the intermediate domain sequence comes from GroEL.

This chimera was cloned into the pTrc plasmid and sequenced to ensure the appropriate swap has been made before being transformed into MGM100 to tests its ability to complement in *E. coli*. The protocol used for the complementation assay is the same as that used in section 4.1.1. According to the original hypothesis, the introduction of the apical domain from GroEL into Cpn60.2 should not affect its ability to complement in *E. coli,* as it has already been shown that both GroEL and Cpn60.2 can function in *E. coli*. As such it was expected that a domain swap between Cpn60.2 and GroEL would also work.

The results that were obtained however were unexpected. MGM100 cells containing the AHC chimera failed to complement in the absence of endogenous GroEL. To ensure this result was not due to lack of expression an SDS-PAGE gel was run. The protein samples were taken from cultures grown in LB with 0.2% arabinose and 0.1 mM IPTG for 5 hours. The SDS-PAGE gels showed that AHC was being over expressed at levels that were comparable to those of GroEL and Cpn60.2. This indicates the lack of complementation was not being caused by lack of expression. It was reasoned that the imprecise nature of the domain swap could be resulting in the lack of complementation. As the intermediate domain of this chimera consisted of partial sequences from both GroEL and Cpn60.2, there is a possibility that it may have affected the chaperonin cycle, as the intermediate domain is vital for the transfer of the allosteric signal between the apical and equatorial domains. A very interesting point to note is that the native gel of the chimera showed a band with comparable size to that

of GroEL (fig 4.7). This is surprising, as it is known that the equatorial domain that is mainly responsible for the stability of the chaperonin structure and not the apical domain (Braig *et al.*, 1994). The results obtained here suggest that a very stable oligomeric structure may not be necessary for function.

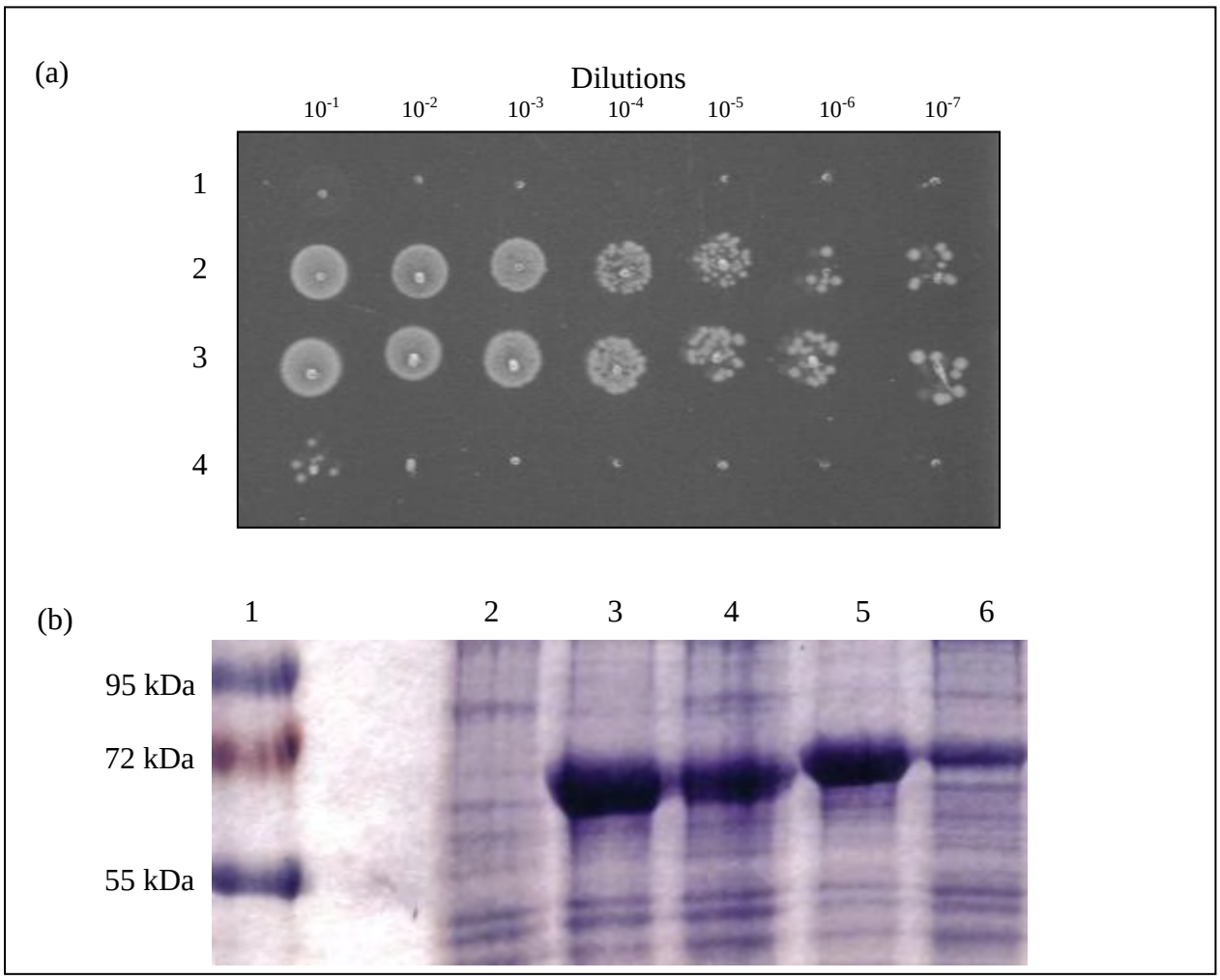


Fig 4.5: Complementation and expression of AHC.

(a) Lack of complementation of AHC in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG. 1- MGM100 pTrc99a, 2- MGM100 pTrc-GroES-GroEL, 3- MGM100 pTrc-GroES-Cpn60.2, 4- MGM100 pTrc-GroES-AHC.

(b) SDS-PAGE gel showing over expression of AEC. Samples were taken from cultures after 5 hours incubation with 0.1 mM IPTG and 0.2% arabinose. Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc99a; Lane 3: MGM100 pTrc-GroES-GroEL with 0.1 mM IPTG, Lane 4: MGM100 pTrc-GroES-Cpn60.2 with 0.1mM IPTG, Lane 5: MGM100 pTrc-GroES-AHC with 0.1 mM IPTG, Lane 6: MGM100 pTrc-GroES-AHC without IPTG.

4.1.4 Chimera GBI

This chimera also consists of an imprecise domain swap between Cpn60.2 and GroEL. In this chimera however, the equatorial domain sequence and a partial segment of the intermediate domain sequence are from GroEL, while the apical domain sequence and the remainder of the intermediate domain sequence comes from Cpn60.2.

As with AHC, it was expected that this chimera would function in *E. coli* MGM100. However, even this chimera was unable to rescue cell growth in glucose (fig 4.6a). Another strange result was that, on SDS-PAGE gels the protein band showed some over expression, but it always showed a band with a lower molecular weight (fig 4.6b). Sequence analysis however, showed that there were no errors in the genetic sequence. The reason for this unclear, but it is possible that these bands were proteolytic fragments. Also, since the chimera's equatorial domain is from GroEL, it was expected that the native gel would show a band at around 840kDa, representing a double ring structure. What was observed however was that, along with the band at around 800kDa, there were also bands present closer to 400kDa (single ring) and a very faint band around 60-70kDa (monomer) (fig 4.7). However, since the chimera had partial sequences of the intermediate domain, it couldn't be claimed that it was entirely due to the apical domain that the oligomeric stability of the chaperonin was being affected.

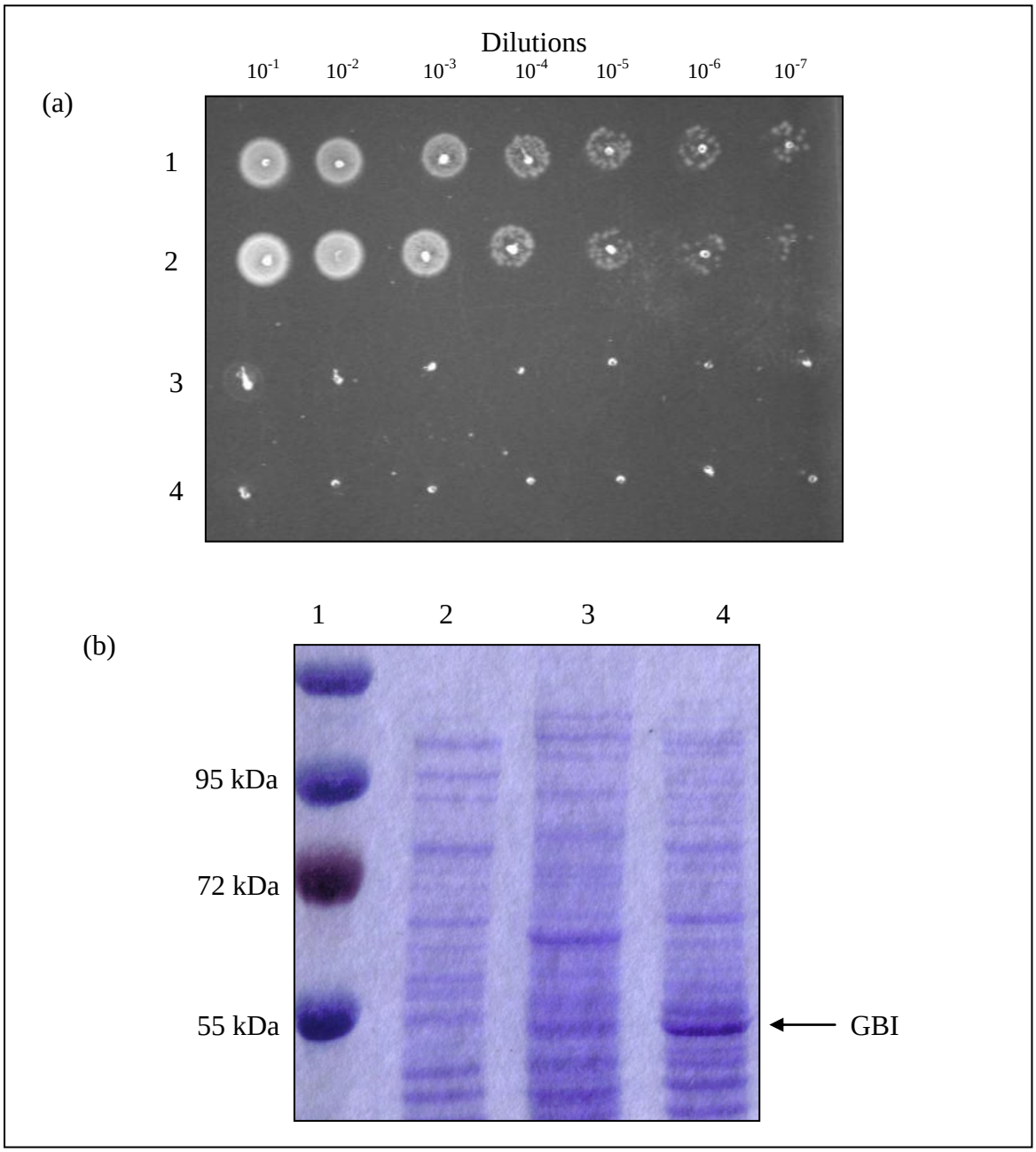


Fig 4.6: Complementation and expression of GBI.

(a) Lack of complementation of GBI in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG. 1- MGM100 pTrc-GroES-GroEL, 2- MGM100 pTrc-GroES-Cpn60.2, 3- MGM100 pTrc99a, 4- MGM100 pTrc-GroES-GBI.

(b) SDS-PAGE gel showing over expression of AEC. Samples were taken from cultures after 3 hours incubation in LB with 0.2% arabinose (with or without 0.1 mM IPTG). Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc99a; Lane 3: MGM100 pTrc-GroES-AHC without IPTG, Lane 4: MGM100 pTrc-GroES-GBI with 0.1 mM IPTG.

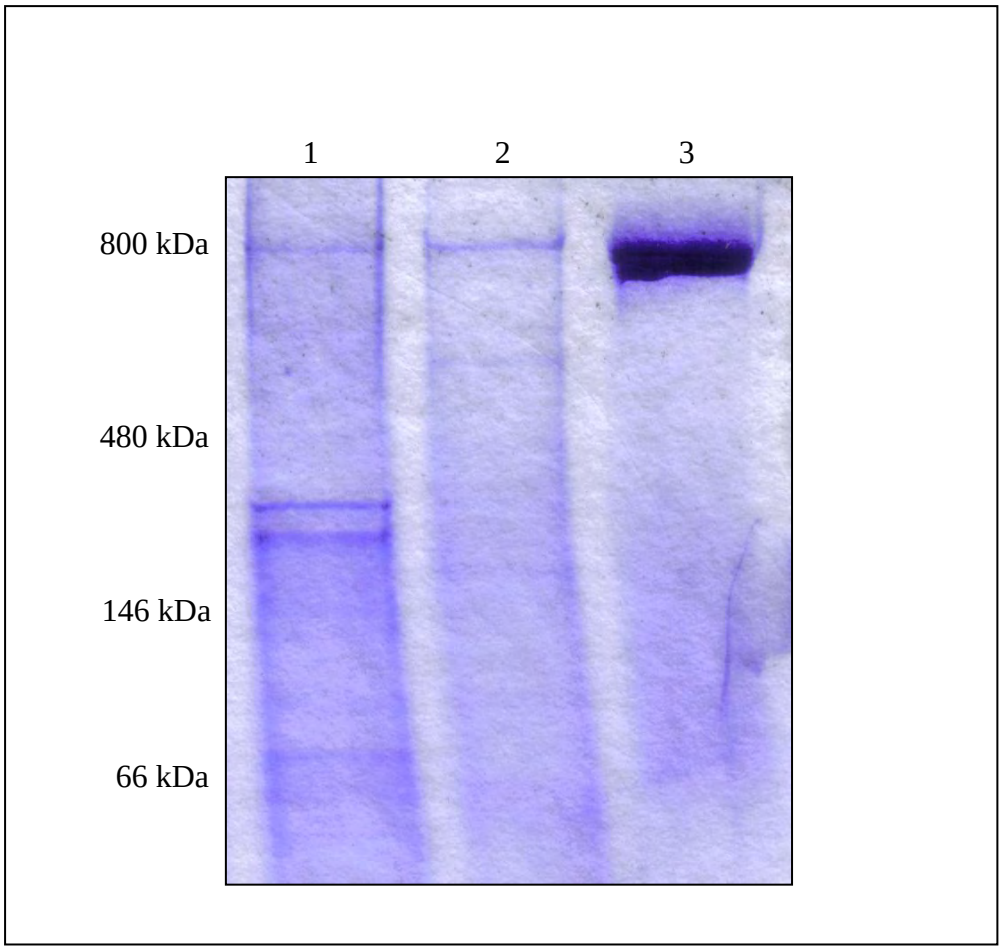


Fig 4.7: Native gel of AHC and GBI. Native samples of AHC and GBI were taken from cells grown in LB with 0.2% glucose and induced with 0.1 mM IPTG for 4 hours. The GroEL native sample was from a purified stock. Lanes are described below:

Lane 1: GroES + GBI

Lane 2: GroES + AHC

Lane 3: GroES + GroEL

4.1.5 Chimeras DHF & GEI

For these chimeras, attempts were made to make imprecise apical domain swaps between GroEL and Cpn60.1. DHF is the chimera where the equatorial domain sequence and a partial segment of the intermediate domain sequence are from Cpn60.1, while the apical domain sequence and the remainder of the intermediate domain sequence come from GroEL. In GEI, the equatorial domain sequence and a partial segment of the intermediate domain sequence are from GroEL, while the apical domain sequence and the remainder of the intermediate domain sequence come from Cpn60.1.

Unfortunately, as was the case with DBF in section 4.1.2, it was not possible to analyse the functional and structural properties of DHF, as the final PCR product could not be cloned into the pTrc plasmid. Attempts at transforming the ligated product into DH5α predominantly caused cell death, while any colonies that were viable after the transformation process usually had frameshift mutations in their chimera sequences. To try and resolve the frameshift issue, site directed mutagenesis was performed using the Quickchange kit (Stratagene), as per the manufacturer's instructions, to reintroduce the missing nucleotide. However, no desirable products were obtained. Again, it can only be speculated that the imprecise nature of the domain swaps may have resulted in this chimera being lethal for the cells. As a result only the clones with non functional chimera were viable.

Chimera GEI could not be analysed as no products were obtained after the final PCR. The reasoning behind this is unclear, as the PCR process was largely successful up to this point. Altering the protocol of the reaction by using a temperature gradient, increasing and reducing extension times, altering the concentration of dNPTs used, and changing the ramping speed had no effect. New primers were also used with no effect. As such, it was decided that the

construction of these chimeras would be reattempted with different junction points, so that only the apical domain was being swapped between them.

Summary of the complementation data of the imprecise domain swaps

A summary of all the complementation data from the imprecise domain swaps has been listed in the table 4.2 below.

	37°C			30°C/26°C		
MGM100	**Arabinose**	**Glucose – IPTG**	**Glucose + IPTG**	**Arabinose**	**Glucose - IPTG**	**Glucose+ IPTG**
GroES + AEC	+	-	-	++	-	-
GroES + DBF	N/A	N/A	N/A	N/A	N/A	N/A
GroES + AHC	+	-	-	++	-	-
GroES + GBI	++	-	-	+++	-	-
GroES + DHF	N/A	N/A	N/A	N/A	N/A	N/A
GroES + GEI	N/A	N/A	N/A	N/A	N/A	N/A
pTrc99a vector only	+++	-	-	+++	-	-
GroES + GroEL	+++	+++	+++	+++	+++	+++
GroES + Cpn60.2	+++	+++	+++	+++	+++	+++

Table 4.2: Complementation data of imprecise domain swapped chimeric proteins in *E. coli* MGM100. The table shows the growth of cells expressing different chimeras with the cochaperonin GroES at different temperatures, with or without 0.1 mM IPTG.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

+ Denotes 0.1 – 1 % survival - Partial complementation

- Denotes <0.1 % survival - No complementation

4.2 Precise domain swaps

As has been discussed throughout section 4.1, the use of imprecise domain swaps could have caused issues in the chaperonin cycle and affected the results of the complementation assays. To ensure that only the apical domain from each chaperonin was swapped and that the intermediate domains remained unaltered, precise domain swapped chimeras were made.

All the primers were redesigned for this section, but the PCR process that was used remained mainly unaltered from the original protocol (section 4.1). In all cases, the PCR products were first run on a 1% agarose gel to confirm fragment sizes and then purified before cloning (not shown). The restriction enzymes remained unaltered from the original protocol.

4.2.1 Chimeras MQO and PNR

In the chimera MQO, the equatorial and intermediate domain is from GroEL while the apical domain is from Cpn60.1. In PNR, the apical domain was from GroEL, while the equatorial and intermediate domains were from Cpn60.1.

Unlike GEI, there were no issues in the PCR process or in the cloning process of MQO. The plasmid was also sequenced to ensure the domain swaps were made at the right junction points before transforming it into MGM100 for the complementation assay. The resultant MGM100 + pTrc strain was grown overnight in arabinose, then sub-cultured and grown in 200 ml LB + 0.2% glucose the following morning for 2 hours. This culture was then serially diluted (in ten-fold steps) to spot on to square petri dishes. 5µl of each dilution was spotted onto the plates with the final dilution factor being $10^{7.}$ The plates were then incubated at 37°C, 30°C, and 26°C. Proteins samples were prepared from the same cultures and run on

SDS-PAGE gels to check expression.

As was the case with the original imprecise domain swaps, it was hypothesised that the apical domain is responsible for substrate recognition, and that the introduction of the apical domain from Cpn60.1 into GroEL would result in this chimera failing to complement for the loss of endogenous GroEL as it is unable to recognise the same substrates as GroEL.

The results obtained from the complementation plates of MQO were as expected. In the presence of arabinose all the cultures grew, but in the presence of glucose when endogenous GroEL expression is suppressed, MQO was unable to complement (fig 4.8a). The expression levels of MQO were examined to ensure that the lack of complementation was not due to a lack in expression. It was observed, on the SDS-PAGE gel, that the expression level of MQO was low but still present (fig 4.8b), and this was further confirmed with MS. Although its expression level was low, it is known that very low numbers of functional chaperonins are needed for normal growth (Lorimer, 1996). On the native gel, a band of approximately 400kDa was observed, which is normally associated with a chaperonin that assembles as a single ring (fig 4.9). This result suggests that the apical domain may also play some role in the oligomeric stability of the chaperonins, as was observed with other chimeric constructs (section 4.1.3 and 4.1.4). However, further experiments would be needed to confirm this observation.

From the results obtained, it was reasoned that if the lack of function was in fact due to the lack of ability to recognise appropriate substrates, then a reciprocal swap between GroEL and Cpn60.1 should yield a fully functional chimeric chaperonin. For this purpose, another chimera was constructed where the apical domain was from GroEL, while the equatorial and

intermediate domains were from Cpn60.1 (referred to as PNR). However, as was the case with DHF, the PCR product could not be cloned into the pTrc plasmid. No viable cells were obtained after transformation into DH5α. It is known, from the experiments done with Cpn60.2 in chapter 3, that the chaperonin constructs on the pTrc plasmid are induced at low levels, even in the absence of IPTG. Since there was a possibility of this chimera being lethal to the cells, even low levels of expression may have been causing cell death. As such, attempts were made to better control its expression using the pBAD24 plasmid. The expression of the chimera from this plasmid can be suppressed by growing the cells on glucose media. Unfortunately, this method was also unsuccessful.

4.2.2 Chimeras JQL and PKR

Chimera JQL is the precise domain swap that has the equatorial and intermediate domain of Cpn60.2 while the apical domain is from Cpn60.1. Chimera PKR has the reciprocal swap with the equatorial and intermediate domains from Cpn60.1, while the apical domain is from Cpn60.2. Both chimeras were made without any issues, and sequenced to ensure no mutations were present. They were then transformed into MGM100 using the protocol described in 2.4.16 to test their functional properties. The complementation method used was unaltered from the previous experiments.

These swaps were made to see if incorporating the apical domain from Cpn60.2 into Cpn60.1 would allow it to function in *E. coli*, and if the reciprocal swap would prevent Cpn60.2 from functioning. It was hypothesised that the apical domain of Cpn60.2 would allow Cpn60.1 to recognise and bind to substrates that would rescue MGM100 growth, while the apical domain of Cpn60.1 would prevent Cpn60.2 from functioning in its regular manner.

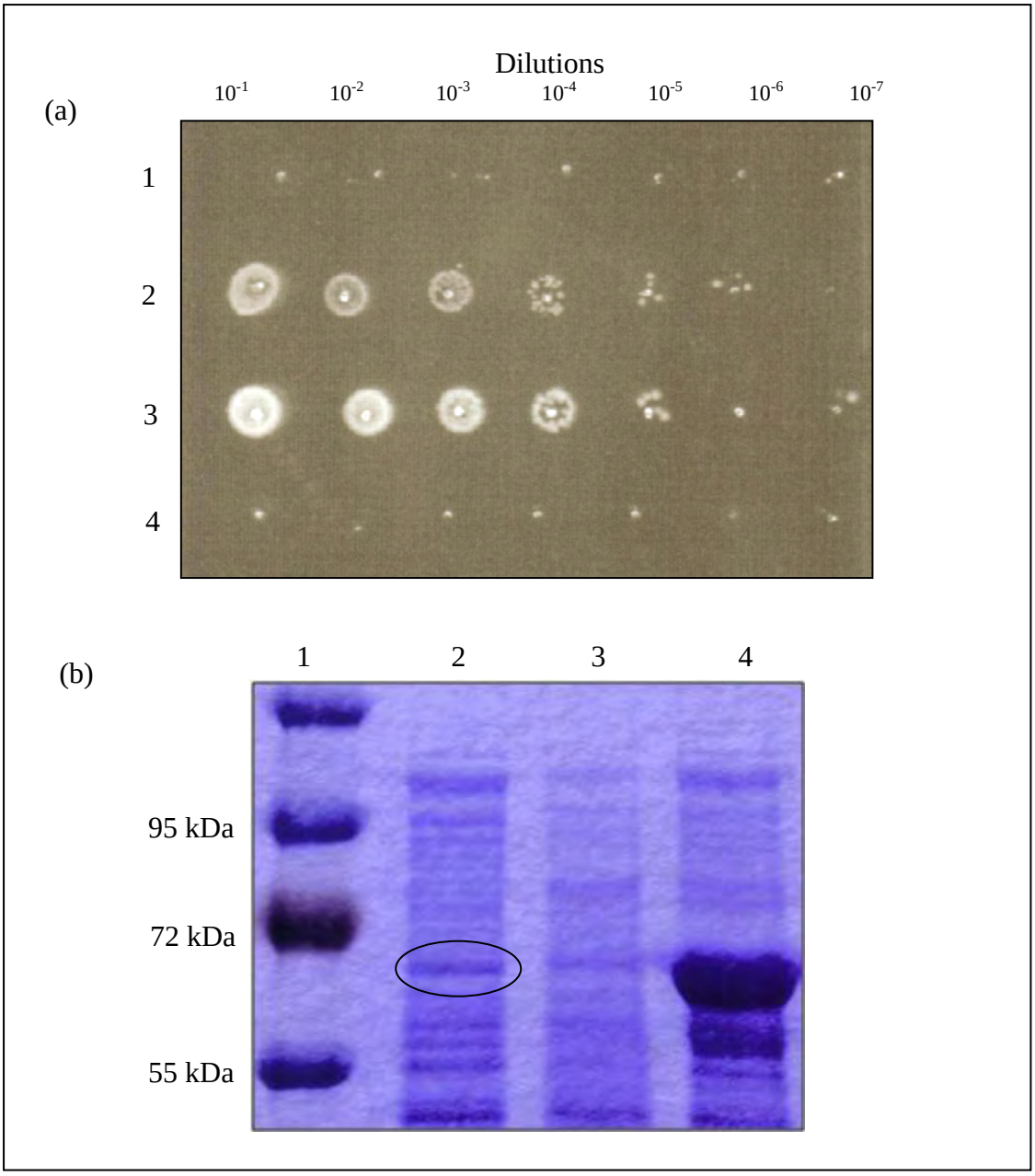


Fig 4.8: Complementation and expression of MQO.

(a) Lack of complementation of MQO in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG. 1- MGM100 pTrc-GroES-MQO, 2- MGM100 pTrc-GroES-GroEL, 3- MGM100 pTrc-GroES-Cpn60.2, 4- MGM100 pTrc99a.

(b) SDS-PAGE gel showing minimal over expression of MQO. Samples were taken from cultures after 4 hours incubation with 0.1 mM IPTG and 0.2% arabinose. Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc-GroES-MQO; Lane 3: MGM100 pTrc99a, Lane 4: MGM100 pTrc-GroES-GroEL.

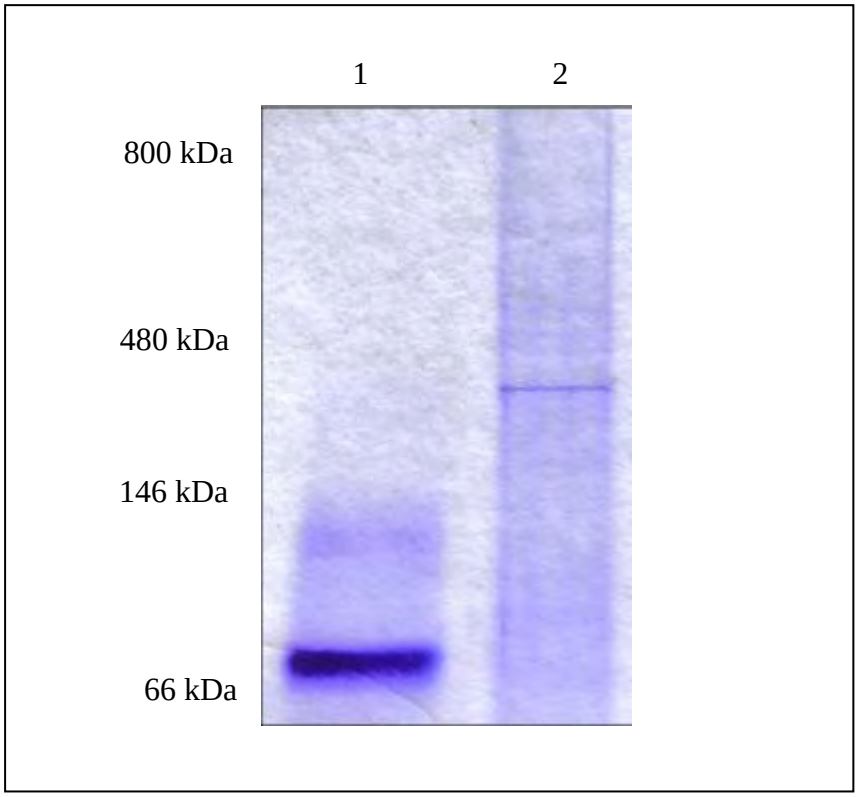


Fig 4.9: Native gel of MQO. Native samples of MQO were taken from cells grown in LB with 0.2% glucose and induced with 0.1 mM IPTG for 4 hours. The Cpn60.2 native sample was from a purified stock. Lanes are described below:

Lane 1: GroES + Cpn60.2

Lane 2: GroES + MQO

The results that were obtained showed that in the presence of arabinose and 0.1 mM IPTG, all MGM100 cultures except those carrying JQL were viable (fig 4.10a). In the presence of glucose when the endogenous GroEL is suppressed, neither of the chimeras was able to complement (fig 4.10b). An SDS-PAGE gel was run to ensure that the lack of complementation was not being caused due to the lack of expression. It is unclear why JQL was causing cell death in the presence of arabinose. A possibility is that JQL may be interacting with GroEL to form oligomers that are unable to function. Due to the results obtained here, it was not possible to either confirm or reject the hypothesis. A native gel of these chimeras was not done.

4.2.3 Chimeras MKO and JNL

One of the original objectives of this chapter was to investigate if it was possible to construct a chaperonin that was not only functional, but also more stable than Cpn60.2 as an oligomer. It is already known that GroEL functions as a stable oligomer. It was also shown in the last chapter that Cpn60.2 can function, but is known to be unstable. So the best choice to obtain a stable and functional chimera was to make domain swaps between GroEL and Cpn60.2. The original attempts that had imprecise domain swaps proved unsuccessful, so precise domain swaps were made. It was predicted that both combinations of chimeras would be able to function in *E. coli* to rescue cell growth, when endogenous GroEL expression is suppressed. It was also predicted that these chimeras would be able to form more stable oligomeric structures owing to the presence of domains from GroEL.

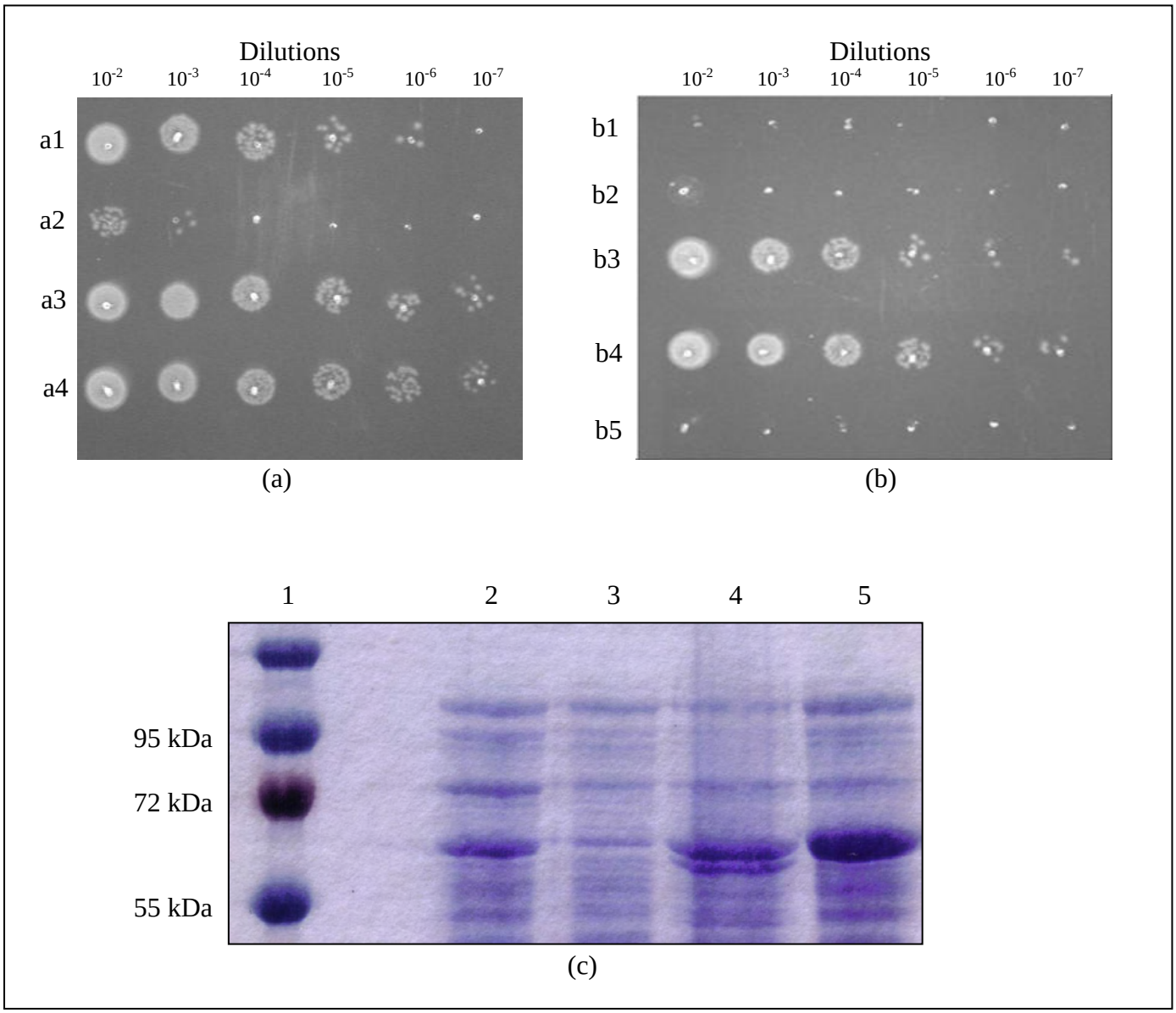


Fig 4.10: Complementation and expression of JQL and PKR.

Complementation results of JQL and PKR in MGM100 at 30°C in the presence of 0.2% arabinose with 0.1 mM IPTG (a) or in 0.2% glucose with 0.1 mM IPTG (b). a1 and b2- MGM100 pTrc-GroES-PKR, a2 and b1- MGM100 pTrc-GroES-JQL, a3 and b3- MGM100 pTrc-GroES-Cpn60.2, a4 and b4- MGM100 pTrc-GroES-GroEL, b5- MGM100 pTrc99a. SDS-PAGE showing levels of over expression of JQL and PKR (c). Samples were taken from cultures after 4 hours incubation with 0.1 mM IPTG and 0.2% arabinose. Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc-GroES-JQL; Lane 3: MGM100 pTrc99a, Lane 4: MGM100 pTrc-GroES-PKR, Lane 5: MGM100 pTrc-GroES-GroEL.

Chimera MKO is the precise domain swap that has the equatorial and intermediate domain of GroEL while the apical domain is from Cpn60.2. Chimera JNL has the reciprocal swap with the equatorial and intermediate domains from Cpn60.2, while the apical domain is from GroEL. Both chimeras were cloned into the pTrc99a plasmid and sequenced to ensure the domain swaps were made at the correct junction points, while also ensuring no mutations were present. These plasmids were then transformed into MGM100 to test their functional properties.

The results obtained here show that both MKO and JNL can complement for the loss of endogenous GroEL in *E. coli* MGM100. While there were no growth issues on arabinose plates, it was noted that JNL was a little better at complementing than MKO in glucose. Interestingly, on SDS-PAGE gels the over expression of MKO was not visible (fig 4.12a). Again, this supports the idea that very small amounts of functional chaperonin are needed to maintain cell growth. Attempts to try and increase the expression of MKO by cloning it into the pBad24 plasmid were unsuccessful. Due to the lack of expression of MKO, a native gel could not be run.

The expression level of JNL was much higher, and as a result its oligomeric properties were analysed on a native gel (fig 4.12). The protein sample was prepared from cells grown in LB with 0.2% glucose and induced with 0.1 mM IPTG for 5 hours. Two distinct bands were visible on the native gel. The first band represented a size of approximately 100kDa, while the second band was approximately 200kDa. When these results are compared with that of Cpn60.2 it can be seen that the 100kDa band approximately correlates with the monomeric/dimeric band of Cpn60.2. The second visible band of JNL however is not seen with Cpn60.2 and could indicate the presence of a slightly more stable structure. However, it

is important to note that this band would only indicate a trimeric/tetrameric structure and not a heptamer. For comparison sake, the tetradecameric GroEL band can also be seen on the native gel in lane 4 (fig 4.12b).

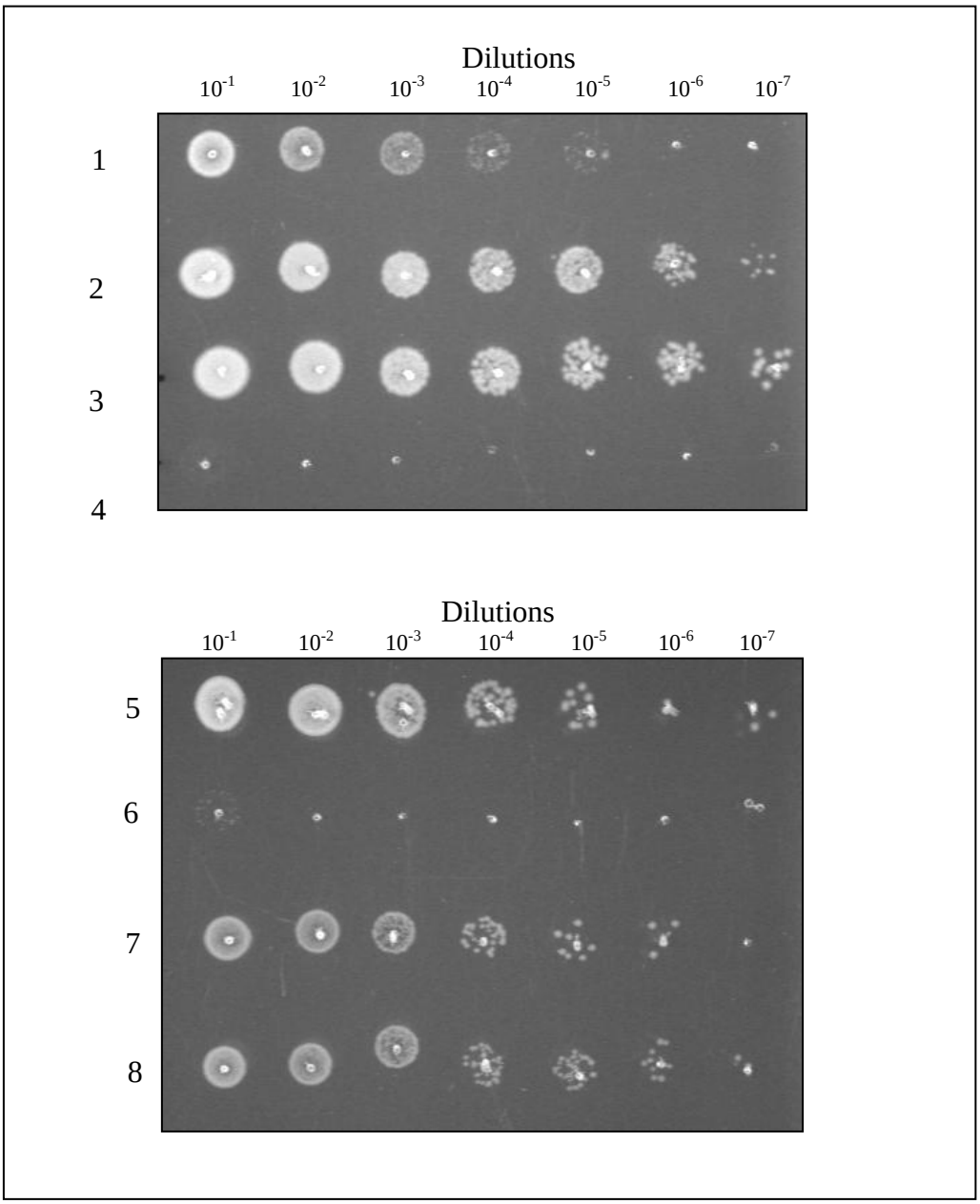


Fig 4.11: Complementation results of MKO and JNL. Complementation results of MKO and JNL in MGM100 at 30°C in the presence of 0.2% glucose and 0.1 mM IPTG.

1- MGM100 pTrc-GroES-MKO
2- MGM100 pTrc-GroES-GroEL
3- MGM100 pTrc-GroES-Cpn60.2
4- MGM100 pTrc99a
5- MGM100 pTrc-GroES-JNL
6- MGM100 pTrc99a
7- MGM100 pTrc-GroES-Cpn60.2
8- MGM100 pTrc-GroES-GroEL

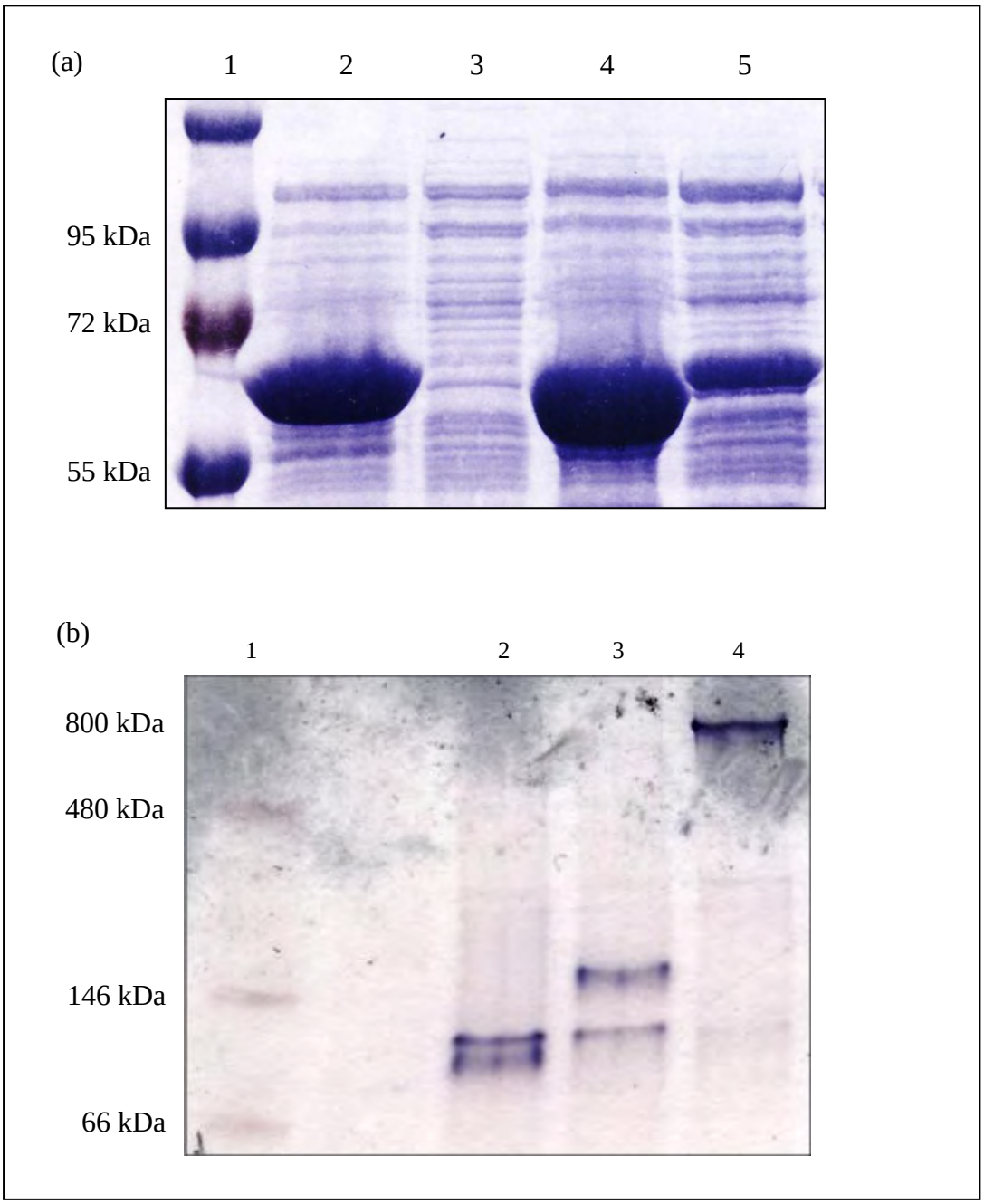


Fig 4.12: SDS-PAGE and native protein gels.

(a) SDS-PAGE gel showing minimal over expression of MKO and JNL. Samples were taken from cultures after 4 hours incubation with 0.1 mM IPTG and 0.2% arabinose. Lane 1- Prestained protein marker (Fermentas), Lane 2- MGM100 pTrc-GroES-Cpn60.2; Lane 3: MGM100 pTrc-GroES-MKO, Lane 4: MGM100 pTrc-GroES-GroEL, Lane 5: MGM100 pTrc-GroES-JNL.

(b) Native gel of JNL. Protein samples were taken from cells grown in LB with 0.2% glucose and induced with 0.1 mM IPTG for 5 hours. The GroEL native sample was from a purified stock. Lane 1: Native protein marker, Lane 2: GroES + Cpn60.2, Lane 3: GroES + JNL, Lane 4: GroES + GroEL.

Summary the complementation data of the precise domain swaps

A summary of the complementation data for the precise domain swaps can be seen below in table 4.3.

	37°C			30°C/26°C		
MGM100	**Arabinose**	**Glucose - IPTG**	**Glucose + IPTG**	**Arabinose**	**Glucose - IPTG**	**Glucose+ IPTG**
GroES + MQO	+++	-	-	+++	-	-
GroES + PNR	N/A	N/A	N/A	N/A	N/A	N/A
GroES + JQL	-	-	-	-	-	-
GroES + PKR	+++	-	-	+++	-	-
GroES + MKO	+++	++	++	+++	++	++
GroES + JNL	+++	+++	+++	+++	+++	+++
pTrc99a Vector only	+++	-	-	+++	-	-
GroES + GroEL	+++	+++	+++	+++	+++	+++
GroES + Cpn60.2	+++	+++	+++	+++	+++	+++

Table 4.3: Complementation data of precise domain swapped chimeric proteins in *E. coli* MGM100. The table shows the growth of cells expressing different chimeras with the cochaperonin GroES at different temperatures, with or without 0.1 mM IPTG.

+++ Denotes 10 – 100 % survival - Complete complementation

++ Denotes 1 – 10 % survival - Partial complementation

+ Denotes 0.1 – 1 % survival - Partial complementation

- Denotes <0.1 % survival - No complementation

Discussion

The objective of this chapter was to try to expand on the knowledge of the functional and oligomeric properties of *M. tuberculosis* chaperonins. In the last chapter the two chaperonins, Cpn60.1 and Cpn60.2, were tested for their ability to function in *E. coli* in the absence of endogenous GroEL. It was shown that while Cpn60.2 can complement for the loss of GroEL, Cpn60.1 could not. As a result, and based on the data obtained from literature, it was accepted that while Cpn60.2 may function as the main house-keeping chaperonin, it can be predicted that Cpn60.1 has evolved to perform other functions (Hu et al., 2008, Kim et al., 2003, Ojha et al., 2005, Qamra et al., 2005, Cehovin et al., 2010).

There could be various reasons as to how both these chaperonins are able to assist in different cellular functions. One of these reasons could be that both of these *M. tuberculosis* chaperonins are able to recognise and fold different groups of substrates. Since Cpn60.2 can also function in *E. coli*, it suggested that Cpn60.2 can also recognise a large number of the GroEL substrates, or at the very least the substrates that are vital for cell survival. A point to note is that, GroEL has very broad substrate specificity as it interacts with a large number of substrates in *E. coli* (Chapman *et al.*, 2006).

It is already known that the apical domain of the chaperonins is where a potential substrate first interacts with the chaperonins. It was reasoned that swapping the apical domain from Cpn60.2 or GroEL into Cpn60.1 might allow it to recognise the same set of substrates as the house-keeping chaperonins. This could then potentially lead to Cpn60.1 functioning in *E. coli*. However, this assumed that Cpn60.1 would be able to fold these new substrates and release them at the rate required by the cell.

The other aspect of this chapter was to see how the oligomeric properties of the *M. tuberculosis* chaperonins are affected by such domain swaps. It is known that both Cpn60.1 and Cpn60.2 do not form very stable oligomeric structures, and that Cpn60.2 crystallises as a dimer (Qamra et al., 2004, Qamra & Mande, 2004). GroEL however, forms much more stable oligomeric structures. As such it was reasoned that making domain swaps between Cpn60.2 and GroEL may yield a functional yet stable oligomeric Cpn60.2 chimera. However, as will be discussed shortly, a lot of the results that were obtained proved to be a little difficult to explain.

Out of the six potential imprecise domain swaps, only three could be cloned and transformed into MGM100, namely AEC, AHC, and GBI. These will be discussed after the chimeras that could not be cloned or made, namely DBF, DHF, and GEI.

Chimeras DBF and DHF could not be cloned into the pTrc plasmid without affecting cell growth and all the attempts that were made to rectify this issue were unsuccessful. It can be speculated that these chimeras were lethal to the cells, and as a result only the strains carrying non functional variants were able to survive. This could have been due to various reasons. Firstly, since the apical domain of DBF and DHF are from Cpn60.2 and GroEL respectively, it is possible that these chimeras were able to recognise and bind to the GroEL substrates, but were unable to fold them. In such a case, the results could lead to cell death. Secondly, it is possible that the imprecise nature of the domain swaps, where parts of the intermediate domain were from one chaperonin while the remainder was from another, could have affected the chaperonin cycle. This is because the intermediate domain is known to be responsible for transmitting the allosteric signals between the apical and equatorial domains to ensure appropriate structural changes are made to fold and release the bound substrates. So if its

structure was altered to the extent where it was unable to fulfil its function, then any substrates that were present in the cavity of the chimeras would not be released. As such, if large numbers of substrates were being trapped within the chimera, it would affect cell growth and could also cause cell death.

The issues encountered with GEI are much harder to explain as I was unable to obtain a final PCR product. A reason for this is unclear. Every aspect of the PCR reaction that could be altered was systematically changed without success, including using temperature gradients, altering the number of cycles, using new primers, altering the concentration of dNTPs, and changing the ramping speed between the temperature ranges.

The imprecise chimeras, which could be constructed and transformed into MGM100 generally showed unexpected results. While the results obtained with AEC were as predicted, those obtained with AHC and GBI (chimeric swaps between GroEL and Cpn60.2) were not. The complementation data of AEC showed that it was unable to complement for the loss of endogenous GroEL. This result was expected, as AEC had the apical domain of Cpn60.1, and it was predicted that, due to this apical domain Cpn60.2 would lose its ability to function in *E. coli*. However, based on the speculations made with DBF and DHF, the lack of function due to the imprecise nature of the swap could not be conclusively rejected.

One of the reasons why the results of the AHC and GBI swaps were unexpected is because both GroEL and Cpn60.2 can function in *E. coli* when expressed along with cochaperonin GroES. As such, it was predicted that a domain swap between them would also function. However, neither of these chimeras was able to complement for the loss of GroEL in MGM100. An SDS-PAGE gel was run to ensure the lack of complementation was not being

caused due to lack of expression. The protein expression analysis of GBI displayed a band lower than that of AHC, although sequence analysis confirmed no mutations were present. It is possible that the band seen for GBI was a proteolytic fragment. Surprisingly, both AHC and GBI showed signs of forming higher oligomeric structures with bands on a native gel that were comparable to that of GroEL, which is known to be a tetradecameric structure. While this result was expected with GBI, as it has the equatorial domain of GroEL and it is known that the equatorial domain is responsible for the majority of the inter-ring and intra-ring interactions, it was not expected with AHC. This result suggests that the apical domain may also play some role in the oligomeric stability of a chaperonin. Even though both these chimeras displayed signs of forming higher oligomeric structures, neither was able to complement for the loss of endogenous GroEL. One possible reason could be that a stable double ringed oligomeric structure is not necessary for substrate folding. Alternatively, it may also simply be because the chimera is not able to properly fold the substrates in this chimeric form even though it assembles as a double ring. A point to note, however, is that purified proteins were not used while running the native gels, so conclusive sizes could not be confirmed. Due to the number of issues being caused by the imprecise nature of the domain swaps, precise apical domain swaps were made.

The precise domain swaps proved to be easier to construct and transform into MGM100. The only cloning issue encountered, was with PNR (which is similar to DHF). Although with DHF, it was speculated that one of the possible reasons could be the imprecise nature of the swap, for PNR this could not be the case. Instead it can be speculated that PNR may be able to recognise and bind to GroEL substrates, since the apical domain is from GroEL, but may not be able to fold it properly in the chamber at a rate that is required by the cells. However this does not explain why no such issue was observed with PKR. Interestingly, a similar

mutant, where the equatorial domain was from Cpn60.1 while the intermediate and apical domains were from GroEL(referred to as $GroEL_{MER}$), was made by Kumar et al., and was shown to be non functional in *E. coli* (Kumar *et al.*, 2009). They also made the reciprocal swap where the equatorial domain was from GroEL, while the intermediate and apical domains were from Cpn60.1 (referred to as $GroEL_{MEF}$), and showed that this chimera could function in *E. coli* and also forms tetradecameric structures. According to their results, Cpn60.1 is capable of recognising the substrates of GroEL, and the only reason it does not function in *E. coli* is because of its lack of oligomerisation. This was a surprising observation as it was shown in the last chapter, that Cpn60.1 cannot complement in *E. coli* at all, and it was suggested that this lack of function was because Cpn60.1 is unable to recognise the same substrates as GroEL or Cpn60.2. Also, based on the data obtained in this chapter, a similar chimera to $GroEL_{MEF,}$ MQO, was shown to express but not complement in MGM100 which supports the hypothesis that Cpn60.1 lacks the ability to recognise GroEL substrates. It would also have been expected that if swapping just the equatorial domain from GroEL to Cpn60.1 allows it to form tetradecameric structures like $GroEL_{MEF}$, then the MQO chimera should have also displayed similar results. However, this was not the case. On native gels, MQO only displayed bands corresponding with single rings and not double rings. A point to note is that the native gel sample of MQO was not purified protein, so it is possible the result may differ with purified MQO.

Although the data obtained here with MQO supports the hypothesis that apical domain of Cpn60.1 cannot recognise the substrates of GroEL, this hypothesis cannot be accepted or rejected because the reciprocal swap of PNR could not be analysed. Similar swaps between Cpn60.1 and Cpn60.2 were done to see if the results obtained with MQO would be repeated. Since both these chaperonins are known to exist as lower oligomers, higher oligomeric

structures were not expected. The results that were obtained showed that neither JQL nor PKR could complement for the loss of endogenous GroEL, in the presence of glucose. It was also noticed that JQL would cause cell death in the presence of arabinose and IPTG, suggesting that when it is over expressed it may be interacting with GroEL to form a non functional GroEL. A possible reason why PKR failed to complement is less apparent. Since the apical domain is from Cpn60.2, it can speculated that it may have been able to recognise and bind to the appropriate substrates, but since the intermediate and equatorial domains are from Cpn60.1, it may not have been able to properly fold these substrates.

The chimeras MKO and JNL were primarily constructed to see if it was possible to improve the oligomeric stability of Cpn60.2. It was also expected that these chimeras would be able to function in *E. coli*. When similar swaps were made with imprecise domains, the resulting chimeras were unable to function in MGM100. As a result, it was speculated that if more precise domain swaps were made where just the apical domain was being swapped, then those chimeras would be able to function. As expected, the complementation data of MKO and JNL showed that both could complement for the loss of endogenous GroEL in *E. coli* MGM100. It was interesting to see that the expression of MKO was extremely low compared with JNL, yet it was able to complement reasonably well, supporting the suggestion that very low amounts of functional chaperonin are needed for cell survival. However, due to the low levels of expression, a native gel of MKO could not be run. An attempt to try and increase its expression levels using an arabinose inducible plasmid, pBAD24, was also unsuccessful. The expression levels of JNL were much better and as a result samples were prepared for a native gel. Based on the original predictions, JNL was not expected to form higher oligomeric structures on the native gel as its equatorial and intermediate domains were from Cpn60.2. However, it was observed that along with a band matching that of Cpn60.2, another band, of

higher molecular weight, was visible on the gel. This second band, approximately 200kDa in size, would normally be associated with a trimeric or tetrameric structure. Although no single or double rings were observed, it was interesting to reconfirm an earlier observation with AHC, that the apical domain may also be playing a role in the oligomeric stability of a chaperonin.

An important point to note here is that while AHC showed signs of forming larger double ringed structures, it was unable to complement in *E. coli*, whereas JNL could complement even though it showed the presence of much smaller structures. This result, along with others in this chapter suggests that forming stable oligomeric structures may not be necessary for substrate folding. However, this suggestion goes against our current understanding on chaperonin function. It is much more likely that JNL does form double ringed structures like AHC but is more unstable.

The aim of this study was to broaden our understanding on the functional and oligomeric properties of the *M. tuberculosis* chaperonins. Results here show that the evolutionary difference between Cpn60.1 and the house-keeping chaperonins Cpn60.2/GroEL is large enough that apical domain swaps alone cannot confer the ability to function in *E. coli*. It also shows that the apical domains may be playing a larger role in stabilising the oligomeric structure of the chaperonins than previously thought. Since all the native gels were run with whole cell protein extracts it was difficult to suggest without a doubt that these chimeras were forming larger oligomers. As such, in the next chapter I went on to test the oligomeric properties of JNL by purifying it and running it in an AUC under different buffer conditions. A table summarising all the data from all our attempted chimeric constructs can be seen in the table 4.4 below.

Name	Construct	Expected to complement?	Did it complement?	Expected to oligomerise?	Did it oligomerise?
	Blue –Cpn60.2; Red - Cpn60.1; Green – GroEL				
AHC		Yes	No	No	Double Ring
GBI		Yes	No	Yes	Double Ring and Single Ring
JNL		Yes	Yes	No	Monomer/Dimer
MKO		Yes	Yes	Yes	Possible Single Ring
DHF		Yes	could not be cloned	No	could not be cloned
GEI		No	could not be made	Yes	could not be made
PNR		Yes	could not be cloned	No	could not be cloned
MQO		No	No	Yes	Possible Single Ring
AEC		No	No	No	Not done
JQL		No	No, Lethal in arabinose	No	Not done
PKR		Yes	No	No	Not done
DBF		Yes	could not be cloned	No	could not be cloned

Table 4.4: Summary of data obtained from apical domain swaps between GroEL, Cpn60.1, and Cpn60.2. The white, orange, and black bars above the chimeric bars represent the approximate location of the equatorial, intermediate, and apical domains respectively.

Chapter 5

Analysis of a chaperonin chimera

Background

As discussed in section 1.5 it is already know from the literature that the accepted mechanism of chaperonin function, involves the formation of cage like structures to trap a non native protein within it. Exactly how the protein is then folded to its native state is still under debate (Lin *et al.*, 2008, England *et al.*, 2008, Horwich *et al.*, 2007, Chakraborty *et al.*, 2010). What is agreed upon however, is that the entrapment of the non native protein allows it to fold without influence from other cellular structures (Ellis, 1994). One of the most widely studied chaperonin is the *E. coli* GroEL. The mechanism by which this chaperonin functions has already been discussed in section 1.5, but briefly, involves the formation of a stable double ringed structure which can bind to and encapsulate its substrate proteins. Along with the help of its cochaperonin GroES, the GroEL chaperonin then goes through a repeating cycle of binding, folding and releasing the substrate in an ATP dependent manner (Horwich *et al.*, 2007).

It was surprising then, that in 2004, Qamra *et al*. showed that the chaperonins of *M. tuberculosis* existed as lower oligomers and crystallised as dimers (Qamra *et al.*, 2004, Qamra & Mande, 2004). Since ATP hydrolysis is an integral part of the chaperonin reaction cycle, it was also surprising to find that these chaperonins displayed very weak ATPase activity (Qamra *et al.*, 2004). They also claimed that the Mycobacterial chaperonins are unable to complement in *E. coli* (Kumar *et al.*, 2009). These findings were unusual, as it suggested the possibility of *M. tuberculosis* chaperonins using a different mechanism to that of GroEL to function. In chapter 3 however, it was shown that Cpn60.2 from *M. tuberculosis* can function in *E. coli*. Although this does not prove the formation of a multimeric structure,

it does support the idea that Cpn60.2 forms higher oligomeric structure *in vivo*, even if it is very unstable. Then in chapter 4 it was shown that the precise apical domain swaps between GroEL and Cpn60.2 resulted in functional proteins. This also supported the hypothesis that Cpn60.2 functions using a similar, if not the same, mechanism than that of GroEL. It was shown that both chimeras JNL and MKO were able to complement well in *E. coli* MGM100, while the chimera AHC failed to do the same. What was interesting however was the observation that, at least on native gels, JNL was a lower oligomer while AHC formed higher oligomeric structures. However, since this observation was based entirely on data from native gels, a better approach was needed.

For this reason, it was decided that JNL (Precise domain swap between Cpn60.2 and GroEL) and AHC (imprecise domain swap between Cpn60.2 and GroEL) would be purified and the protein samples would be analysed using analytical ultracentrifugation (AUC).

AUC is a technique that allows real time monitoring, during sedimentation, of the behaviours of macromolecules in solution in their native state. This method allows the characterisation of proteins by their hydrodynamic and thermodynamic properties in solution, without interaction with any matrix or surfaces. There are two types of experiments that can be performed with AUC, namely, sedimentation velocity and sedimentation equilibrium. Sedimentation velocity is a hydrodynamic technique that measures the rates at which molecules move in a centrifuge as a response to the centrifugal forces that are generated. This method measures the entire time course of sedimentation and provides information on both the shape and the molecular mass of the dissolved molecules. Sedimentation equilibrium on the other hand is a thermodynamic technique that measures concentration distribution of

macromolecules at equilibrium, when sedimentation is balanced by diffusion, and only provides information on the molecular mass of the molecule. In both cases data collection can be done with either absorbance optics or interference optics.

Before analysis on the chimeric proteins was done, extensive work in the lab had shown that under high salt buffer conditions with ATP, *M. tuberculosis* Cpn60.2 displayed the presence of higher oligomeric structures in the AUC (Elsa Zacco, Master's Thesis 2010). MS analysis on this sample by our collaborators, Dr. J Freeke and Dr. J Benesch from the University of Oxford, showed that these higher oligomers included tetradecameric forms of Cpn60.2. As such, it was decided that the chimeras would also be tested under similar conditions. This was done to see if AHC and JNL were also forming higher oligomeric structures, and what affect this had on their ATPase activity. Unfortunately, AHC proved to be very difficult to remake and express. The reason for this is unclear. Due to time constraints it was decided that only JNL would be purified and analysed. It was hypothesised that the oligomeric properties of JNL would be similar to that of Cpn60.2 as its equatorial domain is unchanged. It was also expected that conditions that favour oligomerisation would result in higher ATPase activity by JNL.

Experimental design

JNL was first purified using ion exchange and size exclusion chromatography. The protein sample was then analysed using AUC in low salt buffer without ATP and then high salt buffer with ATP. As discussed above, high salt buffer with ATP results in Cpn60.2 forming higher oligomers. Hence, it was expected that high salt buffer would also result in JNL

forming higher oligomers. The second part of this chapter was to test the ATPase activity of JNL in both buffers. It was expected that the high salt buffer that allows oligomerisation, would also result in an increase in the ATPase activity of JNL. The ATPase activity was tested along with Cpn60.2 and compared against GroEL.

In this chapter various buffers were used and have been listed below:

Buffer 1 – Low salt buffer containing 50 mM Tris, 50 mM NaCl, 1 mM EDTA at pH 6.5

Buffer 2 – High salt buffer containing 50 mM Tris, 500 mM NaCl, 1 mM EDTA at pH 6.5

Buffer 5 – Gel filtration buffer containing 50 mM Tris, 50 mM KCl, 10 mM $MgCl_2$ at pH 7.4

Buffer A- Buffer 5 + 1M $(NH_4)_2SO_4$ and 1 mM ATP

The reason why the buffers are named as such is because buffer 1 to 5 is the usual order in which they are used to purify proteins. Buffers A to F were used for oligomerisation tests with buffer A containing the highest ATP concentration and buffer F containing the lowest. Only the buffers that were used in this chapter have been listed.

5.1 Purification of JNL

5.1.1 Ion exchange purification: DEAE sepharose

For the first purification step, an ion exchange column with DEAE sepharose fast flow chromatography media (GE healthcare) was used. This method separates proteins based on the differences in their net surface charge. The underlying basis of this process involves the interaction of charged particles from proteins of interest with the opposite charge on the column matrix. Since proteins vary in their charge properties, their interaction with the charged chromatography media also varies. The net surface charge of proteins is highly pH dependent and different proteins have different relationship between the net surface charge and the pH. By controlling this property of proteins, it is possible to control the interaction of the proteins with the column matrix to either favour binding or elution.

The protocol used to over-induce and prepare the sample has been described in section 2.6, but briefly, MGM100 cells containing pTrc plasmid encoding both GroES and JNL were over-expressed in 3 x 500ml conical flasks with 200 ml LB, 0.2% glucose, and 0.5 mM IPTG for 4 hours. The samples from each of the flasks was then collected and centrifuged at 6000rpm for 10 minutes at 4°C. The resulting pellet was then resuspended in 6 ml of buffer 1 and then sonicated for a total of 1 minute and 30 seconds with 10 second pulses to break open the cells. This sample was then used in the ion exchange column.

The sample was first loaded into the column and left for approximately 1 hour. This allowed the buffer to flow through along with other unbound molecules into fraction tubes due to gravity. Once all the buffer had been removed, the programme was started with a flow rate of

1ml/min and the elute was collected in the fraction tubes. Each fraction tube collected the end product of the purification step for approximately 10-15 minutes. A sample from every 5th tube was analysed by SDS-PAGE to determine where JNL had eluted. As shown in figure 5.1a, tube 20 contained a clear band corresponding to the size of JNL. As a result, samples from tubes 16-24 were then further analysed by SDS-PAGE (fig 5.1b), and the fractions from the tubes with the highest concentration of JNL (tubes 18-23) were pooled and concentrated. This sample was then used to further purify JNL in a gel filtration column.

5.1.2 Gel filtration purification

The next step in our purification protocol was to try and further purify JNL as the concentrated solution from the ion exchange column still contained a high concentration of smaller proteins. It was decided that the sample would be run through a gel filtration column. This filtration technique separates proteins based on their molecular size and shape. The gel consists of extremely small porous beads that allow smaller molecules to freely enter the internal space while preventing larger molecules from entering. As elution begins, the smaller molecules penetrate the beads, and as a consequence migrate slower than the larger molecules. As a result the larger molecules, which can pass through the column at the same rate as a buffer, are eluted first and collected as fractions. The smaller molecules elute later as they have to travel through a larger volume of solvent. As a result the largest molecules are collected in earlier fraction tubes and a gradual decrease in the size of proteins is observed in later fraction tubes.

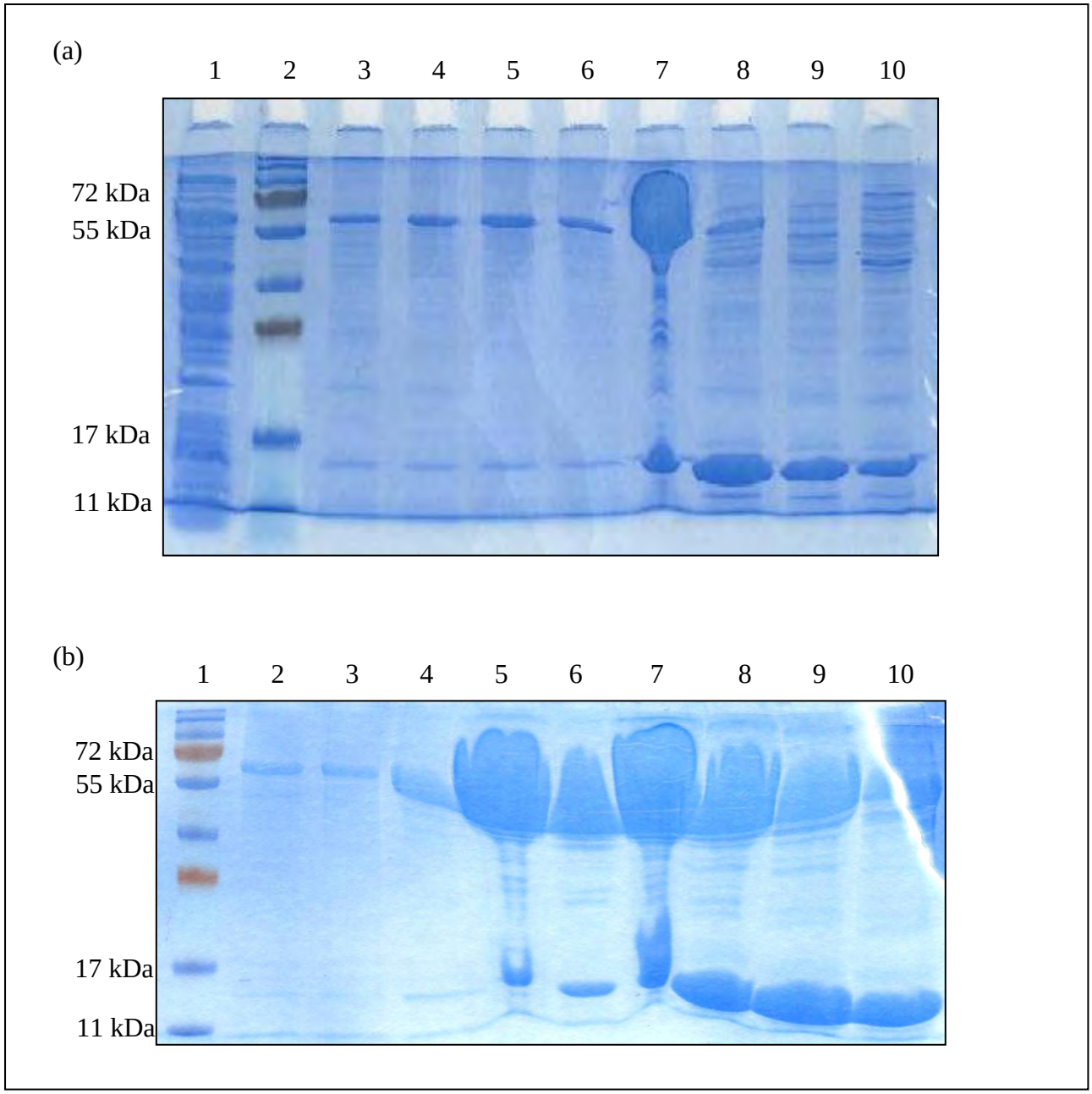


Fig 5.1: SDS-PAGE analysis of JNL from ion exchange column.

(a) Lane 1: Unbound sample; Lane 2: Prestained marker (Fermentas); Lane 3: Tube 1; Lane 4: Tube 5; Lane 5: Tube 10; Lane 6: Tube 15; Lane 7: Tube 20; Lane 8: Tube 25; Lane 9: Tube 28; Lane 10: Tube 30.

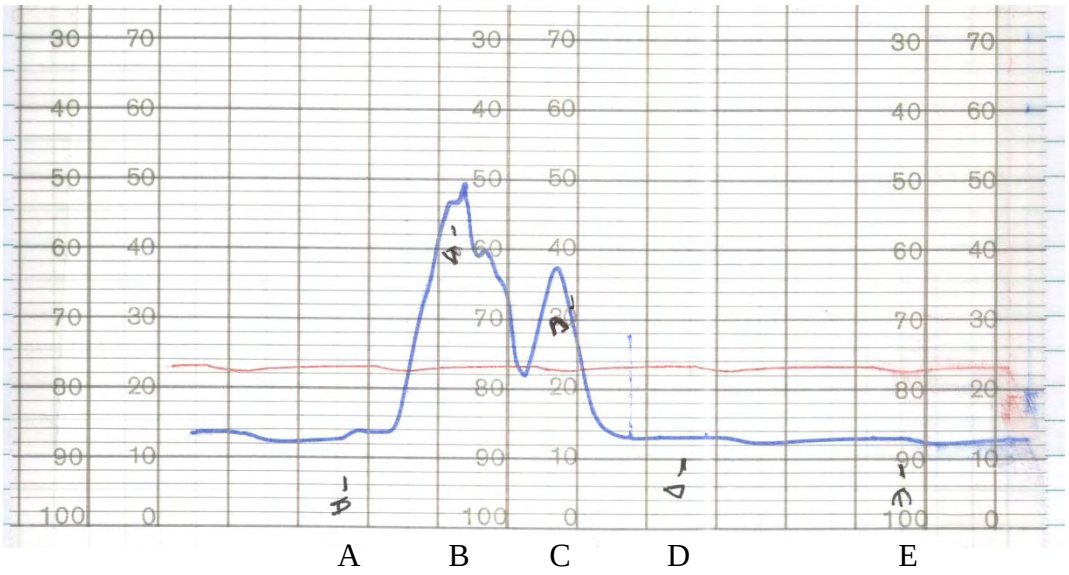
(b) Lane 1: Prestained marker (Fermentas); Lane 2: Tube 16; Lane 3: Tube 17; Lane 4: Tube 18; Lane 5: Tube 19; Lane 6: Tube 20; Lane 7: Tube 21; Lane 8: Tube 22; Lane 9: Tube 23; Lane 10: Tube 24.

The protocol used for gel filtration has been described in section 2.6. The concentrated sample from the ion exchange column was loaded into the gel filtration column and the filtration process was started with a flow rate of 1ml/min. Based on the results obtained from the UV detector (fig 5.2a), the samples between peaks A and D (columns 5-18) were analysed on a SDS-PAGE gel. The results showed that the fractions containing JNL still had a large concentration of smaller proteins present (fig 5.2b).

5.1.3 Second Ion exchange purification: DEAE sepharose

As better separation had been obtained with the ion exchange column, this method was repeated with a shallower buffer gradient. This was done as an attempt to improve the separation of JNL from the smaller proteins. As can be seen in figure 5.3a, this approach was largely successful. Samples from tubes 15 to 18 (lanes 2 to 5 in figure 5.3a) were then collected and pooled together. To ensure JNL had not eluted in fraction tubes prior to tube 15, another SDS-PAGE gel was run and included samples from tubes 10 to 14 (fig 5.3b). This SDS-PAGE gel also had the pooled product from tubes 15 to 18 (purified JNL) as well as running samples from the unbound collection to ensure no JNL was present (5.3b).

(a)



(b)

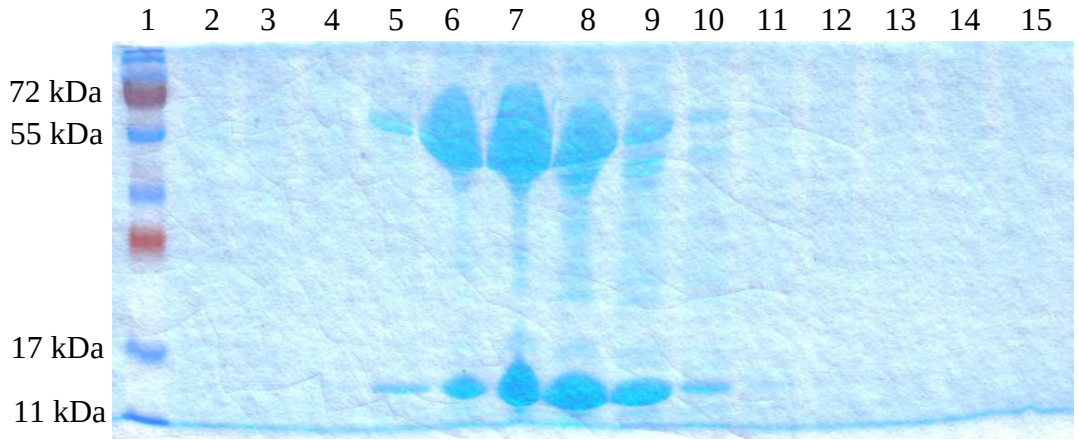


Fig 5.2: UV graph and SDS-PAGE page analysis of JNL from gel filtration column.

(a) UV graph showing regions of protein elution.

(b) SDS-PAGE gel. Lane 1: Prestained marker (Fermentas); Lane 2: Tube 5; Lane 3: Tube 6; Lane 4: Tube 7; Lane 5: Tube 8; Lane 6: Tube 9; Lane 7: Tube 10; Lane 8: Tube 11; Lane 9: Tube 12; Lane 10: Tube 13; Lane 11: Tube 14; Lane 12: Tube 15; Lane 13: Tube 16; Lane 14: Tube 17; Lane 15: Tube 18.

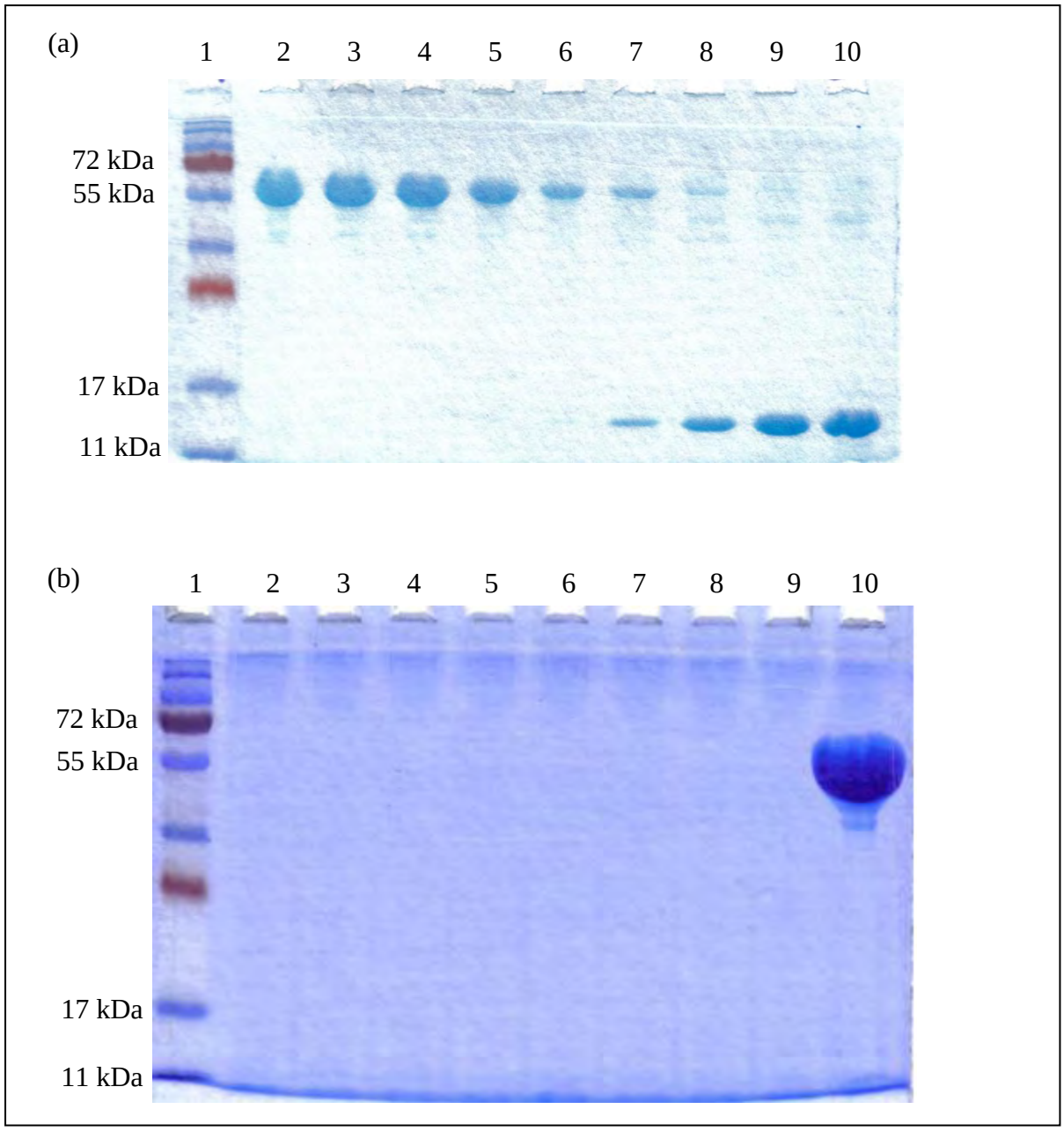


Fig 5.3: SDS-PAGE analysis of JNL from second ion exchange column, and purified JNL

(a) Lane 1: Prestained marker (Fermentas); Lane 2: Tube 15; Lane 3: Tube 16; Lane 4: Tube 17; Lane 5: Tube 18; Lane 6: Tube 19; Lane 7: Tube 20; Lane 8: Tube 21; Lane 9: Tube 22; Lane 10: Tube 23

(b) Lane 1: Prestained marker (Fermentas); Lane 2: Unbound 1; Lane 3: Unbound 2; Lane 4: Unbound 3; Lane 5: Tube 10; Lane 6: Tube 11; Lane 7: Tube 12; Lane 8: Tube 13; Lane 9: Tube 14; Lane 10: Concentrated JNL

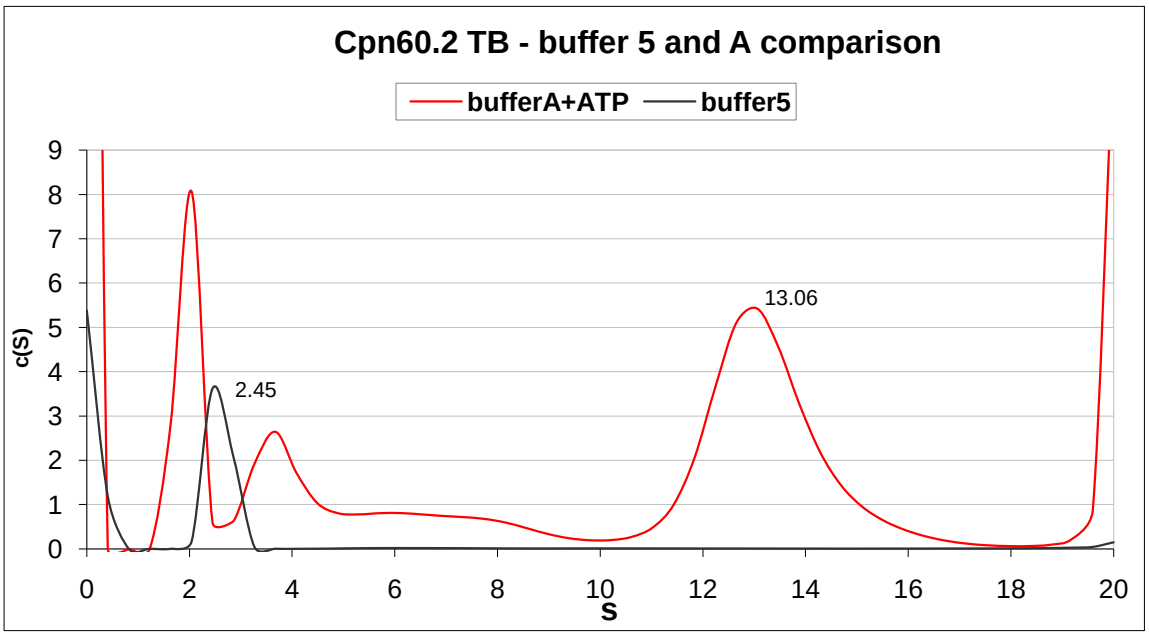
5.2 AUC analysis of JNL

Using AUC (analytical ultracentrifugation), attempts were made to determine the oligomeric state of JNL. The JNL protein sample was run in buffer 5 and buffer A at 20°C and 40,000g for 20 hours. As discussed above, it was shown that under high salt buffer conditions with ATP, *M. tuberculosis* Cpn60.2 displays the presence of higher oligomeric structures in the AUC and this was confirmed to include the presence of tetradecameric structures by MS (fig 5.4a and 5.4b). As such, it was expected that JNL would also oligomerise in buffer A.

On the graphs obtained from the AUC data, the increase in sedimentation coefficient (S) value denotes an increase in oligomeric structures. For Cpn60.2 a sedimentation coefficient of around 13 represents the tetradecameric structure. With that in mind it can be seen from the results obtained here that there is evidence of some large oligomers in buffer A but not in buffer 5 (fig 5.5a and 5.5b). However, since sedimentation velocity experiments depend not only on the molecular mass but also on the shape of a molecule, it becomes difficult to obtain the precise molecular mass measurements from the S value. Although we were unable to analyse the structure further with MS, by analogy with Cpn60.2 we expect this to include tetradecamers. This result was as expected and is consistent with what has been observed with Cpn60.2.

Since it is possible to obtain JNL at higher oligomeric forms, it was decided that ATPase assays would be done to see if its ability to hydrolyse ATP is dependent on its ability to oligomerise. For this reason, ATPase assays were done in buffer A and buffer 5 for JNL, Cpn60.2, GroEL (positive control), and BSA (negative control).

(a)



(b)

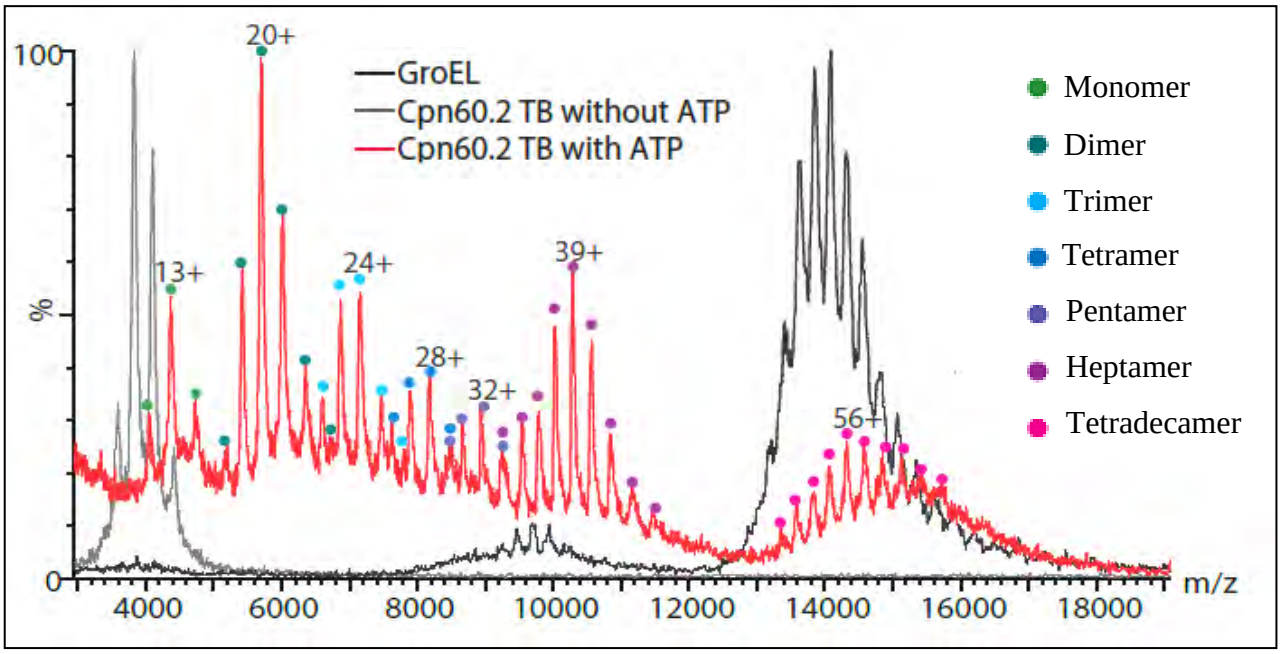
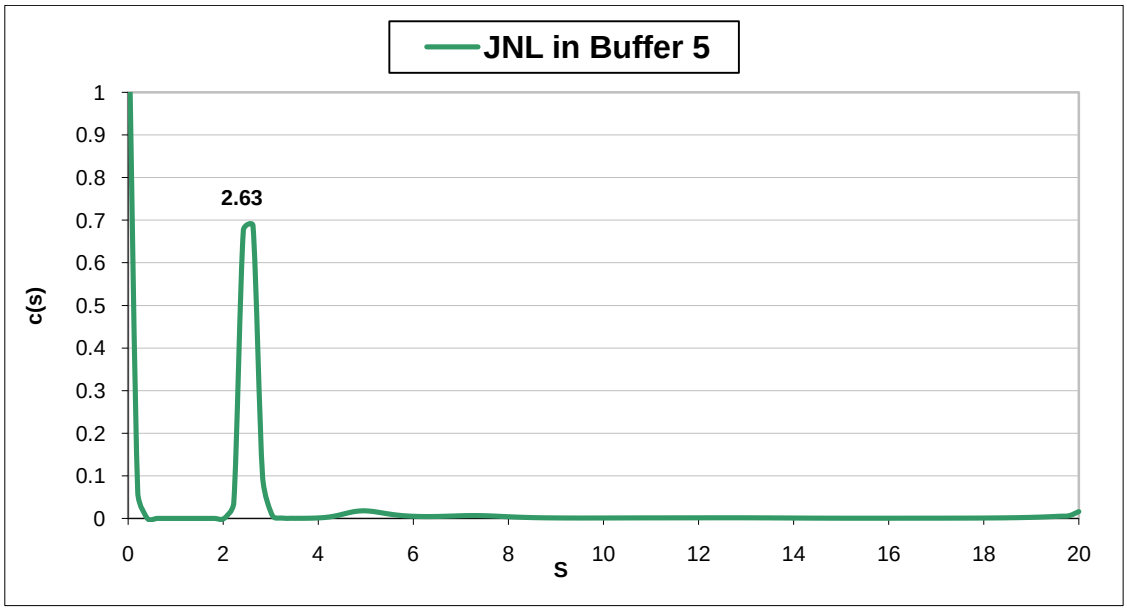


Fig 5.4: Oligomeric states of Cpn60.2. Using AUC and MS the oligomeric states of Cpn60.2 were determined (Adapted from Elsa Zacco, Master's Thesis 2010). (a) AUC analysis showing sedimentation coefficient of Cpn60.2 in buffer A and buffer 5. (b) MS analysis showing the various oligomeric states of Cpn60.2 present in high salt buffer with ATP (result obtained by Dr. J Freeke and Dr. J Benesch from the University of Oxford).

(a)



(b)

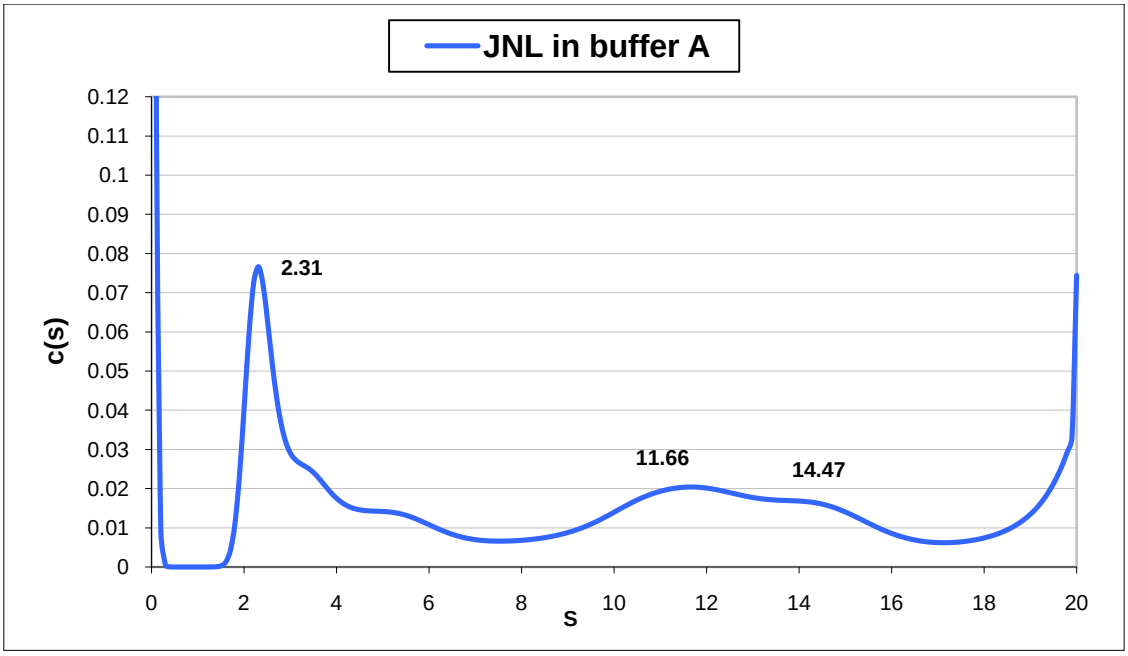


Fig 5.5: AUC data analysis showing sedimentation coefficient of JNL in buffer 5 (a) and buffer A (b).

5.3 ATPase assay of purified JNL

It is known from the literature, and has been confirmed in our lab, that Cpn60.2 has a very weak ATPase activity in standard resuspension buffer (Qamra *et al.*, 2004). However, it was also shown that in buffer A the activity of Cpn60.2 would rise by 3 to 4 folds of normal level (Elsa Zacco, Master's thesis 2010). As the current model of chaperonin function relies on their ability to oligomerise and fold substrates in an ATP dependent manner, it was reasoned that the rise in ATPase activity of Cpn60.2 was due to the presence of higher oligomeric forms of Cpn60.2 in the solution. The aim of this part of the study was to test the hypothesis that oligomerisation plays a role on the ATPase activity of JNL. As it is was shown that higher oligomeric forms of JNL could be obtained in buffer A, it was expected that the ATPase activity in that buffer would also be higher than in buffer 5.

For this experiment, I made use of the P_iColorLock gold ATPase kit (Innova Biosciences), using the protocol described in section 2.6. This enables the use of a simple colorimetric assay on 96 well plates to determine the ATPase activity of a molecule. The assay measures P_i based on the change in absorbance of malachite green in the presence of molybdate (fig 5.6a). The intensity of the colour is directly proportional to the release of P_i in the solution after ATP hydrolysis. An example of this colour change can be seen in figure 5.6b below.

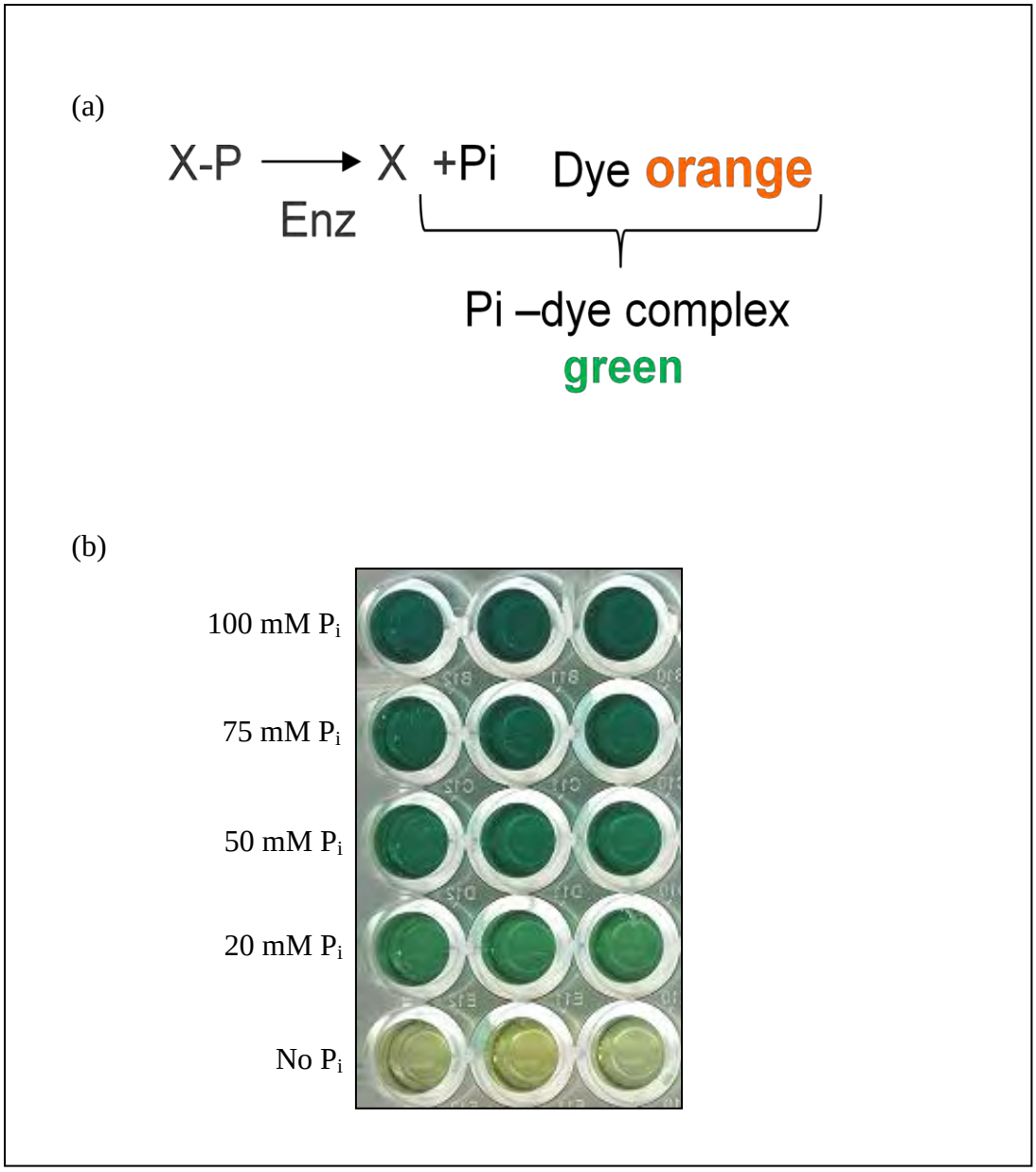


Fig 5.6: (a) Principle of the $P_i$colorlock assay. (b) Example of colour change caused by an increase in free P_i in solution.

Before the ATPase activity of JNL could be assessed, it was necessary to ensure that the assay being used for detecting free phosphate was linear. For this reason a standard curve with different Pi concentrations (0µM, 25µM, 50µM, 75µM, and 100µM) was set up (fig 5.7). The absorbance (620nm) reading after the addition of the gold mix was also tested at different time points, from 0 minutes to 150 minutes, to ensure the stability of the assay (fig 5.7).

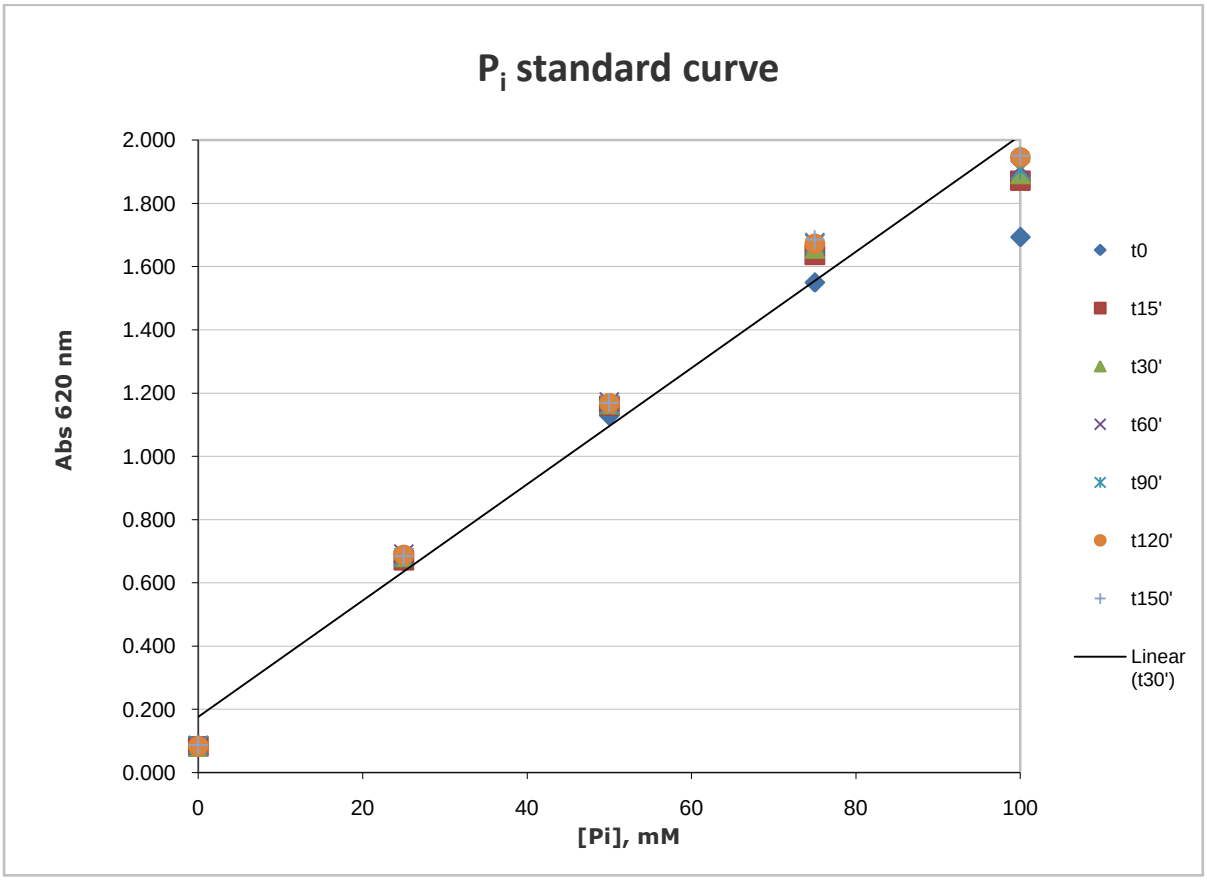
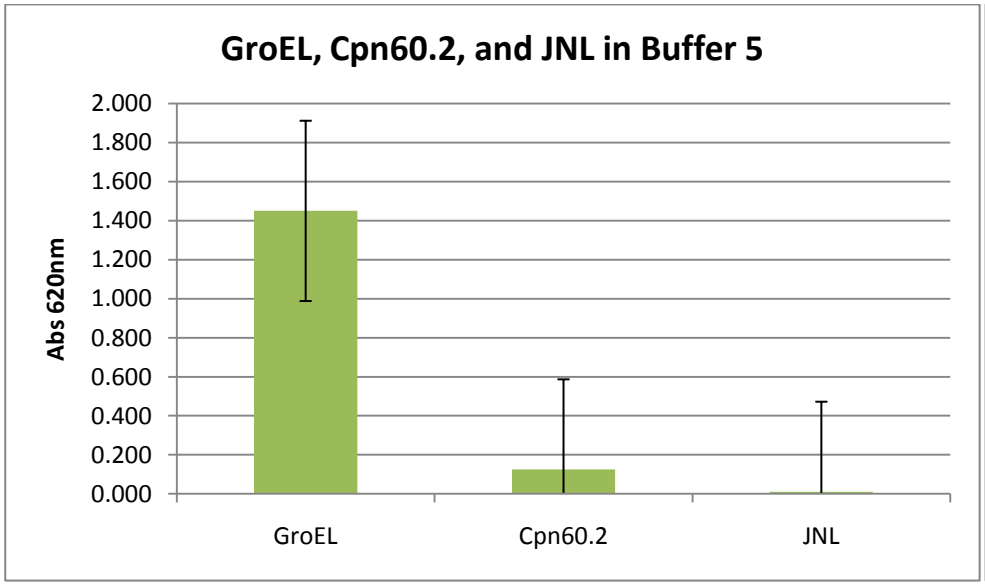


Fig 5.7: Image showing P_i Standard curve at varying P_i concentrations and incubation time.

In this experiment GroEL was the positive control, while bovine serum albumin (BSA) was used to measure the background reading. Cpn60.2 was also included as a comparison for the JNL ATPase activity. The background activity obtained from BSA would be used to correct the observed activity of the chaperonins. In each case, the OD_{620} reading was taken after 30 minutes of incubation with the gold mix.

The results that were obtained showed that the ATPase activity of JNL and Cpn60.2 changed considerably depending on the buffer used. When compared with the activity of GroEL it becomes apparent that the activity of JNL and Cpn60.2 is much lower in buffer 5 than it is in buffer A. Once corrected for the background activity, if GroEL activity is taken as 100%, then the activity of JNL in buffer 5 is almost nonexistent at just 0.49 ± 2.49% of GroEL (fig 5.8). As expected the ATPase activity of Cpn60.2 in buffer 5 is also extremely low at just 8.27 ± 3.82% of GroEL. In buffer A however, both JNL and Cpn60.2 have higher ATPase activity, although still not as high as GroEL. Both Cpn60.2 and JNL are at approximately 32.19 ± 1.66% and 25.86 ± 9.42% of GroEL respectively (fig 5.9). The results of these experiments support the original hypothesis that in conditions where oligomerisation can occur, the ATPase activity of Cpn60.2 and JNL increases.

	1	2	3	Corrected values			Average	% Pi release			Average % Pi release	SD
GroEL	1.984	1.819	1.915	1.684	1.258	1.409	1.450	100.00	100.00	100.00	100.00	0
Cpn60.2	0.512	0.628	0.603	0.212	0.067	0.097	0.125	12.59	5.33	6.88	8.27	3.82
JNL	0.356	0.544	0.499	0.056	-0.017	-0.007	0.011	3.33	-1.35	-0.50	0.49	2.49
BSA	0.300	0.561	0.506									



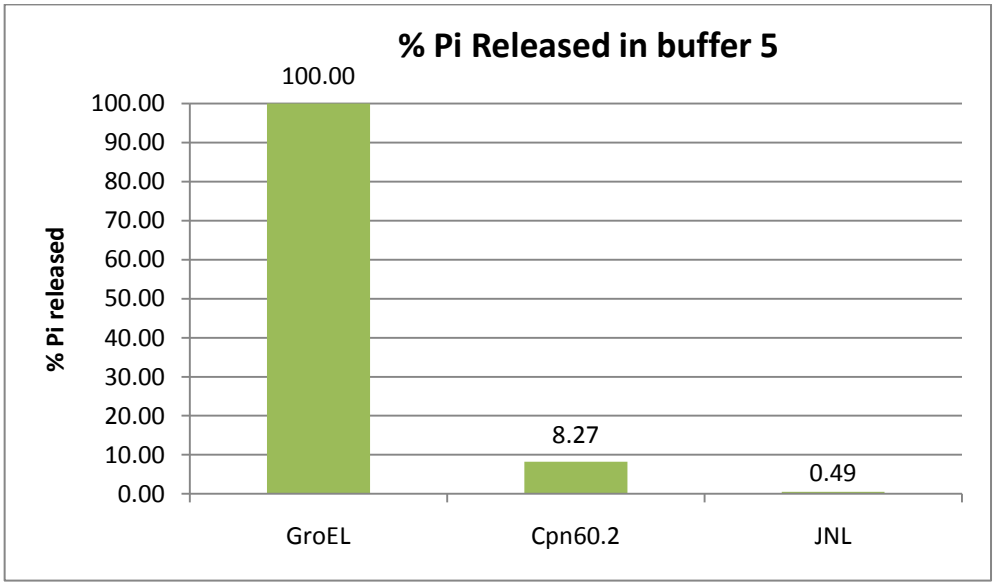


Fig 5.8: ATPase activity of GroEL, Cpn60.2, JNL, and BSA in buffer 5.

	1	2	3	Corrected values			Average	% Pi release			Average % Pi release	SD
GroEL	1.637	1.972	1.960	0.276	0.733	0.671	0.560	100.00	100.00	100.00	100.00	0
Cpn60.2	1.453	1.461	1.510	0.092	0.222	0.221	0.178	33.33	30.29	32.94	32.19	1.66
JNL	1.403	1.482	1.485	0.042	0.243	0.196	0.160	15.22	33.15	29.21	25.86	9.42
BSA	1.361	1.239	1.289									

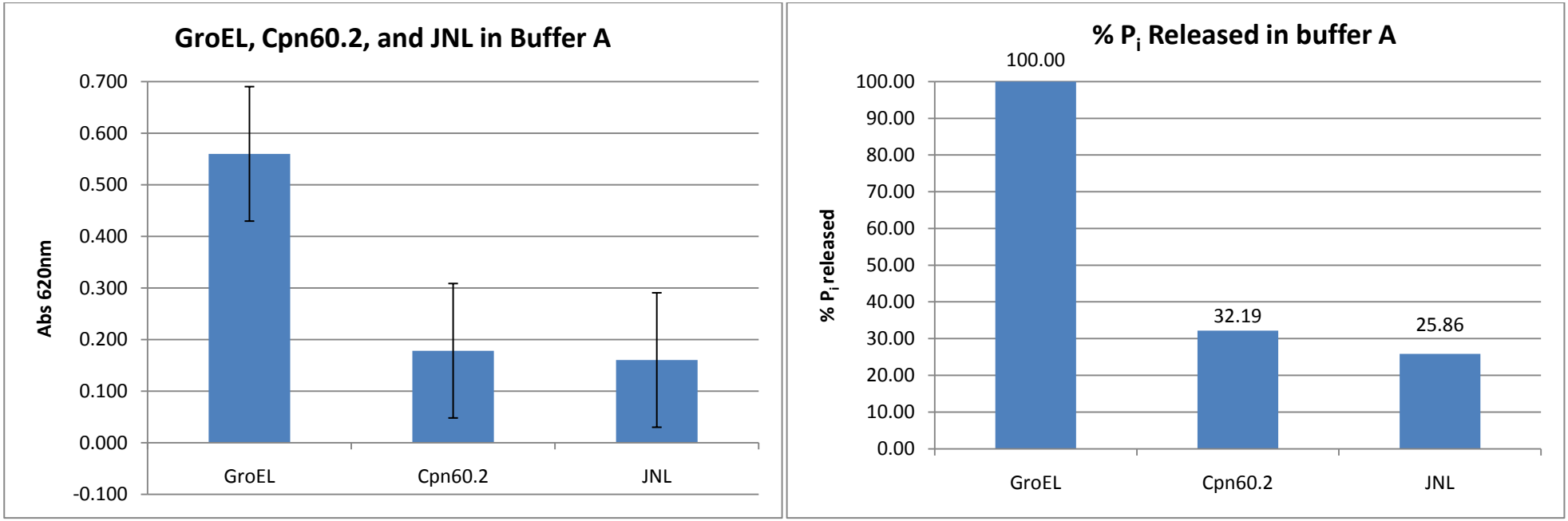


Fig 5.9: ATPase activity of GroEL, Cpn60.2, JNL, and BSA in buffer.

Discussion

As discussed in the introduction, the accepted mechanism by which chaperonins function involves the assembly of stable ringed structures to provide a central cavity for protein folding. Substrate proteins can fold within this central cavity without being influenced by other cellular components (Ellis, 1994). Mycobacterial chaperonins however have been shown to be structurally very unstable and posses very weak ATPase activity (Qamra *et al.*, 2004, Kumar *et al.*, 2009).

The original aim of this section of the project was to analyse and compare the oligomeric and ATP hydrolysing properties of two the Mycobacterial chaperonin chimeras AHC and JNL. AHC however proved to be very difficult to remake and as such only JNL could be further analysed. Once JNL was purified using ion exchange chromatography, samples were run in the AUC using two different buffers. As discussed above, extensive work in the lab had shown that under high salt conditions and in the presence of ATP, Cpn60.2 formed large tetradecameric structures *in vitro* (Elsa Zacco, Master's thesis, 2010). Since the equatorial domain of JNL was from Cpn60.2, it was expected that JNL would also structurally behave like Cpn60.2.

The results that were obtained confirmed this hypothesis. In the presence of low salt and no ATP (buffer 5), JNL was shown to be primarily monomeric. A point to keep in mind is that none of the JNL samples were analysed using MS, and as such the results obtained from AUC can only be compared with those of Cpn60.2. Elsa Zacco had shown that in buffer 5, Cpn60.2 was also monomeric in nature. In the presence of high salt buffer and ATP, as

expected, evidence of larger oligomeric structures of JNL was seen. As MS could not be done, the AUC data was compared with that of Cpn60.2. Due to similar results being obtained with Cpn60.2, it is reasonable to infer that the presence of higher oligomeric structures of JNL also includes tetradecamers.

Interestingly, when the AUC data in buffer 5 of JNL is compared with that of Cpn60.2, there are signs of slightly higher oligomeric structures, possibly trimers or tetramers, present in the JNL sample. This was consistent with what was observed on the native gel. As was speculated, this may be due to the presence of the apical domain of GroEL. Exactly how this interaction takes place is unclear. However it does suggest that the apical domain also plays a role in the structural stability of a chaperonin.

As discussed above the ATPase activity of Mycobacterial chaperonins was shown to be very low when compared with that of GroEL (Kumar *et al.*, 2009, Qamra *et al.*, 2004). The second objective of this study was to analyse the ATPase activity of JNL. As it is known that the protein folding cycle of stable chaperonins such as GroEL is dependent on its ATPase activity, it was expected that in conditions that support oligomerisation, the ATPase activity of JNL would rise. The results that were obtained from the ATPase assays confirmed this hypothesis. It was shown that in buffer 5 when large oligomers are not being formed, the ATPase activity of JNL and Cpn60.2 is extremely low when compared to that of GroEL. However, in buffer A where both Cpn60.2 and JNL show the presence of larger oligomers, their ATPase activity rises considerably. However, even under these conditions the ATPase activity of both Cpn60.2 and JNL is only approximately 30% of that of GroEL.

Based on the results obtained here and from chapters 3 & 4, it becomes apparent that under appropriate conditions, Cpn60.2 and its chimera JNL can not only function in *E. coli,* but also form higher oligomeric structures and hydrolyse ATP. This is in contrast to the data obtained by Kumar *et al.*, and Qamra *et al.* It is however important to note that all the experiments done in this chapter were *in vitro* and used conditions that were in effect forcing the chaperonins to oligomerise. It still does not explain the chaperonin cycle that takes place *in vivo*. It is possible that *M. tuberculosis* chaperonins have evolved to be less energy dependent. This reduced dependency on energy could result in their unstable nature. However, since Cpn60.2 and JNL can both function in *E. coli,* even though they show bands corresponding to monomers on a native gel, it can be speculated that the disassembled and assembled states of the chaperonins are transient. It is possible that under conditions that do not require high concentrations of functional chaperonins, the Mycobacterial chaperonins are primarily in an unassembled monomeric/dimeric state. When the conditions within the cell change to the point where more functional chaperonins are needed, then higher concentrations of oligomeric chaperonins are present. How this is achieved within the cell is unclear. It is possible that the oligomerisation of Mycobacterial chaperonins may be substrate dependent. It can also be speculated that under stressful conditions, such as heat shock, the highly crowded nature of the cells and the over-expression of the heptameric Cpn10 (or GroES in *E. coli*) may provide the Mycobacterial chaperonins with the scaffolding needed to assemble into ringed structures. To prove the presence of large oligomeric structures *in vivo,* further experiments would need to be done. One possibility is cross-linking experiments. Cross-links are ionic or covalent bonds that link interacting polypeptide chains to each other and is used as a probe for protein-protein interaction. In context to Mycobacterial chaperonins, it could be used to see if they are forming tetradecameric structures *in vivo*.

In summary, based on the results obtained here and the results obtained from the previous result chapters, it is very probable that within the cell where there are various crowding agents and different salt concentrations, the Mycobacterial chaperonins form higher oligomeric structures and hydrolyse ATP in the substrate folding reactions.

Chapter 6

Further characterisation of Mycobacterial chaperonins

In this chapter I will be looking to further characterise the chaperonins of *Mycobacterium tuberculosis* in three different sets of experiments. For this reason, the chapter will be split into 3 separate sections so that their background information and results can be discussed separately. The three sections that will be discussed here are as follows:

6.1- Testing the ability of *M. tuberculosis* Cpn60.1 to function in *M. smegmatis*

6.2- Testing the existence of double ring *M. tuberculosis* Cpn60.2 by attempting to make single ring mutants

6.3- Testing the hypothesis that phosphorylation is needed for improved oligomerisation

In section 6.1, since I will be referring to *M. tuberculosis*, *M. smegmatis* and *E. coli* chaperonins, the prefixes Mtb, Ms, or Ec will be used. So, MtbCpn60.1 denotes *M. tuberculosis* Cpn60.1, while MsCpn60.1 denotes *M. smegmatis* Cpn60.1, and EcGroEL denotes *E. coli* GroEL. In all cases, unless mentioned otherwise, the cochaperonin in the experiments was *M. smegmatis* Cpn10. To make it easier to follow, each strain will be denoted with a letter:

Strain A- *M. smegmatis* mc^2 155

Strain B- *M. smegmatis* Δ*cpn60.1*

Strain C- Strain B with pMsGroEL1 plasmid (Ms*cpn10*-Ms*cpn60.1*)

Strain D- Strain B with pMs10-MsCpn60.2 plasmid

Strain E- Strain B with pMs10-MsCpn60.3 plasmid

Strain F- Strain B with pMs10-EcGroEL plasmid

Strain G- Strain B with pMs10-MtbCpn60.1plasmid

Note: The constructs in strains D, E, and F were made by Tara Rao (PhD thesis, 2010, University of Birmingham).

Section 6.1: Testing the ability of *M. tuberculosis* Cpn60.1 to function in *M. smegmatis*

Background

As discussed in the introductory chapter and in chapter 3, it is known that the main house-keeping chaperonin in both *M. tuberculosis* and *M. smegmatis* is Cpn60.2, while Cpn60.1 has evolved to perform alternate functions. Our current understanding of the functions of MsCpn60.1 implicates its role in biofilm maturation for *M. smegmatis* while the MtbCpn60.1 chaperonin has been shown to play a role in granuloma formation during tuberculosis (Hu *et al.*, 2008, Kim *et al.*, 2003, Ojha *et al.*, 2005).

Biofilm formation is an important aspect of many pathogenic bacteria, including *E. coli, V. cholerae, H. influenze,* and *P. aeroginosa* (Hall-Stoodley & Stoodley, 2009, Bjarnsholt *et al.*, 2010, Fux *et al.*, 2005). These bacteria make use of the biofilms to provide them with protection from the diverse environmental and nutrient conditions, while also providing them with the means to protect themselves from the effects of antibiotics (Bjarnsholt *et al.*, 2010, Hall-Stoodley *et al.*, 2004, Cernohorska & Votava, 2010).

In *M. smegmatis,* it has been shown that the loss of biofilm maturation can be caused by deleting the *groEL1* (Ms*cpn60.1*) gene which results in the lack of association between MsCpn60.1 and KasA (Ojha *et al.*, 2005). KasA is a component of the fatty acid synthase (FAS)-II complex which is required for the elongation of mycolic acid chains. So the deletion

of Ms*cpn60.1* can indirectly lead to the non assembly of the FAS-II complex and as a result cause defective mycolic acid biosynthesis. In the same study, Ojha *et al.* also showed that they could reverse this phenotype in *M. smegmatis* by complementing the strain with a plasmid carrying the *M. smegmatis cpn60.1* gene.

As it is known that MtbCpn60.1 has a high degree of sequence similarity (≈90%) and is evolutionarily related to MsCpn60.1, and does not function as the main house-keeping chaperonin in *M. tuberculosis*, I wanted to test the link between sequence homology and functional property to see if it was possible to complement for the loss of MsCpn60.1 in *M. smegmatis* with MtbCpn60.1. It was hypothesised that MtbCpn60.1 should be able to function in *M. smegmatis* and hence be able to reverse the phenotype of the *M. smegmatis* Δ*cpn60.1* knockout strain (strain B). To do this the *M. tuberculosis cpn60.1* gene was expressed in M smegmatis lacking Ms*cpn60.1*, and the biofilm assay repeated.

Experimental design

Ojha *et al*. were kind enough to provide the original complementing plasmid pMsGroEL1, the original *M. smegmatis* strain (strain A), and the Ms*cpn60.1* knockout strain (strain B). To test the ability of MtbCpn60.1 to function in *M. smegmatis*, its gene was first cloned into the complementing plasmid. In doing so, the *M. smegmatis cpn60.1* gene was replaced with the *M. tuberculosis cpn60.1* gene. The plasmid was then sequenced before being transformed into strain B and analysed using the complementation assay used by Ojha *et al*.

Before the experiment was carried out, a colleague was already working on trying to replicate the original results obtained by Ojha *et al.*, but was unable to do so in our lab (Tara Rao, PhD thesis 2010). Further attempts made by me to replicate the results using the same protocol also failed. Tara Rao then made various alterations to all the variables in the protocol including the use of a different growth media, using HPLC grade water, including tween80, changing the iron concentration, and even changing the glassware. Unfortunately none of these resulted in reproducible results. It was only when she went to the lab in Pittsburgh herself, that she was able to obtain the desired results. As such, it was decided that after the construction and transformation of the required plasmid was complete, it would be sent to Ojha *et al*. in Pittsburgh so that the complementation assays could be done by them.

6.1.1 Construction and transformation of the plasmid

To clone the Mtb*cpn60.1* gene into the pMsGroEL1 plasmid a combination of restriction digests and PCRs were used. The original pMsGroEL1 plasmid had a HindIII site both upstream and downstream of the Ms*cpn60.1* gene. As such it was possible to remove the Ms*cpn60.1* sequence by digesting the plasmid with HindIII using the protocol in section 2.4.8, running the resultant fragments on a 1% agarose gel, and then extracting the vector backbone as per the manufacturer protocol (Qiagen gel extraction kit). The vector was then made blunt ended by using Klenow (section 2.4.9). During this process, phusion polymerase (NEB) was used to PCR amplify the Mtb*cpn60.1* gene from the Ptrc-GroES-MtbCpn60.1 plasmid. This results in a blunt ended PCR product. The Vector backbone was then ligated with the PCR amplified Mtb*cpn60.1* and transformed into DH5α. Since blunt ended ligations

can produce products with the wrong orientation, this was checked with restriction digests and PCR screening. The plasmid was then also sequenced to ensure no undesired mutations were present before being transformed into strain B to be sent to Pittsburgh. A schematic representation of the process can be seen in figure 6.1.

6.1.2 Results obtained from America

Since the original complementation assay results could not be replicated in our lab in Birmingham, the construct had to be sent to Pittsburgh for the complementation analysis. However, to check the reliability of the experiments the constructs were first sent blind, such that the collaborators in Pittsburgh were not aware of the identity of the strains. This was done to see if they could correctly identify the constructs based on their phenotypic results alone.

Based on the results that were obtained, they were able to correctly identify all the control strains. Their results showed that both strain A (original strain) and C (complementing Ms*cpn60.1*) produced mature biofilms (fig 6.2a and 6.2c), while strain B (ΔMs*cpn60.1*) lacked any signs of biofilm maturation (fig 6.2b).

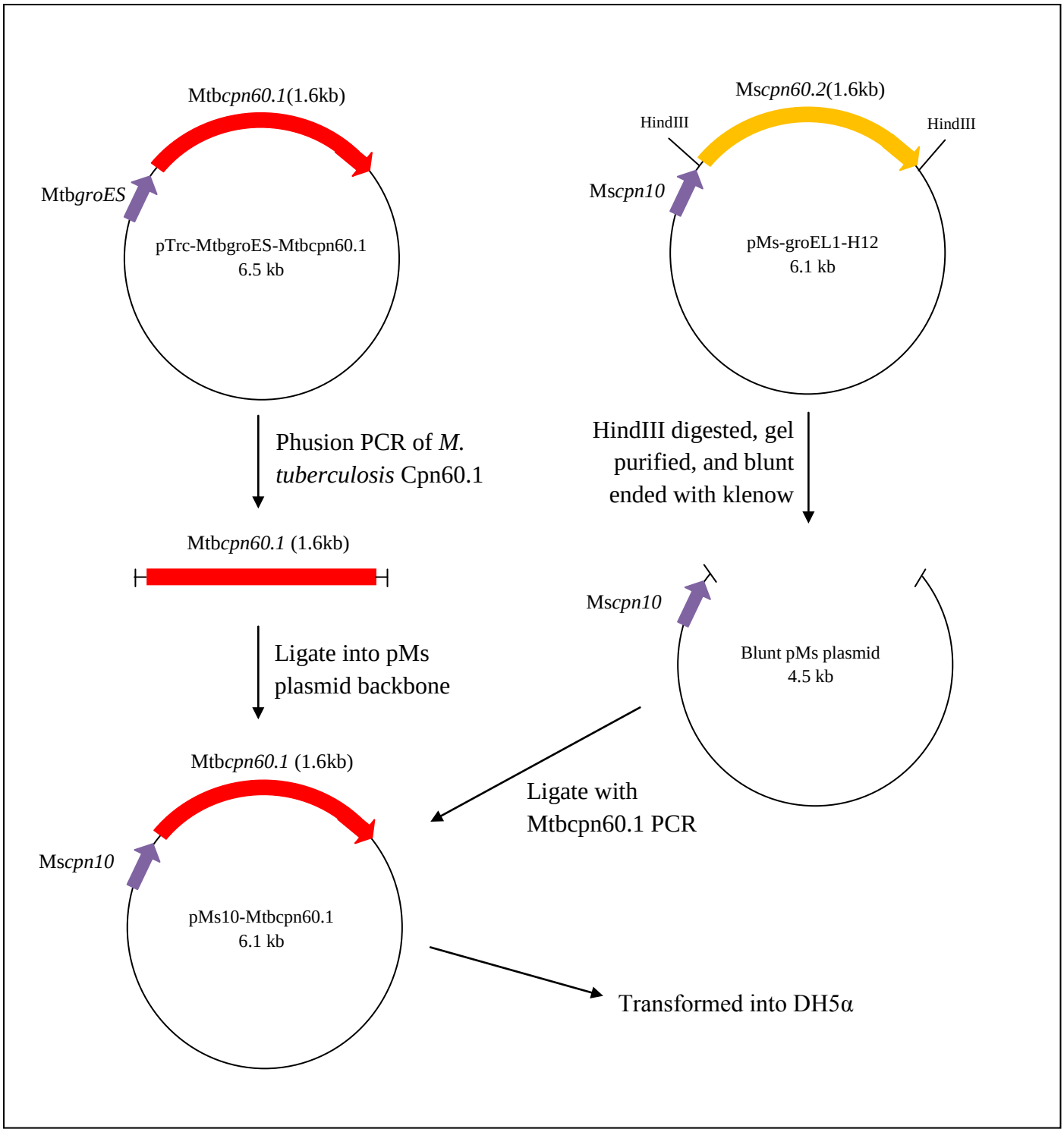


Fig 6.1: A schematic representation of the process used to clone Mtb*cpn60.1* into the pMsGroEL1 plasmid.

As the group in Pittsburgh were able to correctly identify the control strains and obtained expected results, the data from the remaining constructs was analysed. The results that were obtained after 5 days incubation were as expected. According to their analysis, strain D (Ms*cpn60.2*) did not show any major signs of biofilm formation (fig 6.2d); while strains E (Ms*cpn60.3*) and F (Ec*groEL*) showed partial complementation, i.e. the phenotypic biofilm results were in between the positive and negative controls (fig 6.2e and fig 6.2f).

Interestingly, and consistent with the hypothesis, they were also able to show that MtbCpn60.1 (in strain G), was able to complement in the knockout strain and produce a mature biofilm similar to that of Mscpn60.1 (in strain C) (fig 6.2g). This confirmed the original hypothesis that MtbCpn60.1 can substitute for the loss of endogenous MsCpn60.1 in *M. smegmatis*. It should be noted however that visually the complemented strains do not produce as much biofilm as the original strain (strain A). However, upon further inspection it was found that the mass of the biofilm was roughly the same for the complemented strains and the wild type while being lower for the non complemented strains (fig 6.3). Interestingly, the mass of the biofilm from the strain carrying the Ec*groEL* gene was also as good as that of WT.

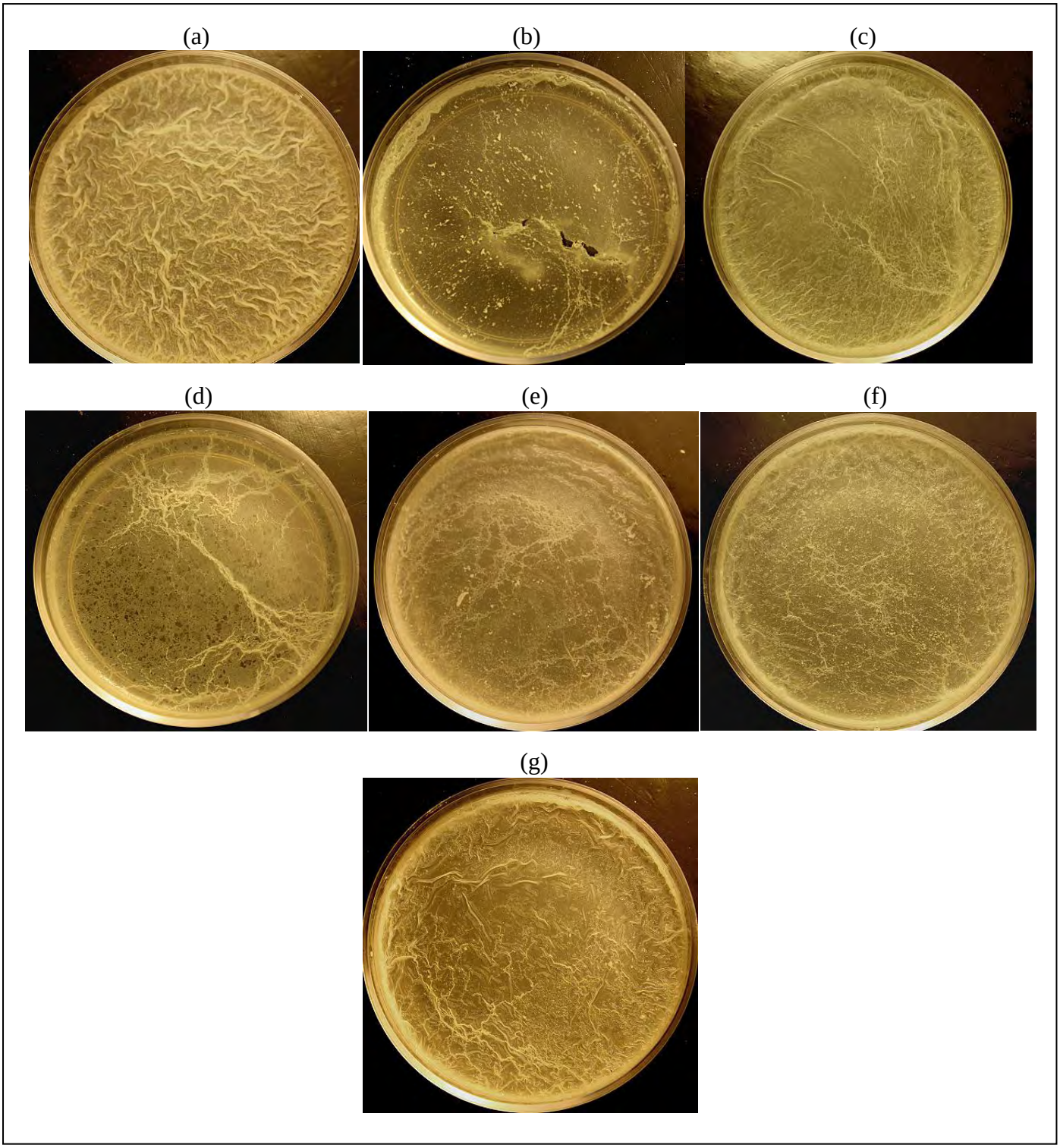


Fig 6.2: Biofilm assay results after 5 days of incubation. The assay was done with strains A (a), B (b), C (c), D (d), E (e), F (f), and G (g).

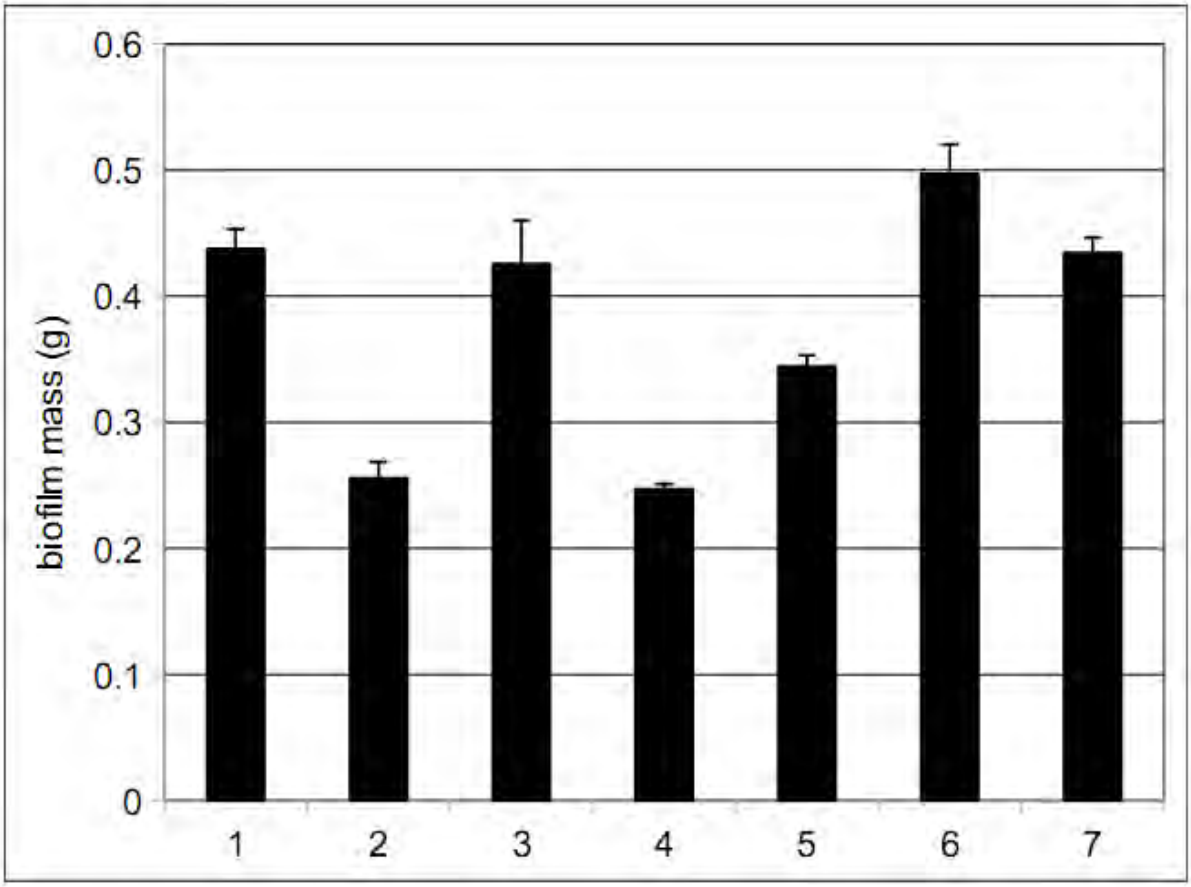


Fig 6.3: Masses of 4 day old biofilms. *M. smegmatis* mc^{2}155 (1), *M. smegmatis* mc^{2}155 Δ*cpn60.1* (2), *M. smegmatis* mc^{2}155 Δ*cpn60.1*/pMs10-cpn60.1 (3), *M. smegmatis* mc^{2}155 Δ*cpn60.1*/pMs10-cpn60.2 (4), *M. smegmatis* mc^{2}155 Δ*cpn60.1*/pMs10-cpn60.3 (5), *M. smegmatis* mc^{2}155 Δ*cpn60.1*/pMs10-EcgroEL (6), *M. smegmatis* mc^{2}155 Δ*cpn60.1*/pMs10-tbcpn60.1 (7). Assays were repeated at least in triplicate. Image obtained from Ojha *et al.*

Discussion

As discussed in the introduction, our understanding of the properties of the chaperonins of Mycobacteria suggests they have different functions within the cell. In *M. tuberculosis*, Cpn60.1 has been shown to be unnecessary for cell viability, but its deletion causes a complete lack of granuloma formation in infected mice and guinea pigs (Hu *et al.*, 2008). Similarly in *M. smegmatis*, Cpn60.1 is not needed for cell survival but plays a role in its biofilm maturation (Ojha *et al.*, 2005). The results obtained in chapters 3 and 4 also support this view, as it was demonstrated that only Cpn60.2 (and chimeras between Cpn60.2 and GroEL) could function as house-keeping chaperonins and not Cpn60.1.

As MtbCpn60.1 and MsCpn60.1 are more closely related to each other than Cpn60.2 (section 1.5.5), it suggested that these chaperonins could be interchangeable between the two organisms. Although the ability of MsCpn60.1 to complement in *M. tuberculosis* has not been tested, the results obtained with *M. smegmatis* enable the prediction that it would be able to complement for the loss of the MtbCpn60.1 chaperonin. The results obtained here show that MtbCpn60.1 can complement for the loss of endogenous Cpn60.1 in *M. smegmatis*. The phenotype observed was also very similar to that seen when the knockout strain is complemented with its own Cpn60.1. A point to note however is that although these constructs could complement in *M. smegmatis;* neither of them seemed to produce biofilms at the same level as that seen in the WT strain. However when the mass of the biofilms was analysed, it was shown that the complemented strains had biofilms that weight roughly the same as that of the WT strain and higher than the un-complemented strains. Another point to remember is that the images that were shown here were taken after a period of 5 days rather

than 7 days, so there is a possibility the biofilms would have matured further, but this would not have affected the overall results.

It would have been interesting to further analyse the properties of these chaperonins. One possibility would have been to make apical domain swaps between the Cpn60.1 and Cpn60.2 chaperonins of *M. smegmatis* to test the hypothesis that the apical domain is responsible for substrate specificity. The resulting chimeric chaperonins would be tested in a similar manner as the complementation tests shown above and in *E. coli* (chapter 3). Another possible experiment could involve making specific mutations between the chaperonins to try and find the regions that are responsible for biofilm maturation. Other experiments, such as testing if MsCpn60.1 could allow *M. tuberculosis* ΔMtb*cpn60.1* strains to cause granuloma in infected mice would also have further expanded our understanding of the specialisation of the Mycobacterial chaperonins.

In summary, the results obtained here support the original hypothesis and show that MtbCpn60.1 can complement for the loss of endogenous Cpn60.1 in *M. smegmatis*. These results are consistent with the phylogenetic data (Figure 1.20c), and with the hypothesis that cpn60 duplicated in the strain that was ancestral to all known existing Mycobacteria, and that separate functions evolved for one of the two copies that have been maintained subsequently (Hughes, 1993, Hu *et al.*, 2008, Lewthwaite *et al.*, 2001).

Section 6.2: Testing the existence of double ring *M. tuberculosis* Cpn60.2 by attempting to make single ring mutants

Background

It has been proposed that Cpn60.2 from *M. tuberculosis* tends to exist as a lower oligomer (Qamra *et al.*, 2004). However, chaperonin function (as currently understood) relies on the ability of chaperonins to form larger oligomeric structures to fold proteins in the central cavity of one of the two rings. The release of this protein from the first ring is then orchestrated by the interaction of ATP and substrate on the second ring. As a result, the unstable nature of these chaperonins was initially surprising, and suggests either that these chaperonins function by a non-canonical mechanism or that oligomers do in fact form under certain defined conditions. The results obtained in the previous chapters and by Elsa Zacco (Master's thesis 2010) have shown that Cpn60.2 can function as a chaperonin in *E. coli*, and supports the hypothesis that it forms higher oligomeric structures under appropriate conditions. These results, along with the fact that Cpn60.2 can function with both Cpn10 and GroES in *E. coli*, both of which form heptameric rings (Hunt *et al.*, 1996, Roberts *et al.*, 2003), further supported the hypothesis that Cpn60.2 forms ringed structures *in vivo*.

The structural stability of GroEL is vital for its function. It has been shown that making specific mutations in the equatorial domain of GroEL causes it to form single ringed structures, referred to as SR1 (Weissman *et al.*, 1995). This single ringed structure, although able to interact with and even provide a chamber for substrate folding, is unable to release the substrate. This occurs because the allosteric signals that are sent from the second ring, which

are required to facilitate the release of the substrate from the first ring, are absent. As a consequence of this, SR1 loses its ability to function as a chaperone (Weissman *et al.*, 1996). A point to note here is that functionality of the SR1 GroEL variant can in fact be resorted by making further specific mutations (Sun *et al.*, 2003). It was shown that the functionality of SRI was restored because GroES was bound very weakly in the presence of ATP, and effectively abolished in the presence of ADP. Also, it is known that there are chaperonins in nature like the mitochondrial chaperonin, Hsp60, that also function as single rings (Nielsen & Cowan, 1998). However, this still requires the formation of at least one ring within which a substrate can fold. The data on Cpn60.2 suggested that they are even more structurally unstable. So it was hypothesised that the *in vivo* oligomeric property of Cpn60.2 could be indirectly analysed by making single ring mutations. If these mutations resulted in Cpn60.2 losing its ability to function, then it would in essence support the idea that Cpn60.2 forms double ringed structures during its substrate folding cycle.

Experimental design

To obtain the single ring variant of GroEL, mutations were made to the residues R452, E461, S463, and V464. Residue 452 was replaced with Glu, while residues 461, 463, and 464 were replaced with Ala (Weissman *et al.*, 1995). To try and replicate these results in Cpn60.2, the Cpn60.2 and GroEL sequences were first aligned (fig 6.4) and then appropriate primers were designed to make the required mutations using site directed mutagenesis (listed in section 2.3.2).

For these experiments the original pTrc-GroES-Cpn60.2 and pTrc-GroES-GroEL plasmids were used. Since it was not known if all four mutations were necessary to obtain a single ring variant (A. Horwich, personal communication), each of the four mutations were initially

made separately. A figure showing the sequence alignment and the residues that were mutated has been shown below in figure 6.4. The next step in these experiments was to analyse the complementation and expression data of the Cpn60.2 single mutants in MGM100. After these mutations were analysed, similar mutations were made in GroEL to see if it can replicate the results obtained with Cpn60.2 and if any one of these mutations result in the formation of SR1. GroEL was used in this experiment because native gels of the Cpn60.2 mutants would not provide any useful information, since WT Cpn60.2 shows up as a monomer/dimer on the gel.

6.2.1 Construction of the plasmids

Site directed mutagenesis was used as per the protocol in section 2.4.13, to change the amino acid sequence of *cpn60.2* to match that of *groEL*. Since four different mutations will be referred to in this section, they will be referred to as Cpn60-K449G (first mutation only), Cpn60-E458A (second mutation only), Cpn60-G460A (third mutation only), and Cpn60-V461A (fourth mutation only). The four sites that were altered including the mutations that were made have been listed in the table below:

Mutation	Old Residue	Old Sequence		New Residue	New sequence
Cpn60-K449G	K449	ccgctgaagcagatc		G449	ccgctggggcagatc
Cpn60-E458A	E458	gggctggagccgggc		A458	gggctggcgccgggc
Cpn60-G460A	G460	gagccgggcgtggtg		A460	gagccggccgtggtg
Cpn60-V461A	V461	ccgggcgtggtggcc		A461	ccgggcgcggtggcc

Table 6.1: List of residues and mutations that were made to obtain the four required Cpn60.2 mutants

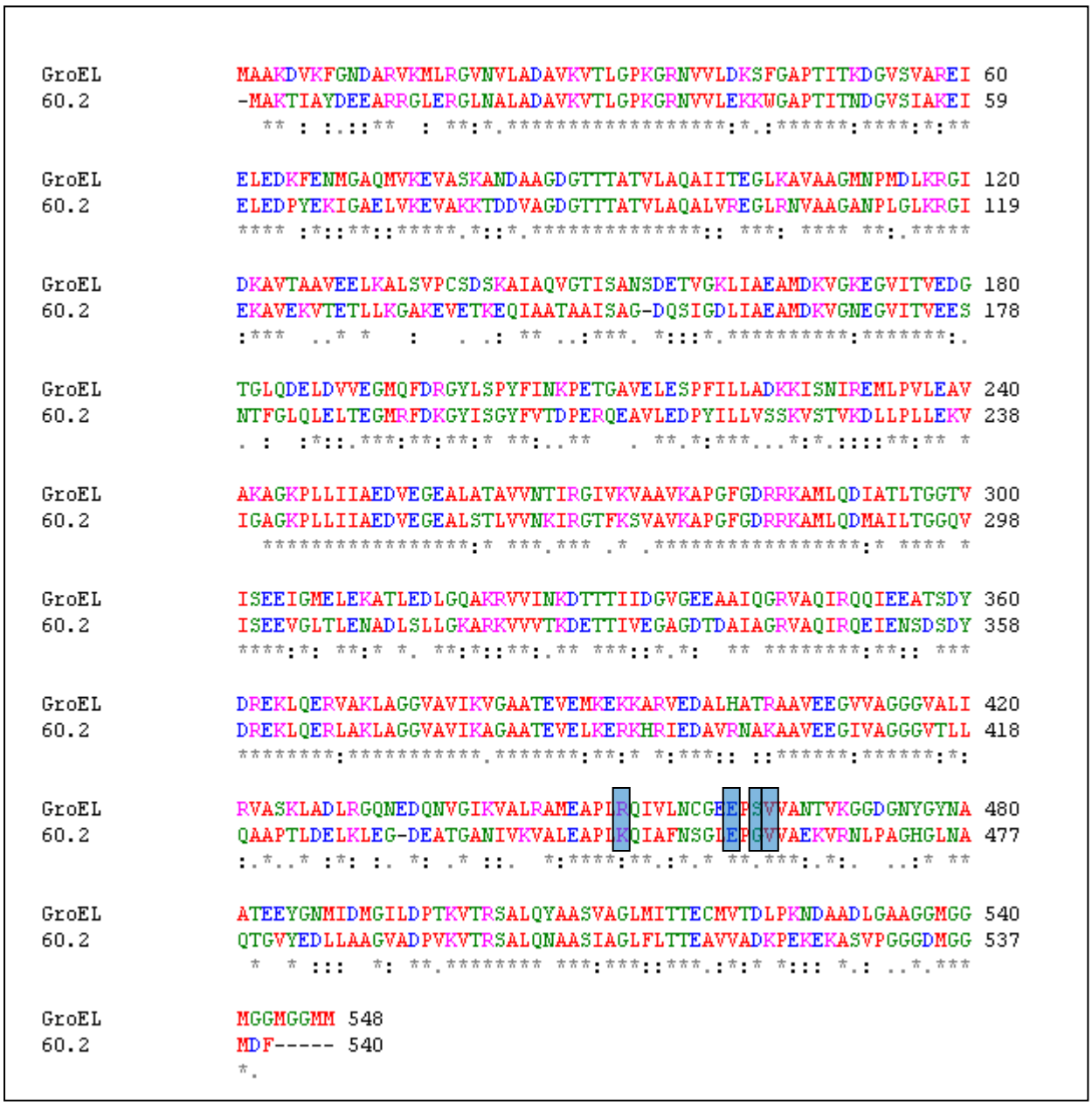


Fig 6.4: Sequence alignment between GroEL and Cpn60.2. The blue regions represent the residues that were selected for mutation.

6.2.2 Complementation and expression results of Cpn60.2 mutants

After all the four plasmids were obtained, they were sequenced to ensure the required mutations were present and that there were no undesired mutations. The constructs were then transformed into MGM100 using the protocol described in section 2.4.16. These strains were then analysed to test their complementation abilities using the protocol described in section 2.5, but briefly, the resultant MGM100 + pTrc strains were grown overnight in arabinose, then sub-cultured and grown in 200 ml LB + 0.2% glucose the following morning for 2 hours, and then serially diluted (in ten-fold steps) to spot on to square petri dishes. 5µl of each dilution was spotted onto the plates with the final dilution factor being 10^7.

The complementation analysis showed that out of the four mutants, only Cpn60-E458A was unable to complement for the loss of endogenous GroEL in MGM100 in the presence of glucose (fig 6.5a). Another point to note is that, while Cpn60-K449G, Cpn60-G460A, and Cpn60-V461A all grew very well in glucose and arabinose (with IPTG), strains with Cpn60-E458A struggled to grow even in arabinose (fig 6.5b). This is not a surprising result, as it was expected that if any of the four mutations were in fact affecting the oligomerisation of Cpn60.2, then that mutant would also lose the ability to function. It was also expected that if this chaperonin was then over-expressed in the presence of IPTG (on glucose or arabinose) then it could be lethal for the MGM100 strain. The reason being, that this mutant chaperonin would interact with and trap its substrates and not be able to release them. Since some of these substrates are essential for cell survival, the lack of ability to fold and release them would result in cell death.

To ensure that the lack of complementation was not due to the lack of expression of Cpn60-E458A, an SDS-PAGE gel was run (fig 6.6). Since Cpn60-E458A was causing cell death in the presence of arabinose as well, it was difficult to induce the cells for long durations of time. As a result, this plasmid could only be induced for 2-3 hours. It was observed that the expression level of Cpn60-E458A was lower than that of Cpn60-K449G, Cpn60-G460A, and Cpn60-V461A. However, even this level of expression should have been enough to allow Cpn60-E458A to complement in MGM100 if it was functional.

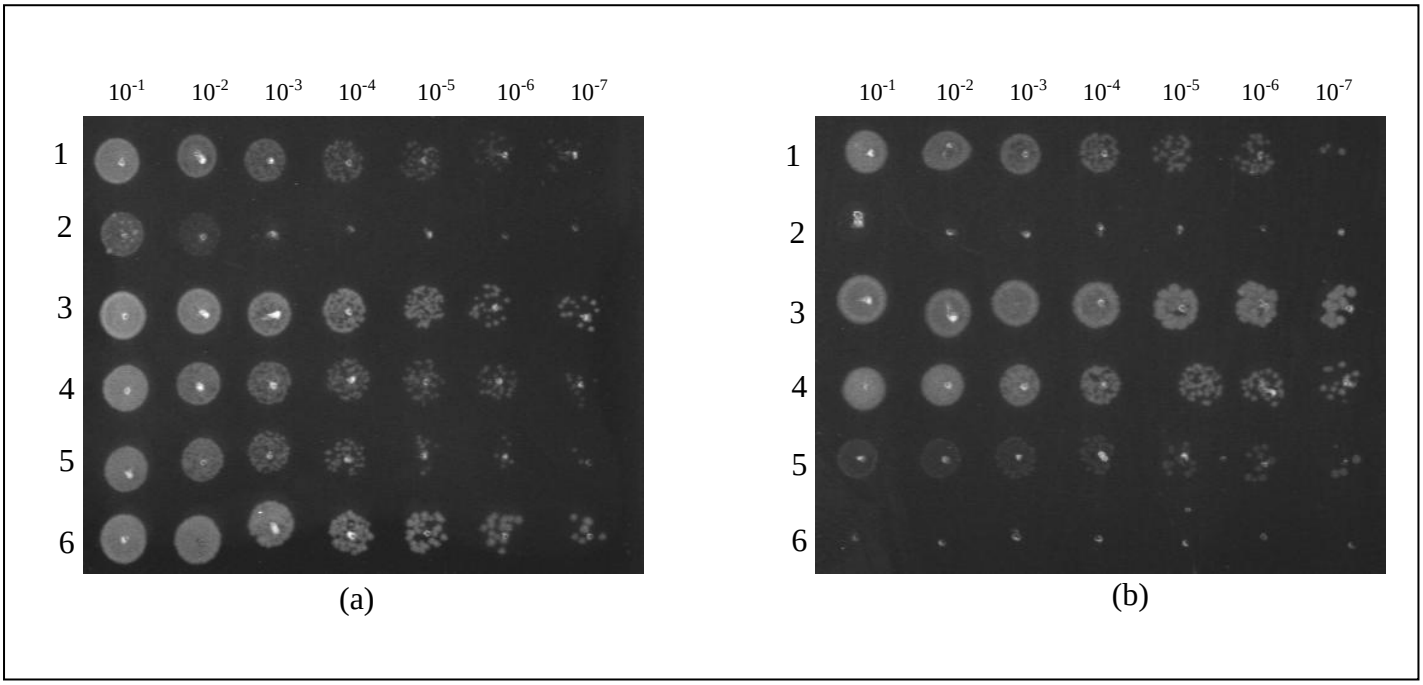


Fig 6.5: Complementation assay of Cpn60.2 mutants in glucose or arabinose. (a) Cells grown on LB plates at 30°C with 0.2% arabinose and 0.1 mM IPTG. (b) Cells grown on LB plates at 37°C with 0.2% glucose and 0.1 mM IPTG.

1- MGM100/ pTrc-GroES-Cpn60-K449G

2- MGM100/ pTrc-GroES-Cpn60-E458A

3- MGM100/ pTrc-GroES-Cpn60-G460A

4- MGM100/ pTrc-GroES-Cpn60-V461A

5- MGM100/ pTrc-GroES-GroEL

6- MGM100/ pTrc99a

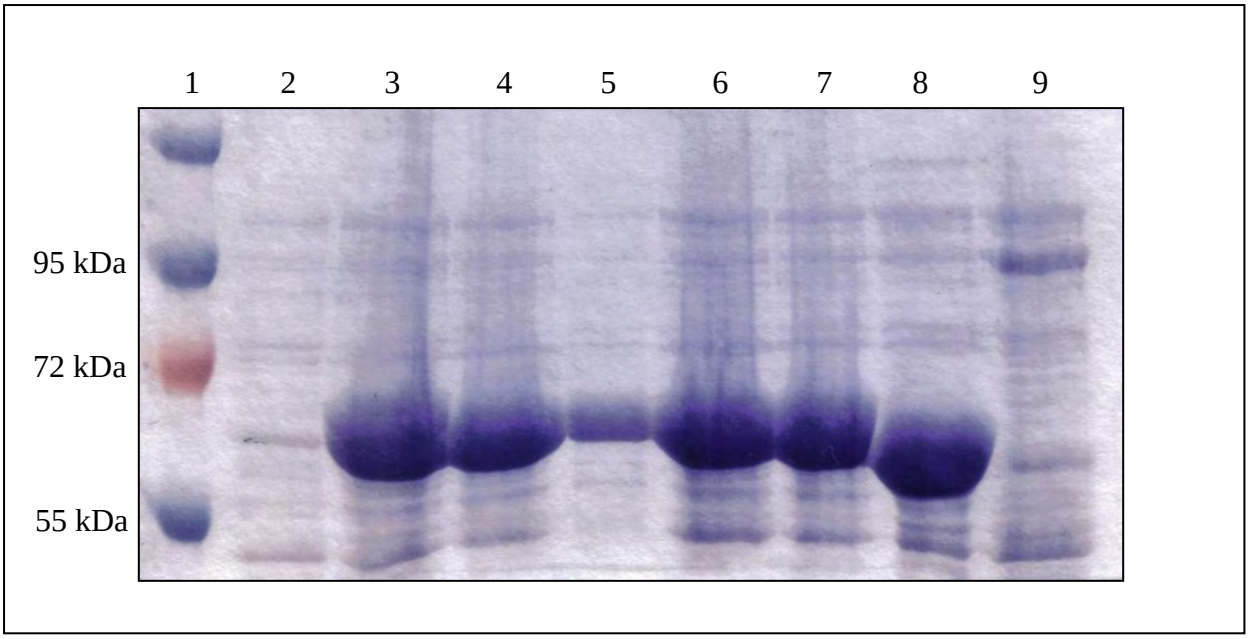


Fig 6.6: SDS-PAGE gel of Cpn60.2 mutants showing over-expression of all four mutants. All the samples, except Cpn60-E458A, were taken from cultures after 5 hours incubation at 37°C with 0.1 mM IPTG and 0.2% arabinose. Cpn60-E458A sample was taken from a culture after 2.5 hours of incubation at 37°C with 0.1 mM IPTG and 0.2% arabinose.

1- Prestained protein marker (Fermentas)

2- MGM100/ pTrc-GroES-MKO

3- MGM100/ pTrc-GroES-JNL

4- MGM100/ pTrc-GroES-Cpn60-K449G

5- MGM100/ pTrc-GroES-Cpn60-E458A

6- MGM100/ pTrc-GroES-Cpn60-G460A

7- MGM100/ pTrc-GroES-Cpn60-V461A

8- MGM100/ pTrc-GroES-GroEL

9- MGM100/ pTrc99a

6.2.3 Complementation analysis of the GroEL mutant

Since the 2nd mutation in Cpn60.2 was enough to render it non functional, attempts were made to see if the same mutation could be made in GroEL, and if it could, then to see if it produced a non functional single ring oligomer on a native gel. The same site directed mutagenesis approach was used for GroEL as was done to make the Cpn60.2 mutants. Primers were designed so that E461 could be changed to A461, the details of which can be seen in the primers list in section 2.3.2. After mutagenesis, the resulting plasmid was sequenced to ensure that the required mutation was present before being transformed into MGM100 to test its functional ability. For the purpose of this thesis, the GroEL mutant will be referred to as GroEL-E461A.

The complementation assay protocol was the same as that used for the Cpn60.2 mutants. The results that were obtained showed that GroEL-E461A does in fact work in *E. coli* MGM100 and complements well for the loss of endogenous GroEL (fig 6.7). No growth issues were encountered in either arabinose or glucose in the presence or absence of IPTG. Since a fully functional chaperonin was obtained, a protein expression gel was not done. Instead, the remaining three possible mutations were attempted to see if they yielded similar results, and if all four mutations were in fact necessary to form SR1. Unfortunately, these mutations could not be made. Although colonies were being obtained after mutagenesis, the only method of screening them was by sequencing. Between 10 and 15 colonies were screened for each of the three mutants without success. Although this is a very small number of colonies, lack of time and financial resources prevented further screening.

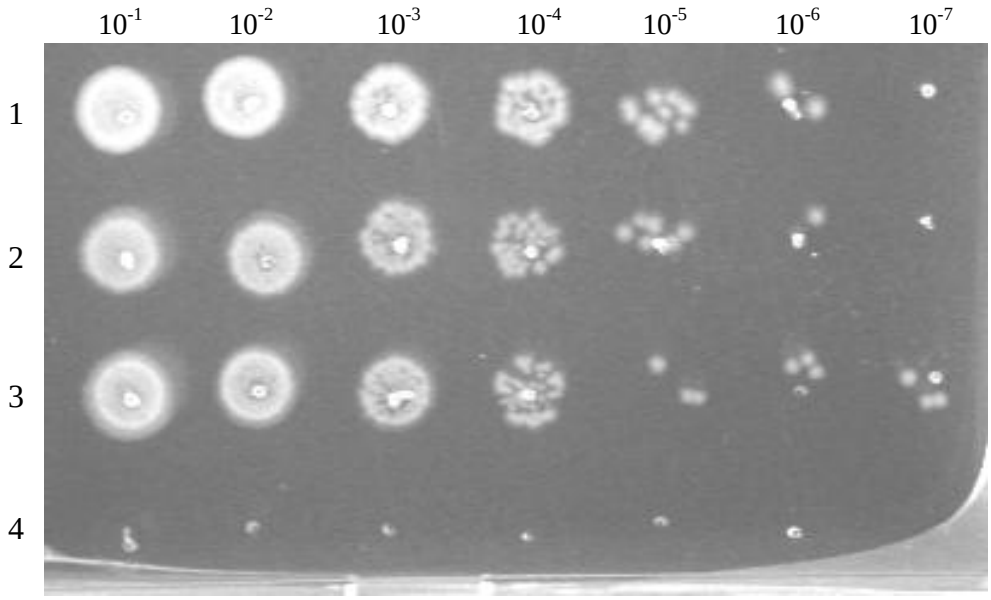


Fig 6.7: Complementation assay of GroEL-E461A mutant in glucose. Cells were grown on LB plates at 30°C with 0.2% glucose and 0.1 mM IPTG.

1- MGM100/ pTrc-GroES-GroEL-E461A

2- MGM100/ pTrc-GroES-GroEL

3- MGM100/ pTrc-GroES-Cpn60.2

4- MGM100/ pTrc99a

Discussion

The double ring structure of chaperonins like GroEL play a major functional role, although single ringed chaperonins are also known to occur in nature that are functional (Nielsen & Cowan, 1998). In either case, a ringed structure (either double or single) is needed to allow substrates to fold within a central cavity. Work done in our lab has already proven the existence of Cpn60.2 tetradecameric structures under certain conditions *in vitro* (Elsa Zacco, Master's thesis 2010, University of Birmingham). It was also shown in the last chapter that JNL forms large tetradecameric structures under the same conditions *in vitro*. However, it becomes a lot more difficult to prove the existence of tetradecameric Cpn60.2 structures *in vivo* due to its unstable nature. The main reason why it is believed Cpn60.2 does form ringed structures *in vivo* is because it can function in *E. coli* MGM100, both with Cpn10 and GroES. As has been stated before, both these cochaperonins are known to exist as heptameric structures (Hunt *et al.*, 1996, Roberts *et al.*, 2003). In section 3.7 this was confirmed by deleting the cochaperonin GroES from the plasmid and showing that Cpn60.2 cannot function in *E. coli* MGM100 without it.

The results that have been obtained here are a little difficult to interpret. It cannot be claimed that Cpn60-E458A proves that the Cpn60.2 forms a double ringed structure, since the same result could not be obtained with GroEL with the same mutation. The GroEL-E461A complementation data was a little surprising. This mutation had no affect on the functional ability of the chaperonin, although similar work done on GroEL where a E461K mutation was made resulted in the cells being heat sensitive (Sewell *et al.*, 2004). However a point to note here is that this mutation alone did not result in the formation of SR1 either. The results obtained here can however support the possibility that the second mutation was able to

destabilise the oligomeric properties of Cpn60.2 further, to the point where it lost its ability to function. It can be suggest that since a single mutation was needed for Cpn60.2 to lose its functional abilities in *E. coli*, a combination of all 4 would also be expected to be non functional. Since we know that GroEL can be made into a single ring with those 4 mutants, which also renders it non functional, it indirectly supports the possibility of Cpn60.2 forming ringed structures *in vivo*.

In summary, the data presented here supports but does not confirm the hypothesis that Cpn60.2 forms large tetradecameric structures *in vivo*. Further work to improve the results obtained here could involve making all 4 mutations in Cpn60.2, rather than assuming what its functional properties would be based on one negative result, and then comparing it with SR1. The Cpn60-E458A mutant could also have been put through the MS under the same conditions that showed Cpn60.2 forms tetradecamers to see if only heptamers were obtained. Other experimental techniques could also include the use of negative stain EM to visually observe if any single rings were present, however the unstable nature of Cpn60.2 may make it difficult to obtain the required results.

Section 6.3: Testing the hypothesis that phosphorylation is needed for improved oligomerisation

Background

Protein phosphorylation is a very important part of many cellular functions. It is involved in changing the properties of many proteins and hence used as a major method of regulating the activity of these proteins.

It is expected that phosphorylation also plays a major role in the cellular processes of *M. tuberculosis* as it has been shown that the *M. tuberculosis* genome encodes 11 Ser/Thr protein kinases (STPKs) (Av-Gay & Everett, 2000, Wehenkel *et al.*, 2008). It has also been shown that phosphorylation promotes the pathogenicity of Mycobacteria (Wehenkel *et al.*, 2008, Walburger *et al.*, 2004). Research on these kinases along with the *M. tuberculosis* Cpn60.1 has revealed that Cpn60.1 is in fact a substrates of these kinases (Canova *et al.*, 2009).

Interestingly, there are potential phosphorylation sites present in the equatorial region of *M. tuberculosis* chaperonins and as a result Kumar *et al.* recently suggested that the oligomerisation of these chaperonins was dependent on the phosphorylation of these sites (Kumar *et al.*, 2009). They showed that they could obtain tetradecameric structures for Cpn60.1, and that only these tetradecameric forms were phosphorylated on Serine 393, and not any of the lower oligomers. They suggested that this should also be true for Cpn60.2.

I found these results and the suggestions that were put forward a little odd. Firstly, M. J. Canova, L. Kremer, and V. Molle, who have done a lot of work on protein phosphorylation, have recently shown that although there are phosphorylation sites in the equatorial domain of Cpn60.1, they are mapped to Thr25 and Thr54 rather than Ser393 (Canova *et al.*, 2009). Secondly, Kumar *et al*., have also stated that the lack of function of Cpn60.2 in *E. coli,* based on their results, may be due to the lack of oligomerisation as a result of the lack of phosphorylation in *E. coli*. While it can be agreed that Cpn60.2 may not be phosphorylated in *E. coli*, as it does not express the kinases required from *M. tuberculosis*; my results however differ in that it has been shown Cpn60.2 can in fact function in *E. coli*.

Recent information with regards to Cpn60.2 suggests that it has three phosphorylation sites present in the equatorial domain, Thr90, Thr92, and Thr127 (V. Molle, personal communication). If phosphorylation was indeed necessary for oligomerisation, then it should be possible to stop or mimic phosphorylation in Cpn60.2 by mutating the phosphorylation sites to either alanine or glutamic acid respectively. These mutants could then be tested to see if phosphorylation does in fact improve oligomerisation. It was hypothesised that changing Thr to Ala or Thr to Glu would not have any effect on the functional properties of the chaperonin but it may improve its oligomeric stability.

Experimental design

For these experiments, Cpn60.2 mutants had their phosphorylation sites either mutated to alanine (referred to as T/A60.2) which would stop phosphorylation, or to glutamic acid (referred to as T/E60.2) which has been known to mimic the affects of phosphorylation. For this purpose I was kindly provided with three pET plasmids by Virginie Molle carrying the WT Cpn60.2, T/A60.2 triple mutant, and T/E60.2 triple mutant. These were first sequenced to ensure no unwanted mutations were present before analysing their complementation properties.

Since the plasmids being used were pET plasmids that make use of the T7 RNA polymerase system to over-express proteins, it was decided that Tab21 should be used to test their complementation abilities. As discussed in chapter 3, Tab21 is a BL21 derivative that expresses T7 RNA polymerase and can be used to over-express proteins from pET plasmids. The chaperonins present on the plasmids provided by V. Molle had N-terminal His-tags, but no cochaperonin genes. Because of this, a plasmid carrying the *groES* gene was transformed into Tab21. This plasmid, pSUEH, which has a chloramphenicol resistance marker, carried both the *groES* and *groEL* genes. As such, restriction digest was used to remove the *groEL* gene before transforming it into Tab21 along with the pET plasmids. Furthermore, the presence of the N-terminal His-tags were causing issues during the complementation assay (section 6.3.2), and as such were also removed by restriction digests.

6.3.1 Modification of GroES inducing plasmid

The plasmid that was picked to over-express GroES was pSUEH. Since it already had both the *groES* and the *groEL* genes it could not be used in the complementation experiments without modification. The plasmid was sequenced to see what restriction sites were available upstream and downstream of GroEL. Unfortunately, this plasmid lacked any appropriate sites that could be used to completely remove GroEL. However, the plasmid did have a HindIII site just downstream of GroEL and an EcoRV site present approximately halfway into the *groEL* gene. So a restriction digest was done with HindIII and EcoRV to remove a large segment of the *groEL* gene, which would render the encoded protein non functional. After the digest, the sample was run on a 1% agarose gel, and the vector backbone was extracted. It was then blunt ended using Klenow, and concentrated with ethanol precipitation before being religated. Using these restriction sites approximately 800bp of the *groEL* gene was removed without affecting the expression of GroES. The entire process has been represented with a schematic figure below (Fig 6.8).

The resulting plasmid, referred to as pSUES, only expresses GroES. This was confirmed by transforming pSUES and pSUEH into Tab21 and then over-expressing them by growing the cell at 30°C in 0.2% arabinose with 0.1 mM IPTG for 3 hours. The SDS-PAGE gel shows clearly that in pSUEH both GroES and GroEL are induced, whereas in pSUES only GroES over-expression can be seen (fig 6.9).

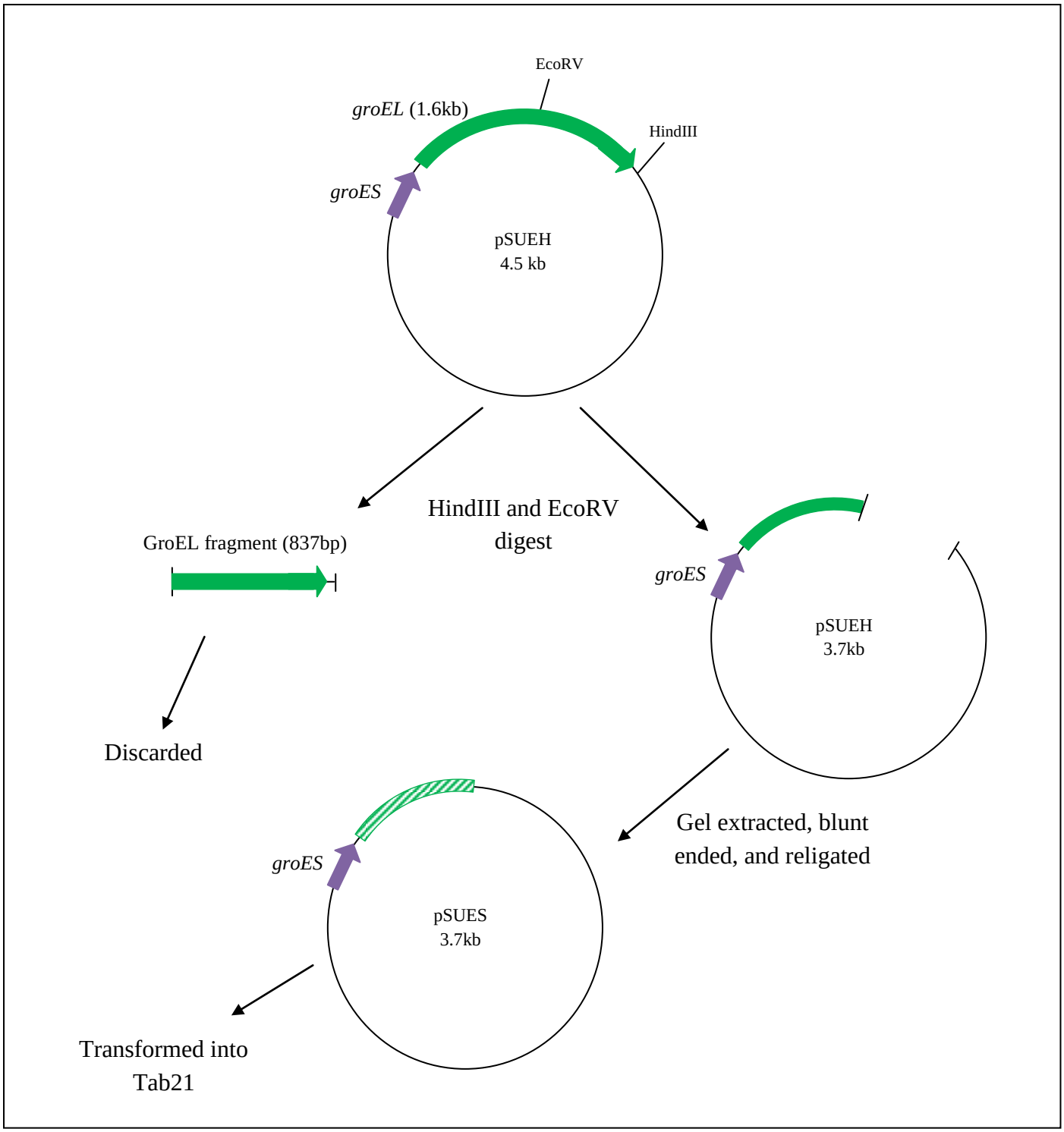


Fig 6.8: A schematic representation of the process used to modify the pSUEH plasmid to obtain pSUES.

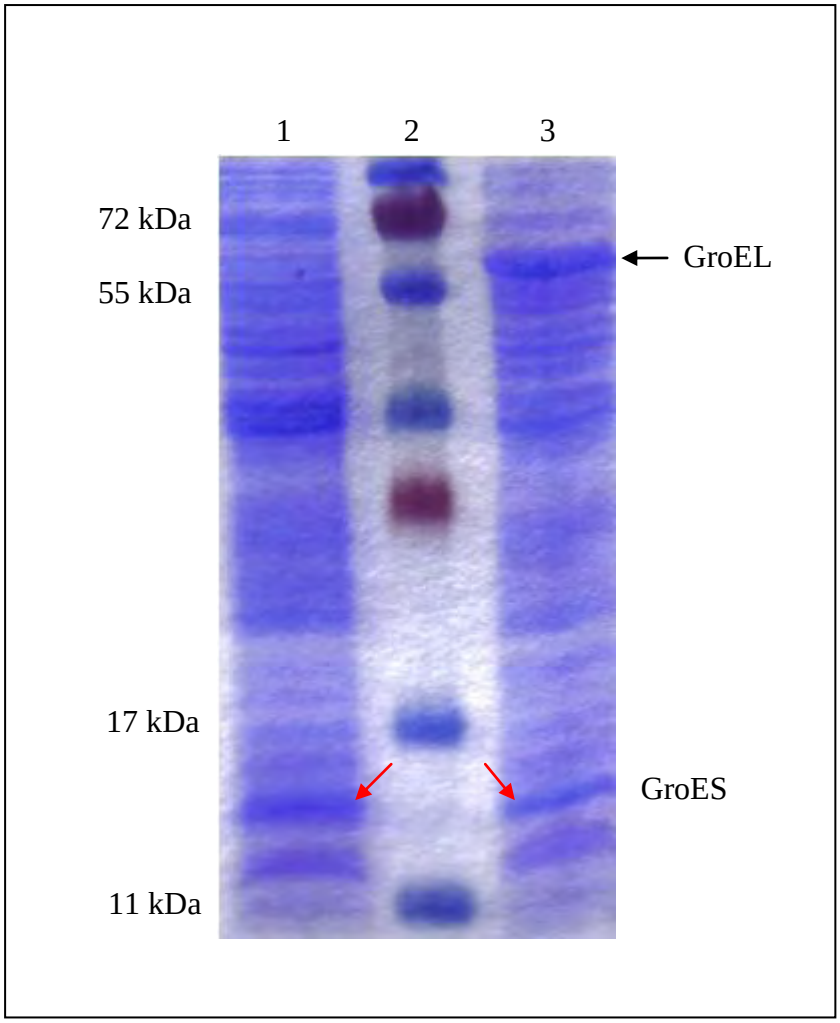


Fig 6.9: SDS-PAGE gel of pSUEH and pSUES. All the samples were taken from Tab21 cultures after 3 hours incubation at 30°C with 0.1 mM IPTG and 0.2% glucose. Red arrows are pointing at GroES.

1- Tab21/ pSUES

2- Prestained protein marker (Fermentas)

3- Tab21/ pSUEH

6.3.2 Complementation results in Tab21

After confirming that only GroES was being expressed from the pSUES plasmid, it was transformed into Tab21 along with the three pET plasmids to test complementation using the complementation protocol described in section 2.5. The resultant Tab21 + pET + pSUES strains were grown overnight in arabinose, then sub-cultured and grown in 200 ml LB + 0.2% glucose the following morning for 2 hours, and then serially diluted (in ten-fold steps) to spot on to square petri dishes. 5µl of each dilution was spotted onto the plates with the final dilution factor being 10^7.

In was noticed that while the cells grew normally in the presence of arabinose (fig 6.10a), the addition of 0.1 mM IPTG saw a drastic reduction in viable colonies (fig 6.10b). On glucose plates no complementation was seen from any of the strains, including WT (fig 6.10c). This was unexpected as it is already known from experiments done in chapter 3 that WT Cpn60.2 does complement in Tab21. To ensure the lack of complementation was not being caused by the lack of protein expression, SDS-PAGE gels were run. These confirmed that all the chaperonins were being expressed at high enough levels to be able to complement (fig 6.11). Native gel analysis by a lab colleague, MingQi Fan, showed that the oligomeric properties of the chaperonins had not changed (Data not shown).

From these results, it was not possible to conclusively say that T/A60.2 and T/E60.2 do not complement, as the WT (positive control) also did not grow. It was reasoned that, since all the Cpn60.2 constructs had His-tags on the N-terminal region, this could be causing issues with the chaperonins ability to function. As a result of this, it was decided that the His-tag would be removed and the complementation experiments would be attempted again.

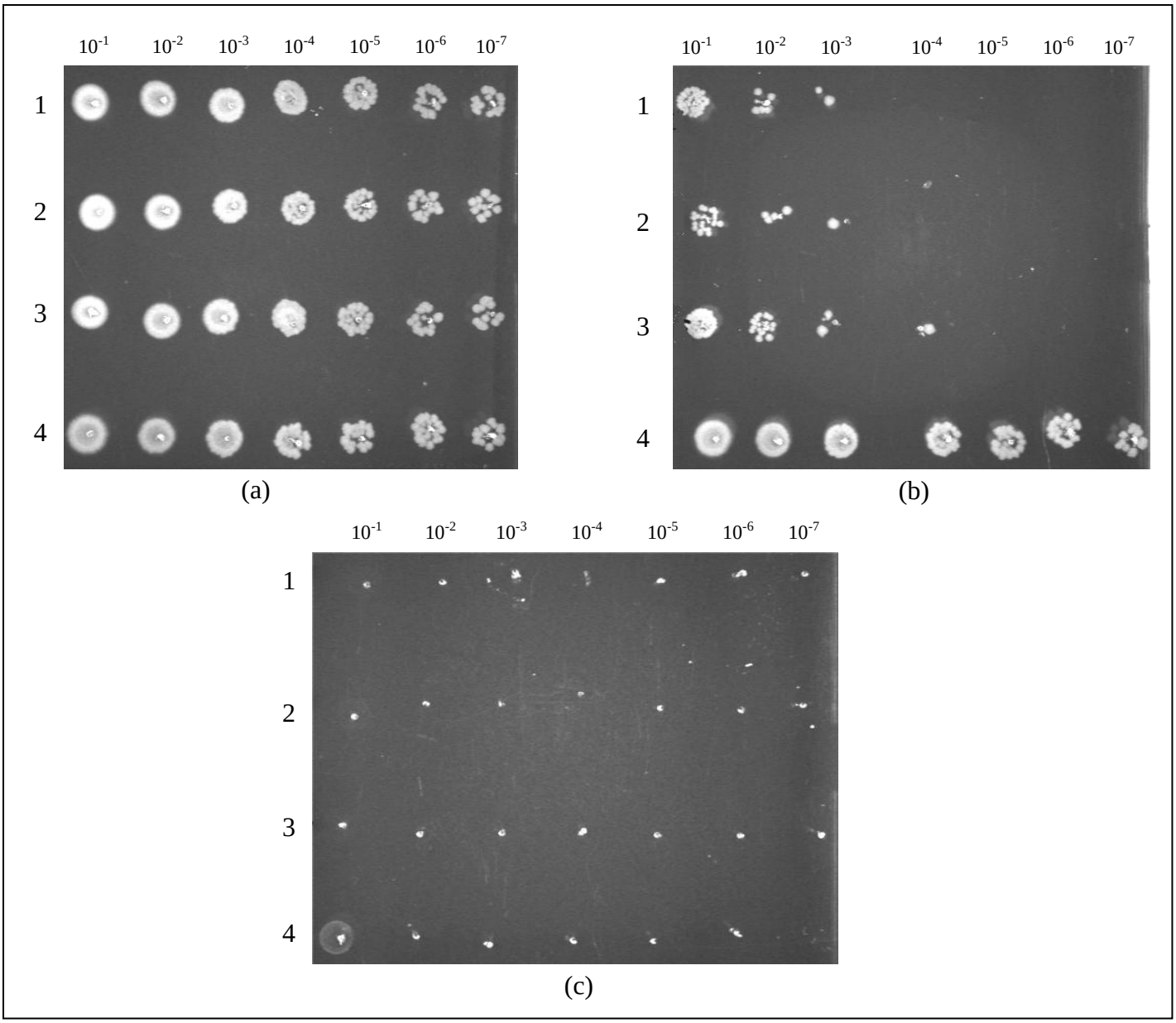


Fig 6.10: Complementation results of the T/A60.2 and T/E60.2 mutants. (a) Plate grown at 37°C in the presence of 0.2% arabinose. (b) Plate grown at 37°C in the presence of 0.2% arabinose and 0.1 mM IPTG. (c) Plate grown at 37°C in the presence of 0.2% glucose and 0.1mM IPTG.

1- Tab21 + pET-Cpn60.2 WT + pSUES

2- Tab21 + pET-T/A60.2 + pSUES

3- Tab21 + pET-T/E60.2 + pSUES

4- Tab21 + pSUES

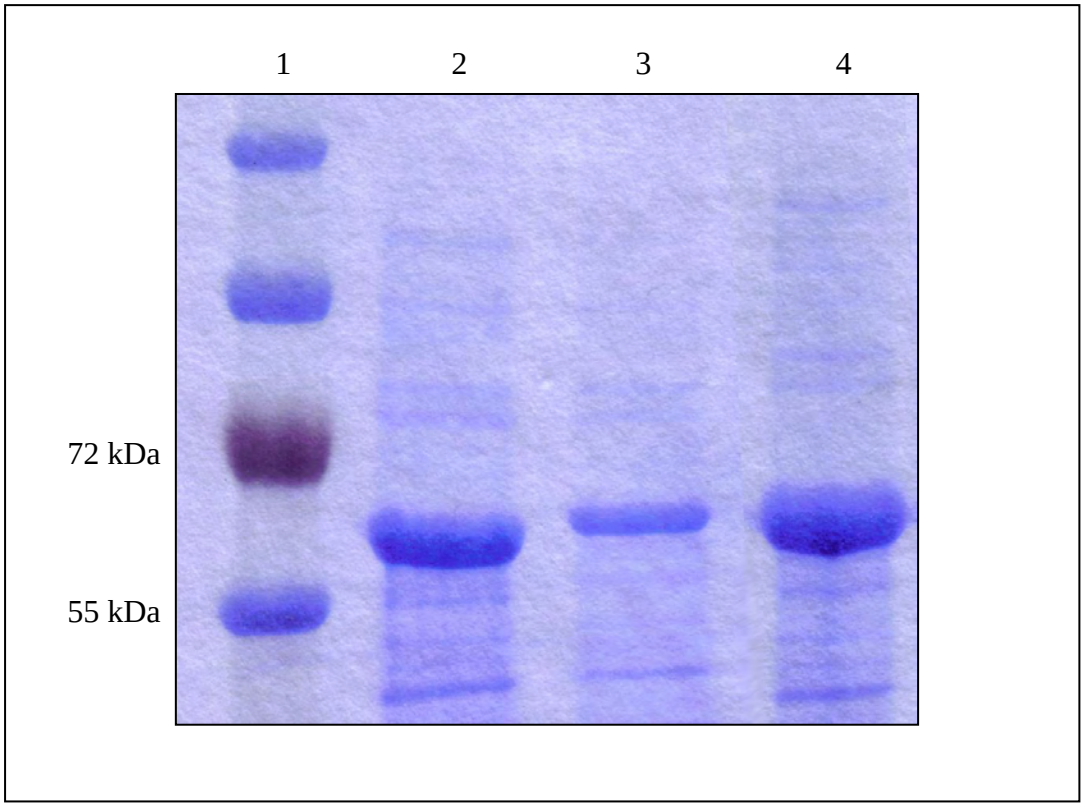


Fig 6.11: SDS-PAGE gel of Cpn60.2 WT, T/A, and T/E mutants showing over-expression. All the samples were taken from Tab21 cultures after 3 hours incubation at 37°C with 0.1mM IPTG.

1- Prestained protein marker (Fermentas)

2- Tab21 + pET-T/A60.2 + pSUES

3- Tab21 + pET-T/E60.2 + pSUES

4- Tab21 + pET-Cpn60.2 WT + pSUES

6.3.3 Complementation results in Tab21 without His-tagged chaperonins

Since His-tags can potentially affect the function of a protein, it was decided that the tags would be removed from all three pET plasmids before re-attempting the complementation assay. To do this restriction digests were used again. From the sequence data, sites that would allow the complete removal of the His-tags without affecting the ribosome binding site (RBS) or the protein sequence were picked. It was noticed that there was an NcoI site upstream from the His-tag but downstream from the RBS and an NdeI site downstream from the His-tag but upstream from the starting codon of the chaperonins. A double digest using the respective restriction enzymes was done to remove the His-tags. The vector backbone was then gel extracted and made blunt ended using Klenow before being religated. These religated plasmids were then re-sequenced to ensure the His-tags had been removed without altering any of the other sequences.

The resulting plasmids were transformed back into Tab21 and their complementation abilities were checked. Unfortunately, the results obtained here were also unexpected. The complementation data remained primarily unaltered with just a small improvement in complementation (fig 6.12). Although an improvement in growth on arabinose + IPTG plates with the WT construct was observed, it clearly showed that cells were still struggling to grow with the T/A60.2 and T/E60.2 triple mutants. On glucose plates there was a slight improvement in growth, but not enough to be regarded as good complementation. A reduction in the levels of over-expression of these chaperonins was also observed (fig 6.13), so there is a possibility that the lack of complementation was caused by the reduction in over-expression.

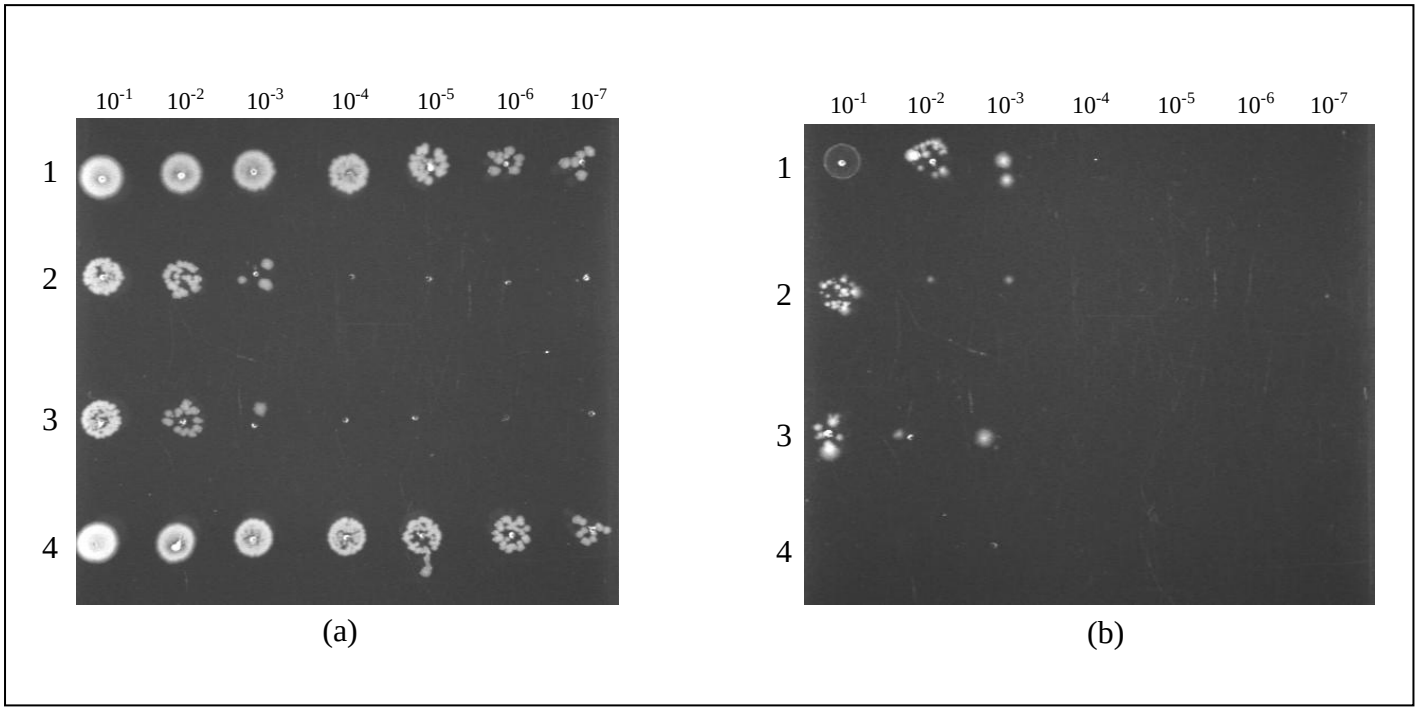


Fig 6.12: Complementation results of the WT Cpn60.2, T/A60.2, and T/E60.2 mutants without His-tag. (a) Plate grown at 37°C in the presence of 0.2% arabinose and 0.1 mM IPTG. (b) Plate grown at 37°C in the presence of 0.2% glucose and 0.1 mM IPTG.

1- Tab21 + pET-Cpn60.2 WT + pSUES (no His-tag)

2- Tab21 + pET-T/A60.2 + pSUES (no His-tag)

3- Tab21 + pET-T/E60.2 + pSUES (no His-tag)

4- Tab21 + pSUES

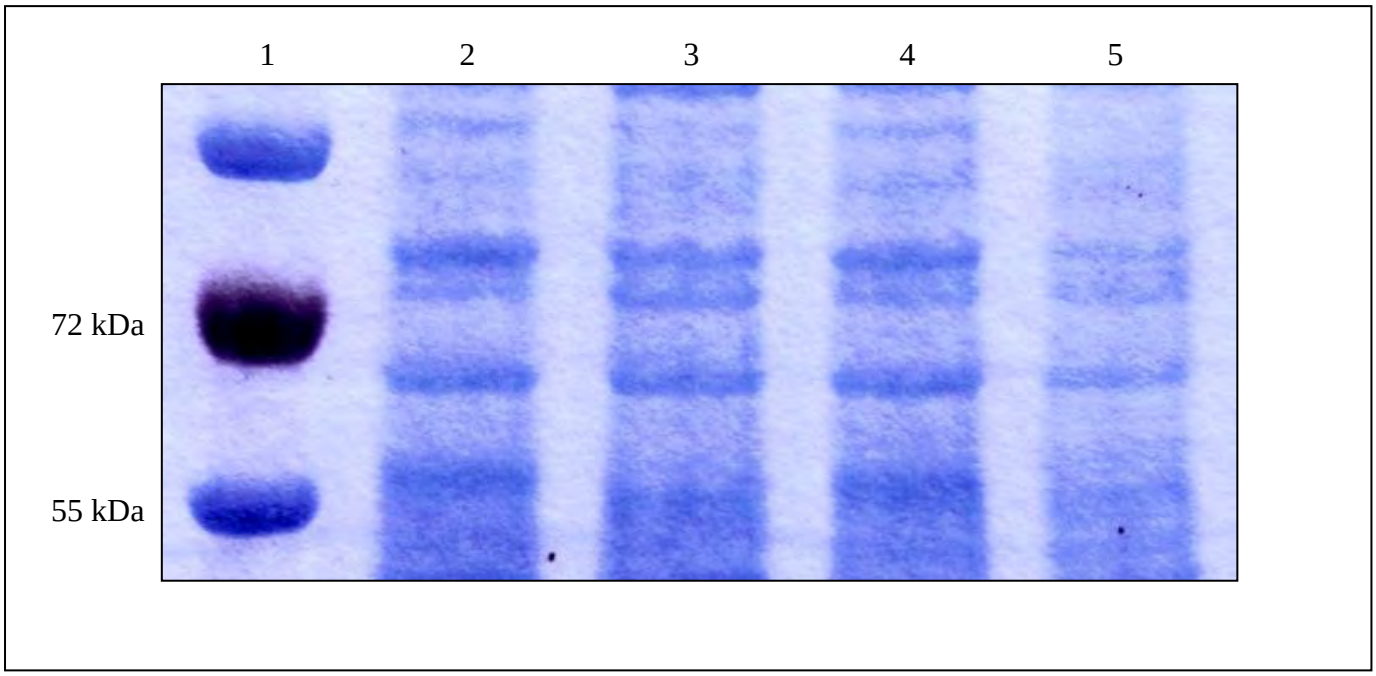


Fig 6.13: SDS-PAGE gel of Cpn60.2 WT, T/A, and T/E mutants without His-tags showing a reduction in expression levels. All the samples were taken from Tab21 cultures after 3 hours incubation in 0.2% arabinose at 37°C with 0.1 mM IPTG.

1- Prestained protein marker (Fermentas)

2- Tab21 + pET-Cpn60.2 WT + pSUES (no His-tag)

3- Tab21 + pET-T/A60.2 + pSUES (no His-tag)

4- Tab21 + pET-T/E60.2 + pSUES (no His-tag)

5- Tab21 + pSUES

Discussion

The results that have been obtained here are quite difficult to interpret. The original objective of these experiments was to see if mimicking phosphorylation resulted in Cpn60.2 oligomerising more readily. Unfortunately these mutants failed to even function in Tab21 although they seemed to be expressing well. While the cells grew fine in the presence of just arabinose, the addition of 0.1 mM IPTG resulted in a drastic reduction in the number of viable cells. This was interesting, as this affect was seen with the WT Cpn60.2 as well, which has already been shown to be able to complement in Tab21 without any issues. In the presence of glucose (with or without IPTG), none of the 3 constructs showed any signs of growth. Since even the WT Cpn60.2 on the pET plasmid failed to complement, it could not be claimed that the mutations on the T/A60.2 and T/E60.2 constructs were responsible for the lack of functionality.

It was initially reasoned that the lack of functionality may have been caused by the presence of the His-tag. However, even after removing the tag from the sequence and ensuring that none of the protein sequence was affected, the levels of complementation were only marginally better. In the presence of arabinose and 0.1 mM IPTG the WT construct grew quite well, but the T/A and T/E triple mutants had the same complementation data as was seen prior to the removal of the His-tag. In the presence of glucose however, there was only a slight improvement in complementation. However, it was observed that the levels of protein expression had fallen. So it is possible that this lack of function could have been due to the lack of protein expression. However, since both the T/A60.2 and the T/E60.2 triple mutants struggled to grow on arabinose plates with IPTG, even after the His-tag was removed it suggests that these mutants are lethal to the cells when their expression levels are raised. It could be speculated that the mutants are in some way hampering the normal function of

endogenous GroEL. The exact mechanism by which this is being done is unclear, but it is possible that the over-expression of these mutants is resulting in the formation of non functional Cpn60.2 which traps the substrates required for cell survival and is unable to release them.

In summary, these results have not provided enough information to come to an appropriate conclusion. Further experiments would need to be done to first try and resolve the complementation issues and then the expression issues before trying to test the oligomeric properties of these mutants. It may be worth trying to re-do the experiments with single or double mutants, as well as attempting the experiments by cloning the mutants into the pTrc vectors such that the cochaperonin is co-expressed.

Summary

The main aim of this study was to better understand the properties of the multiple chaperonins of *Mycobacterium tuberculosis*. There are many reasons why Mycobacterial chaperonins are very interesting and unusual.

One of the reasons why these chaperonins are interesting is that they occur in multiples, while most other bacteria only have a single copy (referred to as GroEL or Cpn60). In *Mycobacterium tuberculosis* there are 2 distinct homologues which encode the chaperonins Cpn60.1 and Cpn60.2. It was initially thought that since *cpn60.1* gene was in the same operon as the cochaperonin *cpn10,* it would be the house-keeping chaperonin. However, it was found that this is not the case, and that Cpn60.1 is not required by Mycobacteria for its survival. Phylogenetic analysis of the Mycobacterial chaperonins has shown that the genes for the duplicated chaperonins diverged a long time ago, such that one of the *groEL* homologues in *M. tuberculosis* is more closely related to one in *M. smegmatis* than it is to the other one in *M. tuberculosis*. This suggested that the two chaperonins in *M. tuberculosis* have evolved to fulfil different functions within the cell.

Although studies have supported the idea that Cpn60.2 is the main housekeeping chaperonin in *M. tuberculosis* (Hu *et al.*, 2008), other studies suggested that it did not function as a regular chaperonin and could not function at all in *E. coli* (Kumar *et al.*, 2009). In this study I have shown that Cpn60.2 both from *M. tuberculosis* and from *M. smegmatis* can complement for loss of GroEL in *E. coli*. Cpn60.1, however, was unable to complement in *E. coli*, but biofilm complementation experiments have shown that Cpn60.1 from *M. tuberculosis* can

complement for loss of the homologue in *M. smegmatis*. These results support the view that both the chaperonins have evolved to fulfil different cellular functions. As the apical domain is responsible for substrate protein binding in GroEL, it was predicted that the difference in function between *M. tuberculosis* Cpn60.2 and Cpn60.1 could be due to a difference in substrate recognition by the apical domain. Unfortunately, attempts to test this by making apical domain swaps did not give unambiguous results.

These chaperonins are also unusual in that they have been shown not to form oligomers under standard conditions, to have very low ATPase activity, and in one case to crystallise as a dimer (Qamra *et al.*, 2004, Kumar et al., 2009, Qamra & Mande, 2004). Most GroEL proteins occur as stable, large complexes with a characteristic double ring structure of 14 subunits. Given that the large structure formed by GroEL and its ATPase activity is vital for its mechanism, it is unclear why the oligomers of Mycobacteria are less stable and less active. It has been proposed that Cpn60.2 does not function in a similar fashion to GroEL and is unable to form oligomers with a central cavity. However, I have shown that Cpn60.2 most probably does function as an oligomer in the cell as it is able to function in *E. coli* with GroES. Although attempts to stabilise the oligomeric nature of Cpn60.2 by making domain swaps were not successful, making apical domain swaps between Cpn60.2 and GroEL resulted in fully functional chimeric chaperonins. This suggests that both chaperonins probably function in a similar manner. The existence of higher oligomeric species of Cpn60.2 and its functional chimera has been studied using analytical ultracentrifugation in this and another project, and in the case of Cpn60.2, mass spectrometry under high salt conditions with ATP. These studies provided evidence of oligomerisation and increased ATPase activity from Cpn60.2 and a chimera derived from it under specific conditions (high salt plus nucleotide). Taken

together, these results strongly suggest that Cpn60.2 from *M. tuberculosis* functions by forming multimeric complexes.

The results that have been obtained in this project have extended and enhanced our understanding of the functional and oligomeric properties of the chaperonins of *M. tuberculosis* and support the hypothesis that Cpn60.2 of *Mycobacterium tuberculosis* functions in a similar manner to GroEL. Cpn60.1, on the other hand, has probably evolved to perform different cellular functions.

List of references

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

Andersson, G. E. & P. M. Sharp, **(1996)** Codon usage in the Mycobacterium tuberculosis complex. *Microbiology* **142 (Pt 4)**: 915-925.

Andra, S., G. Frey, M. Nitsch, W. Baumeister & K. O. Stetter, **(1996)** Purification and structural characterization of the thermosome from the hyperthermophilic archaeum Methanopyrus kandleri. *FEBS Lett* **379**: 127-131.

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

Apetri, A. C. & A. L. Horwich, **(2008)** Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proc Natl Acad Sci U S A* **105**: 17351-17355.

Archibald, J. M., J. M. Logsdon, Jr. & W. F. Doolittle, **(2000)** Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol Biol Evol* **17**: 1456-1466.

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

Barbirz, S., U. Jakob & M. O. Glocker, **(2000)** Mass spectrometry unravels disulfide bond formation as the mechanism that activates a molecular chaperone. *J Biol Chem* **275**: 18759-18766.

Basha, E., K. L. Friedrich & E. Vierling, **(2006)** The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. *J Biol Chem* **281**: 39943-39952.

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

Bergeron, J. J., M. B. Brenner, D. Y. Thomas & D. B. Williams, **(1994)** Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem Sci* **19**: 124-128.

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

Bigotti, M. G. & A. R. Clarke, **(2008)** Chaperonins: The hunt for the Group II mechanism. *Arch Biochem Biophys* **474**: 331-339.

Bjarnsholt, T., T. Tolker-Nielsen, N. Hoiby & M. Givskov, **(2010)** Interference of Pseudomonas aeruginosa signalling and biofilm formation for infection control. *Expert Rev Mol Med* **12**: e11.

Blaha, G., D. N. Wilson, G. Stoller, G. Fischer, R. Willumeit & K. H. Nierhaus, **(2003)** Localization of the trigger factor binding site on the ribosomal 50S subunit. *J Mol Biol* **326**: 887-897.

Bochkareva, E. S., N. M. Lissin, G. C. Flynn, J. E. Rothman & A. S. Girshovich, **(1992)** Positive cooperativity in the functioning of molecular chaperone GroEL. *J Biol Chem* **267**: 6796-6800.

Bose, S., T. Weikl, H. Bugl & J. Buchner, **(1996)** Chaperone function of Hsp90-associated proteins. *Science* **274**: 1715-1717.

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

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

Bryngelson, J. D., J. N. Onuchic, N. D. Socci & P. G. Wolynes, **(1995)** Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**: 167-195.

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

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

Burnett, B. P., A. L. Horwich & K. B. Low, **(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.

Burns, K. E., W. T. Liu, H. I. Boshoff, P. C. Dorrestein & C. E. Barry, 3rd, **(2009)** Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. *J Biol Chem* **284**: 3069-3075.

Burston, S. G., N. A. Ranson & A. R. Clarke, **(1995)** The origins and consequences of asymmetry in the chaperonin reaction cycle. *J Mol Biol* **249**: 138-152.

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

Caplan, A. J., D. M. Cyr & M. G. Douglas, **(1993)** Eukaryotic homologues of Escherichia coli dnaJ: a diverse protein family that functions with hsp70 stress proteins. *Mol Biol Cell* **4**: 555-563.

Carrascosa, J. L., O. Llorca & J. M. Valpuesta, **(2001)** Structural comparison of prokaryotic and eukaryotic chaperonins. *Micron* **32**: 43-50.

Cehovin, A., A. R. Coates, Y. Hu, Y. Riffo-Vasquez, P. Tormay, C. Botanch, F. Altare & B. Henderson, **(2010)** Comparison of the moonlighting actions of the two highly homologous chaperonin 60 proteins of Mycobacterium tuberculosis. *Infect Immun* **78**: 3196-3206.

Cernohorska, L. & M. Votava, **(2010)** Antibiotic resistance and biofilm formation in Staphylococcus saprophyticus strains isolated from urine. *Epidemiol Mikrobiol Imunol* **59**: 88-91.

Chakraborty, K., M. Chatila, J. Sinha, Q. Shi, B. C. Poschner, M. Sikor, G. Jiang, D. C. Lamb, F. U. Hartl & M. Hayer-Hartl, **(2010)** Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**: 112-122.

Chang, Z., T. P. Primm, J. Jakana, I. H. Lee, I. Serysheva, W. Chiu, H. F. Gilbert & F. A. Quiocho, **(1996)** Mycobacterium tuberculosis 16-kDa antigen (Hsp16.3) functions as an oligomeric structure in vitro to suppress thermal aggregation. *J Biol Chem* **271**: 7218-7223.

Chapman, E., G. W. Farr, R. Usaite, K. Furtak, W. A. Fenton, T. K. Chaudhuri, E. R. Hondorp, R. G. Matthews, S. G. Wolf, J. R. Yates, M. Pypaert & A. L. Horwich, **(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.

Chaudhry, C., G. W. Farr, M. J. Todd, H. S. Rye, A. T. Brunger, P. D. Adams, A. L. Horwich & P. B. Sigler, **(2003)** Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *Embo J* **22**: 4877-4887.

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

Chen, B., D. Zhong & A. Monteiro, **(2006)** Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC genomics* **7**: 156.

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

Chuang, S. E. & F. R. Blattner, **(1993)** Characterization of twenty-six new heat shock genes of Escherichia coli. *J Bacteriol* **175**: 5242-5252.

Cong, Y., M. L. Baker, J. Jakana, D. Woolford, E. J. Miller, S. Reissmann, R. N. Kumar, A. M. Redding-Johanson, T. S. Batth, A. Mukhopadhyay, S. J. Ludtke, J. Frydman & W. Chiu, **(2010)** 4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc Natl Acad Sci U S A* **107**: 4967-4972.

Cowing, D. W., J. C. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix & C. A. Gross, **(1985)** Consensus sequence for Escherichia coli heat shock gene promoters. *Proc Natl Acad Sci U S A* **82**: 2679-2683.

Coyle, J. E., F. L. Texter, A. E. Ashcroft, D. Masselos, C. V. Robinson & S. E. Radford, **(1999)** GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat Struct Biol* **6**: 683-690.

Cyr, D. M., **(1995)** Cooperation of the molecular chaperone Ydj1 with specific Hsp70 homologs to suppress protein aggregation. *FEBS Lett* **359**: 129-132.

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

de Jong, W. W., G. J. Caspers & J. A. Leunissen, **(1998)** Genealogy of the alpha-crystallin--small heat-shock protein superfamily. *Int J Biol Macromol* **22**: 151-162.

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

Diaz-Acosta, A., M. L. Sandoval, L. Delgado-Olivares & J. Membrillo-Hernandez, **(2006)** Effect of anaerobic and stationary phase growth conditions on the heat shock and oxidative stress responses in Escherichia coli K-12. *Archives of microbiology*.

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

Dougan, D. A., B. G. Reid, A. L. Horwich & B. Bukau, **(2002)** ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**: 673-683.

Effantin, G., M. R. Maurizi & A. C. Steven, **(2010)** Binding of the ClpA unfoldase opens the axial gate of ClpP peptidase. *J Biol Chem* **285**: 14834-14840.

Ehrnsperger, M., M. Gaestel & J. Buchner, **(2000)** Analysis of chaperone properties of small Hsp's. *Methods Mol Biol* **99**: 421-429.

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

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

Ellis, R. J., **(1995)** Revisiting the Anfinsen cage. *Fold Des* **1**: R9-R15.

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

Ellis, R. J., **(2001)** Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26**: 597-604.

Ellis, R. J., **(2005)** Chaperomics: in vivo GroEL function defined. *Curr Biol* **15**: R661-663.

Ellis, R. J., **(2007)** Protein misassembly: macromolecular crowding and molecular chaperones. *Advances in experimental medicine and biology* **594**: 1-13.

Ellis, R. J. & S. M. Hemmingsen, **(1989)** Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem Sci* **14**: 339-342.

England, J., D. Lucent & V. Pande, **(2008)** Rattling the cage: computational models of chaperonin-mediated protein folding. *Curr Opin Struct Biol* **18**: 163-169.

Erbse, A., O. Yifrach, S. Jones & P. A. Lund, **(1999)** Chaperone activity of a chimeric GroEL protein that can exist in a single or double ring form. *J Biol Chem* **274**: 20351-20357.

Eyles, S. J. & L. M. Gierasch, **(2010)** Nature's molecular sponges: small heat shock proteins grow into their chaperone roles. *Proc Natl Acad Sci U S A* **107**: 2727-2728.

Fayet, O., T. Ziegelhoffer & C. Georgopoulos, **(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. & A. L. Horwich, **(2003)** Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q Rev Biophys* **36**: 229-256.

Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling & N. Ban, **(2004)** Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**: 590-596.

Festa, R. A., F. McAllister, M. J. Pearce, J. Mintseris, K. E. Burns, S. P. Gygi & K. H. Darwin, **(2010)** Prokaryotic ubiquitin-like protein (Pup) proteome of Mycobacterium tuberculosis. *PLoS One* **5**: e8589.

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

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

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

Fux, C. A., J. W. Costerton, P. S. Stewart & P. Stoodley, **(2005)** Survival strategies of infectious biofilms. *Trends Microbiol* **13**: 34-40.

Gatenby, A. A., **(1984)** The properties of the large subunit of maize ribulose bisphosphate carboxylase/oxygenase synthesised in Escherichia coli. *European journal of biochemistry / FEBS* **144**: 361-366.

Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman & C. I. Kado, **(1985)** Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**: 918-921.

Geissler, S., K. Siegers & E. Schiebel, **(1998)** A novel protein complex promoting formation of functional alpha- and gamma-tubulin. *Embo J* **17**: 952-966.

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

Genevaux, P., F. Keppel, F. Schwager, P. S. Langendijk-Genevaux, F. U. Hartl & C. Georgopoulos, **(2004)** In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO reports* **5**: 195-200.

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

Georgopoulos, C. P., R. W. Hendrix, A. D. Kaiser & W. B. Wood, **(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. & B. Hohn, **(1978)** Identification of a host protein necessary for bacteriophage morphogenesis (the groE gene product). *Proc Natl Acad Sci U S A* **75**: 131-135.

Gething, M. J. & J. Sambrook, **(1992)** Protein folding in the cell. *Nature* **355**: 33-45.

Gibson, T. J., (1984) *Studies on the Epstein-Barr virus genome*. PhD thesis, University of Cambridge, UK.

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

Glover, J. R. & J. M. Tkach, **(2001)** Crowbars and ratchets: hsp100 chaperones as tools in reversing protein aggregation. *Biochem Cell Biol* **79**: 557-568.

Goetz, M. P., D. O. Toft, M. M. Ames & C. Erlichman, **(2003)** The Hsp90 chaperone complex as a novel target for cancer therapy. *Ann Oncol* **14**: 1169-1176.

Goldberg, A. L., **(1972)** Degradation of abnormal proteins in Escherichia coli (protein breakdown-protein structure-mistranslation-amino acid analogs-puromycin). *Proc Natl Acad Sci U S A* **69**: 422-426.

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

Gottesman, S., **(1996)** Proteases and their targets in Escherichia coli. *Annu Rev Genet* **30**: 465-506.

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

Graumann, J., H. Lilie, X. Tang, K. A. Tucker, J. H. Hoffmann, J. Vijayalakshmi, M. Saper, J. C. Bardwell & U. Jakob, **(2001)** Activation of the redox-regulated molecular chaperone Hsp33--a two-step mechanism. *Structure* **9**: 377-387.

Gregory A. Petsko , D. R., **(2005)** *Protein structure and function*, p. 180. New Science press.

Gupta, P., S. Mishra & T. K. Chaudhuri, **(2010)** Reduced stability and enhanced surface hydrophobicity drive the binding of apo-aconitase with GroEL during chaperone assisted refolding. *Int J Biochem Cell Biol* **42**: 683-692.

Gur, E. & R. T. Sauer, **(2008)** Recognition of misfolded proteins by Lon, a AAA(+) protease. *Genes Dev* **22**: 2267-2277.

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

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

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

Hall-Stoodley, L., J. W. Costerton & P. Stoodley, **(2004)** Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2**: 95-108.

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

Harris, S. F., A. K. Shiau & D. A. Agard, **(2004)** The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. *Structure* **12**: 1087-1097.

Harrison, C. J., M. Hayer-Hartl, M. Di Liberto, F. Hartl & J. Kuriyan, **(1997)** Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **276**: 431-435.

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

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

Haslbeck, M., S. Walke, T. Stromer, M. Ehrnsperger, H. E. White, S. Chen, H. R. Saibil & J. Buchner, **(1999)** Hsp26: a temperature-regulated chaperone. *Embo J* **18**: 6744-6751.

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

Hendershot, L., J. Wei, J. Gaut, J. Melnick, S. Aviel & Y. Argon, **(1996)** Inhibition of immunoglobulin folding and secretion by dominant negative BiP ATPase mutants. *Proc Natl Acad Sci U S A* **93**: 5269-5274.

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

Herendeen, S. L., R. A. VanBogelen & F. C. Neidhardt, **(1979)** Levels of major proteins of Escherichia coli during growth at different temperatures. *J Bacteriol* **139**: 185-194.

Herman, C., D. Thevenet, R. D'Ari & P. Bouloc, **(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.

Hesterkamp, T., S. Hauser, H. Lutcke & B. Bukau, **(1996)** Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proc Natl Acad Sci U S A* **93**: 4437-4441.

Hickey, T. B., L. M. Thorson, D. P. Speert, M. Daffe & R. W. Stokes, **(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.

Hoffmann, A., F. Merz, A. Rutkowska, B. Zachmann-Brand, E. Deuerling & B. Bukau, **(2006)** Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J Biol Chem* **281**: 6539-6545.

Hoffmann, J. H., K. Linke, P. C. Graf, H. Lilie & U. Jakob, **(2004)** Identification of a redox-regulated chaperone network. *Embo J* **23**: 160-168.

Horst, R., W. A. Fenton, S. W. Englander, K. Wuthrich & A. L. Horwich, **(2007)** Folding trajectories of human dihydrofolate reductase inside the GroEL GroES chaperonin cavity and free in solution. *Proc Natl Acad Sci U S A* **104**: 20788-20792.

Horwich, A. L., A. C. Apetri & W. A. Fenton, **(2009)** The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS Lett* **583**: 2654-2662.

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

Horwich, A. L. & W. A. Fenton, **(2009)** Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys* **42**: 83-116.

Horwich, A. L., W. A. Fenton, E. Chapman & G. W. Farr, **(2007)** Two Families of Chaperonin: Physiology and Mechanism. *Annu Rev Cell Dev Biol*: aheadofprint.

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

Horwich, A. L. & K. R. Willison, **(1993)** Protein folding in the cell: functions of two families of molecular chaperone, hsp 60 and TF55-TCP1. *Philos Trans R Soc Lond B Biol Sci* **339**: 313-325; discussion 325-316.

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

Hu, G., T. Gura, B. Sabsay, J. Sauk, S. N. Dixit & A. Veis, **(1995)** Endoplasmic reticulum protein Hsp47 binds specifically to the N-terminal globular domain of the amino-propeptide of the procollagen I alpha 1 (I)-chain. *J Cell Biochem* **59**: 350-367.

Hu, G., G. Lin, M. Wang, L. Dick, R. M. Xu, C. Nathan & H. Li, **(2006)** Structure of the Mycobacterium tuberculosis proteasome and mechanism of inhibition by a peptidyl boronate. *Mol Microbiol* **59**: 1417-1428.

Hu, Y., B. Henderson, P. A. Lund, P. Tormay, M. T. Ahmed, S. S. Gurcha, G. S. Besra & A. R. Coates, **(2008)** A Mycobacterium tuberculosis mutant lacking the groEL homologue cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. *Infect Immun* **76**: 1535-1546.

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

Hung, G. C. & D. C. Masison, **(2006)** N-terminal domain of yeast Hsp104 chaperone is dispensable for thermotolerance and prion propagation but necessary for curing prions by Hsp104 overexpression. *Genetics* **173**: 611-620.

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

Iizuka, R., T. Yoshida, N. Ishii, T. Zako, K. Takahashi, K. Maki, T. Inobe, K. Kuwajima & M. Yohda, **(2005)** Characterization of archaeal group II chaperonin-ADP-metal fluoride complexes: implications that group II chaperonins operate as a "two-stroke engine". *J Biol Chem* **280**: 40375-40383.

Ingolia, T. D. & E. A. Craig, **(1982)** Four small Drosophila heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proc Natl Acad Sci U S A* **79**: 2360-2364.

Inobe, T. & A. Matouschek, **(2008)** Protein targeting to ATP-dependent proteases. *Curr Opin Struct Biol* **18**: 43-51.

Ito, K., Y. Akiyama, T. Yura & K. Shiba, **(1986)** Diverse effects of the MalE-LacZ hybrid protein on Escherichia coli cell physiology. *J Bacteriol* **167**: 201-204.

Ivic, A., D. Olden, E. J. Wallington & P. A. Lund, **(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.

Jakob, U., H. Lilie, I. Meyer & J. Buchner, **(1995)** Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock in vivo. *J Biol Chem* **270**: 7288-7294.

Jakob, U., W. Muse, M. Eser & J. C. Bardwell, **(1999)** Chaperone activity with a redox switch. *Cell* **96**: 341-352.

Kang, H. J., D. H. Heo, S. W. Choi, K. N. Kim, J. Shim, C. W. Kim, H. C. Sung & C. W. Yun, **(2007)** Functional characterization of Hsp33 protein from Bacillus psychrosaccharolyticus; additional function of HSP33 on resistance to solvent stress. *Biochemical and biophysical research communications* **358**: 743-750.

Kauzmann, W., **(1959)** Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**: 1-63.

Kennaway, C. K., J. L. Benesch, U. Gohlke, L. Wang, C. V. Robinson, E. V. Orlova, H. R. Saibil & N. H. Keep, **(2005)** Dodecameric structure of the small heat shock protein Acr1 from Mycobacterium tuberculosis. *J Biol Chem* **280**: 33419-33425.

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

Keskin, O., I. Bahar, D. Flatow, D. G. Covell & R. L. Jernigan, **(2002)** Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* **41**: 491-501.

Kim, A. I., P. Ghosh, M. A. Aaron, L. A. Bibb, S. Jain & G. F. Hatfull, **(2003)** Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol Microbiol* **50**: 463-473.

Kim, S., K. R. Willison & A. L. Horwich, **(1994)** Cystosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem Sci* **19**: 543-548.

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

Knapp, S., I. Schmidt-Krey, H. Hebert, T. Bergman, H. Jornvall & R. Ladenstein, **(1994)** The molecular chaperonin TF55 from the Thermophilic archaeon Sulfolobus solfataricus. A biochemical and structural characterization. *J Mol Biol* **242**: 397-407.

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

Kramer, G., T. Rauch, W. Rist, S. Vorderwulbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling & B. Bukau, **(2002)** L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**: 171-174.

Kregel, K. C., **(2002)** Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. *J Appl Physiol* **92**: 2177-2186.

Kress, W., H. Mutschler & E. Weber-Ban, **(2007)** Assembly pathway of an AAA+ protein: tracking ClpA and ClpAP complex formation in real time. *Biochemistry* **46**: 6183-6193.

Krukenberg, K. A., F. Forster, L. M. Rice, A. Sali & D. A. Agard, **(2008)** Multiple conformations of E. coli Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure* **16**: 755-765.

Krzewska, J., G. Konopa & K. Liberek, **(2001)** Importance of two ATP-binding sites for oligomerization, ATPase activity and chaperone function of mitochondrial Hsp78 protein. *J Mol Biol* **314**: 901-910.

Kubota, H., G. Hynes, A. Carne, A. Ashworth & K. Willison, **(1994)** Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr Biol* **4**: 89-99.

Kubota, H., G. Hynes & K. Willison, **(1995)** The eighth Cct gene, Cctq, encoding the theta subunit of the cytosolic chaperonin containing TCP-1. *Gene* **154**: 231-236.

Kuehn, M. J., D. J. Ogg, J. Kihlberg, L. N. Slonim, K. Flemmer, T. Bergfors & S. J. Hultgren, **(1993)** Structural basis of pilus subunit recognition by the PapD chaperone. *Science* **262**: 1234-1241.

Kumar, C. M., G. Khare, C. V. Srikanth, A. K. Tyagi, A. A. Sardesai & S. C. Mande, **(2009)** Facilitated oligomerization of mycobacterial GroEL: evidence for phosphorylation-mediated oligomerization. *J Bacteriol* **191**: 6525-6538.

Kusukawa, N. & T. Yura, **(1988)** Heat shock protein GroE of Escherichia coli: key protective roles against thermal stress. *Genes Dev* **2**: 874-882.

Lai, E., T. Teodoro & A. Volchuk, **(2007)** Endoplasmic reticulum stress: signaling the unfolded protein response. *Physiology (Bethesda)* **22**: 193-201.

Lakshmipathy, S. K., S. Tomic, C. M. Kaiser, H. C. Chang, P. Genevaux, C. Georgopoulos, J. M. Barral, A. E. Johnson, F. U. Hartl & S. A. Etchells, **(2007)** Identification of nascent chain interaction sites on trigger factor. *J Biol Chem* **282**: 12186-12193.

Landick, R., V. Vaughn, E. T. Lau, R. A. VanBogelen, J. W. Erickson & F. C. Neidhardt, **(1984)** Nucleotide sequence of the heat shock regulatory gene of E. coli suggests its protein product may be a transcription factor. *Cell* **38**: 175-182.

Langer, T., C. Lu, H. Echols, J. Flanagan, M. K. Hayer & F. U. Hartl, **(1992)** Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**: 683-689.

Levinthal, C., **(1968)** Are there pathways for protein folding? *Extrait du Journal de Chimie Physique* **65**: 44-45.

Lewis, V. A., G. M. Hynes, D. Zheng, H. Saibil & K. Willison, **(1992)** T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature* **358**: 249-252.

Lewthwaite, J., A. Skinner & B. Henderson, **(1998)** Are molecular chaperones microbial virulence factors? *Trends Microbiol* **6**: 426-428.

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

Li, Y., X. Gao & L. Chen, **(2009)** GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side. *J Biol Chem* **284**: 4324-4331.

Li, Y., Z. Zheng, A. Ramsey & L. Chen, **(2010)** Analysis of peptides and proteins in their binding to GroEL. *J Pept Sci*.

Lill, R., E. Crooke, B. Guthrie & W. Wickner, **(1988)** The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. *Cell* **54**: 1013-1018.

Lin, C. Y., Y. S. Huang, C. H. Li, Y. T. Hsieh, N. M. Tsai, P. J. He, W. T. Hsu, Y. C. Yeh, F. H. Chiang, M. S. Wu, C. C. Chang & K. W. Liao, **(2009)** Characterizing the polymeric status of Helicobacter pylori heat shock protein 60. *Biochemical and biophysical research communications* **388**: 283-289.

Lin, Z., D. Madan & H. S. Rye, **(2008)** GroEL stimulates protein folding through forced unfolding. *Nat Struct Mol Biol* **15**: 303-311.

Lindner, R. A., T. M. Treweek & J. A. Carver, **(2001)** The molecular chaperone alpha-crystallin is in kinetic competition with aggregation to stabilize a monomeric molten-globule form of alpha-lactalbumin. *Biochem J* **354**: 79-87.

Liou, A. K. & K. R. Willison, **(1997)** Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *Embo J* **16**: 4311-4316.

Llorca, O., J. Martin-Benito, J. Grantham, M. Ritco-Vonsovici, K. R. Willison, J. L. Carrascosa & J. M. Valpuesta, **(2001)** The 'sequential allosteric ring' mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *Embo J* **20**: 4065-4075.

Llorca, O., E. A. McCormack, G. Hynes, J. Grantham, J. Cordell, J. L. Carrascosa, K. R. Willison, J. J. Fernandez & J. M. Valpuesta, **(1999)** Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature* **402**: 693-696.

Lorimer, G. H., **(1996)** A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J* **10**: 5-9.

Lund, P. A., **(2009)** Multiple chaperonins in bacteria--why so many? *FEMS Microbiol Rev* **33**: 785-800.

Lupas, A. N. & J. Martin, **(2002)** AAA proteins. *Curr Opin Struct Biol* **12**: 746-753.

Maguire, M., A. R. Coates & B. Henderson, **(2002)** Chaperonin 60 unfolds its secrets of cellular communication. *Cell Stress Chaperones* **7**: 317-329.

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

Marchenkov, V. V. & G. V. Semisotnov, **(2009)** GroEL-assisted protein folding: Does it occur within the chaperonin inner cavity? *Int J Mol Sci* **10**: 2066-2083.

Martin-Benito, J., J. Boskovic, P. Gomez-Puertas, J. L. Carrascosa, C. T. Simons, S. A. Lewis, F. Bartolini, N. J. Cowan & J. M. Valpuesta, **(2002)** Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *Embo J* **21**: 6377-6386.

Martin-Benito, J., J. Grantham, J. Boskovic, K. I. Brackley, J. L. Carrascosa, K. R. Willison & J. M. Valpuesta, **(2007)** The inter-ring arrangement of the cytosolic chaperonin CCT. *EMBO reports* **8**: 252-257.

Martin, A., T. A. Baker & R. T. Sauer, **(2007)** Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. *Mol Cell* **27**: 41-52.

Martinez-Hackert, E. & W. A. Hendrickson, **(2009)** Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* **138**: 923-934.

Maurizi, M. R. & D. Xia, **(2004)** Protein binding and disruption by Clp/Hsp100 chaperones. *Structure* **12**: 175-183.

Mayer, M. P., H. Schroder, S. Rudiger, K. Paal, T. Laufen & B. Bukau, **(2000)** Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat Struct Biol* **7**: 586-593.

McLennan, N. & M. Masters, **(1998)** GroE is vital for cell-wall synthesis. *Nature* **392**: 139.

McLennan, N. F., S. McAteer & M. Masters, **(1994)** The tail of a chaperonin: the C-terminal region of Escherichia coli GroEL protein. *Mol Microbiol* **14**: 309-321.

Minami, Y., J. Hohfeld, K. Ohtsuka & F. U. Hartl, **(1996)** Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *J Biol Chem* **271**: 19617-19624.

Mogk, A. & B. Bukau, **(2004)** Molecular chaperones: structure of a protein disaggregase. *Curr Biol* **14**: R78-80.

Mogk, A., C. Schlieker, C. Strub, W. Rist, J. Weibezahn & B. Bukau, **(2003)** Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. *J Biol Chem* **278**: 17615-17624.

Morimoto, R. I., M. P. Kline, D. N. Bimston & J. J. Cotto, **(1997)** The heat-shock response: regulation and function of heat-shock proteins and molecular chaperones. *Essays Biochem* **32**: 17-29.

Morita, M., M. Kanemori, H. Yanagi & T. Yura, **(1999)** Heat-induced synthesis of sigma32 in Escherichia coli: structural and functional dissection of rpoH mRNA secondary structure. *J Bacteriol* **181**: 401-410.

Motohashi, K., Y. Watanabe, M. Yohda & M. Yoshida, **(1999)** Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones. *Proc Natl Acad Sci U S A* **96**: 7184-7189.

Nakai, A., M. Satoh, K. Hirayoshi & K. Nagata, **(1992)** Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J Cell Biol* **117**: 903-914.

Nakamoto, H. & L. Vigh, **(2007)** The small heat shock proteins and their clients. *Cell Mol Life Sci* **64**: 294-306.

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

Nathan, D. F., M. H. Vos & S. Lindquist, **(1997)** In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. *Proc Natl Acad Sci U S A* **94**: 12949-12956.

Nemoto, T., Y. Ohara-Nemoto, M. Ota, T. Takagi & K. Yokoyama, **(1995)** Mechanism of dimer formation of the 90-kDa heat-shock protein. *European journal of biochemistry / FEBS* **233**: 1-8.

Nielsen, K. L. & N. J. Cowan, **(1998)** A single ring is sufficient for productive chaperonin-mediated folding in vivo. *Mol Cell* **2**: 93-99.

Nissen, P., J. Hansen, N. Ban, P. B. Moore & T. A. Steitz, **(2000)** The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920-930.

Obermoeller, L. M., I. Warshawsky, M. R. Wardell & G. Bu, **(1997)** Differential functions of triplicated repeats suggest two independent roles for the receptor-associated protein as a molecular chaperone. *J Biol Chem* **272**: 10761-10768.

Ohki, M., F. Tamura, S. Nishimura & H. Uchida, **(1986)** Nucleotide sequence of the Escherichia coli dnaJ gene and purification of the gene product. *J Biol Chem* **261**: 1778-1781.

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

Onuchic, J. N., H. Nymeyer, A. E. Garcia, J. Chahine & N. D. Socci, **(2000)** The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv Protein Chem* **53**: 87-152.

Pan, A., C. Dutta & J. Das, **(1998)** Codon usage in highly expressed genes of Haemophillus influenzae and Mycobacterium tuberculosis: translational selection versus mutational bias. *Gene* **215**: 405-413.

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

Patzelt, H., G. Kramer, T. Rauch, H. J. Schonfeld, B. Bukau & E. Deuerling, **(2002)** Three-state equilibrium of Escherichia coli trigger factor. *Biol Chem* **383**: 1611-1619.

Pearce, M. J., J. Mintseris, J. Ferreyra, S. P. Gygi & K. H. Darwin, **(2008)** Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science* **322**: 1104-1107.

Pearl, L. H. & C. Prodromou, **(2006)** Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem* **75**: 271-294.

Pereira, J. H., C. Y. Ralston, N. R. Douglas, D. Meyer, K. M. Knee, D. R. Goulet, J. A. King, J. Frydman & P. D. Adams, **(2010)** Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. *J Biol Chem* **285**: 27958-27966.

Pickart, C. M., **(1997)** Targeting of substrates to the 26S proteasome. *FASEB J* **11**: 1055-1066.

Powers, M. V. & P. Workman, **(2006)** Targeting of multiple signalling pathways by heat shock protein 90 molecular chaperone inhibitors. *Endocr Relat Cancer* **13 Suppl 1**: S125-135.

Pratt, W. B. & D. O. Toft, **(2003)** Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp Biol Med (Maywood)* **228**: 111-133.

Prodromou, C., S. M. Roe, P. W. Piper & L. H. Pearl, **(1997)** A molecular clamp in the crystal structure of the N-terminal domain of the yeast Hsp90 chaperone. *Nat Struct Biol* **4**: 477-482.

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

Qamra, R., S. C. Mande, A. R. Coates & B. Henderson, **(2005)** The unusual chaperonins of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* **85**: 385-394.

Qamra, R., V. Srinivas & S. C. Mande, **(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.

RA, M., **(2004)** 'Molecular Chaperones'. *Encyclopaedia of Molecular Cell Biology and Molecular Medicine (second edition)*.

Radford, S. E., **(2006)** GroEL: More than Just a folding cage. *Cell* **125**: 831-833.

Ranson, N. A., S. G. Burston & A. R. Clarke, **(1997)** Binding, encapsulation and ejection: substrate dynamics during a chaperonin-assisted folding reaction. *J Mol Biol* **266**: 656-664.

Ranson, N. A., H. E. White & H. R. Saibil, **(1998)** Chaperonins. *Biochem J* **333 (Pt 2)**: 233-242.

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

Reyes, D. Y. & H. Yoshikawa, **(2002)** DnaK chaperone machine and trigger factor are only partially required for normal growth of Bacillus subtilis. *Bioscience, biotechnology, and biochemistry* **66**: 1583-1586.

Rivenzon-Segal, D., S. G. Wolf, L. Shimon, K. R. Willison & A. Horovitz, **(2005)** Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. *Nat Struct Mol Biol* **12**: 233-237.

Roberts, M. M., A. R. Coker, G. Fossati, P. Mascagni, A. R. Coates & S. P. Wood, **(2003)** Mycobacterium tuberculosis chaperonin 10 heptamers self-associate through their biologically active loops. *J Bacteriol* **185**: 4172-4185.

Rommelaere, H., M. De Neve, R. Melki, J. Vandekerckhove & C. Ampe, **(1999)** The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* **38**: 3246-3257.

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

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

Saibil, H. R., **(2008)** Chaperone machines in action. *Curr Opin Struct Biol* **18**: 35-42.

Satoh, M., K. Hirayoshi, S. Yokota, N. Hosokawa & K. Nagata, **(1996)** Intracellular interaction of collagen-specific stress protein HSP47 with newly synthesized procollagen. *J Cell Biol* **133**: 469-483.

Schirmer, E. C., J. R. Glover, M. A. Singer & S. Lindquist, **(1996)** HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci* **21**: 289-296.

Schoehn, G., E. Quaite-Randall, J. L. Jimenez, A. Joachimiak & H. R. Saibil, **(2000)** Three conformations of an archaeal chaperonin, TF55 from Sulfolobus shibatae. *J Mol Biol* **296**: 813-819.

Schubert, U., L. C. Anton, J. Gibbs, C. C. Norbury, J. W. Yewdell & J. R. Bennink, **(2000)** Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**: 770-774.

Sethna, K. B., N. F. Mistry, Y. Dholakia, N. H. Antia & M. Harboe, **(1998)** Longitudinal trends in serum levels of mycobacterial secretory (30 kD) and cytoplasmic (65 kD) antigens during chemotherapy of pulmonary tuberculosis patients. *Scand J Infect Dis* **30**: 363-369.

Sewell, B. T., R. B. Best, S. Chen, A. M. Roseman, G. W. Farr, A. L. Horwich & H. R. Saibil, **(2004)** A mutant chaperonin with rearranged inter-ring electrostatic contacts and temperature-sensitive dissociation. *Nat Struct Mol Biol* **11**: 1128-1133.

Sharp, P. M., E. Bailes, R. J. Grocock, J. F. Peden & R. E. Sockett, **(2005)** Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* **33**: 1141-1153.

Shearstone, J. R. & F. Baneyx, **(1999)** Biochemical characterization of the small heat shock protein IbpB from Escherichia coli. *J Biol Chem* **274**: 9937-9945.

Shiau, A. K., S. F. Harris, D. R. Southworth & D. A. Agard, **(2006)** Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**: 329-340.

Shtilerman, M., G. H. Lorimer & S. W. Englander, **(1999)** Chaperonin function: folding by forced unfolding. *Science* **284**: 822-825.

Siegert, R., M. R. Leroux, C. Scheufler, F. U. Hartl & I. Moarefi, **(2000)** Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**: 621-632.

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

Singh, S. K., R. Grimaud, J. R. Hoskins, S. Wickner & M. R. Maurizi, **(2000)** Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP. *Proc Natl Acad Sci U S A* **97**: 8898-8903.

Smith, D. F., L. Whitesell, S. C. Nair, S. Chen, V. Prapapanich & R. A. Rimerman, **(1995)** Progesterone receptor structure and function altered by geldanamycin, an hsp90-binding agent. *Mol Cell Biol* **15**: 6804-6812.

Spiess, C., E. J. Miller, A. J. McClellan & J. Frydman, **(2006)** Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. *Mol Cell* **24**: 25-37.

Strickland, E., B. H. Qu, L. Millen & P. J. Thomas, **(1997)** The molecular chaperone Hsc70 assists the in vitro folding of the N-terminal nucleotide-binding domain of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **272**: 25421-25424.

Striebel, F., F. Imkamp, M. Sutter, M. Steiner, A. Mamedov & E. Weber-Ban, **(2009a)** Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat Struct Mol Biol* **16**: 647-651.

Striebel, F., W. Kress & E. Weber-Ban, **(2009b)** Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Curr Opin Struct Biol* **19**: 209-217.

Stroh, C. & K. Schulze-Osthoff, **(1998)** Death by a thousand cuts: an ever increasing list of caspase substrates. *Cell Death Differ* **5**: 997-1000.

Stromer, T., M. Ehrnsperger, M. Gaestel & J. Buchner, **(2003)** Analysis of the interaction of small heat shock proteins with unfolding proteins. *J Biol Chem* **278**: 18015-18021.

Summers, D. W., P. M. Douglas, C. H. Ramos & D. M. Cyr, **(2009)** Polypeptide transfer from Hsp40 to Hsp70 molecular chaperones. *Trends Biochem Sci* **34**: 230-233.

Sun, Z., D. J. Scott & P. A. Lund, **(2003)** Isolation and characterisation of mutants of GroEL that are fully functional as single rings. *J Mol Biol* **332**: 715-728.

Szabo, A., R. Korszun, F. U. Hartl & J. Flanagan, **(1996)** A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *Embo J* **15**: 408-417.

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

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

Tehver, R. & D. Thirumalai, **(2008)** Kinetic model for the coupling between allosteric transitions in GroEL and substrate protein folding and aggregation. *J Mol Biol* **377**: 1279-1295.

Teter, S. A., W. A. Houry, D. Ang, T. Tradler, D. Rockabrand, G. Fischer, P. Blum, C. Georgopoulos & F. U. Hartl, **(1999)** Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* **97**: 755-765.

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

Thomas, J. G. & F. Baneyx, **(2000)** ClpB and HtpG facilitate de novo protein folding in stressed Escherichia coli cells. *Mol Microbiol* **36**: 1360-1370.

Thulasiraman, V., C. F. Yang & J. Frydman, **(1999)** In vivo newly translated polypeptides are sequestered in a protected folding environment. *Embo J* **18**: 85-95.

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

Tobe, T., N. Kusukawa & T. Yura, **(1987)** Suppression of rpoH (htpR) mutations of Escherichia coli: heat shock response in suhA revertants. *J Bacteriol* **169**: 4128-4134.

Todd, M. J., G. H. Lorimer & D. Thirumalai, **(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.

Trent, J. D., E. Nimmesgern, J. S. Wall, F. U. Hartl & A. L. Horwich, **(1991)** A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**: 490-493.

Tyagi, N. K., W. A. Fenton & A. L. Horwich, **(2009)** GroEL/GroES cycling: ATP binds to an open ring before substrate protein favoring protein binding and production of the native state. *Proc Natl Acad Sci U S A* **106**: 20264-20269.

van Eden, W., R. van der Zee, A. G. Paul, B. J. Prakken, U. Wendling, S. M. Anderton & M. H. Wauben, **(1998)** Do heat shock proteins control the balance of T-cell regulation in inflammatory diseases? *Immunol Today* **19**: 303-307.

VanBogelen, R. A., P. M. Kelley & F. C. Neidhardt, **(1987)** Differential induction of heat shock, SOS, and oxidation stress regulons and accumulation of nucleotides in Escherichia coli. *J Bacteriol* **169**: 26-32.

Walburger, A., A. Koul, G. Ferrari, L. Nguyen, C. Prescianotto-Baschong, K. Huygen, B. Klebl, C. Thompson, G. Bacher & J. Pieters, **(2004)** Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* **304**: 1800-1804.

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

Wang, K. & A. Spector, **(2001)** ATP causes small heat shock proteins to release denatured protein. *European journal of biochemistry / FEBS* **268**: 6335-6345.

Warrens, A. N., M. D. Jones & R. I. Lechler, **(1997)** Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186**: 29-35.

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

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

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

Weissman, J. S., Y. Kashi, W. A. Fenton & A. L. Horwich, **(1994)** GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78**: 693-702.

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

Whitesell, L., E. G. Mimnaugh, B. De Costa, C. E. Myers & L. M. Neckers, **(1994)** Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. *Proc Natl Acad Sci U S A* **91**: 8324-8328.

Wiedmann, B., H. Sakai, T. A. Davis & M. Wiedmann, **(1994)** A protein complex required for signal-sequence-specific sorting and translocation. *Nature* **370**: 434-440.

Wild, J., W. A. Walter, C. A. Gross & E. Altman, **(1993)** Accumulation of secretory protein precursors in Escherichia coli induces the heat shock response. *J Bacteriol* **175**: 3992-3997.

Wolf, S., I. Nagy, A. Lupas, G. Pfeifer, Z. Cejka, S. A. Muller, A. Engel, R. De Mot & W. Baumeister, **(1998)** Characterization of ARC, a divergent member of the AAA ATPase family from Rhodococcus erythropolis. *J Mol Biol* **277**: 13-25.

Wolynes, P. G., J. N. Onuchic & D. Thirumalai, **(1995)** Navigating the folding routes. *Science* **267**: 1619-1620.

Yam, A. Y., Y. Xia, H. T. Lin, A. Burlingame, M. Gerstein & J. Frydman, **(2008)** Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat Struct Mol Biol* **15**: 1255-1262.

Yamamori, T. & T. Yura, **(1982)** Genetic control of heat-shock protein synthesis and its bearing on growth and thermal resistance in Escherichia coli K-12. *Proc Natl Acad Sci U S A* **79**: 860-864.

Ying, B. W., H. Taguchi, M. Kondo & T. Ueda, **(2005)** Co-translational involvement of the chaperonin GroEL in the folding of newly translated polypeptides. *J Biol Chem* **280**: 12035-12040.

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

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

Zako, T., Y. Murase, R. Iizuka, T. Yoshida, T. Kanzaki, N. Ide, M. Maeda, T. Funatsu & M. Yohda, **(2006)** Localization of prefoldin interaction sites in the hyperthermophilic group II chaperonin and correlations between binding rate and protein transfer rate. *J Mol Biol* **364**: 110-120.

Zhang, J., M. L. Baker, G. F. Schroder, N. R. Douglas, S. Reissmann, J. Jakana, M. Dougherty, C. J. Fu, M. Levitt, S. J. Ludtke, J. Frydman & W. Chiu, **(2010)** Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463**: 379-383.

Zhao, R., M. Davey, Y. C. Hsu, P. Kaplanek, A. Tong, A. B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili & W. A. Houry, **(2005)** Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**: 715-727.

Zhou, Y. N., N. Kusukawa, J. W. Erickson, C. A. Gross & T. Yura, **(1988)** Isolation and characterization of Escherichia coli mutants that lack the heat shock sigma factor sigma 32. *J Bacteriol* **170**: 3640-3649.

Zylicz, M., T. Yamamoto, N. McKittrick, S. Sell & C. Georgopoulos, **(1985)** Purification and properties of the dnaJ replication protein of Escherichia coli. *J Biol Chem* **260**: 7591-7598.