Functional analysis of Group 2 chaperonins from archaeal species in *E. coli*

By

Riddhi Shah

A thesis submitted to the

University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

The University of Birmingham

September 2013

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

The chaperonin proteins form a ubiquitous family of molecular chaperones and are absolutely required for correct folding or assembly of a subset of proteins in the cell. They are divided into two groups based on phylogeny: Group 1 found in bacteria and eukaryotic-organelles and Group 2 found in eukaryotic-cytoplasma and archaea. The two groups share a significant degree of conservation but differ in structure and mechanism and are believed to have evolved to serve specific client proteins. Using archaeal chaperonin from *M. maripaludis* (MmCCT) as a representative, we report here for the first time that a Group 2 chaperonin can partially replace the function of a Group 1 chaperonin from *E. coli* (GroEL). We have also identified and characterized two functional variants of MmCCT that show better GroEL complementation and have utilized them for a preliminary mutational analysis of potential client binding residues of MmCCT. We further demonstrate an initiative using a tagging approach for identification of bacterial proteins that interact with MmCCT *in vivo.* We suggest that our findings provide a novel platform for genetic dissection of MmCCT using a comparatively simple host, *E. coli,* which in turn can help identify properties of this archaeal chaperonin and provide insights for structure-function co-relations of Group 2 chaperonins in general.

Dedicated to Sir Kenneth Murray

(30 December 1930 – 7 April 2013)

For his vital initiative, the Darwin Trust of Edinburgh that funded my PhD

Acknowledgements

I owe my first mention in this section to my mother Dr. Alka Shah. I thank her for instilling in me and nurturing throughout my curiosity for science, especially biology. In the same breadth I thank my father Mr. Bhupendra Shah for being a firm believer, a motivator and for being my strongest support for these years and more to come.

I thank my teachers Dr. Sudeshna Menon and Dr. Saby Vadakan for playing a major role in keeping biology live, active and interesting in me during my undergraduate degree. I also thank my MRes supervisor Dr. Owen Jones for being a zealous teacher, a great scientist and a very strong inspiration for my PhD.

This brings me to thank the most important person for my PhD, Dr. Peter Lund for first giving me the opportunity to work in his lab and then for being a guide, a teacher and a constant support for these years. I remember his words during our very first supervisor-student meeting "these years will pass before you know it"! Well, you were right. We have not been very expressive, (rather I have not been) but I have learnt a lot from you and thanking for that is not enough.

I would also like to acknowledge Dr. Andrew Large for his help during my first year and for the long conversations on pleasantly weird topics. A big thank you to all the members of T101 lab for their support, Matt for his fast-paced accent (after understanding him, I could understand everyone) and his pranks, Amanda for keeping the lab live and aloud, Tim for his perfect coffee breaks, Faye for sharing 'the PhD student-stress' for last one and half years, Denisse for solving my queries, Cathy for keeping us hale and hearty, and Yana for being an advisor, a motivator, and a very caring friend.

I am also thankful to Dr. Jörg Martin for the opportunity to work at MPI, Tübingen and SGM for supporting my visit to Germany. I thank Dr. Astrid Ursinus and Eva Scultz for helping me during my time at MPI.

I also thank Dr. Kiran Bhatt and Dr. Apporva Bhatt for their warmth and guidance during my Ph.D.

Also a special mention goes to all my friends for making these years memorable, Fa for being my chatter-confession box and a stress-evaporator, Yash for some funniest moments and for helping me understand the 'norms', Nawsheen from whom I have learnt a lot (for life and otherwise), Jaanika for her help with the statistics, and more for making us adventurous and taking us to the hikes, Fernanda and Sonia for the lunch-long-life-talks and Jainy for being that person.

I humbly acknowledge my family here Meera masi, Chandrakant masa and Jaimin for their immense support and love that kept me rooted, and gave me strength during difficult times. I also thank my brother Satvik for patiently answering my dim-witted physics or maths (or life-) questions with ideal references even in the middle of the night.

Last but not the least I thank Kushal for his love, patience and understanding, for being there during the toughest of the times and for letting me be and for being him. Thank you.

TABLE OF CONTENTS

List of figures

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of permissions for applicable figures

Figure	Reference	Publisher	Permission obtained from	Order Licence Id
Figure 1.1	Hartl and Hayer-hartl, 2009	Nature publishing group	Copy right clearance centre	3251400950327
Figure 1.2	Horwich *et al.,* 2007	Annual review of cell and developmental biology	Copy right clearance centre	3251391329952
Figure 1.4	Horwich *et al.,* 2007	Annual review of cell and developmental biology	Copy right clearance centre	3251391329952
Figure 1.5	Hunt *et al.,*1996	Nature publishing group	Copy right clearance centre	3251400950327
Figure 1.8	Pereira et al., 2010	American Society for Biochemistry and Molecular Biology	Letter from publisher	N/a
Figure 5.10	Pereira et al., 2010	American Society for Biochemistry and Molecular Biology	Letter from publisher	N/a

List of tables

List of abbreviations

ADP   Adenosine diphosphate

Amp   Ampicillin

ATP   Adenosine triphosphate

APS   Ammonium persulphate

CCT   Chaperonin-containing TCP-1

DNA   Deoxyribonucleic acid

dNTP   Deoxyribonucleotide triphosphate

EDTA   Ethylenediamine tetra-acetic acid

Hsp   Heat shock protein

IPTG   Isopropyl-ı-D-thiogalactoside

Kan   Kanamycin

kDa   Kilodalton

LB   Luria-Bertani

M   Molar

OD   Optical density

PAGE   Polyacrylamide gel electrophoresis

PBS   Phosphate buffered saline

PCR   Polymerase chain reaction

RNA   Ribonucleic acid

rpm   Revolutions per minute

SDS   Sodium dodecyl sulphate

TAE   Tris-acetate EDTA

TCP-1   T-complex polypeptide-1

TEMED   N.N.N', N'-tetramethylethylenediamine

Tris   2-amino-1-(hydroxymethyl)-1, 3-propanediol

Chapter 1
Literature review

1.1 Molecular chaperones

1.1.1 Protein folding

Almost every biological process depends upon proteins. Protein functionality is inevitably associated with its specific three-dimensional (3D) conformation. Knowledge of how protein folding occurs, hence, is central to understanding of the nature of living systems.

1.1.1.1 'Sequence determines structure'

The groundwork for protein folding was laid in early 1960s with classic studies carried out by Christian Anfinsen and colleagues. Bovine pancreatic ribonuclease, containing four disulfide linkages in its native structure, was reduced and denatured using urea and β-mercaptoethanol giving eight sulfahydral (-SH) moieties. When this denatured protein was allowed to oxidise using molecular oxygen under physiological milieu, the enzymatic activity was regained slowly. Once freed from the denaturing agent, almost all of the protein was recovered with no changes in activity as compared to the native protein. The study established for the first time that the information needed for folding of a protein into its native state is inherently present in its own sequence (Anfinsen *et al.*, 1961). Further experimentation was carried out to study what guides the selection of the unique set of four disulphide bonds out of 105 possible combinations. Different non-native forms of RNAse that were similar in physical properties to native forms were made using pH changes, varying denaturing conditions or using modified amino-acids. The re-oxidation of such molecules to their native state was studied under various conditions (Haber and Anfinsen, 1962) and it was conceived that the process is driven entirely by free energy of conformation (Epstein *et al.,* 1963). This series of experiments

gave rise to the thermodynamic hypothesis which states that "the native conformation of the protein under physiological conditions is that in which the Gibbs free energy of the whole system is the lowest; that is the structure is determined by the totality of inter-atomic interactions and hence by the amino acid sequence in a given environment" (Anfinsen, 1973).

1.1.1.2 Levinthal's paradox and the principle of self-assembly of proteins

Cyrus Levinthal pointed out that if a protein was to attain its correctly folded state by sequentially sampling all possible conformations, it would require an astronomical amount of time to reach its correct native conformation (Levinthal, 1968). For example a polypeptide chain consisting of 100 amino acid residues would have to go through at least 2^{100} or in an order of 10^{30} different conformations (assuming that there are only two possible conformations per residue) before attaining its native conformation. Assuming that a protein can explore new conformations at the same rate that bonds can reorient (10^{11} structures/second), it would take nearly 10^{19} seconds or 10^{11} years for the polypeptide chain to go through all the possible conformations. However, it takes less than a minute for a protein to fold *in vitro* to its native state. Levinthal speculated a kinetic control over protein folding considering the biological time scale and the available phase space. In answer, he put forward a model wherein protein folding is speeded by the rapid formation of partially stable intermediates that in turn determine the further folding of the peptide (Levinthal, 1968). The concept that a protein folds into a structure dictated by the fastest folding pathway and not necessarily by its state of lowest free energy slowly emerged (Levinthal, 1969).

A number of studies hence followed in search of pathway-intermediates and investigation of the underlying folding kinetics (Ikai and Tanford, 1971; Tsong *et al.*, 1971; Wetlaufer, 1973; Creighton, 1974; Kim and Baldwin, 1982). These studies described various models explaining the proceedings of protein folding. However the next landmark in protein folding was not until the late 1980's when two groups independently published papers describing amalgamation of thermodynamic and kinetic concepts (Dill, 1985; Bryngelson and Wolynes, 1987). According to their proposed model, the folding is directed within funnel-like energy landscape (Figure 1.1). The width of funnel indicates the entropy associated with a particularly folded form of the protein while the depth is related to the free energy difference between the unfolded and fully folded forms of protein. The number of possible intermediates gets fewer as the funnel gets narrower, finally attaining the desired stable native conformation (Wolynes *et al.*, 1995; Dill and Chan, 1997; Onuchic *et al.*, 2000).

The present view for protein folding has revealed the ruggedness of the then accepted smooth energy landscape (Onuchic and Wolynes, 2004; Jahn and Radford, 2005). It suggests that some intermediates along with being the on-pathway stepping stones towards the native state are also susceptible to formation of non-native yet more stable conformations. Such intermediates typically expose at least some hydrophobic residues and regions of unstructured polypeptide backbone, that are buried in the native state. By interacting amongst themselves, the intermediates can either form inactive non-soluble amorphous structures or highly ordered toxic amyloid aggregates. These structures are non-native but thermodynamically more stable and form the so-called 'kinetic trap' in the energy landscape (Figure 1.1) (Hartl and Hayer-Hartl, 2009). Hence

the folding pathways for different proteins vary and depend on the fate of the folding intermediates.

Nonetheless, Anfinsen's hypothesis remains correct stating that all the information for a protein chain to reach its native structure is encoded in its amino acid sequence and the folding process is driven thermodynamically even though more stable states do exist than the native form. The folding proceeds by formation of well-defined partially structured intermediates, some of which are transient while others are significantly populated before converging into the final correct conformation (Dobson and Karplus, 1999), the concept on the whole accepted as the 'principle of self-assembly'.

1.1.1.3   Folding and assembly of some proteins requires assistance

It was pointed out that the conclusions drawn from *in vitro* studies, perhaps, cannot be generalized to the *in vivo* proteins (Creighton, 1983). *In vivo* protein folding differs in many ways from the *in vitro* folding. Firstly, folding takes place within a complex environment crowded with high concentrations of wide variety of molecules. This increases the probability of unwanted intra-molecular interactions between unfolded/ partly folded peptides resulting into formation of aggregates. Aggregation occurs when folding intermediates become trapped in partially misfolded states that themselves associate to form stable and less soluble oligomeric complexes (Ben-Zvi and Goloubinoff, 2001; Ellis, 2007). Secondly, stable conformation requires at least complete polypeptide chain to be present. With the ~30 aminoacids restricted to the ribosomes during the polypeptide synthesis, aggregation-sensitive nascent chains form a large population in the actively dividing cell (Hartl, 1996; Ellis, 2007).

Moreover *in vivo*, some proteins have to be transported to distant regions from ribosomes bypassing membranes that, in some cases, cannot let large 3D protein structures pass through. The fact that proteins have to be temporarily maintained in a non-native state until translocated was reported long ago (Rothman, 1989), raising the question of what holds the proteins from spontaneous folding until they reach their destination. Furthermore, the cells *in vivo* are exposed to constant physiological stresses like temperature fluctuations, pH variations, osmotic changes etc. threatening the integrity of the already folded protein molecules. How and with what mechanism, the stress-exposed proteins are protected, renatured or degraded allowing the cell to regain its original functioning was an intriguing question unresolved by any *in vitro* studies (Hartl and Hayer-Hartl, 2002). All these facts accounted for anticipation that protein folding in the cell might be 'aided' by some additional means, the inkling that eventually led to the discovery of molecular chaperones.
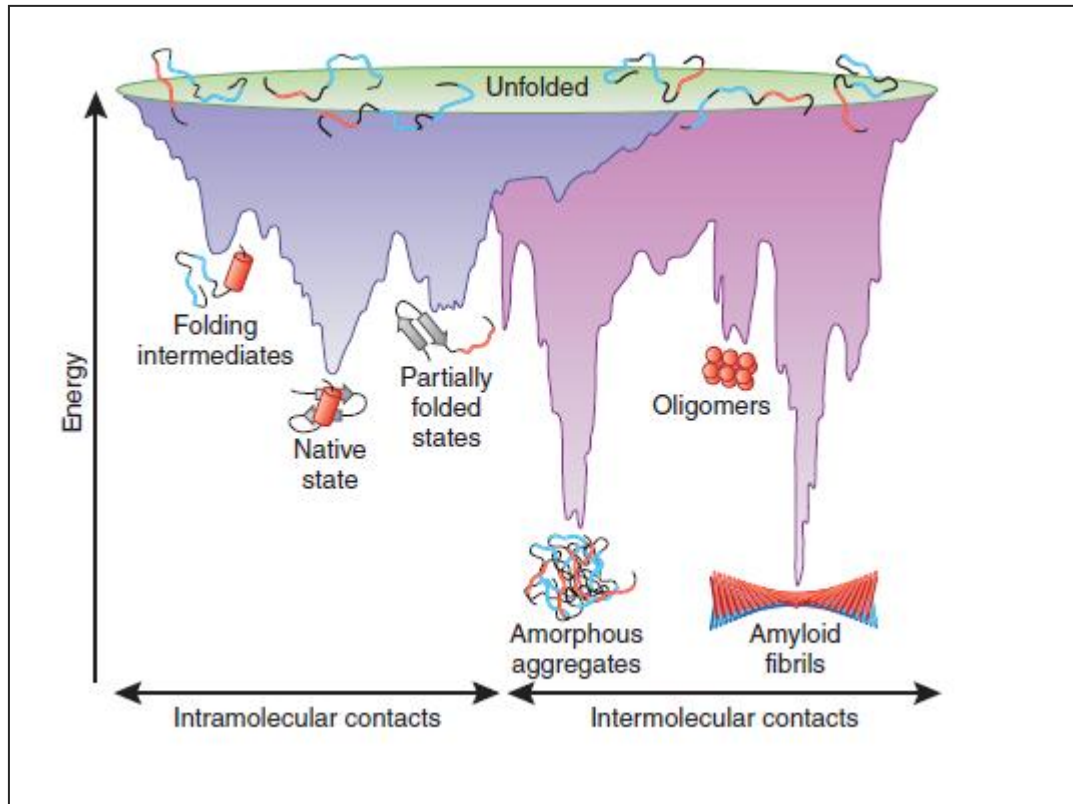
Figure 1.1
The energy landscape scheme for protein folding and aggregation.
(Hartl and Hayer-Hartl, 2009)

The figure represent hypothetical energy landscape scheme for protein folding. The purple area (left) shows the ensemble of conformations that are funnelled to the native state via intra-molecular contacts. The pink area (right) denotes the conformations that can culminate into formation of non-native aggregates formed by inter-molecular interactions of the intermediates, more stable than the native conformations. Both parts of the energy surface overlap.

1.1.2 Molecular chaperones

1.1.2.1 Emergence of the 'molecular chaperone' concept

The thought that the proteins are perfectly capable of folding on their own both *in vitro* and *in vivo* was believed until the starting years of 1980s. However, in 1980, it was reported that the enzyme RuBisCO (Ribulose-1,5-bisphosphate carboxylase oxygenase) was dependent on a chloroplast-binding-protein (now identified as a HSP60 chaperone) for correct assembly (Barraclough and Ellis, 1980). In another study, it was shown that a protein named BiP (now identified as a HSP70 chaperone) transiently binds to immunoglobulins in the endoplasmic reticulum (Haas and Wabl, 1983). Both the studies indicated a potential role of additional proteins that helped in the assembly of the client protein. At about the same time, a paper reported the mandatory role of an acidic nuclear protein, nucleoplasmin in assembly of histones and DNA to form nucleosomes in extracts of *Xenopus eggs*. The role of nucleoplasmin was described as the role of 'chaperone' (literal meaning: escort or safeguard) (Laskey *et al.*, 1978). Such studies prompted ideas about some sort of protein function in the cell that could possibly be associated with assembly of other proteins (Musgrove and Ellis, 1986).

The idea was further strengthened by a study that described the role of heat shock proteins (HSPs), the then identified families HSP70 and HSP90, in the cell. These proteins were required in increased amounts when cell proteins were damaged by stress, to help unscramble the aggregates of impaired proteins and to prevent further damage by binding to exposed hydrophobic surfaces (Pelham, 1986). Gathering the above discussed facts under one umbrella, John Ellis in 1987 proposed the notion of molecular chaperones as a generalized concept to describe "a class of proteins that

functions to ensure that the folding of other polypeptide chains and their assembly into oligomeric structures occurs correctly" (Ellis, 1987).

1.1.2.2   Molecular chaperones

The chaperone concept, although maintaining Anfinsen's principle of self assembly, clearly originated the picture of 'assisted folding' *in vivo.* Undoubtedly a protein folds spontaneously according to the information encoded in its sequence; however, chaperones ensure that proteins *can* do so by overcoming the *in vivo* threats posed to the non-native proteins such as molecular crowding, misfolding and aggregations during *de novo* protein synthesis, constant stress, and need of conformational flexibility for transport among others. The emergence of the concept elicited enormous research on discovering the chaperoning proteins, their functioning and chaperone-assisted folding mechanisms in general. At present over 50 types of molecular chaperones and chaperone cofactors are known; the functions of these broadly falling into one or more of the following categories (Bukau and Horwich, 1998; Leroux and Hartl, 2000; Ellis, 2006):

1. Transient stabilization of the polypeptide while being synthesized.

2. Assistance in *de novo* protein folding and assembly.

3. The unfolding and refolding of the proteins during the transport across membranes.

4. Prevention of protein aggregation and misfolding.

5. Unscrambling of the aggregates and renaturation of proteins under stress conditions.

6. Facilitation of the apoptotic activity of the cell by guiding the unstable proteins to degradation.

Most of the chaperones were discovered by their increased synthesis after the cell has undergone a stress like a heat shock and further by finding the homologues in the same species or other related/unrelated species. Such proteins were initially named as heat shock proteins (HSPs), however, it is worth noting that many chaperones are not HSPs and many HSPs are not chaperones. Subsequently, with the emergence of a horde of chaperone proteins, they were classified into various families.

1.1.3 Families of Molecular chaperones

Although chaperone proteins differ in size, shape, mechanisms and abundance, all of which can form the basis of classification, the most widely accepted classification of chaperones is based on sequence similarity. Presently, based on phylogenetic similarity, chaperones that are involved in protein folding are classified into ten major families as shown in Table 1.1. Discussing about every chaperone in detail is outside the scope of this assignment and also disrupts the narrative flow of the story of the *chaperonins* class which is the subject of this project, nonetheless key functions are outlined in the Table 1.1.

Table1.1: Representative classes of molecular chaperones

Note: 'm' stands for mitochondria, 'c' for cytosol, 'er' for endoplasmic reticulum

Chaperone Family	Organism	Chaperone (compartment)	Function	References
HSP100 Caseinolytic protease (Clp) family	Bacteria Eukaryotes	ClpA, ClpB, ClpC, ClpX, HslU Hsp78, mcxI (m) Hsp101, Hsp104 (c)	Prevention of aggregation, disassembly of aggregates, refolding of denatured proteins, maintenance and propagation of Prions (yeast)	Schirmer *et al.*, 1996; Santagata *et al.*, 1999; Maurizi and Xia, 2004; Hodson *et al.*, 2012
HSP90	Bacteria Eukaryotes	HtpG Grp94 (er) TRAP1 (m) Hsp90, Hsp82, Hsp83 (c)	In Bacteria: thermotolerance, cold acclimation, swarming, regulation of chemotaxis In eukaryotes: Essential proteins involved in activation and stabilization of proteins, organelle integrity, cellular homeostasis and assistance in protein degradation Potential drug targets for neurodegenerative diseases and cancer	Pearl and Prodromou, 2006; Richter and Buchner, 2006; Salminen *et al.*, 2011; Altieri, 2013

HSP70	Bacteria	DnaK	Versatile chaperonins involved in a variety of protein folding processes like prevention of aggregation of unfolded/newly translated/newly imported proteins, assembly and disassembly of oligomeric complexes, regulation of stability and activity of certain natively folded proteins, regulation of heat shock response among others	Hartl and Hayer-Hartl, 2002; Mayer and Bukau, 2005; Meimaridou et al., 2009
	Eukaryotes	BiP (er) SSc (m) Hsc (c)		
HSP60	Bacteria	GroEL	Essential proteins for specific protein folding of newly translated and newly imported proteins by encapsulation mechanism.	Subject of this study
	Archaea	Thermosomes, Cpn60,		
	Eukaryotes	TRiC (c) Cpn (m)		
HSP40	Bacteria	DNAJ	Essential co-chaperones for Hsp70: Stimulation and regulation of ATPase activity of Hsp70 proteins and transient stabilization of aggregation-prone proteins	Cheetham and Caplan 1998; Qiu et al., 2006
	Eukaryotes (Absent in many bacterial species)	HSP40, Ydj1 (c)		

HSP10	*E. coli* Eukaryotes	Hsp10 Cpn10 (m)	Essential for co-chaperoning function for HSP60 proteins	This study
Small HSPs	All organisms	sHSPs	Stabilisation of unfolded proteins, prevention of protein aggregation. Potential role in protection from Alzheimer's amyloid formation.	Jakob *et al.*, 1993; Van Montfort *et al.*, 2001; Haslbeck *et al.*, 2005
Trigger Factor	*E. coli*	TF	Binds and assists folding of nascent ribosome-bound polypeptides using hydrophobic interactions	Maier *et al.*, 2005; Hoffmann *et al.*, 2010; O'Brien *et al.*, 2012
Nascent polypeptide complexes	Archaea Eukaryotes	Nascent polypeptide associated complex (NAC) Ribosome associated complex (RAC)	Similar function as the trigger factor in *E. coli,* however absence of sequence homology between the two families of chaperones.	Rospert *et al.*, 2002; Wegrzyn and Deuerling, 2005

1.2 Chaperonins

Chaperonins are one of the ubiquitous, well-studied, sequence-related class of chaperones. The term chaperonin was coined by Hemmingsen *et al*, to represent a specific class of proteins that helped in assembly of other proteins (Hemmingsen *et al.*, 1988). They are present in all organisms studied to date (with the exception of certain strains of Mycoplasma and Ureaplasma) and are found to be essential in all cases tested (Fayet *et al.*, 1989; Stoldt *et al.*, 1996; Kapatai *et al.*, 2006). The chaperonins are phylogenetically divided into two main groups by alignment of chaperonin sequences from three domains of life. The sequence identity between Group 1 and Group 2 chaperonins is in the order of 20-25% and sequence similarity in the order of 40% (Horwich *et al.*, 2007)

A. GROUP 1 chaperonins

These chaperonins are found in all bacteria except certain strains of Mycoplasma and Ureaplasma (Lund, 2009). They are also found in organelles of endosymbiotic origins i.e. in chloroplast (Ellis, 1990), mitochondria (Reading *et al.*, 1989) and related organelles such as hydrogenosomes and mitosomes (Bui *et al.*, 1996).

B. GROUP 2 chaperonins

These chaperonins are found in the eukaryotic cytosol and in archaea, with archaeal proteins being phylogenetically distinct from the eukaryotic ones but significantly more similar to them than they are to the Group 1 chaperonins (Trent *et al.*, 1991; Frydman *et al.*, 1992).

1.2.1 Nomenclature

Group 1 - Bacterial chaperonins are generally referred to as chaperonin60 or Cpn60 (Coates *et al.*, 1993). However, they are analogously also often named as GroEL (which actually refers to *E. coli* chaperonin). The proteins that help chaperonins to carry out their function are referred to as co-chaperonins, GroES (in *E. coli*) or Cpn10. The mitochondrial and chloroplast chaperonins are usually called Hsp60 and Rubisco-binding protein respectively.

Group 2- Eukaryotic chaperonins were originally called TCP-1 (for tailless complex polypeptide1) (Silver *et al.*, 1980), which actually referred to only one of its subunits. Hence, now they are more generally referred to as TRiC (for TCP-1 containing ring complex) or CCT (chaperonin containing TCP-1). Archaeal chaperonins are mostly termed thermosomes as they were first studied in *Thermoplasma* species (Phipps et al, 1993); however, they are sometimes also referred to as CCT or archeosomes. They have also been described as rosettasomes for the typical arrangement reported in *S. shibatae* species (Trent *et al.*, 1991).

In this report the convention used refers eukaryotic chaperonins as **TRiC**, the archaeal chaperonins as **CCT**, and the *E. coli* chaperonins as **GroEL**. The substrate proteins that the chaperonin refolds are called **clients** because the natural substrate of chaperonins is ATP.

1.2.2 General chaperonin architecture

The domain architecture of the chaperonin complex is largely conserved across both the groups. They are large protein assemblies (800-1000 kDa) containing two rings stacked back to back giving rise to a hollow barrel like structure (Figure 1.2) (Hendrix, 1979). Each ring is made-up of seven to nine subunits each approximately weighing 60kDa and encloses a central cavity of expandable volume (Braig *et al.*, 1994; Xu *et al.*, 1997). The chaperonin complex may be homo-oligomeric or hetero-oligomeric. Unfolded client proteins are encapsulated into the chaperonin cavity with the help of either a co-chaperonin (in the case of Group 1 chaperonins) or by extensions of the chaperonin complex itself (in the case of Group 2 chaperonins). Once inside the cavity, the client protein is shielded from the risks faced by a non-native protein in the crowded environment of the cells, and thus can fold productively. The mechanisms of folding are discussed at length individually for both the groups later in this section.

Each chaperonin subunit consists of three domains (Braig *et al.*, 1994; Ditzel *et al.*, 1998) (Figure 1.2 and Figure 1.3). These domains undergo extensive conformational rearrangements to function as a highly allosteric folding machine for the client proteins.

Apical domain: This domain forms the part of the complex that harbours the client binding site and hence interacts with the client proteins and co-chaperonins (Figure 1.3, red coloured domain) (Sigler *et al.*, 1998). In Group 1 chaperonins the apical domains expose the hydrophobic residues at the rim of the central cavity and these residues bind to both clients and the co-chaperonin GroES (Fenton *et al.*, 1994; Buckle *et al.*, 1997; Xu *et al.*, 1997; Chen and Sigler, 1999). In Group 2 chaperonins the function of co-chaperonin is replaced by the flexible helical protrusion that expands from the tip of the

apical domain and serves as a cap to the cavity when a client protein is enclosed (Figure 1.3) (Klumpp *et al.*, 1997; Ditzel *et al.*, 1998; Iizuka *et al.*, 2004). The client binding sites for Group 2 chaperonins have not been explicitly defined however the regions analogous to GroEL client-site have been proposed to play a role in client interaction (Gomez-Puertas *et al.*, 2004; Pereira *et al.*, 2010).

The equatorial domain: This is the largest domain compose of amino acid residues from both N-terminal and C-terminal of the protein (Figure 1.3, green coloured domain) . The domain is mainly involved in binding and hydrolysing ATP thereby bringing about conformational changes in the complex. The nucleotide binding site (Sequence: GDGTTT) is located at the top of the domain and is conserved in almost all chaperonins with a few identified exceptions like chaperonin homologues of *Chlamydia trachomatis* (Cpn60.2 and Cpn60.3) that are thought to have diverged for other functions (Lund, 2009). The domain also forms a site for functional inter-ring and intra-ring interactions (Fenton *et al.*, 1994) .

The intermediate domain: This domain is the most flexible region of the chaperonin complex and lies between the apical and the equatorial domain (Figure 1.3, blue coloured domain). The two hinges of the intermediate domain relay the allosteric signal generated by the nucleotide binding to the equatorial domain to the apical domain. This signal is important in driving conformational changes in the chaperonin oligomer which in turn influence the binding and encapsulating of clients (Kawata *et al.*, 1999).
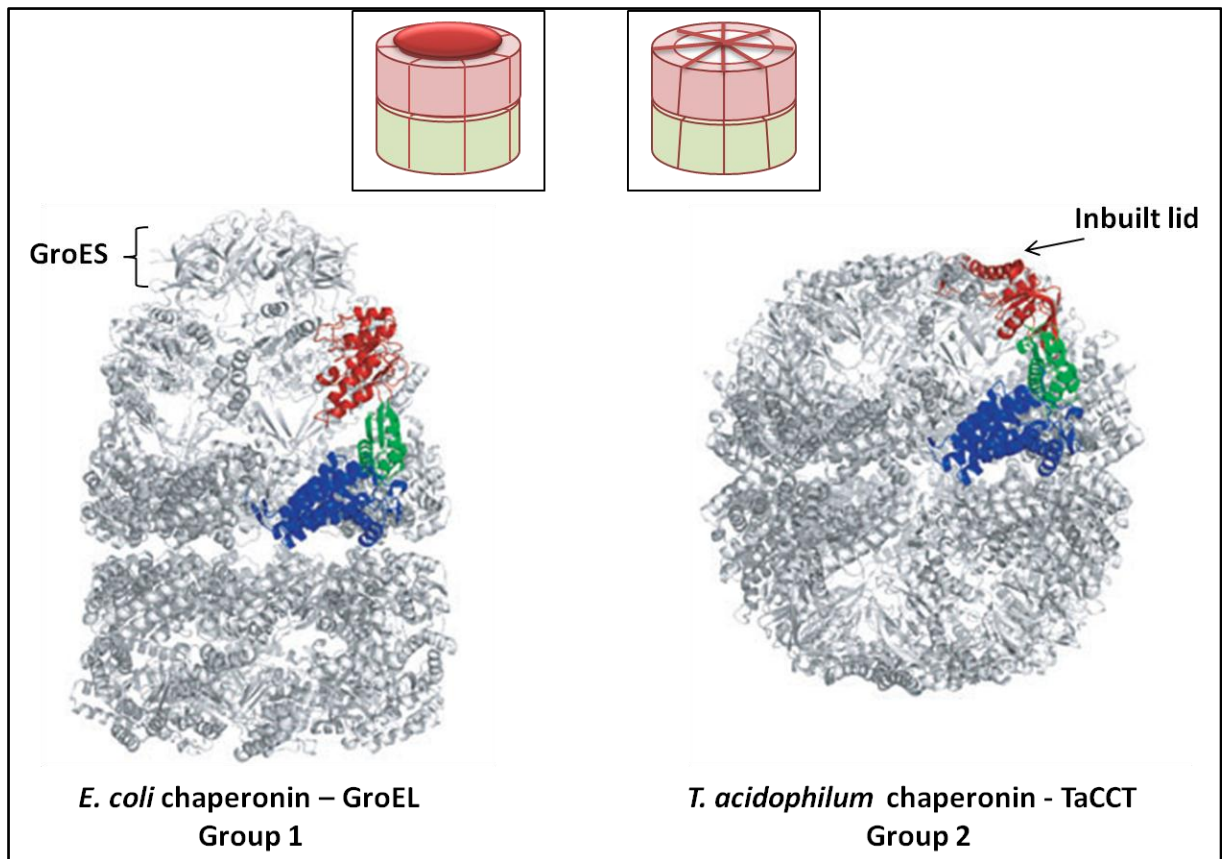
Figure 1.2
General architecture of Group 1 and Group 2 chaperonins.

Crystallographic ribbon diagrams of Group 1 chaperonin GroEL from *E. coli* (left) and Group 2 chaperonin TaCCT from archaeon *T. acidophilum* (right) are shown. A single subunit is coloured in each chaperonin. The insets represent a schematic of general chaperonin architecture - back to back stacked two rings (shaded red and green) capped by co-chaperonin GroES for Group 1 and by inbuilt helical protrusions for Group 2 chaperonins.

Note: Figure adapted from Horwich *et al.,* 2007.

E. coli chaperonin - GroEL   *M. maripaludis* chaperonin - MmCCT
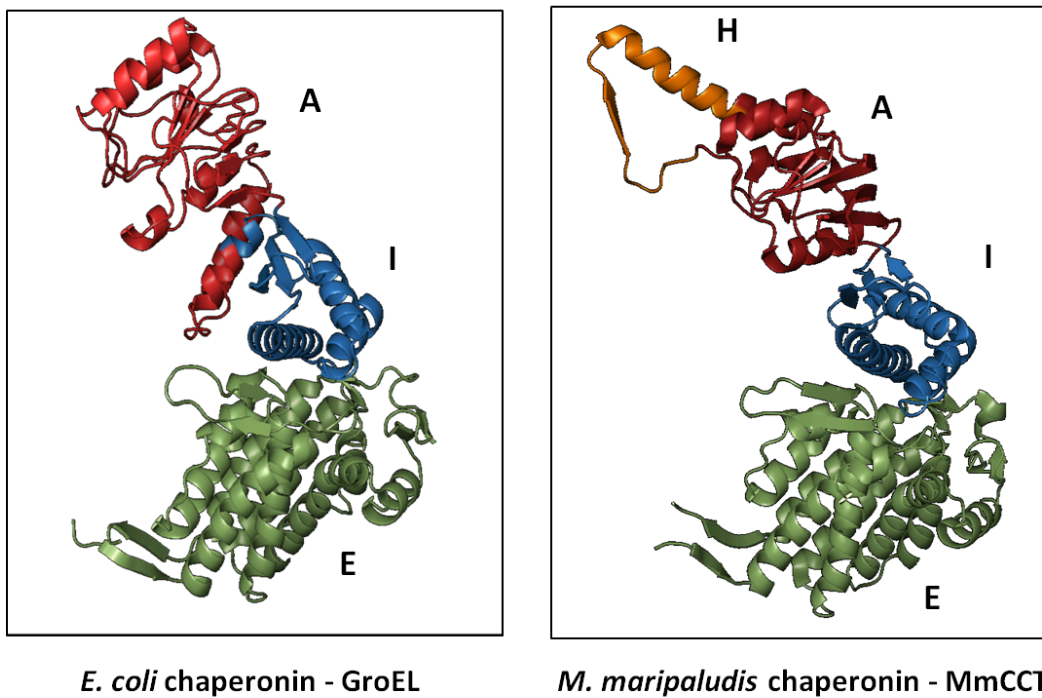
A: Apical substrate binding domain
I: Intermediate domain
E: Equatorial ATP binding domain

Figure 1.3

Domain architecture of single subunits of Group 1 and Group 2 chaperonins.

Ribbon diagrams of single subunits of Group 1 chaperonin GroEL from *E. coli* (left) and Group 2 chaperonin MmCCT from archaeon *M. maripaludis* (right) are shown. The apical domain is shown in red (A), the intermediate domain in blue (I), and the equatorial domain in green (E). The helical protrusion in case of MmCCT is coloured orange (H). Diagrams are drawn in PYMOL software using crystal structures of GroEL (PDB ID- 1AON) and MmCCT (PDB ID-3KFK) submitted to protein data bank.

1.2.3 Group 1 chaperonins

The chaperonin GroEL and its co-chaperonin GroES from *E. coli* have been extensively studied and used as common models to study Group 1 chaperonins. Hence, to understand the general architecture and mechanism of the Group 1 chaperonins, archetypal GroEL-ES system is described. Both GroEL and GroES are essential for cell viability under all conditions (Fayet *et al.*, 1989).

1.2.3.1   Structure of the GroEL-GroES complex

The first crystal structure of the GroEL was solved at 2.8 Å in an unliganded state (Braig *et al.*, 1994; Braig *et al.*, 1995). It established a near-atomic model of the double-ring architecture of GroEL previously observed by cryo-electron microscopy and negative staining (Langer *et al.*, 1992; Braig *et al.*, 1993; Chen *et al.*, 1994). The GroEL complex is a cylindrical protein-assembly comprising of two rings stacked back to back as previously described for all chaperonins (Figure 1.4). Principle dimensions of the double-ring cylinder are roughly 150 Å in height and 140 Å in width enclosing an approximately 50 Å wide central cavity. Each ring is composed of 7 homologous subunits (543 amino acids) arranged with nearly exactly 7-fold rotational symmetry and characteristically folded into three distinctive subunits (Braig *et al.*, 1994). The crystal structures of GroEL complexed with non-hydrolyzable ATP analogue recognized the details of highly conserved ATP-binding pocket in the equatorial domain and the resulting conformational shifts, thought to facilitate protein folding (Boisvert *et al.*, 1996).

The crystal structure of GroES was solved around the same time and the details of its architecture were presented (Hunt *et al.*, 1996). GroES is also a heptamer formed of approximately 10 KDa subunits and displaying a 7-fold rotational symmetry. Each subunit is folded into a single domain and has a core β-barrel structure with two β-hairpin loops on the top and bottom of the barrel (Figure 1.5). A long mobile loop extends out from the bottom β-hairpin (containing residues Glu 16 to Ala 32). The mobile loop segments are precisely visualized in GroEL-bound complex but are hidden in all but one subunit in a stand-alone GroES heptamer. These segments have been confirmed structurally and by mutagenesis to interact with GroEL and form the GroEL-GroES interface (Landry *et al.*, 1993; Xu *et al.*, 1997)
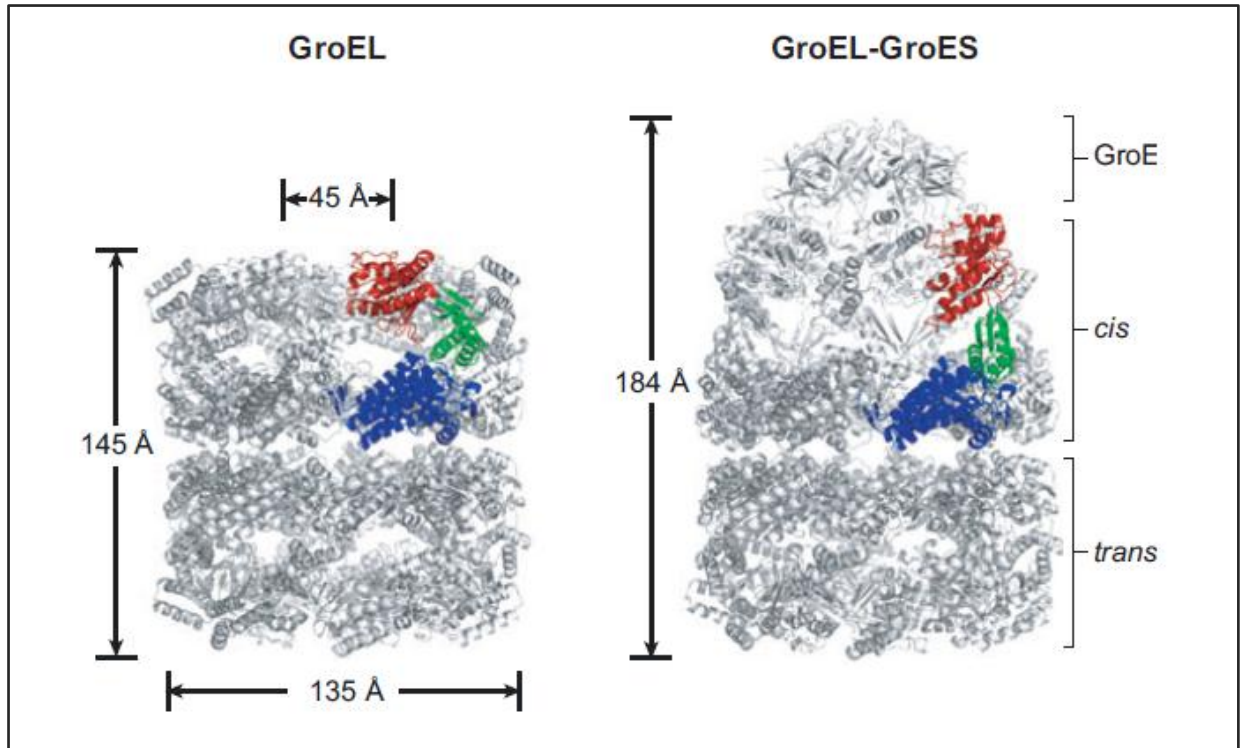
Figure 1.4

Architecture of Group 1 chaperonins: Crystallographic models of *E. coli* GroEL and GroES.

GroEL structures shown in open (left) and GroES-bound closed (right) conformations determined by Xray crystallography (Braig *et al.*, 1994). One subunit has been coloured: apical domain in red, intermediate domain in green and equatorial domain in blue.

Note: Figure has been taken from Horwich *et al.,* 2007.
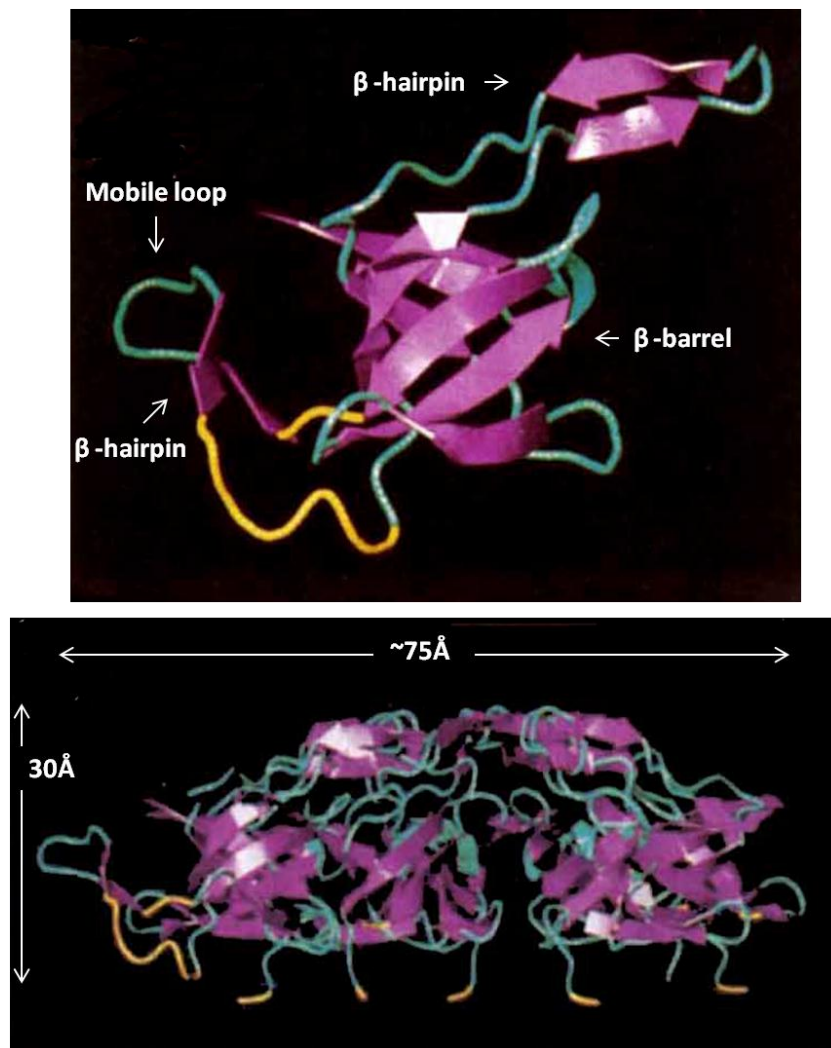
Figure 1.5
Structure of co-chaperonin GroES.

Ribbon diagrams of single subunit of GroES (above) and GroES-heptamer (below) are shown. The residues that link the β-hairpin of the mobile loop and the core barrel are coloured in yellow to describe the positioning of the mobile loop in all but one subunit in the GroES-heptamer.

Note: Figure has been adapted from Hunt *et al.,* 1996.

1.2.3.2 General mechanism of working for Group 1 chaperonins

The mechanism of GroEL-GroES machinery has been extensively studied by structural and mutational analysis (Martin *et al.*, 1993; Braig *et al.*, 1994; Fenton *et al.*, 1994; Boisvert *et al.*, 1996; Wang and Boisvert, 2003; Ranson *et al.*, 2006) and widely reviewed (Sigler *et al.*, 1998; Walter, 2002; Horwich *et al.*, 2007; Hartl and Hayer-Hartl, 2009; Saibil *et al.*, 2013). The basic reaction cycle is thus well established and is discussed stepwise below (Figure 1.6):

1. Polypeptide binding:

The non-native polypeptide binds to one of the GroEL rings in the open state (the ring now becomes the *cis* ring) at the hydrophobic amino acids rich patch exposed around the central rim of the cavity. Also, at physiologic concentrations of ATP (~1 mM) the seven subunits of *cis* ring bind to ATP cooperatively in a concerted manner. It was believed that the client binding of the complex is followed by ATP binding (Hartl and Hayer-Hartl, 2009), however, a recent study reports that the ATP binds before the client polypeptide and in turn enables more efficient capture of the client (Tyagi *et al.*, 2009).

2. Polypeptide encapsulation:

The ATP binding to the *cis* ring brings about a small degree of elevation and a counter clockwise twist. Such a conformational change triggers GroES association and client disassociation from the apical domain. The hydrophobic sites on GroEL apical domains, required for client binding, are also involved in GroES binding (Chen *et al.*, 1994; Fenton *et al.*, 1994). The binding of GroES produces upward and outward rigid whole body movement in the chaperonin expanding the volume of the central cavity, attaining the closed conformation of the complex. The movements bury the hydrophobic residues of

the binding site causing the bound peptide to be released. Because the GroES now caps the cavity, the bound peptide moves towards the inside. Some changes also take place in the internal surface of the cavity (e.g. making it more hydrophilic) and have been stipulated to facilitate the folding of the client peptide (Horwich and Fenton, 2009).

3. Polypeptide folding:

The polypeptide now folds within the confines of the *cis* ring till the bound ATP is hydrolyzed (usually 8-10 seconds). ATP hydrolysis is triggered by ATP-binding to the opposite *trans* ring that in turn lowers the affinity for ATP in *cis* ring. Hydrolysis of ATP weakens the interaction between GroEL and GroES and prepares the *cis* ring for releasing GroES.

4. Polypeptide release:

The GroES-release leads to a conformational change in the complex that brings the ring back into its open conformation. Subsequently the polypeptide is released either in fully folded condition, or partially folded states that may undergo further cycles of binding and release. The *trans* ring now becomes the new *cis* ring and the cycle repeats.
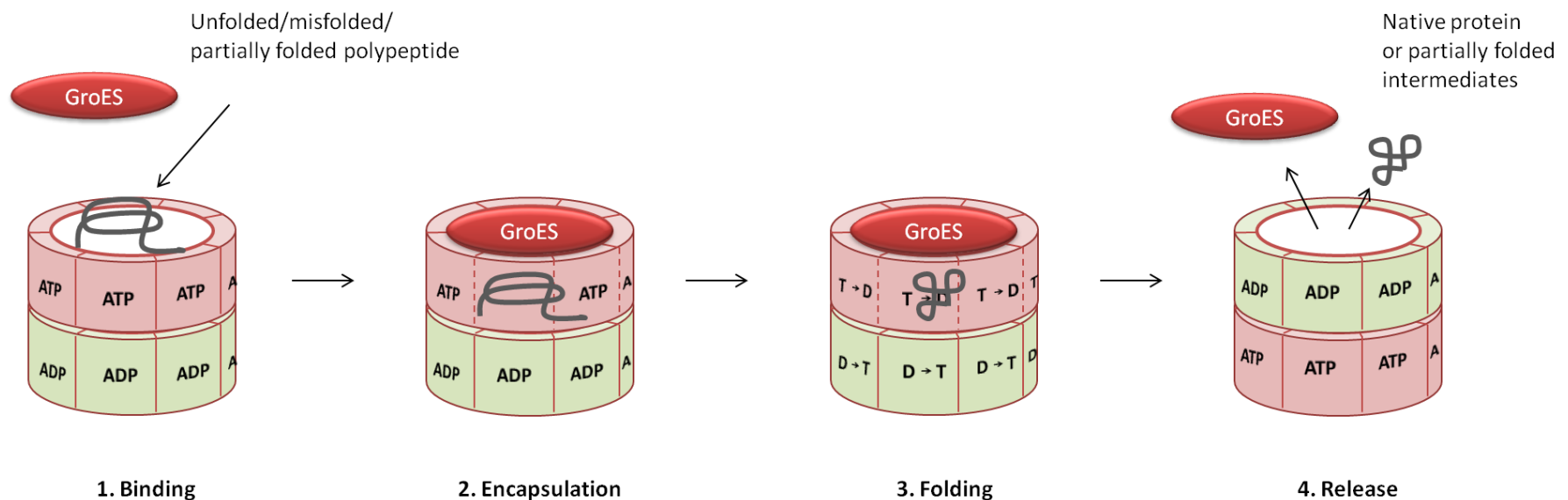
Figure 1.6: Illustration of a basic GroES-GroEL reaction cycle.

1) **Binding:** Polypeptide is normally accepted into the open ring of the asymmetric complex (the ring now is called *cis* ring). Whether ATP binding is synchronous, ahead or later to polypeptide binding is still controversial.

2) **Encapsulation:** GroES binding to the ATP-occupied open ring is associated with large rigid body movements and release of polypeptide into the cavity.

3) **Folding:** The folding of the polypeptide is triggered in the encapsulated chamber. This is longest step of the reaction cycle and lasts for approximately 10 s.

4) **Release:** ATP hydrolysis in the *cis* ring weakens the *cis* complex. This gates entry of ATP into the *trans* ring, which in turn triggers release of the *cis* ligands (GroEL and native or partially folded intermediates) in less than 1 s. This at the same time enables binding of GroES to the other ring forming a new folding-active *cis* complex and the cycle continues.
   Note: T denotes ATP and D denotes ADP

26

1.2.3.3   Different models of folding

Three different models have been put forward for the folding mechanism of GroEL (Horwich *et al.*, 2007; Horwich and Fenton, 2009; Jewett and Shea, 2010). They are described as follows:

1. Anfinsen cage model:

According to this model the client is encapsulated in the cavity. The cavity acts as a passive nanocage – an isolated microenvironment that provides a 'non-stick' surface to the clients allowing spontaneous folding without any general or client-specific involvement (Rye *et al.*, 1999; Brinker *et al.*, 2001; Horwich *et al.*, 2009). Hence the native state is reached by following the intrinsic folding pathway according the Anfinsen's principle of folding.

2. Iterative annealing model (Forced unfolding model):

A range of GroEL client proteins consists of aggregation-prone proteins that can form irreversible kinetically trapped states. Such client proteins are unable to refold spontaneously. A number of studies have suggested that client proteins that are kinetically trapped in a non-native state are unfolded by GroEL before undergoing a folding cycle. Such an effect gives the aggregation-prone proteins a fresh opportunity to fold, starting from a higher point in the energy landscape of the nascent chain (Todd *et al.*, 1996; Shtilerman *et al.*, 1999; Sharma *et al.*, 2008). It is believed that GroEL exerts a stretching force, presumably due to allosteric transitions resulting from ATP binding, that brings about forced unfolding of the client (Shtilerman *et al.*, 1999; Thirumalai and Lorimer, 2001). It has also been suggested that the hydrophobic region of the apical domain around the client-GroEL interaction site catalyzes the partial unfolding of the

clients (Lin *et al.*, 2008). Thus the transient ternary complex of GroEL-ATP-client can have two fates – 1) release of the partially unfolded client into the cytosol (bulk solution) for another cycles 2) release inside the cavity for a passive folding. The iterative annealing model stipulates that repeated cycles of binding and release free the proteins from the kinetic traps and hence further allows spontaneous folding (Todd *et al.*, 1996). The more the number of cycles, the faster is the folding. Such a model also explains GroEL-mediated folding of the higher molecular weight clients that cannot be encapsulated inside the cavity. The concept has been further extended to 'stationary iterative annealing' which means that even when the bound client protein proceeds towards passive folding inside the cavity, it is not released inside the cavity (Jewett and Shea, 2010). This is supported by a recent study that suggests that client remains bound to the apical domain even when it has proceeded for folding inside the cavity (Motojima and Yoshida, 2010). In sum, the model puts forward a mechanism for kinetically-trapped clients and clients that cannot encapsulate inside the cavity; however whether every GroEL-client undergoes forced unfolding before encapsulating and whether complete folding can occur without encapsulation still remains a speculation especially when the importance of encapsulation has been shown biophysically (Jewett and Shea, 2008).

3. Active cage model:

This model suggests that the cavity actively contributes to accelerate the folding process by at least two ways – 1) Nature of the residues facing towards the inside and 2) confinement of the clients in the folding chamber. The inside of the GroEL cavity is hydrophilic and consists of a higher proportion of negatively charged residues (Horwich *et al.*, 2007; Cong *et al.*, 2010). It was suggested that a hydrophilic inner could favour the

burial of hydrophobic surfaces in the encapsulated client. Although experimental evidence for such a proposition is lacking but it has been shown that changing six negatively charged residues of the cavity-inside to positively charged residues has considerable negative effects on folding of the client protein Rubisco (Tang *et al.*, 2006; Tang *et al.*, 2008). Next, it is proposed that the closed confinement of the client inside the cavity reduces the number of possible intermediates of the energy landscape and hence speeds up folding thermodynamically. For example closed confinement would not allow extended conformations and elimination of such will help the protein to reach its native state, faster (Brinker *et al.*, 2001). Experiments with extended C-terminal-tail variants of GroEL that hypothetically should make the cavity more compact indeed resulted in 50% acceleration of folding of rhodanese (33KDa) suggesting an active involvement of cavity and optimization for accelerated folding. The results have been disputed however, where the same effect could not be reproduced for malate dehydrogenase protein, also 33KDa (Farr *et al.*, 2007). Current view of folding continues to remain debated.

1.2.3.4 Clients interacting with GroEL

GroEL recognizes interacts with ~10% of the cellular proteins and binds them mostly post-translationally (Ewalt *et al.*, 1997). Using radio-labelling, 2D gel analysis and co-immunoprecipitation studies, clients that bind to GroEL were identified *in vivo* (Houry *et al.*, 1999). The sizes of majority of these proteins were <60 KDa. Proteins of this range of sizes can be encapsulated in the cavity. However, larger proteins e.g., MetE (82 kDa), RNA polymerase β (151 kDa), phosphate acetyltransferase (77 kDa), and DNA gyrase A subunit (97 kDa) were also identified. The same research group further determined the range of client proteins bound to GroEL using affinity purification of his-tagged GroES-

client complexes and mass spectrometry (Kerner *et al.*, 2005). They reported that ~250 of the total ~2400 cytosolic proteins interact with GroEL-GroES and classified them into three classes: class I (38 proteins) that interact but do not require GroEL for folding, class II (126 proteins) that prefer GroEL and GroES but can also fold with assistance of DnaK system and class III (84 proteins) that stringently require GroEL and GroES for folding. More recently a study that characterized GroEL-clients by decreasing the level of GroE to under 10% modified the classification. They found that not all class III proteins were dependent on GroEL for folding and introduced a fourth class that contained 49 of class III proteins plus eight others that absolutely required GroEL and GroES for folding (Fujiwara *et al.*, 2010).

It was observed that most of client proteins had complex topologies more prone to kinetic-trapping and requiring long-range interactions for stabilization barrel fold (Kerner *et al.*, 2005). Such an observation indicates the involvement of GroEL in binding the proteins post-translationally and prevention of aggregates rather than folding of newly formed peptides.

1.2.4 Group 2 chaperonins

Group 2 chaperonins are found in the eukaryotic cytosol and in archaea. The eukaryotic TRiC complex is composed of eight different yet structurally related subunits (Rommelaere *et al.*, 1993) while in the archaeal thermosomes, each ring is made up of eight (Andra *et al.*, 1996) or nine subunits (Knapp *et al.*, 1994) that are products of one to five genes (Large and Lund, 2009).

Despite the similar architectural outline with the Group 1 chaperonins, the sequence similarity between the two groups is largely limited to the equatorial domain. Significant differences are seen in the sequences of the apical domain and intermediate domains (Kim *et al.*, 1994; Dekker *et al.*, 2011b). An important difference between the two classes lie in the apparent absence of a GroES-like co-chaperonin to seal off the cavity formed by the stacked double rings. Instead, a hydrophobic stretch of 27 residues in the apical domain forms a 25 Å long flexible helix-turn-helix motif that extends as a helical protrusion and acts as a lid in capping the central cavity (Klumpp *et al.*, 1997; Reissmann *et al.*, 2007) (Refer to Figure 1.3). The lid region is dispensible for polypeptide binding and encapsulation, however is essential for folding of the client proteins (Iizuka *et al.*, 2004; Reissmann *et al.*, 2007)

1.2.4.1   Group 2 chaperonins: Eukaryotic view

The eukaryotic TRiC is made up of hetero-oligomeric rings containing eight different albeit homologous subunits (30% identity) (CCT α, β, γ, δ, ε, ζ, η, and θ) in each of its rings and hence attaining eight fold symmetry. All the eight subunits are encoded by eight distinct and independent genes (Rommelaere *et al.*, 1993). Moreover, these genes have been proved to individually essential, at least in yeast (Stoldt *et al.*, 1996).

Phylogenetic analysis suggests a common ancestor to all these genes and they are thought to have evolved through multiple gene duplications very early in the eukaryotic lineage (Archibald *et al.*, 2000; Fares and Wolfe, 2003). This suggests that the complexity in the structure is a parallel consequence with the complexity in structure attained from the transition from prokaryotes to eukaryotes.

The crystal structure of eukaryotic chaperonins was not available until very recently when Dekker *et al.* resolved by crystallography the closed form of yeast CCT at 3.8 Å (Dekker *et al.*, 2011a). However, previously various electron microscopy and biochemical studies provided structural insights for TRiC complex. The specific arrangement of the individual subunits within the TRiC complex has remained controversial. Based on biochemical data and cryo-EM constructions, it was shown that TRiC complex exhibited an asymmetric organization with a specific sequence of its subunits in the two rings (Liou and Willison, 1997; Martin-Benito *et al.*, 2007). This was argued by a study which, on the basis of high-resolution cryo-EM maps of bovine TRiC, proposed a different arrangement with a 2-fold symmetry between the subunits (Figure 1.7A) (Cong *et al.*, 2010). The group with the former view however solved the crystal structure of yeast CCT at 3.8 Å that displays the same sequence of subunits as they initially proposed albeit now with a 2-fold symmetry (Figure 1.7B) (Dekker *et al.*, 2011a).
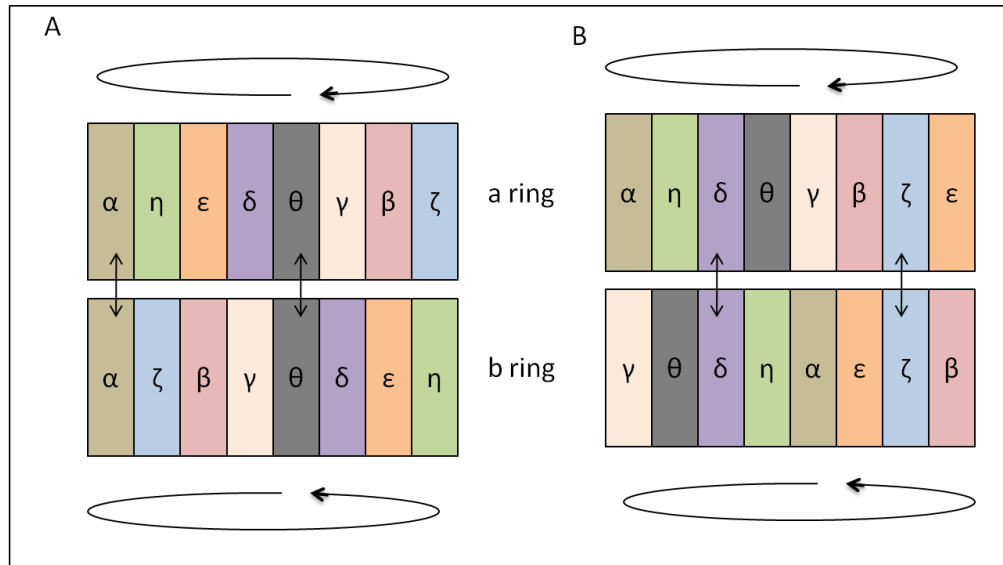
Figure 1.7

Specific arrangement of subunits within each ring of the eukaryotic chaperonin complex-TRiC

The arrangement of subunits within each ring remains controversial with two different models: (A) proposed by (Cong *et al.*, 2010) using cryo-EM structure of bovine TRiC and (B) proposed by (Dekker *et al.*, 2011a) using crystal structure of yeast CCT. Circular arrows show the orientation of the rings and double arrows highlight the difference between the homotypic interations between the two proposed models.

As compared to GroEL and archaeal chaperonins, eukaryotic chaperonins are relatively less abundant. Also, they are generally not heat induced as GroEL or archaeal chaperonins; basal levels are always found to be present in the cell (Horwich *et al.*, 2007). This observation correlates with the finding that TRiC is required to fold ~10% of newly synthesized cellular proteins (Thulasiraman *et al.*, 1999). Initial data of co-immunoprecipitation studies with chaperone specific antibodies identified actin and tubulin as the major clients for TRiC complex (Gao *et al.*, 1992; Yaffe *et al.*, 1992). This was supported by the mutational studies in *cct* genes that lead to cytoskeletal and other defects in a variety of organisms (Ursic *et al.*, 1994; Vinh and Drubin, 1994). Although, it was first thought that only the cytoskeletal proteins are major clients for TRiC complex, a number of putative TRiC clients involved in a range of cellular activities were gradually identified. This list includes cyclin E (Won *et al.*, 1998), SMRT-histone deacetylation complex (Guenther *et al.*, 2002), Von Hippel-Lindau (VHL) tumor suppressor (Feldman *et al.*, 1999), anaphase-promoting complex (Camasses *et al.*, 2003), G-beta folding intermediates (Kubota *et al.*, 2006) among others.

Proteome-wide analysis to identify *bona fide* interactors of yeast CCT and mammalian TRiC complex further identified a range of potential clients (Dekker *et al.*, 2008; Yam *et al.*, 2008). Yam *et al.,* showed that 90% of the TRiC interactome consisted of proteins belonging to homo or hetero oligomeric complexes and 40% of them are essential proteins. Their sizes ranged from 40-75 kDa which is in agreement with the predicted normal range for the chaperonin cavity (Ditzel *et al.*, 1998). However, certain larger proteins with even >100KDa of size were also found, prompting the idea that some proteins are not fully encapsulated in the complex during their folding. Indeed this has been the case for certain GroEL client proteins and mechanisms of folding-without-

encapsulation or folding in the bulk solution have been proposed (Farr *et al.*, 2003; Motojima and Yoshida, 2010). An enrichment of hydrophobic sequences was also noted amongst the potential TRiC-interactors (Yam *et al.*, 2008). Such observations along with oligomeric nature of the clients reinforces the proposition that eukaryotic chaperonins mainly fold proteins belonging to oligomeric assemblies and may serve as a reservoir to stabilize them against aggregation or degradation prior to complex formation (Spiess *et al.*, 2004; Yam *et al.*, 2008).

1.2.4.2   Group 2 chaperonins: Archaeal view

The general architecture of archaeal chaperonins is similar to that of the eukaryotic TRiC or GroEL. They possess one to five subunits (Archibald and Roger, 2002b; Large and Lund, 2009) that assemble as octameric (Ditzel *et al.*, 1998) or nonameric (Trent *et al.*, 1991) complexes of repeating subunits as compared to eight distinct subunits of eukaryotic TRiC.

1.2.4.2.1 Structural information about archaeal chaperonins

The first crystal structure for archaeal chaperonins was solved for CCT from thermophilic archeaon *Thermoplasma acidophilum* (Ditzel *et al.,* 1998)*.* The *T. acidophilum* CCT complex is hexadecameric with alternating α and β subunits and forms a spherical complex as compared to rather cylindrical GroEL-GroES complex (Refer to Figure 1.2) (Nitsch *et al.*, 1997). The volume of the cavity is ~25% smaller than that of GroEL-ES, thereby reducing the upper size limit of the clients to be encapsulated. The domain arrangement resembles to that of the GroEL-ES *cis* ring except the helical protrusions from the apical domain that form the lid. There is a 46% of sequence similarity between the *T. acidophilum* CCT and GroEL, mainly in the regions of equatorial

and intermediate domains (Waldmann *et al.*, 1995a). The nucleotide binding site (containing the GDGTTT motif) in the equatorial domain is conserved suggesting a common mechanism of ATP hydrolysis between two groups of chaperonins. These conclusions were further validated and generalized by a study that presented crystal structure of an archaeal chaperonin from thermococcus strain KS-1 (Shomura *et al.*, 2004). However, the structures derived from both the studies were in nucleotide-bound closed form of the chaperonins restricting the study of client binding and the conformational changes that take place during the folding cycle.

A significant amount of work was further done to understand the various folding-conformations of archaeal chaperonins from the species *Sulfolobus shibatae* (Schoehn *et al.*, 2000a; Schoehn *et al.*, 2000b) and *Methanococcus maripaludis* (Clare *et al.*, 2008) using cryo-electron microscopy and 3D-reconstructions. However the crystal structures of archaeal chaperonins in their open conformation were available only recently when two groups independently presented their studies for CCTs from species *M. maripaludis* (Pereira *et al.*, 2010) and *Acidianus tengchongenesis* (Huo *et al.*, 2010). Because the archaeal chaperonin from *M. maripaludis* (MmCCT) has been used as a model organism for this thesis, the conclusions from the former study are discussed in more detail below.

The open forms of MmCCT were solved using its lidless-variant (MmCCT-Δlid) because of the disorder that the highly flexible helical protrusions have been reported to cause in the open state (Heller *et al.*, 2004). MmCCT-Δlid retain the ability to bind clients, hydrolyze ATP and has been shown to achieve a functionally open state (Reissmann *et al.*, 2007). For comparison, closed forms of MmCCT with and without lid were also derived. The closed states of MmCCT and MmCCT-Δlid display a height of 162 Å and 148

Å respectively, and a diameter of 161 Å (Figure 1.8) (Pereira *et al.*, 2010), similar to previously obtained for other archaeal chaperonins (Ditzel *et al.*, 1998; Shomura *et al.*, 2004). However the open form of MmCCT-Δlid shows an increase in height to 178 Å while maintaining the diameter of 159 Å (Figure 1.8). A model reconstruction of full length wildtype MmCCT protein in its open conformation shows a remarkable height of 204 Å. Hence the volume of folding chamber in its open form is much higher compared to the closed form explaining the ability of Group 2 chaperonins to fold relatively higher molecular weight clients. The change in the dimensions between the two forms occurs due to radical domain movements: The equatorial domain rotates outwards at an angle of 35 degrees when transitioning from an open state to a closed state, while the apical domain moves inwards (Figure 1.8). The entire subunit rotates anti-clockwise as a rigid body around a pivot point identified near the residue P41 at the ATP binding site in the equatorial domain (Pereira *et al.*, 2010). The functional implications of these changes are discussed at length in the next section of mechanisms of folding of Group 2 chaperonins.
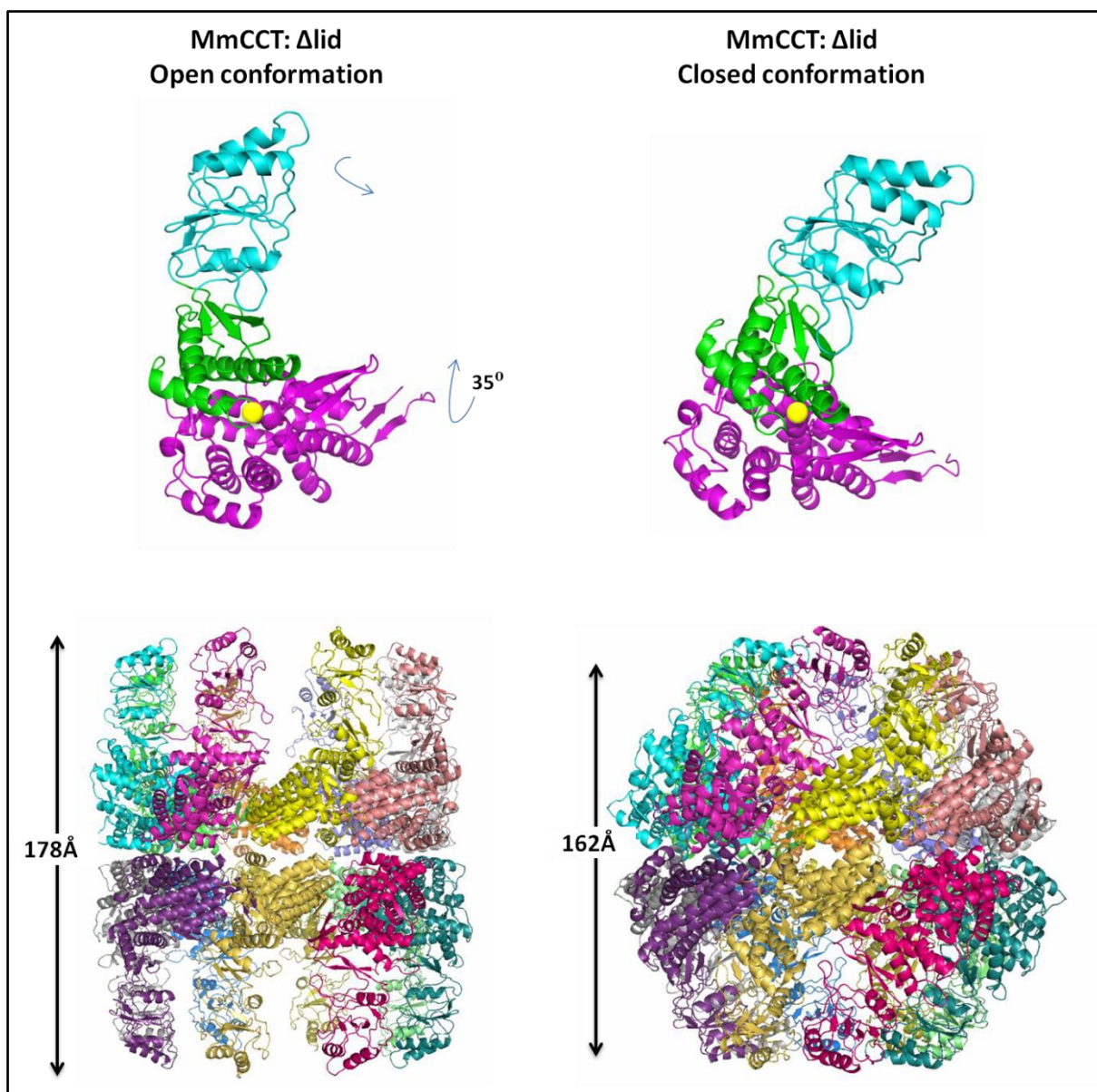
Figure 1.8

Overall architecture of lidless version of *M. maripaludis* chaperonin (MmCCT)

Ribbon diagrams of single subunit (top panel) and the whole oligomeric complex (bottom panel) for lidless versions of MmCCT in their open (left panel) and closed (right panel) states are shown. The conformational change for a single subunit from the open state to the closed state can be seen as a rigid body movement by ~35^0 anti-clockwise around a pivot point (coloured in yellow).

Note: Figure adapted from Pereira *et al.,* 2010.

1.2.4.2.2 Clients interacting with archaeal chaperonins

The knowledge about the client range of archaeal chaperonins was limited until recently when Hirtreiter *et al.* in 2009 elucidated an interactome for archaeal chaperonins from *Methanosarcina mazei* . *M. mazei* has copies of both Group 1 (MmaGroEL) and Group 2 chaperonins (MmaCCT-α, β and γ) (Klunker *et al.,* 2003) and hence provides an illustrative model to compare the client range for both groups simultaneously. Using co-immunoprecipitation assays and mass spectrometry of client-chaperonin complexes, they found that the archaeal chaperonins interacted with about 13% of the total cellular proteins i.e. ~333 proteins out of ~2500 soluble proteins in *M.mazei*. Out of these, 182 proteins (10 of them being essential) interacted with MmaGroEL-GroES chaperonin and 252 proteins (14 of them being essential) interacted with the MmaCCT chaperonins.

While a fraction of client proteins showed an exclusive interaction with either MmaGroEL-GroES or MmaCCTs, almost 100 proteins (out of total 333) were found to be interacting with both types of chaperonins (Figure 1.9). Additionally, most of the overlapping proteins were found to be equally abundant in the isolated fractions of MmaGroEL-GroES and MmaCCTs. Indeed certain proteins like citrate synthase were found to be enriched in fractions of MmaCCTs compared to MmaGroEL-GroES suggesting a gradual transition to becoming chaperonin-specific clients. The molecular mass of most of the interacting proteins was in the range of 10-50 KDa, however certain high molecular proteins (in the range of 80-210 KDa) were detected and interestingly, were found to be interacting with MmaCCTs and not MmaGroEL-GroES. This observation is consistent with the suggestion that large multi-domain proteins are more compatible with Group 2 chaperonins (Yam *et al.*, 2008). Another difference was that the MmaGroEL-interactors were relatively more hydrophobic and less charged

compared to the interactors of MmaCCTs. This supports the proposition that client-binding for Group 2 chaperonins, unlike Group 1, might not be solely governed by hydrophobic interactions (Gomez-Puertas *et al.*, 2004). Such an observation however differs from the TRiC-interactome data where a higher hydrophobicity quotient has been reported (Yam *et al.*, 2008) and with studies that show hydrophobic interactions to be prime for client binding in Group 2 chaperonins (Spiess *et al.*, 2006; Pereira *et al.*, 2010). Also it was observed that the MmaGroEL-selective clients shared more complex topologies and were relatively slow evolving than the structurally diverse MmaCCT-selective clients, speculating folding to be a more pronounced effect of MmaCCT than preventing aggregation. Lastly, it was reported that the client-selectivity was independent of the phylogenetic origin of the clients. Approximately 16% of *M. mazei* genes have been acquired by horizontal gene transfer, and hence were expected to be MmaGroEL-selective. Surprisingly though, 93% of these proteins including MmaGroEL were found to be interacting with MmaCCTs for folding, speculating an influence of structural topology rather than an evolutionary control over differential client-specificity (Hirtreiter *et al.*, 2009).

.
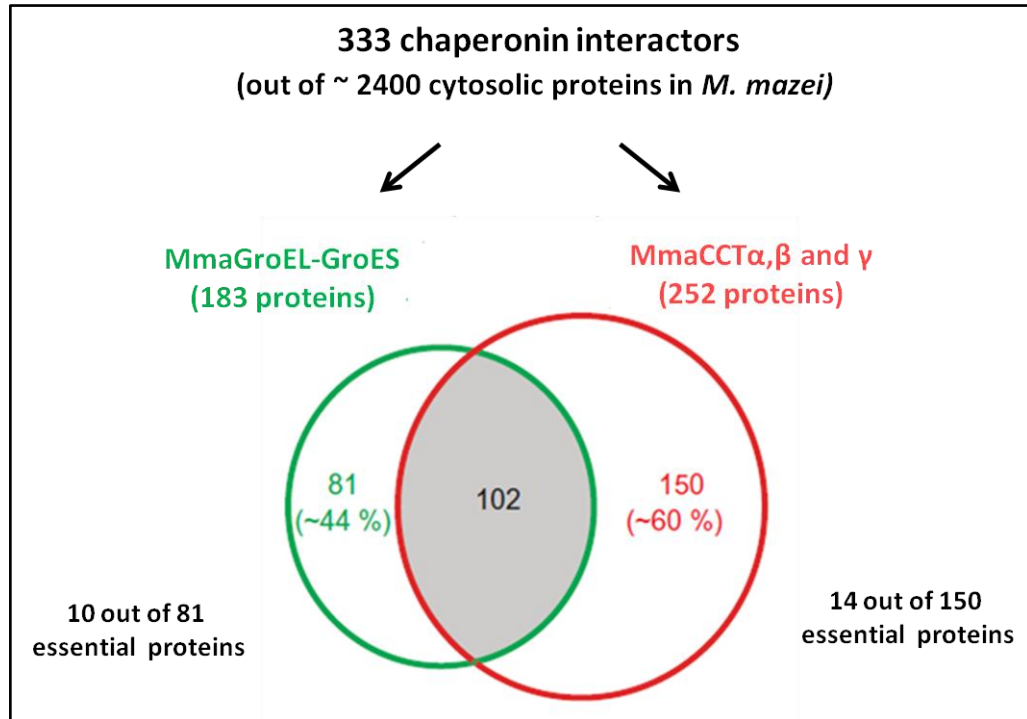
Figure 1.9 Relative distributions of *M. mazei* chaperonins' interactors.

Figure adapted from Hirtreiter *et al.,* 2009

1.2.5 Mechanisms of folding: Group 1 Vs Group 2 chaperonins

The mechanisms of folding for Group 1 and Group 2 are thought to follow a similar basic flow that is: binding to the client proteins through a client-binding site, ATP binding, client encapsulation and ATP hydrolysis, folding of the client and finally the release. A number of differences however exist at each level making Group 1 and Group 2 chaperonins distinct in their folding methodology. For ease of understanding, the varying aspects of folding mechanism at each stage have been described in a step-wise manner below:

1.2.5.1 Client binding

As discussed previously, the interactions of Group 1 clients with their chaperonin subunits are largely suggested to be hydrophobic and mostly non-specific (Fenton *et al.*, 1994; Chen and Sigler, 1999). However, the same for Group 2 remains to be confirmed. The Group 2 residues homologous to Group 1 client-binding hydrophobics are largely polar and charged (Horwich *et al.*, 2007). Structural data analysis for TRiC complexes with actin and tubulin have suggested a patch of positively charged polar residues in the apical domain to be involved in client binding in a subunit-specific manner (Llorca *et al.*, 1999; Pappenberger *et al.*, 2002; Dekker *et al.*, 2011a). This has been further supported by biochemical and phylogenetic analysis (Hynes and Willison, 2000; Ritco-Vonsovici and Willison, 2000) In contrast, using site-directed mutagenesis and cryo-EM constructions of VHL-TRiC interactions, hydrophobic residues in helix 10 and 11 were proposed for client binding in TRiC (Spiess *et al.*, 2006). The group further identified more clearly a conserved hydrophobic patch at the interface of helix 10 and 11 of archaeal chaperonin, MmCCT, by crystallography studies and has suggested it to be the

potential client binding site, (Pereira *et al.*, 2010), however the same remains to be confirmed by genetic or biochemical studies.

1.2.5.2 Intra-ring and Inter-ring allostery

ATP-induced allosteric transitions form the functional basis of all chaperonins. Each ring is in equilibrium between two states: tensed state, with low affinity for ATP and high affinity for un/misfolded clients; and a relaxed state, with high affinity for ATP and low affinity for un/misfolded protein clients (Yifrach and Horovitz, 1996). Thus the two rings exhibit negative co-operativity i.e. when one ring encapsulates the other ring remains free. This property has been observed for both, Group 1 chaperonin GroEL (Yifrach and Horovitz, 1995) and Group 2 chaperonins, TRiC (Kafri *et al.*, 2001) and MmCCT (Kusmierczyk and Martin, 2003a). The inter-ring negative co-operativity is thought to enhance the folding efficiency by enabling the chaperonin to function as a 'two-stroke' motor to bind kinetically trapped clients with high affinity in one ring and release productively folded substrates from the other (Horwich *et al.*, 2007). Nonetheless, studies have shown that single rings are sufficient to mediate productive folding of clients (Nielsen and Cowan, 1998). Moreover, mutants have been also reported that allow GroEL to function more effectively as a single ring (Sun *et al.*, 2003).

While a negative inter-ring co-operativity is seen in all chaperonins, intra-ring allostery differs for Group 1 and Group 2 chaperonins. In GroEL, binding of ATP to subunits within a ring is co-operative i.e. the binding of ATP to one subunit increases the affinity of the remaining subunits for the nucleotide. In TRiC however, weak intra-ring positive co-operativity is seen and the binding of ATP is sequential rather than concerted (Rivenzon-Segal *et al.*, 2005). A sequential nature of binding can allow domain-to-

domain interactions of hetereologous subunits of TRiC with multi-domain proteins. Such an arrangement can also support co-translational folding consistent with the idea that the TRiC client proteins are newly formed polypeptides primarily of multi-domain nature (Spiess *et al.*, 2006). The CCT from *T. acidophilum* on the other hand, does not show positive co-operativity (Bigotti and Clarke, 2005). However, the homo-oligomeric chaperonin complex from *M. maripaludis* exhibits weak positive co-operativity like TRiC (Kusmierczyk and Martin, 2003a). Hence, co-operativity within a ring does not seem to be a generalized feature in thermosomes.

1.2.5.3   Folding cycle

Client folding via Group 2 chaperonins is thought to be similar to Group 1 in the context that clients bind to the apical domains followed by ATP binding (or ATP binding precedes client binding). However, once the client and ATP have bound to the chaperonins, further mechanism of lid closure differ between Group 1 and 2 chaperonins as schematically shown in Figure 1.10. As discussed in previous section, in GroEL, binding of GroES brings large conformational changes that rearranges apical domain and releases client into the cavity. Lid closure precedes ATP hydrolysis. However, for Group 2 chaperonins the closing of the lid is triggered by the transition state of ATP hydrolysis. Single particle cryo-EM analysis of open and closed forms of bovine TRiC identified molecular motions that govern lid closure and established that unlike GroEL, ATP hydrolysis and not ATP binding is responsible for lid closure (Booth *et al.*, 2008). In view to further expand the theory, the same group have recently reported high resolution cryo-EM structures of ATP-free open state, ATP-hydrolysed closed state of Δlid variants (Zhang *et al.*, 2010) and of a D396A variant of MmCCT that binds to ATP but does not hydrolyze it (Zhang *et al.*, 2011). Comparisons of all three

structures reveal significant structural differences between the open, pre-hydrolysis and the closed state. The data suggests that like GroEL, a transition from ATP-free to ATP-bound pre-hydrolysis state does not involve radical changes in the structure. Instead, only a small shrinkage of the folding chamber entrance and a 45^0 of counter-clockwise rotation of the apical domain was observed (Figure 1.11A and B). At the same time, the lower surface of the intermediate domain is shifted towards the nucleotide binding pocket to optimise the structure for hydrolysis. A rocking motion occurs in the structure when the ATP bound pre-hydrolysis state traverse to the ATP hydrolysed closed state including in the helical protrusions that now cover the lid completely (Figure 1.11C and D). The potential client binding hydrophobic sites (Pereira *et al.*, 2010) are buried in the closed state indicating that the client should be released into the cavity. Thus ATP hydrolysis is implicated in dual function of lid closure and client encapsulation in Group 2 chaperonins (Pereira *et al.*, 2012).

The release of the folded client from the chaperonin occurs presumably because of the ATP binding to the opposite ring that in turn, through intra-ring communication, passes a signal for opening of the lid. The precise nature of such interactions remains to be established. Of note, client folding inside the GroEL cavity occurs during the ATP hydrolysis step and the release primarily is mediated by the conformational changes occurring due to ATP hydrolysis (Xu and Sigler, 1998). Certainly ATP hydrolysis is a result of ring's lower affinity for ATP caused due to ATP binding to the opposite ring. However, from the above data for Group 2 chaperonins, it is suggested that folding occurs only during the transition from already hydrolysed closed conformation to the open form. Such a mechanism can explain the sequential ATP binding in Group 2 chaperonins that can slow down the process giving more time to the client for folding.
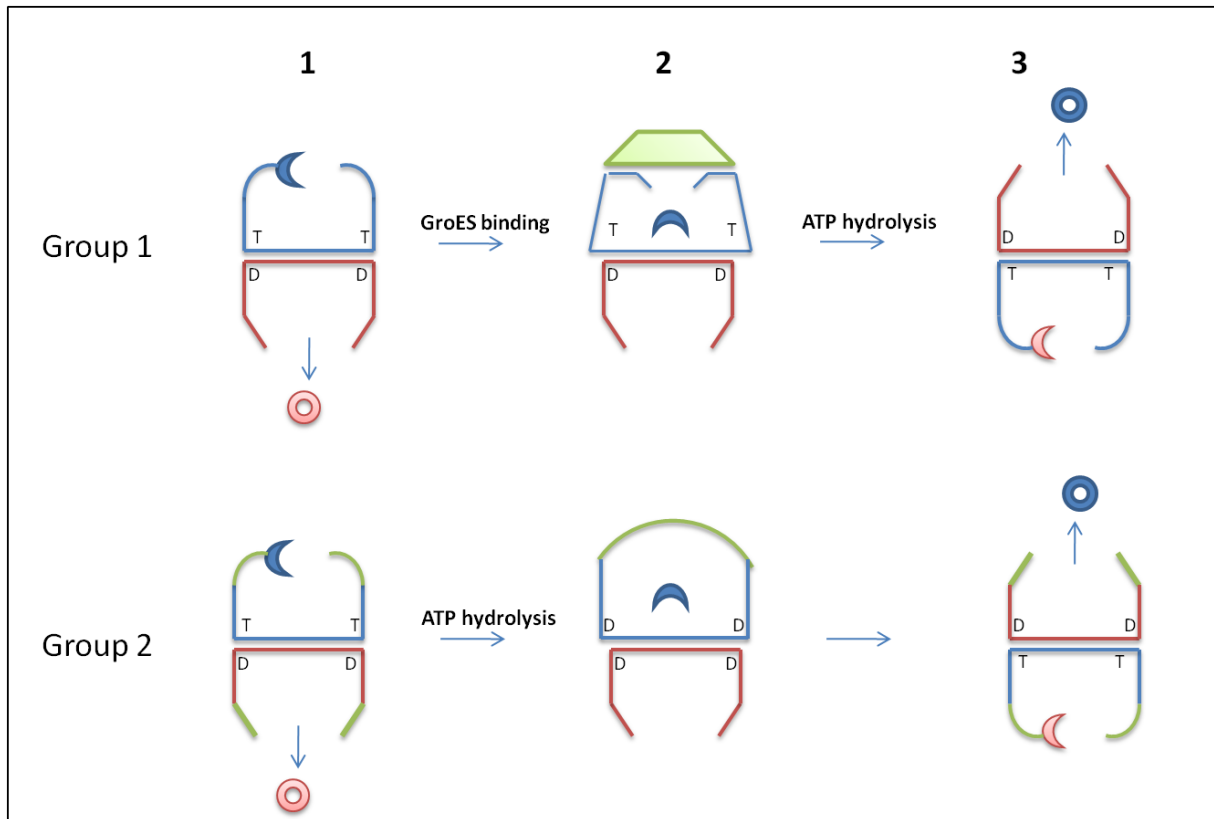
Figure 1.10

A cartoon representation of different folding mechanisms in Group 1 and Group 2 chaperonins.

For **Group 1** chaperonins (upper panel), ATP binding brings a small conformational change (blue) that enhances binding to unfolded clients (semi-circle) and co-chaperonin (Green triangle). Binding of co-chaperonin induces a large rotational movement closing the cavity. The folding occurs while ATP is hydrolysed leading to release of the folded client (circled). The two rings work alternatively.

For **Group 2** chaperonins (lower panel), ATP binding brings a small conformational change, similar to Group 1, which leads to binding of the unfolded client. In contrast, ATP hydrolysis leads to closing of the cavity. How the cavity opens and releases the client is still under question.
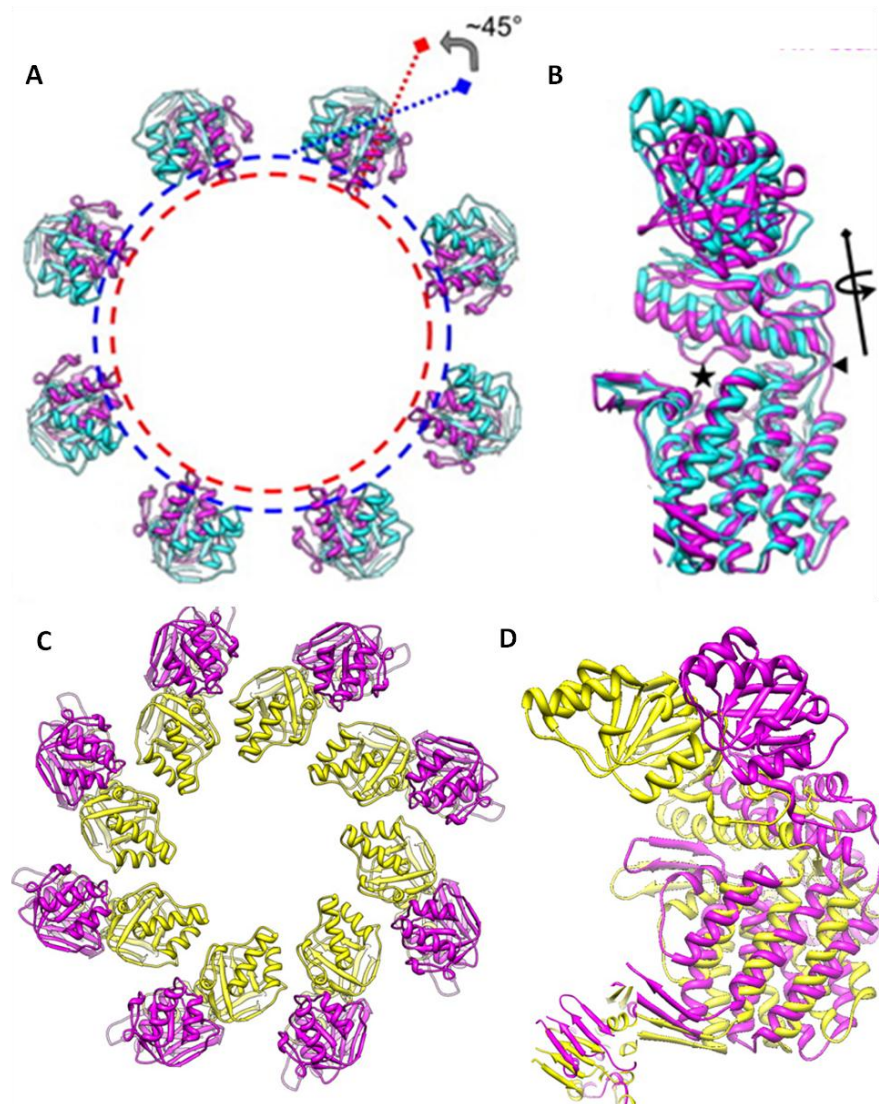
Figure 1.11

Structural changes occurring during the ATP binding and hydrolysis cycle of MmCCT

(A and C) Top-view of apical domains of MmCCT complex in ATP bound and ATP hydrolysis stages respectively.

(B and D) Conformational changes in single subunit of MmCCT complex during ATP bound and ATP hydrolysis.

The dotted blue and red circles touch the innermost part of the apical domains of ATP-free and ATP-bound states. The hinge point is marked by a black solid triangle. The black star labels the ATP-binding pocket.

Note: Figure taken from Zhang *et al.,* 2011

1.2.5.4 Folding models:

The passive cage model was initially accepted for Group 2 chaperonins also (Thulasiraman *et al.*, 1999). However, currently both the forced unfolding and active cage models have been suggested for Group 2 chaperonins. Analysis of live folding of actin on yeast-CCT and similar other studies have shown that the clients are not released in the chamber during folding. Instead the client remains bound to the apical domains and undergo folding in two steps: 1) ATP binding causes the apical domain with their bound clients to move and this movement mechanically force client folding 2) ATP hydrolysis liberates client and brings back the chaperonin to open state (Gomez-Puertas *et al.*, 2004; Stuart *et al.*, 2011). On the other hand, for archaeal chaperonins, encapsulation has been suggested to be an essential step for folding. For example, studies using Δlid variants of MmCCT show that the lid and hence complete encapsulation is indispensible for folding (Meyer *et al.*, 2003; Reissmann *et al.*, 2007). Moreover, using D386A variants of MmCCT it has been shown that ATP binding alone is ineffective for productive folding of their stringent clients rhodanese or malate dehydrogenase (Douglas *et al.*, 2011).

On the basis of such observations and recent crystal structures of open and closed MmCCT (Pereira *et al.*, 2010; Zhang *et al.*, 2011), an active-cage model has been proposed for archaeal chaperonins wherein the client is encapsulated upon lid closure and the cavity contributes to accelerated folding. This model is similar to that proposed for GroEL where closed confinement and the nature of the residues inside the cavity are thought to be the basic reasons for accelerated folding. Interestingly though, the inside of the cavity as seen for closed conformations of GroEL, thermosome and TRiC differ in their charge distribution and hydrophobicity (Figure 1.12) (Cong *et al.*, 2010). As seen

in the figure, the charge distribution for thermosome and GroEL is symmetrical while asymmetry is observed in TRiC complex; nonetheless, both TRiC and thermosome show higher distribution of positively charged residues as compared to the relatively negatively charged patches in the GroEL cavity. Both the inside of TRiC and thermosome are highly hydrophilic consistent with the active cage model. Notably, they display a lower proportion of hydrophobic residues as compared to GroEL.

1.2.5.5 Co-factors for Group 2 chaperonins

Although the Group 2 chaperonins do not require assistance of a co-chaperonin like GroES in Group 1 chaperonins, certain co-factors have been reported to be modulating the folding process of Group 2 chaperonins. One of such co-factors is the heterohexameric protein called prefoldin. Prefoldin also called as GimC is not phylogenetically related to GroES and unlike GroES, functions directly by binding to the client proteins. In the cases studied, this hetero-oligomeric protein complex is known to interact with the non-native substrates (typically actin and tubulin in eukaryotes) and deliver them to the chaperonin system by directly interacting with the chaperonin complex (Vainberg *et al.*, 1998). Prefoldin homologes are found both in eukaryotes and in the archaea alongside Group 2 chaperonins though their mechanism of client binding has been shown to be different between the two domains (Martin-Benito *et al.*, 2002). In addition to prefoldin, recently, the interactions of conserved phosducin-like proteins (Plp1 and Plp2) have been reported to be playing a role in TRiC mediated folding of actin and tubulin (Stirling *et al.*, 2007; McCormack *et al.*, 2009). Such proteins have also been reported for productive folding of G-proteins (Willardson and Howlett, 2007).
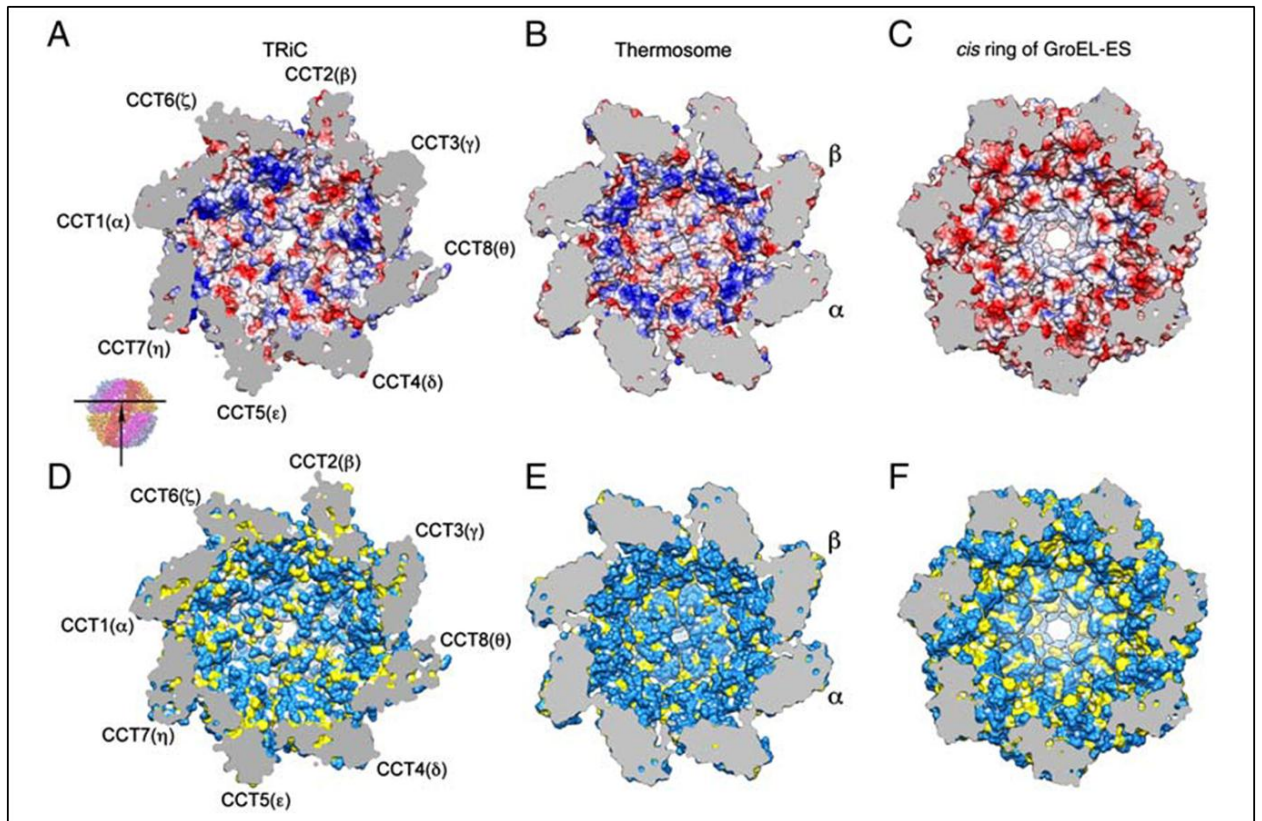
Figure 1.12

Surface properties of the central cavity for TRiC, thermosome from *T. acidophilum* and GroEL

(A-C) Charge distribution of inner cavity of TRiC, thermosome and GroEL: blue represents positively charged patches, red negatively charged patches, and white neutral patches. (D-F) Degree of hydrophobicity of the cavities: hydrophilic (sky blue), hydrophobic (yellow), and main chain (white)

Note: Figure taken from Cong *et al.,* 2010.

1.3   Archaea as models to study Group 2 chaperonins

1.3.1 Archaea

Till late 1980s the general accepted classification of life was a five kingdom classification placing bacteria and archaea (then known as archaebacteria) in the same kingdom, Monera. The basis of Monera as a different kingdom was rather negative as organisms lacking similarities with eukaryotic organisms. However, the molecular sequencing revolution identified unusual and definitely diverging characteristics in archaea discrete from bacteria and eukaryotes; one of the characteristic being the distinguishing sequence of small subunit rRNA. On the grounds of dividing the living world on the basis of phylogenetic characteristics, at the molecular sequence level rather than at a cellular or phenotypic level, Woese introduced a new taxon of 'domain' above the kingdom level and described a three domain classification that classified archaea as a separate domain of life along with the domains Bacteria and Eukarya (Woese et al., 1990) (Figure 1.13).

The archaea are divided into two main classes according to rRNA comparisions, the Crenarchaeota and Euryarchaeota. Crenarchaeota mostly consist of hyperthermophiles (e.g. *Sulfobus*) while Euryarchaeota is a more diverse group and includes methanogens (e.g. *Methanosarcina*), halophiles (e.g. *Haloferax*), thermophilic methanogens (e.g. *Methanococcus*) and hyperthermophiles (e.g. *Thermoplasma*). Two more phylums have been further added and are referred to as Korararchaeota (Barns *et al.*, 1996) and Nanoarchaeota (Huber et al., 2003). Archaeal species belonging to these classes are relatively rare and have peculiar characteristics for example nanoarchaeota equitans has a significantly smaller genome size of 490 kb only (Huber *et al.*, 2003) .
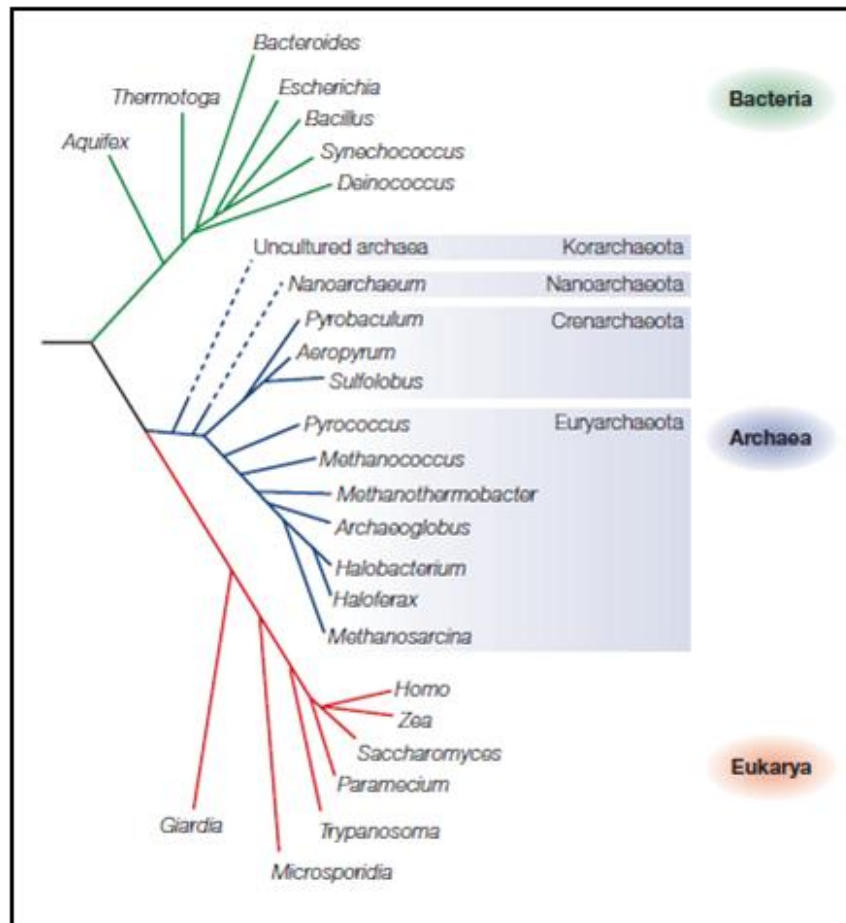
Figure 1.13

Universal phylogenetic tree using rRNA as molecular chronometer

The figure shows the division of living world in three domains and relationship of archaea with both Bacteria and Eukarya. The domain Archaea is split into two main kingdoms: The most diverse - Euryarchaeota having the euryarchaeal type of rRNA comprise extreme halophiles and methanogens, whereas Crenarchaeota consists of thermoacidophiles. Nanoarchaeota has one known member (Nanoarchaeum equitans), and so far Korarchaeota are indicated only by environmental DNA sequences. The positions of these phyla on the rRNAtree are uncertain (indicated by dashed branches).

Note: (Figure taken from Allers and Mevarech, 2005)

It is known that many of the archaeal features are either similar to bacteria or eukarya. For example bacteria and archaea have similar cellular framework i.e. similar general cell sizes, lack of nuclear membrane and organelles, the presence of a large circular chromosome and similar operational machinery of housekeeping genes (Cohen et al., 2003). However, the archaeal information-processing systems like DNA replication, transcription and translation resemble to those in eukaryotes (Grayling et al., 1996; Bell and Jackson, 2001; Hickey et al., 2002). Because only a subset of the eukaryotic proteins is found in archaea, the archaeal system is simpler and more amenable to analysis. Archaea have served a streamlined model for understanding complex mechanisms of eukaryotes in many fields such as DNA replication (Grabowski and Kelman, 2001).

Moreover, thermophilic and halophilic archaeal proteins have been constantly used for elucidation of structures for example the first crystal structure for a ribosome was obtained using the large ribosomal subunit from *Haloarcula marismortui* (Ban et al., 2000). Thermostable enzymes of archaea are routinely used in laboratories for DNA amplification by PCR (for example, Pfu DNA polymerase from *Pyrococcus furiosus*). Also, large number of genetic manipulation techniques for archaea are being developed permitting exploration of archaeal models to better understand complex eukaryotic systems (Allers and Mevarech, 2005).

1.3.2 Why study archaeal chaperonins?

There are three main reasons to study chaperonins in archaea (Large and Lund, 2009; Lund, 2011):

1. Archaeal chaperonins are excellent models to study general properties of Group 2 chaperonins. As discussed in the section 1.2, it becomes apparent that while a lot of information has been convincingly established for the working of Group 1 chaperonins, mechanisms of Group 2 chaperonins are less understood. Archaeal chaperonins contain fewer unique subunits as compared to eight distinct subunits of TRiC. This allows easier recombinant expression in *E. coli*. As a result large scale purification and further use for biochemical and biophysical methods is enabled. Data interpretation is simplified which then can be extended to Group 2 chaperonins by targeted experiments.

2. Although archaeal chaperonins are phylogenetically more related to Group 2 chaperonins, they resemble to Group 1 chaperonins in a number of aspects. For e.g. archaeal chaperonins are induced by heat shock like GroEL. Also, they exhibit at least some homology between their subunits i.e. not all subunits are encoded by distinct genes as for eukaryotic chaperonins. It can be speculated that archaeal chaperonins represent a link between Group 1 and Group 2 chaperonins and hence, may hold a key to understanding the divergence between the two groups.

3. They are of interest in their own right to understand the process of protein folding in archaea. Archaea can thrive under extreme conditions e.g. high temperature and high salinity. It is intriguing to understand how these organisms sustain such conditions that otherwise could have profound effects on protein folding.

1.3.3 Project flow

The ultimate aim for this project is to lead to a better understanding of properties of archaeal chaperonins and to further comprehend the basis of divergence between the Group 1 and Group 2 chaperonins. This study is a functional analysis of the chaperonin from archeaon *M. maripaludis* (MmCCT) using *E. coli* as a host organism.

The first part of the project is targeted at a study where a Group 2 chaperonin, MmCCT, was introduced in a Group 1 host, *E. coli*, when the native Group 1 chaperonins, GroEL and GroES, are limited or absent. This analysis allows an understanding of whether or not and if yes to what extent, can an archaeal chaperonin replace the function of GroEL. Using mutagenesis approach, further experiments were performed to isolate functional variants of MmCCT that can more effectively replace GroEL function (Chapter 3). MmCCT along with the functional variants were characterized biochemically for their *in vitro* activity to understand their differential GroEL substituting ability (Chapter 4). The variants of MmCCT with an improved function in GroEL-depleting *E. coli* cells provided a platform to dissect the properties of archaeal chaperonins using *E. coli* as host. Thus using the generated system, preliminary studies have been initiated to elucidate clients that interact with MmCCT in *E. coli* and to test the validity of potential client binding residues of MmCCT (Chapter 5). To test whether the observed phenomenon holds for other archaeal chaperonins also, the study has been extended to chaperonins from various archaeal species such as *Methanosarcina mazei*, *Thermoplasma acidophilum* and *Haloferax volcanii* (Chapter 6). An initiative has also been described to develop an *in vivo* genetic system in archeaon *H. volcanii* that can allow direct dissection of the properties of archaeal chaperonins (Chapter 7).

CHAPTER 2
Materials and Methods

2.1 Strains

2.1.1 *E. coli* strains

Table 2.1: List of *E. coli* strains used in the study

Strain	Genotype/Features	Reference(s)
DH5α	F$^-$ *gyrA96 (Nal^r) recA1 relA1 endA1 thi-1 hsdR17(r$_K$$^-$ m$_K$$^+$) glnV44 deoR Δ(lacZYA-argF) U169 nupG [Φ80d Δ (lacZ) M15] λ–*	Hanahan, 1983, Gibco-BRL
JM110	F' *traD36 proA$^+$B$^+$ lacI^q Δ (lacZ) M15I rpsL (Str^r) thr leu thi lacY galK gal Tara fhuA dam dcm glnV44Δ(lac-proAB)*	Yanisch-Perron *et al.*, 1985
MGM100	Modified MG1655 *kan^r-pBAD -groESL*	McLennan and Masters, 1998
BL21 (DE3)	F– *ompT hsdS*B(rB– mB–) *gal dcm* (DE3)	Novagen
TAB21	BL21 containing *groESL* under *pBAD* promoter	Established in the lab
AI90	Modified TG1 Δ*groEL::kan^r*	Ivic *et al.*, 1997

2.1.2 *H. volcanii* strains

Table 2.2: List of *H. volcanii* strains used in the study

Strain	Description	Reference(s)
H. volcanii DS70	Wild-type strain cured from endogeneous pHV2 plasmid.	Wendoloski *et al.*, 2001
H. volcanii H26	Δ*pyrE2* deletion in DS70 background	Bitan-Banin *et al.*, 2003
H. volcanii H26:Δ*cct2*	H26 strain carrying a deletion of *cct2* (created by using the *pyrE2* knockout method.)	Kapatai *et al.*, 2006
H. volcanii H26:Δ*cct3*	H26 strain carrying a deletion of *cct3* (created by using the *pyrE2* knockout method.)	Kapatai *et al.*, 2006
H. volcanii H26: Δ*cct1/Δcct3*	Double knockout H26 strain carrying a deletion of *cct1* and *cct3* genes (created by using the *pyrE2* knockout method.)	Kapatai *et al.*, 2006
H. volcanii H26:Δ*cct2/Δcct3*	Double knockout H26 strain carrying a deletion of *cct2* and *cct2* genes	This study

2.2 Plasmids

Table 2.3: List of plasmids used in the study

Plasmid	Features	Reference(s)
pOK12	Low copy number pUC-type vector with a p15A replicon and a *lacZ* gene downstream the multiple cloning site (MCS).	Vieira and Messing, 1991
pET21b	A pET expression plasmid containing T7 promoter, lac operator, ampicillin resistance marker (Amp^R), N-terminal T7 tag and a C-terminal His tag.	Novagen
pET22b	Similar to pET21 except for lacking the N-terminal T7 tag.	Novagen
pET27b	Similar to pET21 except for containing a Kanamycin resistance marker (Kan^R) instead of an Amp^R	Novagen
pTrc99A	An expression plasmid driven by the p*trc* promoter, IPTG inducible.	Amann *et al.*, 1988
pACYC184	Low copy number expression plasmid containing the p15A origin of replication along with chloramphenicol (Cam^R) and tetracycline (Tet^R) resistance markers.	
pCRPrtneo	A pCR2.1TOPO derived vector containing *M. maripaludis hpt* gene under constitutive hmv promoter and a Neomycin resistance marker (Neo^R)	Moore and Leigh, 2005

pTA131	pBluescript II with fragment containing *H. volcanii pyrE2* gene under ferridoxin (fdx) promoter of *H. salinarum*	Allers *et al.*, 2004
p354	*E. coli/H. volcanii* shuttle vector derived from pTA131 with an additional fragment containing pHV1/4 origin of replication	Norais *et al.*, 2007
p131-5'3'flank*cct2*	p131 containing upstream and downstream flanking sequences of *H. volcanii-cct2*-gene	This study
pOK-PTC2	*ptna* driven *cct2* gene and *pyrE2* gene inside the *cct1* gene lockers inserted in pOK12	This study

2.3 Primers

Table 2.4: Primers for mutagenesis

Name of the oligo	Sequence (5' to 3')	Length (in bases)
MmCCT-K216n-F	GGTGTTTTAGTTGACnnnGAAAGAGTAAGCGCTCAAATGCC	41
MmCCT-K216n-R	GGCATTTGAGCGCTTACTCTTTCnnnGTCAACTAAAACACC	41
MmCCT-K216E-F	GGCATGATCATTGACGAGGAAAGAGTCCACAC	32
MmCCT-K216E-R	GTGTGGACTCTTTCCTCGTCAATGATCATGCC	32
MmCCT-D386A-F	GCAAGAGCAGTAGACGCTGCTGTTGGTGTAGTTG	34
MmCCT-D386A-R	CAACTACACCAACAGCAGCGTCTACTGCTCTTGC	34
MmCCT-G61C-F	GACGTTGTTGTTACAAACGACTGTGTTACAATCTTAAGAG	40
MmCCT-G61C-R	CTCTTAAGATTGTAACACAGTCGTTTGTAACAACAACGTC	40
MmCCT-V273n-F	AATGTTAAAAGACATGnnnGCTGAAATCAAAGCAAGCGGTGC	42
MmCCT-V273n-R	GCACCGCTTGCTTTGATTTCAGCnnnCATGTCTTTTAACATT	42
MmCCT-Y297n-F	CGATGACTTAGCACAACACnnnTTAGCTAAAGAAGGAATCGTAG	44
MmCCT-Y297n-R	CTACGATTCCTTCTTTAGCTAAnnnGTGTTGTGCTAAGTCATCG	44
Mma-alphaM226I-For	GGAAAGAGTCCACACAAATATCCCTGAGAAAGTAAAGGATGC	42

Mma-alphaM226I-Rev	GCATCCTTTACTTTCTCAGGGATATTTGTGTGGACTCTTTCC	42
Mma-gammaM228I-For	CGCGTCCACACCGGCATCCCGGAAGTTGTGAAG	33
Mma-gammaM228I-Rev	CTTCACAACTTCCGGGATGCCGGTGTGGACGCG	33
Xba1_MmCCTinternal_FOR	GCCCCGTCTAGAGAAACAGTAAGATCAACACTCGGTCCA	39
MmCCTinternal_Not1_REV	CTAAAAGCGGCCGCAACTTCTGTTGAACCTCCACCAGCTAC	41

Table 2.5: Primers for sequencing

Name of the oligo	Sequence (5' to 3')	Length (in bases)
T7 promoter primer	TAATACGACTCACTATAGGG	20
T7 terminator primer	GTAAGCTGCAACCTTGTCATCCATG	25
MmCCT-seq1	GATACATGGGAAGAGATGCTC	21
MmCCT-seq2	AGAGTAAGCGCTCAAATGCC	20
MmCCT-5'-Rev	GTTGGGTGTACATTTTGGTCC	21
MmCCT-Mid-F	CAAAGCAAGCGGTGCAAAC	25
MmCCTgene-F	CTTAACGGATTCCTGAAATGAGCTGTTGACAATTAATC	38
MmCCTgeneSalI-R	GGTTAAGTCGACTGAAAATTACATCATTCCTGGCATTC	38
AI90-GroEL-For	CGTAAAATTCGGTAACGACGCTCGTGT	27
AI90-GroEL-Rev	GTAGCCACGGTCGAACTGCATACCTTC	27
Mma-alpha-mid-For	CGGATACAGACTGGCAGCC	19
Mma-alpha-mid-Rev	GTTCTCCCACATGTCAACGACC	22
cct2gene-For	ACAATCAAGGCCTGCATATGAGCCAGCGAA	30
cct2gene-Rev	CCCTCGGACTGTCGGGAATTCGTAGTCGGC	30

Table 2.6A: Primers for for inserting CBP tag in *Mmcct* by overlap extension PCR

Name of the oligo	Sequence (5' to 3')	Length (in bases)
MmCCTgeneFOR	AGA-TAT-ACA-TAT-GGC-ACA-ACC-AGG-AGT-TTT-ACC-TGA-A	37
MmCCTgeneREV	GAA-TTC-GGA-TCC-TGA-AAA-TTA-CAT-CAT-TCC-TGG-CAT-TCC	39
CBP-A-REV	GAT-TCG-CCG-CGC-TCA-CCG-CAA-TAA-AAT-TTT-TTT-TCC-AGC-GGC-GTT-TGC-CGC-TGC-CTT-TTTCAA-ACT-GCG-GAT-GGC-TCC-AGC-CGC-TGC-CGC-TGC-CAG-GGT-GTT-TGC-ATT-CTT-CTA-CC	115
CBP-B-FOR	TGC-GGT-GAG-CGC-GGC-GAA-TCG-CTT-TAA-AAA-AAT-TAG-CAG-CAG-CGG-CGC-GCT-GGG-CAG-CGGCCA-TCA-TCA-TCA-TCA-TCA-TCA-TCA-TGG-CAG-CGG-CAA-AGC-AGT-TAC-AAT-GCT-CAT-C	115

Table 2.6B: Primers for amplifying flanking regions of *cct2* gene in *H. volcanii*

Name of the oligo	Binding site on the gene	Sequence (5' to 3')	Length (in bases)
5'flankingcct2-FOR	474bp upstream of *cct2* gene	GGAGACGTTCGAATTCGACGTGC	23
5'flankingcct2-REV	26bp upstream of *cct2* gene	AGACAGGCCTGGATCCTTTGTGATTC	26
3'flankingcct2-FOR	25bp downstream of *cct2* gene	AAATCGGATCCCGCCTTCACACCCC	25
3'flankingcct2-REV	524bp downstream of *cct2* gene	CGTTGACCCTCTAGATGCCCGCGG	24

2.4   General microbial techniques: *E. coli*

2.4.1 Growth media and supplements

Lysogeny-broth (LB) media: 1% w/v tryptone, 0.5% w/v yeast extract and 1% NaCl dissolved in deionised water and sterilized by autoclaving. For LB agar plates 15 g/L of agar was added to the liquid media and then sterilized by autoclaving.

Supplements: The media supplements are listed in Table 2.7 with their stock and working concentrations. All supplements were supplied from Sigma-Aldrich, UK. The stock solutions were prepared in ultrapure MilliQ water (Merck Millipore) unless stated, and filtered using 0.2 μm diameter Acrodisc syringe filter (Pall Life Sciences).

2.4.2 Growth of *E. coli* strains

Overnight liquid cultures were grown in LB medium with appropriate supplements in a shaking incubator (180-200 rpm) at 37^{0}C unless otherwise stated. Growth of liquid cultures was monitored by measuring optical density at a wavelength of 600 nm (OD_{600}) using a spectrophotometer. Plate cultures were grown on LB agar plates containing suitable additives with incubation at various temperatures.

2.4.3 Glycerol Stocks

Glycerol stocks *of E. coli* strains were made by mixing 875 μl of a fresh overnight culture with 125 μl of 80% sterile glycerol, after which they were stored at -80^{0}C.

Table 2.7: List of additives used in bacterial culturing with their concentrations

Supplement	Stock concentration	Working concentration
Ampicillin (Amp)	100 mg/ml	100 µg/ml
Kanamycin (Kan)	50 mg/ml	50 µg/ml
Chloramphenicol (Cam) (Solvent: Ethanol)	50 mg/ml	50 µg/ml
L-Arabinose (Ara)	20% w/v	0.2% (w/v)
D-Glucose (Glu)	20% w/v	0.2% (w/v)
IPTG	1 M	1 mM
Sucrose	50% w/v	5% w/v

2.4.4 Plasmid isolation

Small scale plasmid DNA preparations from *E. coli* cultures were carried out by using the Qiagen Qiaprep spin Plasmid Miniprep kit as per manufacturer's instructions or by alkaline lysis method as described below. The large scale preparations were carried out using Qiagen Plasmid Midiprep kit as per recommended guidelines.

Alkaline lysis method:

Single colonies were inoculated into 5 ml of sterile LB broth (with required additives) and allowed to grow overnight with shaking at 37^0C. Cells from overnight cultures were pelleted at 13,000 rpm in a bench top microcentrifuge for 2 minutes. 200 µl of sterile GTE (50 mM glucose, 25 mM TRIS pH 8.0, 10mM EDTA) was added and the cells were resuspended by vortexing or pipetting. Next, 400 µl of freshly made lysis solution (1% SDS, 0.2 M NaOH) was added and the samples were mixed by inversion and left on ice for 5 minutes (lysis is observed by clearing of the turbid solution). Further, 300 µl of 3 M sodium acetate pH 5 adjusted by glacial acetic acid was added and the solution, inverted several times to mix and incubated at room temperature until a white precipitate developed. The tubes were then spun for 10 minutes at 13,000 rpm. The resulting supernatant free from the protein precipitates was transferred to a clean eppendorf tube.

To precipitate nucleic acids, 600 µl of isopropanol was added and mixed thoroughly by inverting tube several times. The tubes were then spun for 20 minutes at 13,000 rpm to get a small white pellet of plasmid DNA. The pellet was washed first with 500 µl of 70% ethanol to remove salts and then with 500 µl of 100% ethanol. Lastly the pellet was dried for 10 minutes in a 37^0C incubator and resuspended in 50 µl of sterile TE buffer

(10 mM TRIS pH 8.0, 1 mM EDTA) to obtain the plamid DNA. The plasmids were stored at -20°C freezer.

2.4.5 Preparation of chemically competent *E. coli* cells

Reagents required:

TfbI	294.5 mg KAc, 121 mg RbCl$_2$, 220 mg CaCl$_2$, 1g MnCl$_2$ and 15 ml of glycerol were dissolved in 65 ml of MilliQ water. The pH was adjusted to 5.8 using glacial acetic acid. The solution was made up to 100 ml with DDW water, filter sterilised and stored at -20^{0}C.
TfbII	209 g MOPS, 1.65 g CaCl2, 121mg RbCl$_2$was dissolved in 60 ml DDW water. To this solution 15 ml of glycerol was added and the pH adjusted to 6.5 using 1 M KOH. The solution was made up to 100 ml, filter sterilized and stored at -20^{0}C.

Method:

A flask with 100 ml of pre-warmed LB was inoculated with 1 ml of overnight culture and grown in a shaking incubator at 37^{0}C. The growth of the culture was monitored every 30 minutes until the OD$_{600}$ reached 0.4. Care was taken to ensure that the cultures do not over-grow. The flask was then placed on ice for 5 minutes following which the cells were harvested by centrifuging at 5000 rpm for 10 minutes at 4^{0}C. The supernatant was discarded and the pellet was gently resuspended in 50 ml of TfbI. This suspension was incubated on ice for 10 minutes and then spun down at 6000 rpm for 5 minutes. The resultant pellet of cells was then resuspended in 2 ml of TfbII and incubated on ice for 2 hours. Aliquots (100 µl) of the solution were made and frozen at -80^{0}C.

2.4.6 Transformations

To 100 μl of competent cells (thawed on ice) approximately 10-100 ng of plasmid DNA or appropriate amounts of ligation reactions was added and incubated on ice for 45 minutes. The cells were then subjected to heat shock by placing them in a 42^0C water bath for 90 seconds, followed by incubation on ice for 5 minutes. To this, 900 μl of LB media was added and incubated at 37^0C with shaking for 1 hour. The cells were then plated on LB plates containing appropriate antibiotics.

2.5 General microbial techniques: *H. volcanii*

2.5.1 Growth media

Modified Growth Medium (MGM): *H. volcanii* grows at high salt concentrations. For standard procedures, *H. volcanii* was cultured on 18% modified growth medium (MGM) as described (Table 2.8; Dyall-Smith, Protocols of haloarchaeal genetics). The ingredients were dissolved and pH adjusted to 7.5 using 1 M Tris Base. The volume was adjusted to 1000 ml with pure water and the media was sterilized by autoclaving. For pouring plates, 15 g of Bacto-Difco agar was added to liquid MGM and dissolved by boiling to 100^{0}C for 15 minutes. This was then sterilized by autoclaving and poured into plates containing appropriate supplements.

Minimal medium: The *pyrE2* cured strain *H. volcanii* H26 is a uracil auxotroph. To provide a medium for selection, a minimal medium was used (Table 2.9). For solid media 2% of bacto-difco agar was added and dissolved prior to autoclaving. Uracil and 5-fluroorotic acid (5-FOA) were added to cultures for *pyrE2* based selection.

Table 2.8: Modified growth medium for culturing *H. volcanii*

Ingredient	18% MGM (per 1 L)
Salt water (30% stock) (240 g NaCl, 30 g $MgCl_2.6H_2O$, 35 g of $MgCl_2.7H_2O$, 7 g of KCl, in 1 L of MilliQ water. After mixing, 5 ml of 1 M $CaCl_2.2H_2O$ was added and pH adjusted to 7.5 with 1 M Tris Base)	600 ml
Pure water	367ml
Peptone (Oxoid)	5g
Yeast extract	1g

Table 2.9: Minimal media composition for culturing *H. volcanii*

Ingredient	Minimal medium (per 100 ml)	Final concentration
Salt water (30% stock)	67 ml	20%
Pure water	23 ml	-
500 mM Alanine	5 ml	25 mM
Mixed and pH set to 7; sterilization by autoclaving		
10% Succinate	4 ml	0.4%
1 M $CaCl_2$	0.5 ml	0.5g/ml
0.5 M (KH)PO_4 pH 7.4	0.3 ml	
Trace metals	0.1 ml	Trace
*Uracil in DMSO 50 mg/ml	0.1 ml	50 µg/ml
*5 FOA in DMSO 50 mg/ml	0.1 ml	50 µg/ml

2.5.2 Growth of *H. volcanii* strains

Liquid cultures were grown in MGM or minimal media inoculated with either a single colony or a loop equivalent of glycerol stocks of H. volcanii at 37^0C or 42^0C on a shaking incubator (180rpm). It took about 2 days in MGM and 3 days in minimal media for the cultures to reach log phase. The growth of liquid culture was monitored by measuring the optical densities of the culture at 650 nm (OD650). For plate cultures, the inoculum was spread on MGM or minimal media plates and incubated at 37^0C or 42^0C. The plates were kept in a sealed box with wet filters to keep the atmosphere moist. The colonies took about seven days to come up on MGM plates and ten days to come up on minimal media plates. The growth was faster at 42^0C with 4-5 days on MGM plates and 7-8 days on minimal media plates.

2.5.3 Transformations

Polyethylene glycol (PEG) mediated transformation procedure was used for *H. volcanii* (Cline *et al.*, 1989).

Reagents required:

Buffered spheroplasting solution (BSS)	58.5 g NaCl (1 M) , 2.01 g KCl (27 mM), 50 ml of 1 M Tris-HCl pH 8.2 (50 mM), and 150 g sucrose (15%) were dissolved in 1 L of milliQ water and sterilized.
Unbuffered spheroplasting solution (USS)	58.5 g NaCl (1 M), 2.01 g KCl (27 mM) and 150 g sucrose (15%) were dissolved, pH adjusted to 7.5 using 1 M NaOH, volume made up to 1 L.
Spheroplast dilution solution	15 g of sucrose (15%) dissolved in 76 ml of 30% saltwater (23%). The final volume was made upto 100 ml with milliQ water and the solution was autoclaved. After cooling, 0.75 ml of 0.5 M $CaCl_2$ was added.

Regeneration solution	60% w/v of 30% Saltwater (18%), 0.5% yeast extract, 0.1% peptone, 0.1% cas amino acids, 15% sucrose, pH to 7.5 and autoclaved, 3 ml of 1 M CaCl$_2$ after autoclaving
60% PEG solution	600 µl of PEG$_{600}$ in 400 µl of USS

*All solutions were sterilized by autoclaving except for PEG solution

Method:

Single colonies were inoculated into 5 ml of MGM (or minimal medium depending upon the experiment) and allowed to grow for two days (OD$_{650}$ between 0.8 and 1). 1 ml of the culture was used to set up a 50 ml culture in the same medium and the culture was allowed to reach log phase (O.D$_{650}$ between 0.4 and 0.8) (~24 hours, depends upon the strain). 10 ml of the culture was harvested by centrifuging at 5000 rpm for 10 minutes at 4°C. The pellet obtained was resuspended in 2 ml of BSS and re-pelleted. After the washing step, the pellet was again resuspended in 600 µl of BSS out of which 150 µl was used per transformation. 15 µl of 0.5 M EDTA (pH 8.0) was added from the side of the tube and the solution mixed by gentle inversion. Mixture was allowed to stand at room temperature (RT) for 10' to allow the spheroplast (cells without cell wall) formation. Because archaeal cells have a chemically more resistant form of cell wall, conventional transformation methods like heat-shock does not allow entry of DNA into the cells. EDTA removes the S-layer of cell wall making it susceptible to DNA-intake. Next, the DNA was added (not more that 2 µg in concentration and 7.5 µl in volume) to the spheroplast solution and allowed to stand for 5 minutes at RT. Note that DNA used here was prepared from a dam$^-$ strain of *E. coli,* JM110*.* Non-methylated DNA was important to use as the restriction system of *Haloferax volcanii* degrades methylated DNA,

specifically methylated A residues at GATC sites and reduced transformation efficiency by 1000 folds (Allers and Mevarech, 2005). 60% PEG solution was then added to the mixture (185 μl), the solution mixed gently until homogeneous and incubated for 30 minutes to 1 hour at RT. PEG aids in delivering plasmid DNA to the cell surfaces. Further, 1 ml of spheroplast dilution solution was added, left at RT for 2 minutes and then centrifuged (8000 rpm for 5 minutes) to pellet the spheroplasts. The supernatant was gently removed and 1 ml of regeneration solution was added without disturbing the pellet. Incubation at 37^0C without disturbing the pellet allows the spheres to grow a new cell wall around them. The pellet was resuspended after 3 to 4 hours and left overnight at 37^0C to regenerate. Next day the regenerated cultures were plated onto appropriate plates containing selectable markers and if necessary, with dilutions.

2.6   General molecular techniques

2.6.1 Restriction digestions

For the restriction digest of a plasmid, procedures from New England Biolabs (NEB; MA, USA) were followed. All enzymes used were from NEB, the combinations determined using the NEB guidelines. Approximately 500 ng of plasmid DNA was used for all digests. The buffer and the enzyme concentrations were used according to the manual. Incubations were done at 37^0C unless other temperatures required for specific enzymes.

2.6.2 Dephosphorylation

The digested plasmids were dephosphorylated using alkaline phosphatase enzyme (NEB) at 37^0C for one hour. Enzyme activity was inactivated by incubating the samples at 65°C for 10 minutes. The digested and dephosphorylated DNA was then purified using DNA clean and concentrator kit (Zymo research, CA, USA) when required for downstream analysis like PCR, ligations or transformations.

2.6.3 Polymerase chain reaction (PCR)

PCR was used to amplify desired regions of a gene from a plasmid or chromosomal DNA for cloning or downstream screening purposes. All primers were designed using the web-based oligoanalyzer primer design programme (Intergrated DNA technologies). A pre-made Reddymix PCR solution was used in general for all PCR reactions (Abgene). 50 µl reaction mixtures were set up containing 45 µl of reddymix solution, 25 pmol of forward and reverse primers (2.5 µl each of 10 pmol stocks) and 50-100 ng of DNA. The reaction cycle was used as shown in Table 2.10. Reddymix employs an ultra-stable *Taq* polymerase for elongation. Hence, the pre-denaturation and the denaturation

temperatures were set to 95^{0}C as per manufacturers' guide. The annealing temperature was varied from 55^{0}C to 70^{0}C depending upon the melting temperatures of the primers. A temperature-gradient feature of the thermocycler was also employed under certain circumstances where in a number of replicates were simultaneously subjected to PCR at a specific range of temperature. The expected band sizes were calculated theoretically.

The PCR reactions employing reddymix were robust and reproducible however had a higher chance of mutation-rate because of the low-fidelity enzyme *taq* polymerase. Thus, in specific cases of cloning, Phusion High-Fidelity PCR Kit was used. This kit employs *phusion* DNA polymerase, a commercially made high-fidelity enzyme. Reactions were performed following the exact protocol by the manufacturer and as briefly outlined in Table 2.11.

Table 2.10: PCR reaction-cycle details when using PCR-reddymix

Step	Temperature	Time	Cycles
Initial denaturation	95^{0}C	1 minute	1
Denaturation	95^{0}C	30 seconds	
Annealing	x^{0}C	30 seconds	30
Extension	72^{0}C	x minutes (1min/kb)	
Final extension	72^{0}C	6-10 minutes	1

Table 2.11: PCR reaction details when using Phusion High-Fidelity PCR Kit

For a 50µl reaction mix		
Component	**Volume**	**Final Concentration**
MilliQ water	Add to make 50 µl	
5X Phusion buffer	10 µl	1X
10mM dNTPs	1 µl	200 uM
Primer A	2.5 µl (1:10 stock)	0.5 uM
Primer B	2.5 µl (1:10 stock)	0.5 uM
Template DNA	X µl	1 pg to 10 ng
DMSO (Optional)	(1.5 µl)	(1%)
Phusion DNA pol	0.5 µl	0.02 units/µl

Cycle step	2-step protocol		3-step protocol		Cycles
	Temp.	**Time**	**Temp.**	**Time**	
Initial denaturation	98^0C	30 s	98^0C	30 s	1
Denaturation	98^0C	5 - 15 s	98^0C	5 - 10 s	25-35
Annealing	-	-	X°C	10 - 30 s	
Extension	72^0C	15 -30 s/kb	72^0C	15 - 30 s/kb	
Final extension	72^0C	5-10 min	70^0C	5-10 min	1
	4^0C	hold	4^0C	hold	

Note: 2-step protocol was used when the melting temperature (Tm) for both the primers was more than 70^0C. If the Tm for both the primers was less than 70^0C, 3-step protocol was used (Primer length >20 bp : Annealing temp = Tm+3^0C of the lower Tm primer; Primer length <20bp: Annealing temp = Tm of the lower Tm primer). Tm of the primer always calculated from the online Finenzyme calculator (http://www.finnzymes.com/tm_determination.html)

2.6.4 Site-directed mutagenesis

Site directed mutagenesis was performed using QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, CA, USA) as per manufacturer's protocol. Identical primer pairs consisting of 25-45 bases were designed such that the desired point mutation is in the middle of the fragment. The length, GC content and the melting temperatures of these oligonucleotides were set according to manufacturer's conditions. When subjected to PCR, the primers hybridize to homologous regions of corresponding strands on the plasmid (Figure 2.1). The high-fidelity enzyme *pfu* DNA polymerase of the kit then extends the complete strand of circular DNA with a nick at the 5' end of the primer. At the end of the first PCR cycle, two products result, one containing the mutation and second that is identical to the parental DNA (Figure 2.1). The process is then repeated for 18 cycles producing both mutated DNA and the parental DNA (Figure 2.1). The cycle number is less as compared to normal PCR reactions to reduce the amount of wildtype DNA produced. The methylated and hemi-methylated parental DNA is digested by *Dpn1* endonuclease enzyme. The resultant product is transformed into XL-Blue competent cells (a part of the kit) and transformants are selected using the selectable marker present on the plasmid that is mutagenised.
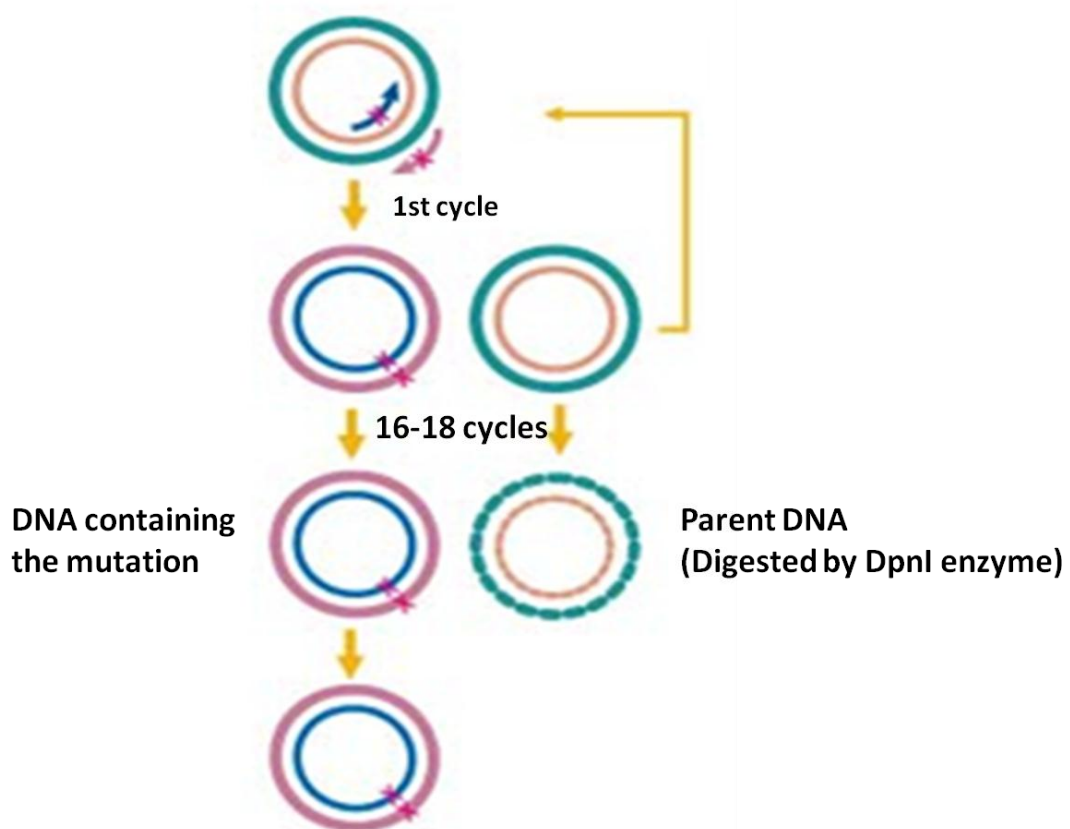
Figure 2.1
Schematic diagram to explain site-directed mutagenesis using QuikChange Kit
(Picture adapted from QuikChange mutagenesis protocol from Agilent technologies)

2.6.5 Random mutagenesis using XL1-mutator strain

Random mutagenesis was performed using XL-1 mutator strain (Agilent technologies). This *E. coli* strain is deficient in three genes of error prone mismatch repair system (*mutD5, mutS* and *mutT*) and hence, according to manufacturer, has a ~5000 fold higher mutation rate than the wildtype. The plasmid to be randomly mutagenised was transformed into XL-1 red mutator strain and grown on LB agar with appropriate selectable markers. The colonies obtained were mixed together by resuspending in 500 μl LB on the plate using a sterilized and bent glass rod and were further cultured in 100 ml of LB. After two rounds of further sub-culturing, the plasmids were isolated that now should contain random mutations. Appropriate strains were then transformed with these plasmids for selection of mutants.

2.6.6 Ligations

For ligations, the vector back-bone and the DNA insert (obtained either by digestion or PCR) were usually mixed in a molar ratio of 1:3 and treated with T4 DNA ligase enzyme along with appropriate buffer (New England Biolabs) according to manufacturer's instructions. Ligations were carried out at RT for 2 hours or overnight at 4^0C.

2.6.7 Colony PCR

To screen for colonies containing the desired insert, individual colonies were picked on pipette tips streaked onto labelled agar plates and simultaneously swirled into PCR tubes containing the PCR reaction mixture. On heating at 95^0C in the first step itself in a thermocyler, plasmid DNA is released and serves as a template for PCR reactions with primers designed to identify the desired region inserted in the vector sequence. The reaction recipe and the cycle were same as described above in section 2.6.3.

2.6.8 Agarose gel electrophoresis

DNA samples including plasmids, digests and PCR fragments were resolved on 1% or 0.8% agarose gels. SYBR safe dye (Invitrogen, CA, USA) was added while preparing the gels. Loading buffers and DNA markers employed for the electrophoresis were from Bioline (Bioline, MA, USA). Hyperladder I and lV were generally used. The gel was run in a tank containing 1X TAE (Tris-acetate EDTA buffer) at 100 V until appropriate migration of DNA was indicated by the loading dye. The bands were visualized under UV lamp and photographed.

2.6.9 Sequencing

DNA samples were sequenced by the Sanger Sequencing method at the Genomics laboratory, School of Biosciences, University of Birmingham (Sanger *et al.*, 1977). Oligonucleotides were synthesized by Alta Biosciences, University of Birmingham or Eurogentec, GBR. The samples were prepared by mixing 100 to 200 ng of plasmid DNA or 1 to 10 ng of PCR fragments with 3.2 pmol of primer in a total volume of 10 μl adjusted with water.

2.7 Complementation analysis (spot test)

The required strains were grown overnight at 37^0C. The cultures were adjusted to an OD_{600} of 1 and serially diluted from neat to a dilution of 10^{-6} or 10^{-7}. A maximum of 10 μl of each dilution was then spotted onto a labelled square plate and allowed to dry. Once the spots had dried, the plated were sealed with parafilm and placed at various temperatures (usually 26^0C, 30^0C, and 37^0C) to test for complementation at those temperatures.

2.8 Protein Analysis

2.8.1 SDS gel electrophoresis and staining

Reagents required:

Resuspension buffer A	30 mM Tris-HCl pH 7.2, 50 mM NaCl and 5 mM $MgCl_2$
4X SDS loading dye	0.25 M pH 6.8 TRIS/HCl, 8% SDS, 40% glycerol, 1% Coomassie brilliant blue R250 and 3% β-mercaptoethanol added fresh
1X SDS Running buffer	25 mM Tris pH 8.3, 192 mM glycine and 0.1% SDS
Coomassie staining solution	0.25g of Coomassie brilliant blue, 45% methanol, 10% glacial acetic acid in DDW.
Destaining solution	30% methanol, 10% glacial acetic acid in DDW

Preparation of samples:

From E. coli:

1 ml of log phase or early stationary phase cell culture (in case of induction, cultures induced with IPTG for at least 4 hours) was centrifuged at 13,000 rpm for 3 minutes. The cell pellet was then resuspended in 150-200 µl of resuspension buffer A. These cell suspensions could be stored in the -20^0C freezer for 2 to 3 days without having any effect on the resultant protein expression analysed by SDS-PAGE. However if these samples were further to be used for isolation of the soluble fractions for Native-PAGE, long term storage is not recommended. For running SDS-PAGE, 10 µl from the resuspended mix was drawn into a new eppendorf already containing 2.5 µl of SDS-

loading dye. The samples were heat inactivated at 60°C for ten minutes on a heat block before loading onto the gel.

From H. volcanii:

1 ml of log phase culture (0.8 OD) was centrifuged for 5 minutes at 13,000 rpm. The pellet was resuspended in 1 ml of 2 M NaCl containing 50 mM EDTA and incubated for 15 minutes at RT. The resuspended cells were sonicated 3 times with 10 seconds bursts and again incubated at RT for 10 minutes. The soluble fraction was harvested by spinning the samples at 13,000 rpm for 10 minutes and the proteins were precipitated using 30% TCA solution. Following incubation on ice for 10 minutes, the precipitated proteins were then isolated by spinning at 13,000 rpm for 20 minutes at 4°C. The proteins were observed as a tiny white pellet. This pellet was washed thrice with double distilled water (DDW) and then SDS-loading dye was added. The samples were boiled for 10 minutes and placed on ice for 2 minutes before loading.

SDS-PAGE:

The samples were subjected to sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) using 10% Tris-glycine SDS-polyacrylamide gels prepared according to the recipe shown in Table 2.12. The gels were run in a Bio-Rad apparatus using SDS running buffer. Approximately 10-15 µl of protein samples were loaded onto separate wells along with protein markers (Figure2.2; PageRuler prestained protein ladder, Fermentas). The gels were run at 120 V for first 10 minutes (for stacking) and further at 180 V for about 1 hour (visually seen by the successful run of the protein ladder).

Detection: After completion of electrophoresis, gels were stained using the staining solution for about 1 hour with shaking (HS250 Basic, IKA, Germany). The stained gels were then destained using the destaining solution for 20-30 minutes, again with shaking. Once satisfactory resolution of proteins was obtained, the gels were kept in water until dried. For storage, the gels were air-dried by sandwiching them between two cellophane sheets using the GelAir-Drying accessories from Bio-Rad.

Silver staining: When higher sensitivity was desired for detection of low concentration proteins, silver-staining method was used. The SilverQuest™ Kit for Mass spectrometry-compatible silver staining from Invitrogen was employed for the same. The fast-staining protocol was exactly followed as given in the manufacturer's guidelines.

Table 2.12: Recipe for preparing SDS-PAGE gel
(Equivalent for making two gels in a Bio-Rad apparatus)

10% Resolving gel (10 ml)	Ingredient	5% Stacking gel (5 ml)
4 ml	DDW	3.4 ml
3.3 ml	30% acrylamide mix	830 µl
2.5 ml	(pH 8.8) 1.5M TRIS (pH 6.8)	630 µl
100 µl	10% SDS	50 µl
100 µl	10% Ammonium persulphate	50 µl
4 µl	TEMED	5 µl
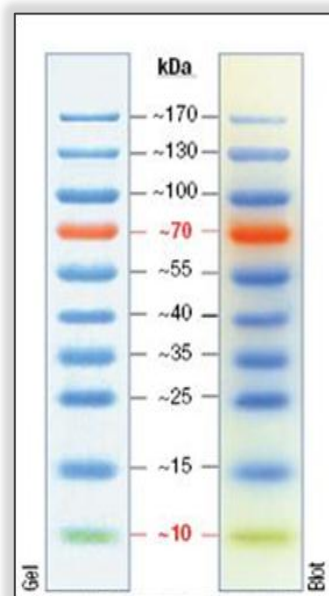
Figure 2.2 PageRuler prestained protein ladder (Fermentas)

2.8.2 Native gel electrophoresis and staining

To test the assembly of large oligomeric chaperonins either linear or gradient native polyacrylamide gel electrophoresis (native-PAGE) techniques were used.

2.8.2.1 Native-PAGE

Reagents required:

Resuspension buffer B	50 mM Tris HCl (pH 7.5), 5 mM EDTA (pH 8.0), 100 mM KCl, 500 µg/ml lysozyme (Stratagene), 100 µg/ml DNAse (Stratagene) and 50 µg/ml of protease inhibitor cocktail (Roche)
4X Native gel loading dye	0.25 M pH 6.8 TRIS/HCl, 80% glycerol, 0.04% bromophenol blue
1X Native gel running buffer	25 mM Tris pH 8.3, 192 mM glycine

Preparation of samples:

3 ml of overnight culture was harvested at 13000 rpm for 5 minutes and resuspended in 350 µl of resuspension buffer B. The solution was allowed to stand for 10 minutes at RT and then sonicated 3 times with 10 second bursts on ice. After incubation at RT for a further 10 minutes, the solution was spun at 13000 rpm and 4^0C for 10 minutes. The supernatant was transferred into a new eppendorf tube containing 100 µl of native loading dye to obtain the sample preparation for Native-PAGE.

Native-PAGE

The samples were loaded on native gels prepared using the recipe shown in Table 2.13. Electrophoresis was carried out at 220 V for the first hour and then at 250 V for next 6

hours (to resolve the ~900kDa archaeal chaperonins) using native gel running buffer. The gel tank was placed in ice to avoid heating up due to long gel runs. Detection was done in the same way as described above using the Coomassie blue staining method.

2.8.2.2 Gradient Native gel electrophoresis

Gradient Native gel electrophoresis provided a better resolution of high molecular weight proteins and was quicker to use.

Reagents required:

Freeze thaw buffer (FT buffer)	50 mM Tris-HCl (pH 7.5), 5 mM EDTA (pH 8.0), 5 mg lysozyme, 1 mg of DNAse and 50 µg/ml of protease inhibitor cocktail
4X Native gel sample buffer	400 mM MOPS/KOH (pH 7.2), 50% glycerol, 0.05% bromophenol blue
1X Native gel running buffer	100 mM MOPS/KOH (pH 7.2)

Preparation of samples:

Cell cultures with OD_{600} equivalent to 2 were harvested by spinning at 13000 rpm for 5 minutes. The pellets were resuspended in 100 µl of FT buffer and incubated on ice for 5 minutes. The resulting cell resuspensions were frozen in a -25^{0}C freezer (usually takes 7 – 10 minutes) and thawed at RT. One cycle of freezing and thawing proved both necessary and sufficient to disrupt the cells and at the same time to protect the integrity of large protein structures. The soluble fraction containing proteins was obtained by centrifuging the thawed samples at 13000 rpm for 10 minutes at 4^{0}C. 10 µl of the

soluble fractions were mixed with 4 µl of native gel sample buffer and loaded onto the gel.

Gradient Native-PAGE:

The gradient native gels were prepared using 3% and 10% acrylamide gel solutions using the recipe shown in Table 2.14. Firstly the 10% gel solution was prepared and poured into the gel cassette to make approximately 1 cm layer. The layer was levelled by overlaying with water and was allowed to set. The remaining 10% gel solution was placed on ice to delay polymerization. Subsequently 3% gel solution was made on ice. After the 10% gel had set the excess water was removed by decanting. To introduce a gradient, firstly 2.6 ml of 3% gel solution was drawn in a 5 ml glass pipette followed by drawing of 2.6 ml of 10% gel solution. An air bubble was introduced in the pipette by suctioning the pipettor in air. While traversing towards the top, the air bubble develops a 3 to 10% gel gradient in the pipette. This gel solution was now poured on the top of the already set 10% gel. Combs were introduced and the gel was allowed to set after overlaying with excess 3% gel solution (usually took around 30 minutes). Once the gel had set, the gel cassettes were soaked in a wet tissue and covered in silver foil to store them for up to two weeks (4^0C) or the gels were directly used for Native-PAGE. Electrophoresis was carried out at a constant of 80 V for four hours at RT. The resolved gels were stained and destained by Coomassie staining method as discussed earlier.

Table 2.13: Recipe for preparing a native (non-denaturing) gel
(Equivalent for making two gels in a BioRad apparatus)

7.5 % Resolving gel (10 ml)	Ingredient	5 % Stacking gel (5 ml)
2.8 ml	DDW	2.5 ml
2.5 ml	40% acrylamide-bisacrylamide mix	850 µl
2.5 ml	(pH 8.8) 1.5 M TRIS (pH 6.8)	625 µl
2.0 ml	80% glycerol	1.25 ml
100 µl	10% ammonium persulphate	50 µl
8 µl	TEMED	5 µl

Table 2.14: Recipe for preparing a gradient native gel
(Equivalent for making six gels in a Bio-Rad apparatus)

10% gel (15 ml)	Ingredient	3% gel (15 ml)
7 ml	DDW	10.5 ml
3.75 ml	40% acrylamide-bisacrylamide mix	1.5 ml
450 µl	2% bis-acrylamide	-
3 ml	0.5 M MOPS/KOH (pH 7.2)	3 ml
75 µl	10% ammonium persulphate	75 µl
7.5 µl	TEMED	7.5 µl

2.8.3 Western blotting

For detecting specific proteins in samples, western blotting was employed. For this technique, an SDS PAGE was run as described before. The gel thus containing the resolved proteins was transferred on a membrane for blotting of desired proteins using antibodies specific to that proteins in a process as described below.

Transfer: The gel once obtained after the SDS-PAGE run, was taken out and immersed in pre-chilled transfer buffer (10mM CAPS and 10% methanol in DDW, pH adjusted to 11 using conc. NaOH). 4 strips of whatmann filter paper and two fiber pads along with a PVDF membrane cut to the gel size (Immobilon P-0.45 μm, Sigma) were also immersed into transfer buffer for at least 5 minutes. Soaking the membrane in methanol for ~5 minutes prior to immersing in the transfer buffer increased the transfer efficiency. A sandwich of these components was assembled in the following order (from anode to cathode): Fibre pad, two strips of filter paper, gel, PVDF membrane, two strips of filter paper, and a fibre pad. The sandwich was placed in a Bio-Rad blotting apparatus with an ice-pack and immersed in a tank filled with pre-chilled transfer buffer. The entire assembly was placed in a container of ice and run for an hour at a constant current of 300mA.

Blocking: The membrane now containing the proteins was washed with 1X TPBS (Phosphate Buffer saline; 2g KH_2PO_4, 14.48g NaH_2PO_4, 80g NaCl and 2g of KCl in 1000ml of DDW to make 10X PBS; 0.1% of Tween20 in 1X PBS to make 1X TPBS) and incubated in blocking solution (4% dried milk powder in 1X TPBS) overnight with gentle shaking.

Blotting: The next day the blocking solution was poured off and the membrane was vigorously washed in TPBS (3 times over a period of 15 minutes). Next, the membrane

was incubated in the primary antibody (1:1000 dilution of anti-thermosome antibody in blocking solution, with 0.05% sodium azide; provided by Dr G. Bosch, Baumeister lab) for an hour with shaking. Further, the membrane was again washed with TPBS (3 times over a period of 15 minutes). The secondary antibody solution (1:50000 dilution of anti-rabbit HRP conjugated IgG in blocking solution; GE healthcare) was then added to the membrane and incubated for an hour with gentle shaking. This was followed by washing the membrane twice with TPBS over a period of 15 minutes, the process culminating in one last vigorous wash with TBS (Tris buffer saline; 10mM Tris-HCl pH 7.5 and 0.9% NaCl). The membrane was dried by pressing gently on a tissue and further subjected to detection.

Detection: The ECL plus kit (Amersham) was used for the detection. The solutions A and B were mixed in 1:1 ratio to make 5 ml of volume and spread over the membrane for 1 minute using a Pasteur pipette. The reagent was removed by again drying the membrane on tissue. Next, the membrane was covered in cling film and placed in a film cassette and taken to the dark room. It was exposed to the photo-sensitive film (Kodak Biomax, Sigma) for 1 to 5 minutes and the film was then developed using the XO-Graph machine using manufacturer's instructions.

2.8.4 Protein purification

Reagents required:

Lysis buffer	30 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, 15% ethanol (v/v), 1 mM PMSF, 1 mM protease inhibitor cocktail (Roche), small amount of DNAase
Column buffer A (Anion exchange)	30 mM Tris-HCl (pH 7.5), 5 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT
Column buffer B (Anion exchange)	30 mM Tris-HCl (pH 7.5), 1 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT
Column buffer C (Gel sizing)	30 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT

Method:

Cultures of *E. coli* BL21 cells carrying the pET plasmid-constructs of MmCCT gene or its mutants were grown from an overnight culture in LB media at 37^0C until OD_{600} of 0.4 was achieved. Protein expression was induced by addition of 1 mM IPTG and the cultures were further grown for four hours. Subsequently, the cells were harvested by centrifugation for 10 minutes at 4^0C and 4000 rpm in a Sorvall SLC-4000 rotor. The pellets at this stage could be stored in a -70^0C freezer. These pellets were then resuspended in lysis buffer and the cells were lysed by three passages through a French pressure cell. The lysates were centrifuged for 20 minutes at 4^0C and 16000 rpm in Sorvall SS-34 rotor and the supernatant was filtered through a 0.2 μ filter. The clear supernatant was now applied to a Q high performance (QHP) self packed anion exchange column (40 ml bed volume) pre-equilibrated with column buffer A. Elution was achieved with a linear gradient of 0-50% of column buffer B over 200 ml with a flow

rate of 2 ml/min and a fraction size of 4 ml. 10 µl of the appropriate fractions as seen from the peaks obtained on the resultant chromatogram were resolved by 12% SDS-PAGE to gradient Native-PAGE. Running a Native-PAGE gel at this stage was essential to differentiate between the fractions that have MmCCT and the ones that have GroEL, as MmCCT and GroEL single subunits share similar molecular weights that cannot be resolved by SDS-PAGE. The fractions containing MmCCT were pooled and diluted 4 fold into column buffer A and applied to a MonoQ 16/10 anion exchange column (10 ml bed volume) (Pharmacia Biotech, Sweden). Elution was achieved in a similar manner as described above and the relevant fractions were identified by running on 12% SDS-PAGE. These fractions were pooled and precipitated with 70% ammonium sulphate, resuspended in column C buffer and applied to S300 HiPrep 26/60 column (Pharmacia, Sweden) pre-equilbrated with buffer C. The eluates were obtained as a result of gel-sizing. The flow rate was adjusted to 1 ml/min and the fraction size to 6 ml. Again the relevant fractions as checked by 12% SDS-PAGE and gradient Native-PAGE were pooled and concentrated up to 10 mg/ml using Amicon Ultra-15 Centrifuge devices. Protein concentration was determined with a NanoDrop Spectrophotometer ND-1000 (Peqlab, Erlangen, BRD) using the specific extinction co-efficient of MmCCT or MmCCT mutants. The pure preparations were supplemented with 10% glycerol and stored at -70^{0}C.

2.8.5 CBP-tagging and pull downs

2.8.5.1 CBP-tagging by overlap extension PCR

A streptavidin-cbp-his tag was used for tagging. The sequence for the same is GSGSGWSHPQFEKGSGKRRWKKNFIAVSAANRFKKISSSGALGSGHHHHHHHHGSG (Pappenberger *et al.*, 2006) where red letters represent streptavidin tag, brown letters represent cbp tag and green letters represent his tag. The sequence was reverse translated by Sequence Manipulation Suite Copyright © 2000, 2004 Paul Stothard, using codon table from *E. coli* B strain. The tag was incorporated using overlap extension PCR (as described in Figure 5.5, Chapter 5) with primers specifically designed such that they contain the entire tag sequence with overlapping ends (Table 2.5A, underlined). The first two PCR steps for this technique were carried as described in Table 2.10 however with a lesser time period for denaturation (20 sec). For the final overlap extension PCR step the cycle remained the same, however the primers were added at the end of the fifth cycle instead of in the initial mix, to allow the annealing of the tag overlapping region (Figure 5.5, Chapter 5).

2.8.5.2 Pull downs of CBP-tagged MmCCT and its interactors

The protocol for pull downs of CBP tagged MmCCT was developed with trial experiments first and was finalized and described as below. The reagents and their concentrations are noted in Table 2.15.

- The TAB21 (BL21 strain containing GroEL under pBAD promoter) cells expressing cbp-tagged MmCCT-wt and/or mutants under pET21 vector were grown in a 200 ml LB culture containing 0.2% arabinose, 50 μg/ml Kan (for TAB21 selection), 100 μg/ml Amp (for pET21 selection) and 50 μg/ml Chloromphenicol (for pACYC

selection). Protein expression was induced by addition of 1 mM IPTG when the cultures reached an OD_{600} of 0.4. The cells were further grown for 4 hours at 37^0C.

- The cells were harvested by centrifugation (SLC-4000 rotor: 6000 rpm at 4^0C for 10 min or Eppendorf 5504R: 5000 rpm at 4^0C for 10 min). The pellets could be stored at least for a week in -70^0C freezer at this stage.

- The pellets were resuspended in 6 ml of lysis buffer. DNAase and PMSF (phenylmethanesulfonylfluoride) were added separately after even resuspension. Lysis was carried out either by passing the resuspensions through a French pressure cell at 1600 psi for three rounds or by sonication with 5 rounds of 30 sec bursts and 30 second pauses. Protease inhibitor solution was added at the end of the first of round in both the cases.

- The soluble fractions containing proteins were obtained by centrifugation (SS-34 rotor: 15000 rpm at 4^0C for 20 min or Eppendorf 5504R: 11000 rpm at 4^0C for 20 min). The step was repeated to remove additional debris.

- **Calmodulin resin equilibration:** The settled calmodulin resin (Agilent Tech. cat no. 214303) was gently mixed to suspend evenly. 50 μl of the resin was drawn in a 1.5-ml eppendorf and the pellet containing the beads was obtained by spinning at 1000 rpm for 2 mins in a microcentrifuge. The storage buffer was removed carefully. The calmodulin affinity resin was equilibrated in $CaCl_2$-containing equilibration buffer. 200 μl of equilibration buffer was added to the resin and mixed gently. This was followed by spinning at 1000 rpm for 2 minutes and removal of the buffer with the help if a pipette. The step was repeated four times to ensure correct equilibration of the resin beads.

- Additional $CaCl_2$ to achieve a final concentration of 5 mM was added to 800 μl of cleared lysate. This was now mixed with equilibrated calmodulin resin beads and incubated overnight at 4^0C on a mechanical rotator.

- The reaction mix was centrifuged (1500 rpm, 3 minutes, 4^0C) to separate the beads with the bound protein from the unbound material. The unbound supernatant was retained for analyzing on SDS-PAGE.

- The resin beads were washed three times with 300 μl of equilibration buffer and three times with 300 μl of wash buffer (Washing with wash buffer containing lower amounts of $CaCl_2$ prepares for the next 10mM EDTA step). The bound proteins were eluted by 100 μl of elution buffer. 10 μl of all the final protein preparation along with wash samples, unbound sample and cleared lysate (diluted ten-fold) were analysed by SDS-PAGE and Native-PAGE.

Table 2.15: Constituents of CBP tagging and pulldowns

Constituent	Stock Concentration	Working concentration	Volume (For 100 ml)
Lysis buffer			
Hepes pH 7.2	1 M (Filter sterilize; Fridge)	300 mM	30 ml
NaCl	2 M	200 mM	10 ml
Glycerol	80 % (Autoclaved)	35 %	43 ml
EDTA pH 8	0.5 M	1 mM	200 µl
PMSF	0.1 M (Solvent: Ethanol; Store -20^{0}C)	1mM	Directly to lysate (For e.g. 100 µl per 10 ml)
Protease Inhibitor (EDTA-free Roche)	Tablet	1 tablet/10ml	
Equilibration buffer			
Hepes pH 7.2	1 M	20 mM	2 ml
NaCl	2 M	1 mM	50 ml
Glycerol	80 %	20 %	25 ml
DTT	1 M (Store -20^{0}C)	2mM	200 µl
Triton-X100	100 %	0.01 %	10 µl
CaCl$_2$	0.1 M (Autoclaved)	2mM	2 ml
Wash buffer			
Hepes pH 7.2	1 M	20 mM	2 ml
NaCl	2 M	1 mM	50 ml
Glycerol	80 %	20 %	25 ml
DTT	1M	2mM	200 µl
Triton-X100	100 %	0.01 %	10 µl
CaCl$_2$	0.1 M	0.2 mM	200 µl
Elution buffer			
Hepes pH 7.2	1 M	20 mM	2 ml
NaCl	2 M	1 mM	50 ml
Glycerol	80 %	20 %	25 ml
DTT	1 M	2 mM	200 µl
Triton-X100	100 %	0.01 %	10 µl
EDTA pH 8	0.5 M	10 mM	2 ml

2.8.6   Chaperonin assays

2.8.6.1 Citrate synthase aggregation assay

Citrate synthase (CS) is a thermolabile protein that denatures at temperatures above 42^0C (Minuth *et al.*, 1998).   The folding intermediates formed after denaturation aggregate due to non-specific hydrophobic interactions.   The aggregated particles scatter light that allows the process to be followed by spectrophotometric measurement at 320 nm. Chaperonins are able to recognize and bind folding intermediates in turn minimizing the local concentration of folding intermediates in solution.   Consequently the binding ability of chaperonins can be measured by suppression of light scattering.

Reagents:
- **CS assay buffer:** 30 mM MOPS-KOH (pH 7.5), 300 mM KCl, 5 mM $MgCl_2$

- **CS enzyme (from porcine heart, Sigma Catalogue No. C3260):** 1 μM

Method:
0.5 μM of MmCCT protein was mixed in 1 ml of CS assay buffer containing 0.5 μM of CS. The molecular weight of MmCCT was considered as 931KDa to obtain 0.5 μM concentration. The reaction mixture was incubated in a pre-heated water-bath at 50^0C. The extent of aggregation was followed by measuring the light scattering at 320 nm over a period of 30 minutes. Assay buffer containing only CS and no protein was used as a positive control.

2.8.6.2 Malachite green ATPase assay

The ATP hydrolysis activity of chaperonins can be determined by measuring the liberated P_i by malachite green assay. The assay is based on the principle that phosphate reacts with molybdate and forms a green complex with malachite green that can be detected photometrically at 640 nm.

Reagents:

- **ATPase assay buffer:** 30 mM Tris-HCl (pH 7.5), 300 mM KCl, 5 mM $MgCl_2$, 1 mM DTT

- **Malachite green (MG) solution:** 45 gm of MG (Sigma, Catalogue No. 213020) and 1.09 gm of ammonium molybdate (Sigma, Catalogue No. A7303) were mixed in 100 ml 1 M HCl by stirring for 1 hr at RT. Further, 1% Tween-20 was added and stirring was continued overnight at 4^0C. The solution was centrifuged for 10 minutes at 4000 rpm and 4^0C to obtain a cleared MG solution that was stored at 4^0C.

- **1 mM ATP:** 1 ml 100 mM stock solution of ATP (disodium salt, bacterial source, Sigma) was prepared in DDW and 100 µl aliquots were stored at -20^0C. The ATP solution was diluted in a ratio of 1:1 before use. 1 µl of the 1:1 diluted solution was added to a 50 µl reaction mixture to obtain a final concentration of 1 mM.

- **10 mM EDTA:** 1 µl of 0.5 M EDTA stock solution was added to 50 µl of reaction mixture.

- **34% citric acid:** 34 gms of citric acid (Fisher, Catalogue No. 624053) was mixed in 100 ml of water and stored at 4^0C.

- **1 to 5 nM of KH_2PO_4:** A 100 mM stock solution of KH_2PO_4 (stored at -20^0C) was freshly diluted 100 fold to obtain 1 mM. This solution was further diluted to obtain nanomolar concentrations.

Method:
10-15 µg of protein and 1 mM ATP were mixed in ATPase assay buffer (50 µl reaction volume) and incubated at 37^0C on a heat block for 15 minutes for nucleotide hydrolysis reaction. The reaction was stopped by addition of 10 mM EDTA followed by a short-spin in a microcentrifuge. 800 µl of MG solution was then added into the reaction mix and the resulting solution was incubated at RT for 10 minutes. The colour formation was

stopped by addition of 100 μl of 34% citric acid. The intensity of the colour was measured photometrically at 640 nm. The reaction mix without proteins was used as a negative control to determine the amount of ATP self-hydrolysis. The OD_{640} measured for the test sample was correlated with the concentration of released P_i by a calibration curve. For the calibration curve 1 to 5 nmol of KH_2PO_4 were mixed with 50 μl of ATPase assay buffer, 800 ul of MG solution was added followed by addition of 34% citric acid after 10 min. The colour intensities were measured at 640 nm and plotted against Pi concentration to obtain the calibration curve.

2.8.6.3 Rhodanese refolding assay

Rhodanese is a 33 kDa mitochondrial sulphur-transferase enzyme that catalyzes the conversion of toxic cyanide into thiocynate which in turn can conjugate with ferric ions to form a red-orange complex. The amount of active rhodanese can be assayed by photometric estimation of this complex. If rhodanese is subjected to chemical denaturation, it loses its ability to refold spontaneously. Chaperonins can act upon the denatured rhodanese and refold it and the extent of refolding can be determined by photometric assay described above. Thus the enzyme is used as a model system to measure the refolding ability of chaperonins.

$S_2O_3^-$ + CN^- → SCN^- + SO_3^{2-} → $Fe^{3+}(SCN)_3$ (red-orange complex)
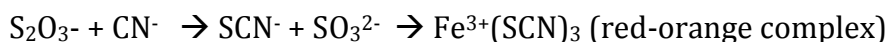
Reagents:
- **Rhodanese enzyme solution:** 25 mM MOPS (pH 7.2), 10 mM KCl, 20 mM $Na_2S_2O_3$ or $K_2S_2O_3$ and 3% glycerol.

- **Rhodanese enzyme:** 10 mg/ml aliquots of Rhodanese enzyme (Sigma Catalogue No. R1756) were prepared in rhodanese enzyme solution and stored at -20^0C.

- **Guanidinium HCl (GuHCl):** 6 M solution of GuHCl was prepared in 20 mM MOPS (pH 7.2). This solution was always prepared fresh.

- **Rhodanese assay buffer:** 20 mM MOPS (pH 7.2), 150 mM KCl, 5 mM MgCl₂, 5 mM DTT, 20 mM $Na_2S_2O_3$ or $K_2S_2O_3$.

- **Fe(NO₃).9H₂O solution:** 1 gm of $Fe(NO_3).9H_2O$ (Sigma) was added to 10 ml of H_2O. 2 ml of 65% HNO_3 was further added and mixed. The final volume was adjusted to 15 ml and the solution was stored at 4^0C.

Method:

The process was divided into following steps:

- **Rhodanese denaturation:** 2 μl of rhodanese was mixed with 14 μl of 6 M GuHCl and 3 mM DTT and incubated for 30 minutes at RT.

- **Binding of rhodanese to protein:** 35 μg of protein (1-10 μl volume) was mixed with 55 μl of rhodanese assay buffer and 0.8 μl of denatured rhodanese. The reaction mix was incubated at 25^0C for 15 minutes. A negative control containing the denatured rhodanese and buffer only was also simultaneously added to the experiment to determine the amount of spontaneous refolding.

- **Refolding of rhodanese:** 1 mM ATP was added and the samples were incubated at RT for 30 minutes. The refolding was stopped by adding 10 mM EDTA and allowing the samples to stand at RT for 5 minutes.

- **Rhodanese enzymatic reaction:** The samples were diluted by adding 120 μl of buffer. The enzymatic reaction was initiated by addition of 30 μl of 1:1:1 solution of KCN, $Na_2S_2O_3$ or $K_2S_2O_3$ and KH_2PO_4 and incubation at 30^0C for 10 minutes. This reaction was stopped by adding 15% folmaldehyde. 300 μl of $Fe(NO_3)$ solution was

then added to the reaction mixture followed by an incubation at RT for 20 minutes. The intensity of the resultant coloured complex was measured at 420 nm.

2.8.6.4 Proteinase K sensitivity assay

Purified MmCCT proteins and its mutants were respectively incubated in 50 µl of assay buffer (30 mM MOPS-KOH (pH 7.5), 300 mM KCl, 5 mM $MgCl_2$, 5 mM $MgCl_2$ and 1 mM DTT). The reaction was started by addition of 1 mM ATP or ADP. In case where metal fluoride complexes were to be generated, 5 mM of $Al(NO_3)_3$ and 30 mM KF (all dissolved in H_2O) were added to the mix before the addition of ATP or ADP. Next, 1 µg/ml Proteinase K was added and the samples were incubated for 15 min at 30^0C. 5 mM PMSF (100% v/v in ethanol) was added to inhibit protease activity and the samples were immediately mixed with SDS loading dye to be analyzed by 12% SDS PAGE.

2.9   Mass spectrometry (MS)

MS analysis was carried out at the Advanced Mass Spectrometry facility at School of Biosciences, University of Birmingham. The samples were prepared by cutting out specific bands of proteins resolved on a Coomassie or silver stained Native-PAGE gel. In a case where multiple proteins were to be analyzed, the bands were cut vertically to accommodate more proteins as shown in Figure 2.3. The slicing of gels was carried out in a sterile environment and the gel fragments were stored in -20^0C freezer before handing over to the facility for in-gel trypsin digestion and further MS analysis.

The method employed at the facility has been briefly outlined below: The gel fragments are further cut into as small as 2 mm^3 cubes and destained using organic solvent acetonitrile and buffer salt ammonium bicarbonate. Gel pieces are then reduced at 60^0

for 15 min in ammonium bicarbonate solution containing DTT to achieve unfolding of tertiary structures. The unstable –SH groups thus formed are alkylated and converted into stable cysteine derivatives by incubating with iodoacetamide in dark. In-gel trypsin digestion (20 μg of trypsin) is then performed using shaking for 30 minutes and a further overnight incubation after dilution with ammonium bicarbonate. The peptides are extracted using acetonitrile and formic acid, dried and further resuspended in a 0.1% formic acid and water solution to concentrate and separate using HPLC (Dionex, Sunnyvale, CA, USA) (linear gradient of 0.1% formic acid in acetonitrile solution). The peptides are eluted directly into the LTQ Orbitrap Velos ETD mass spectrometer (ThermoFisher Scientific, Germany) via a Triversa Nanomate nanospray device (Advion Biosciences, NY). LTQ-Orbitrap-Velos uses enhanced electrospray ionization for ion sourcing, dual pressure cells and quadruple mass filter for linear ion trapping (LTQ) and Orbitrap as the mass analyzer (Olsen *et al.*, 2009). Because protein modifications were not to be studied here, collision induced dissociation (CID) was used for product-ion dissociation. The obtained spectra was evaluated using MASCOT software (Perkins *et al.*, 1999) with standard MudPIT scoring against BL21 (DE3) database.
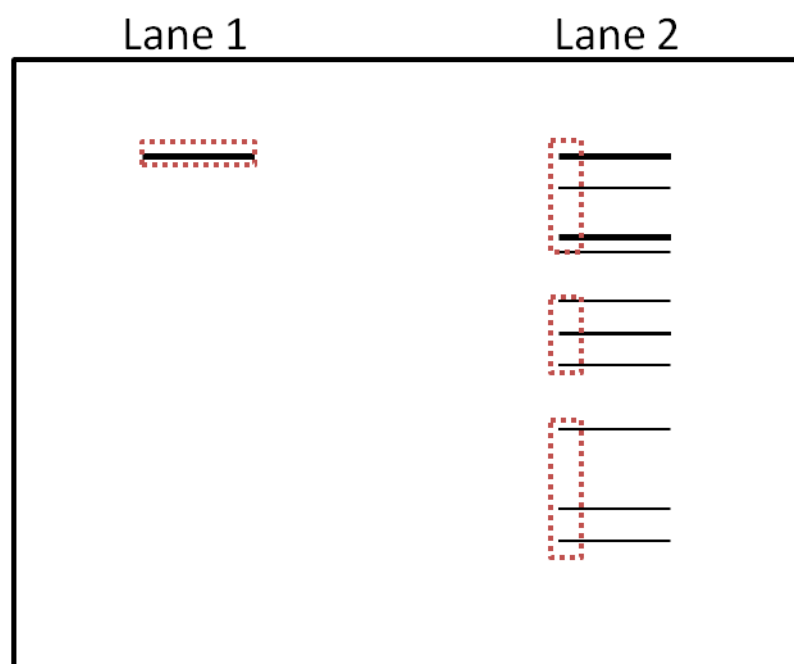
Figure 2.3
A schematic diagram to show how band excision for MS analysis was done.

CHAPTER 3
Can an archaeal chaperonin replace GroEL function in *E. coli*?

3.1 Background

The chaperonins function to fold a wide variety of client proteins starting from very simple prokaryotic proteins to components of large eukaryotic protein complexes. As discussed in Chapter 1, Group 1 and Group 2 chaperonins share similar structures however there are significant structural and functional differences. These differences have been occasionally attributed to the adaptation of chaperonins to their clients in respective host organisms. For example, the transition from homo-oligomeric chaperonin rings in Group 1 to a mostly hetero-oligomeric distribution in Group 2 has been suggested to have occurred in parallel with the evolution of complex proteins in higher organisms (Kubota *et al.*, 1994; Dekker *et al.*, 2011b). Indeed, Group 2 chaperonins have been reported to interact with faster-evolving proteins with greater structural variety, in contrast to comparatively more-conserved, slower-evolving clients for Group 1 chaperonins (Kerner *et al.*, 2005; Hirtreiter *et al.*, 2009). The intra-ring allosteric shift of ATP hydrolysis from *concerted* in Group 1 chaperonins to *sequential* in Group 2 has been implicated in ordered release of for instance, multi-domain proteins bound specifically to selected subunits in contrast to simultaneous release from all subunits in Group 1 chaperonins (Horovitz and Willison, 2005; Horwich *et al.*, 2007). The varying surface properties of the chaperonin cage in Group 1 and Group 2 chaperonins have also been proposed to be associated with differential client specificity (Tang *et al.*, 2006; Cong *et al.*, 2010).

A study has indeed shown that GroEL from *E. coli* or its mitochondrial homologue, is unable to fold eukaryotic TRiC clients, actin and tubulin, *in vitro* (Tian et al., 1995). Conversely it has also been reported that bacterial proteins that require GroEL

assistance are unable to fold in the eukaryotic cytosol (Kerner et al., 2005). Moreover in the archaeon *Methanosarcina mazei*, where both Group 1 and Group 2 chaperonins co-exist, a non-overlapping set of clients that exclusively interact with either of the groups was found in addition to a group of common clients (Hirtreiter *et al.*, 2009). All of these studies propose client-optimization to be one of the drivers of divergence between Group 1 and Group 2 chaperonins. However, substantial biochemical data is still needed to support this hypothesis. Moreover, the precise nature of such deviation between the groups at a structural level and its contribution to changes in the mechanisms of action is not fully resolved, one of the limitations being the higher degree of structural complexity in most Group 2 chaperonins and difficult *in vivo* analysis in archaeal and eukaryal expression systems.

On the other hand, it may well also be considered that over 80% of the TRiC interactors have been shown to be recognized by GroEL (albeit not folded) (Yam *et al.*, 2008), and that, GroEL could at least bind and release, if not fold, obligate TRiC-clients, actin and tubulin (Tian *et al.*, 1995). Such results implicate a broader range of specificity probably reflecting the common elements for client recognition or binding in Group 1 and Group 2 chaperonins. It can be speculated hence that the apparent client specificity of Group 2 chaperonins is more pronounced at a protein-folding level than client recognition. Moreover, while the subunit-specific functionalization is complete in eukaryotic chaperonins, archaeal chaperonins do not present absolute sub-functionalization (Ruano-Rubio and Fares, 2007). It is plausible to propose thus, that the Group 1 and Group 2 chaperonins may share a certain degree of functional overlap that is more likely to be prominent in simple thermosomes than for complex eukaryotic chaperonins.

We thought that if a Group 2 archaeal chaperonin in its native form or its improved functional mutants can function in a Group 1 host in complete absence of its native Group 1 chaperonin, the system hence generated will help the understanding of the divergence between the groups and will also allow genetic dissection of properties of Group 2 chaperonins using simple Group 1 hosts like *E. coli.*

Using archaeal chaperonins as Group 2 representatives and GroEL from *E. coli* as the default Group 1 chaperonin, we proposed a chaperonin swap study to test whether an archaeal chaperonin can replace the functioning of GroEL in *E. coli*.

Archaeal chaperonin from an obligate mesophile, *Methanococcus maripaludis* **(MmCCT)** was selected as it has following properties: 1) The MmCCT complex has identical subunits and hence is a simpler system to use. For hetero-oligomeric chaperonins, assembly into authentic particles with correct subunit combination remains a concern. Also, the homo-oligomeric property of MmCCT facilitates recombinant expression and mutagenesis studies in bacterial hosts. 2) The optimum temperature of growth for this mesophile is 35^0C-40^0C i.e. the chaperonin should exhibit full functionality at bacterial growth temperature optima. 3) This chaperonin has already been recombinantly expressed in *E. coli* and characterized for ATPase and *in vitro* folding properties (Kusmierczyk and Martin, 2003b). The system used for testing the hypothesis is outlined below.

3.1.1 Experimental Design: TAB21 expression system

If GroEL expression is repressed in a strain of *E. coli*, growth ceases, as GroEL is essential for survival. When such a strain is complemented by another chaperonin, growth will only occur if this chaperonin can functionally replace GroEL. Thus, a repressible

expression system, where the native *groESL* operon can be switched on and off, was established (McLennan and Masters, 1998). In the BL21 (DE3) strain of *E. coli,* the native *groESL* operon in *E. coli* was swapped with externally introduced *groESL* operon under the control of repressible/inducible *pBAD* promoter along with a Kanamycin resistance (KAN[R]) cassette. This gave rise to BL21 (DE3) strain containing *pBAD-*groESL-KAN[R] called as **TAB21** (Table 2.1) (Figure 3.1B). The *pBAD* promoter is an arabinose induced and a glucose repressed promoter. Hence the cells grow only when arabinose is supplemented in the medium, by turning on the *groES* and *groEL* genes. Alternatively, when glucose is supplied in the medium, the levels of GroES and GroEL gradually deplete as the cells grow and divide and eventually reduce to levels that are too low to sustain growth.

The BL21 (DE3) strain is engineered to have λ-prophage that has T7 RNA polymerase controlled by *lac* regulatory construct (*lac* promoter and a *lac* operator). The expression of T7 RNA polymerase is hence activated in presence of lactose or its analogue (IPTG in this case). The strain enables the use of pET expression plasmids that contains a T7 promoter and the *lac* operator upstream of the polylinker site (Figure 3.1A). When IPTG is present inside the cell, it displaces the *lac* repressor protein, both from the chromosomal genome of BL21 and the pET plasmid. This activates the expression of T7 RNA polymerase which in turn transcribes the gene cloned downstream of T7 promoter on the plasmid. The system put together ensures rapid and efficient production of the desired protein in the strain (Figure 3.1).
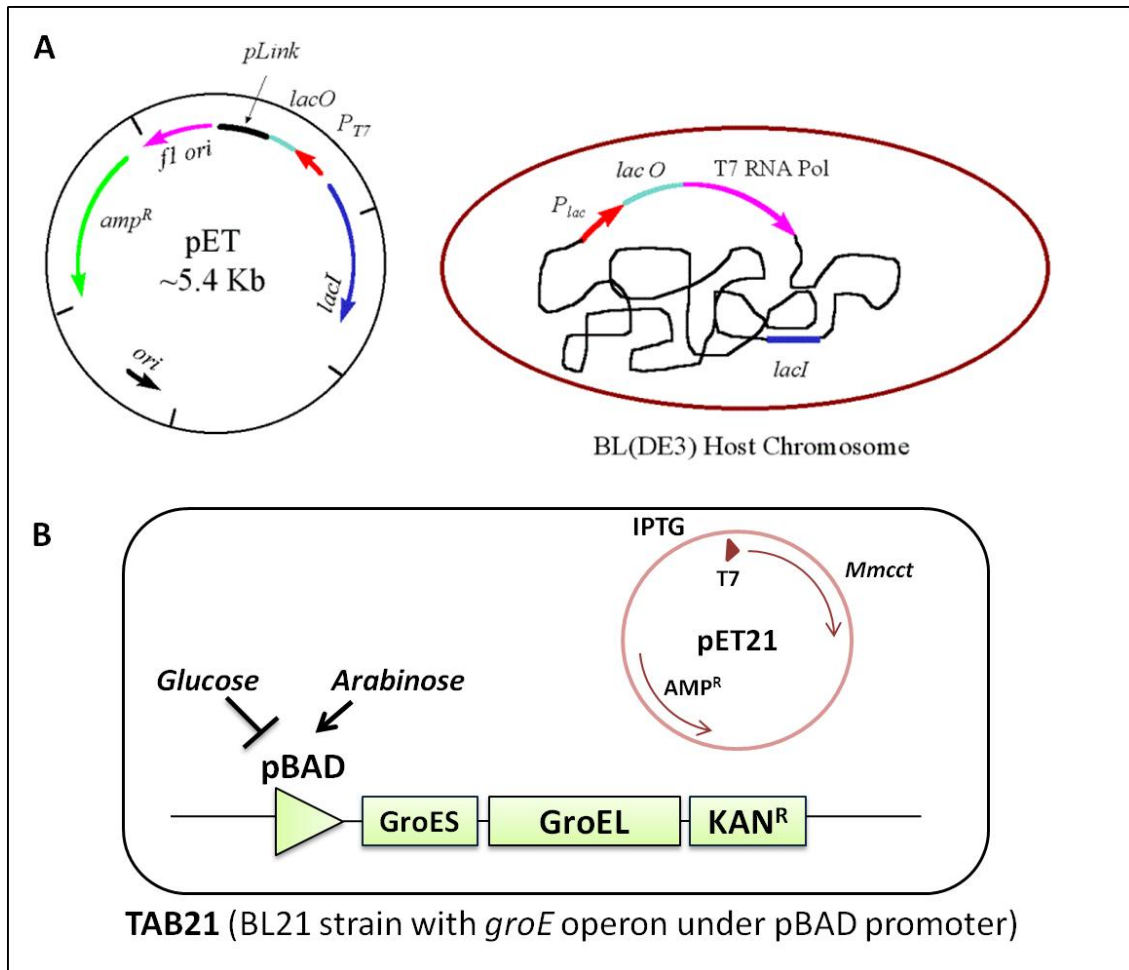
Figure 3.1

Schematic representation of TAB21 expression strain with pET vector

(A) Components of a pET vector and *E. coli* BL21 strain

(B) A schematic diagram representing the TAB21 experimental system for studying the effects of introducing archaeal *cct* genes in *E. coli*

3.1.2 Previous work

It has been previously found in our lab that archaeal chaperonin MmCCT can express and assemble in TAB21 and weakly sustain growth at 26^0C under *groEL* repressed conditions (Figure 3.2) (Bevan Lin, unpublished results and P. Gowarinathan, PhD thesis). To determine whether mutants can be isolated with improved function in *E. coli*, *Mmcct* was subjected to random mutagenesis using XL-1 red mutator strain (Section 2.6.5, Materials and Methods). It was observed that a single point mutation from methionine at the 223 position of the *Mmcct* gene to a bulky isoleucine residue (M223I) enhanced the ability of MmCCT to complement for the (work by former post-doctoral fellow, Dr. Andrew Large) (Figure 3.2).

In this chapter I have built on the observations made by Dr. Andrew Large and Dr. Preethy Gowarinathan by first replicating their experiments and then expanding on the results. In line with these findings and our hypothesis, this part of the project had following objectives:

1. To find additional variants of *Mmcct* with better function in *E. coli*.

2. To characterize these mutant proteins *in vivo* for their expression and assembly.

3. To determine the complementation ability of these mutants in *groEL* knock-out strain.
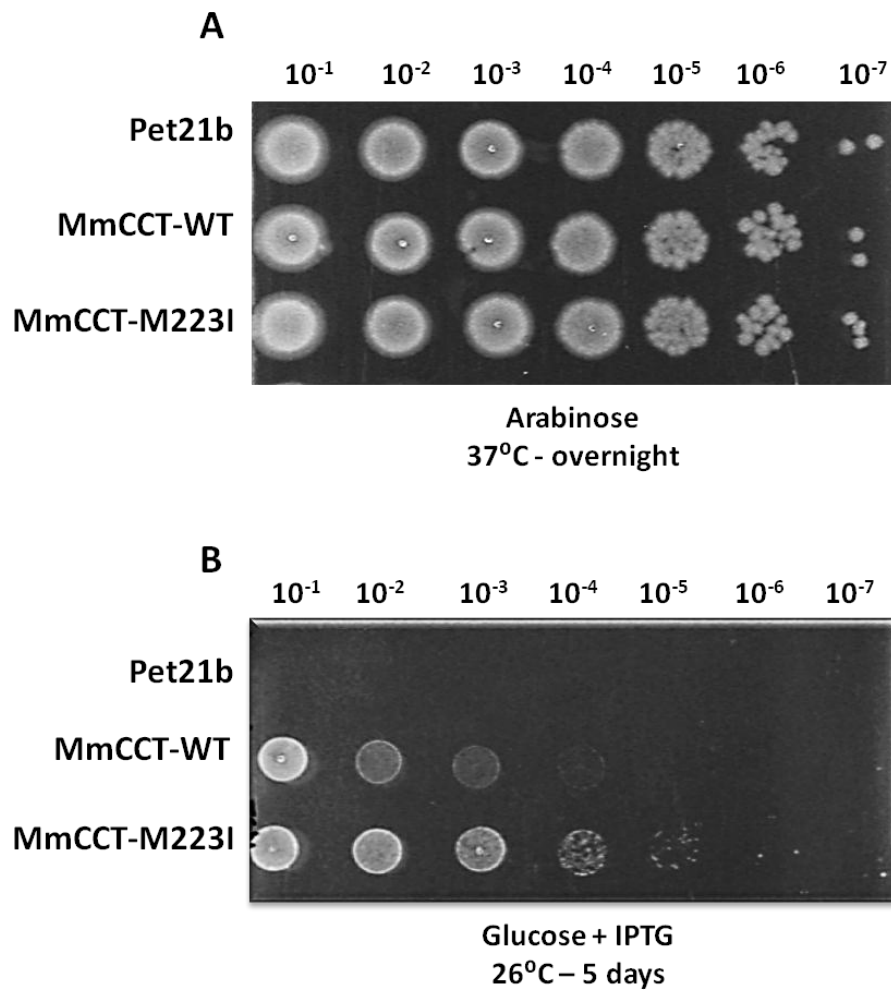
Figure 3.2

***M. maripaludis* chaperonin MmCCT-WT and a point mutant M223I can rescue growth of *E. coli* TAB21 under GroESL limiting conditions**

TAB21 cells containing plasmids pET21, pET21-*Mmcct*-WT, pET21-*Mmcct*-M223I grown on **(A)** arabinose (*groE* induction) at 37^0C and on **(B)** 0.2% glucose (*groE* repression) and 1 mM IPTG (*Mmcct* induction) at 26^0C.

3.2 Results

3.2.1 Random Mutagenesis of *Mmcct* using XL1-Red mutator strain

Before proceeding to the characterization of MmCCT-M223I mutant, random mutagenesis was carried out again to determine whether any additional *Mmcct* mutants could be isolated with improved functionality in *E. coli.* Previous attempts using different methods including error prone PCR and chemical mutagenesis had shown that the use of XL-1 mutator strain gave the best results in terms of isolating mutants (Work done by Dr. Andrew Large). Hence XL-1 red strain was used for random mutagenesis. This *E. coli* strain is deficient in three genes of error prone mismatch repair system (*mutD5, mutS* and *mutT*) and hence according to manufacturer has a ~5000 fold higher mutation rate than the wildtype strain.

The pET21-*Mmcct-wt* plasmid was transformed into XL-1 red strain and grown on LB agar containing Amp. All the colonies obtained were mixed together by resuspending in 500 μl LB on the plate and were cultured overnight at 37^{0}C in 100 ml of LB containing Amp. For a higher mutation rate, these cultures were sub-cultured and grown overnight for two cycles, after which the plasmids were isolated that now should contain mutations at random positions. The plasmids (confirmed to contain *Mmcct* gene using restriction digestions) were transformed into TAB21 and grown on plates containing LB agar with Kan, Amp, 0.2% glucose and 1 mM IPTG (referred to as KAGI plates) at 26^{0}C, 30^{0}C and 37^{0}C. It was already known that cells where the *Mmcct*-wt gene was expressed and *groEL* gene was repressed, took up to five days to come up as colonies on the plates, however in this case a few faster growing colonies were observed at 26^{0}C and 30^{0}C. These colonies were distinctly bigger than the rest of the colonies. No growth was

observed on the plates at 37^0C. The faster growing colonies obtained on 26^0C and 30^0C plates could be a result of either of the following:

1. A mutation in *pBAD* promoter or *AraC* gene (responsible for regulation of *pBAD* promoter by AraC protein) on the chromosome, arising due to random events occurring in the cells. In either case, repression could be lifted allowing faster growth due to *groEL* expression.

2. A mutation in the T7 promoter on the pET plasmid that increases the expression of MmCCT or a mutation on the plasmid that increases its copy number and hence the amount of MmCCT protein in the cell. Higher amounts of MmCCT protein might help in better complementation.

3. A mutation in *Mmcct* gene itself that allows the protein to efficiently replace GroEL function.

To exclude the first possibility, it was necessary to confirm that the phenotype of such colonies segregated with the plasmid. The fast growing colonies from the plates grown at both 26^0C and 30^0C were selected by eye and cultured overnight on LB containing arabinose. The plasmids were isolated from these cultures and retransformed into TAB21. Individual colonies from the resultant plates were cultured in LB containing arabinose for $\sim$ 3 hours (until OD_{600} =1) and serially diluted from 10^{-1} to 10^{-7} dilutions. 10 µl of each dilution were spotted on KAGI plates along with dilutions of cells carrying pET21 empty vector, pET-*Mmcct-wt* and pET-*Mmcct*-M223I as controls at appropriate temperatures.

Two representative plates each for 26^0C and 30^0C are shown in Figure 3.3. Some of the colonies grew as fast as those containing pET-*Mmcct*-M223I (b, c and f in the figure), while some grew a little faster than those carrying pET-*Mmcct*-WT (a, d and e), both suggesting that the improved phenotype was carried with the plasmid. The plasmids of the six potential candidates were sequenced using four different *Mmcct* gene primers with overlapping ends (Table 2.4, Materials and Methods). The sequence of whole gene was constructed from the best read regions of all four primers. The final gene sequences obtained for all candidates were aligned with the *Mmcct-wt* sequence (Xbase ID: MMP1515) using ClustalW multiple alignment software. The gene sequences were translated into corresponding protein sequences using ExPASy translate tool and aligned against the MmCCT-WT protein sequence (GenBank ID: NP_988635.1).

In accordance with the previous finding Methionine to Isoleucine substitution at position 223 was obtained for two out of six candidates (b and c in Figure 3.3A). Additionally for one candidate, a substitution from Lysine at position 216 to Glutamic acid (K216E) was found (f in Figure 3.3A). No mutations were detected for the rest of the candidates (a, d, and e). The improved phenotype in such candidates could be due to mutations in T7 promoter or rest of the plasmid as mentioned in the second point above. To exclude the possibility of such a process also happening in the obtained mutants, the T7 promoter region was sequenced using an internal *Mmcct* gene reverse primer. No mutations were found here. The aspect of increase in copy number of the plasmid and hence an increase in the amount of total MmCCT protein in the cell was checked by comparing the protein expression profiles of the mutants and the wildtype. No significant increase in protein levels was observed making the possibility of higher amounts of protein to have an effect on the phenotype, less likely. To completely ensure

that the improved phenotype was a result of the point mutations, the M223I and K216E mutations were independently incorporated in the *Mmcct-wt* gene by site-directed mutagenesis. Growth for both the mutants was indeed observed to be faster than the wildtype in GroEL limiting conditions. This part of experiment will be explained in more detail in section 3.2.2.3.

Both K216 and M223 reside in the apical domain of the MmCCT complex and attain different positions in open and closed conformations (Figure 3.4 and 3.5). The M223 residue lies in a loop beneath helix 11 near to the helical protrusion and is seen to be embedded on the outer surface of the chaperonin in a surface representation of closed MmCCT complex (Figure 3.6). The K216 residue faces towards the inner of the cavity and contributes to the hydrophilic nature of the cavity as described recently (Pereira et al., 2010) (Figure 3.7). Both K216 and M223 are highly conserved in all archaeal species. When tested among the archaeal chaperonins of 20 different species, K216 and M223 were found to be 100% and 95% conserved respectively (Figure 3.8). Because mutant M223I was found earlier than K216E, this order will be maintained throughout the document.
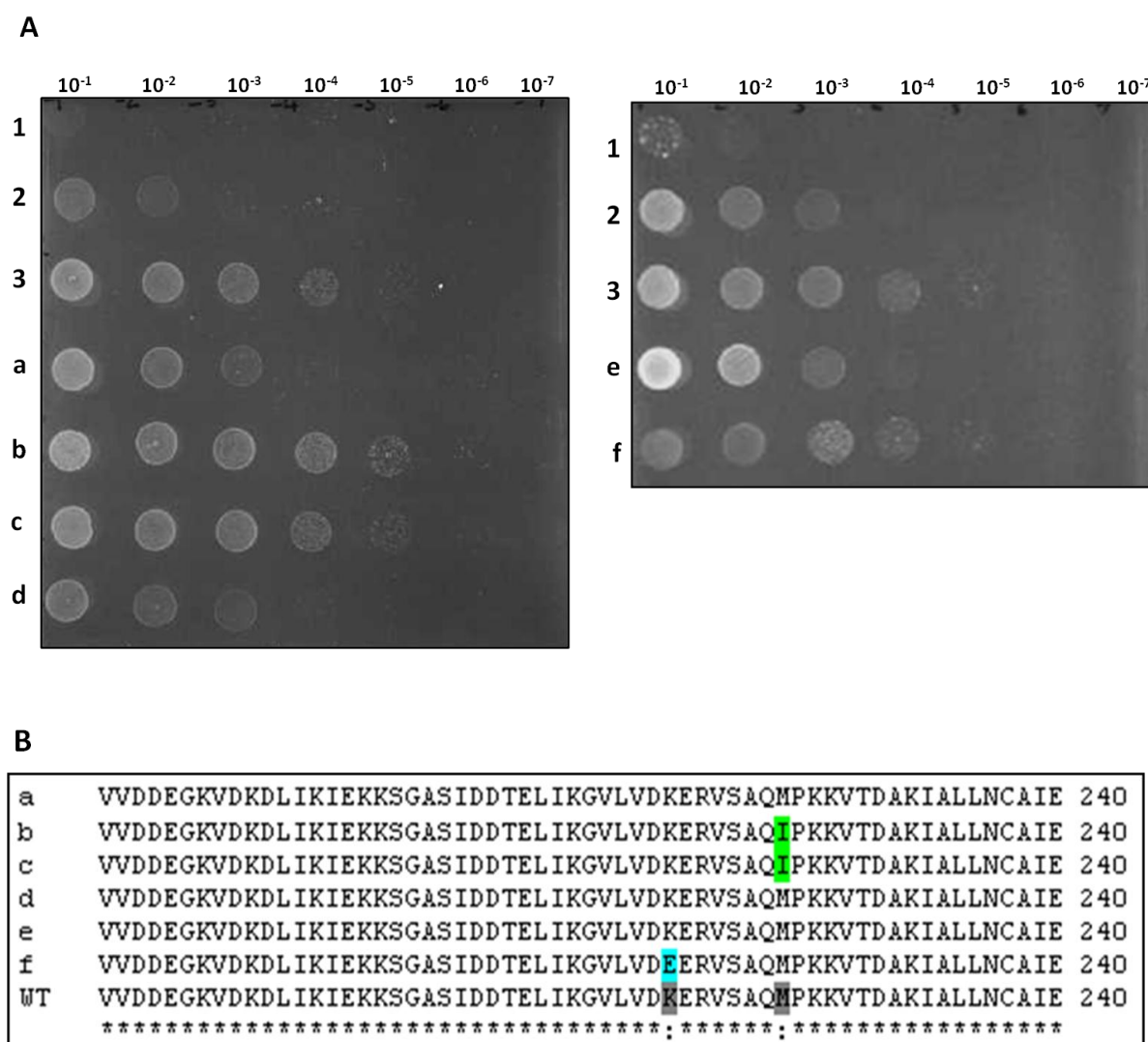
Figure 3.3

Random mutagenesis of *Mmcct* yielded mutants MmCCT-M223I and MmCCT-K216E with improved phenotype in *E. coli* under GroEL and GroES depleting conditions.

(A) Six fast growing clones of TAB21 (a-f) obtained as a result of random mutagenesis along with controls TAB21/pET21b (1), TAB21/pET21-Mmcct-wt (2) and TAB21/pET21-Mmcct-M223I (3) were cultured on arabinose, serially diluted and spotted on LB agar plates containing 0.2% glucose and 1mM IPTG at 26°C (left) and 30°C (right).

(B) Alignment of the protein sequences derived from the *Mmcct* gene sequences of the potential mutant candidates against the wildtype protein by ClustalW software.
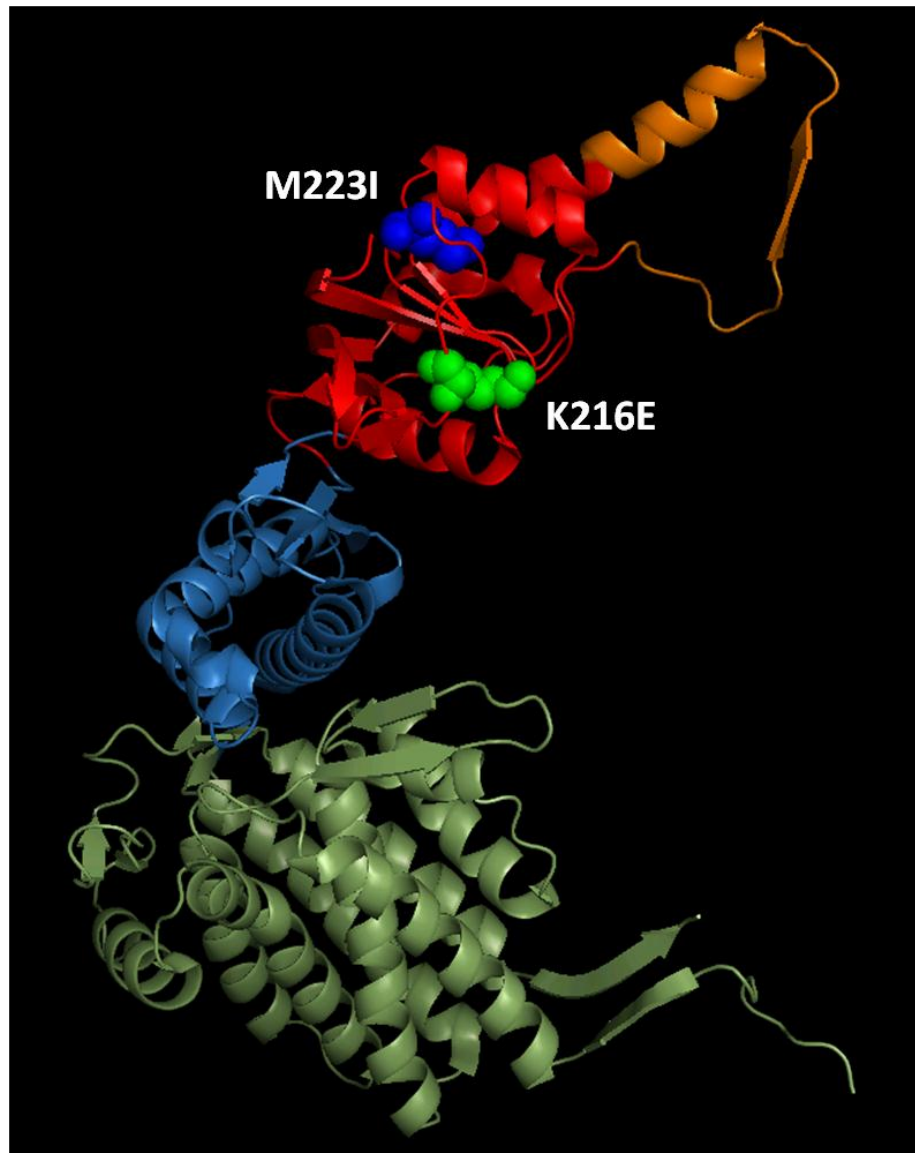
Figure 3.4

Location of M223 and K216 residues on MmCCT subunit

MmCCT subunit showing M223 and K216 residues highlighted as blue and green respectively. The subunits were visualized in PyMOL from structures submitted to Protein Data Bank (PDB ID: 3KFK) by Pereira *et al.*, 2010. Red: Apical domain, Blue: Intermediate domain, green: Equatorial domain and Orange: Helical protrusion.
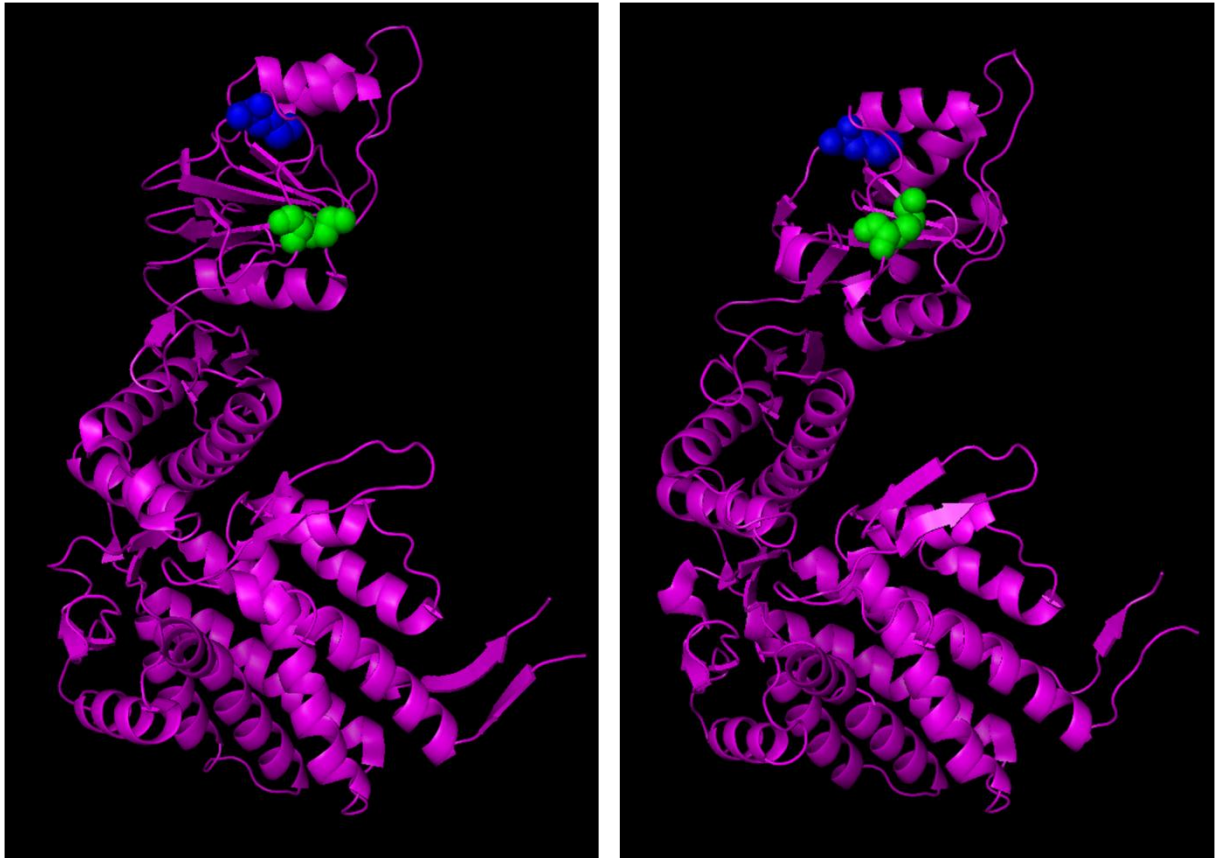
Figure 3.5

Location of M223 and K216 residues on open and closed subunits of MmCCT

MmCCT subunit showing M223 (blue) and K216 (green) residues on lidless-closed (left) and lidless-open (right) conformations. The images were derived from PDB codes 3KFB and 3KFE respectively (Pereira *et al.*, 2010) and visualized in PyMOL.
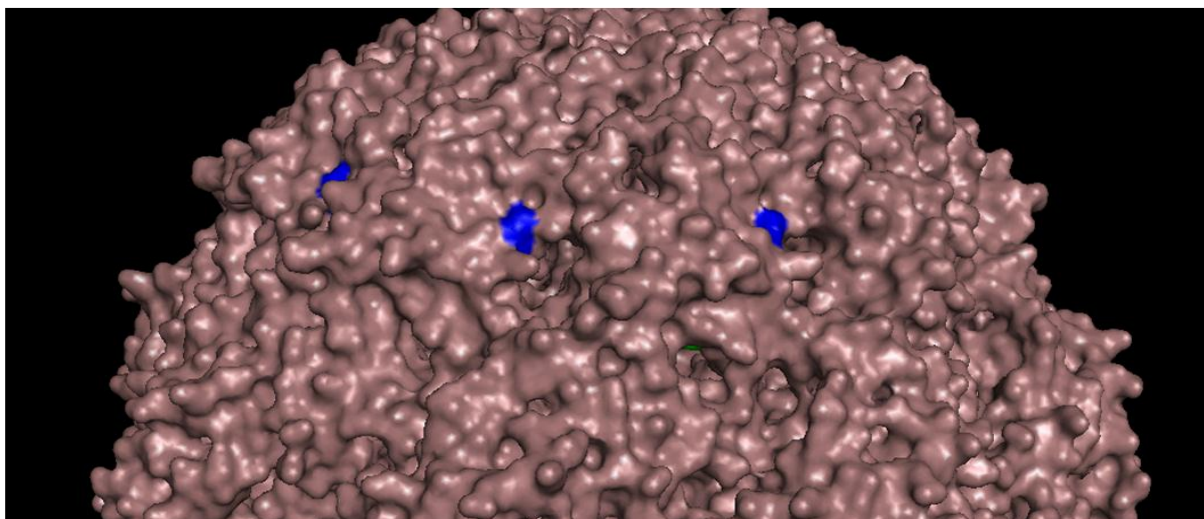
Figure 3.6

Location of M223 residues on a surface representation of MmCCT closed complex

Side view of closed conformation of MmCCT complex highlighting M223 residues (blue). The images were drawn in PyMOL using structures derived from PDB (ID: 3KFK) (Pereira *et al.,* 2010).
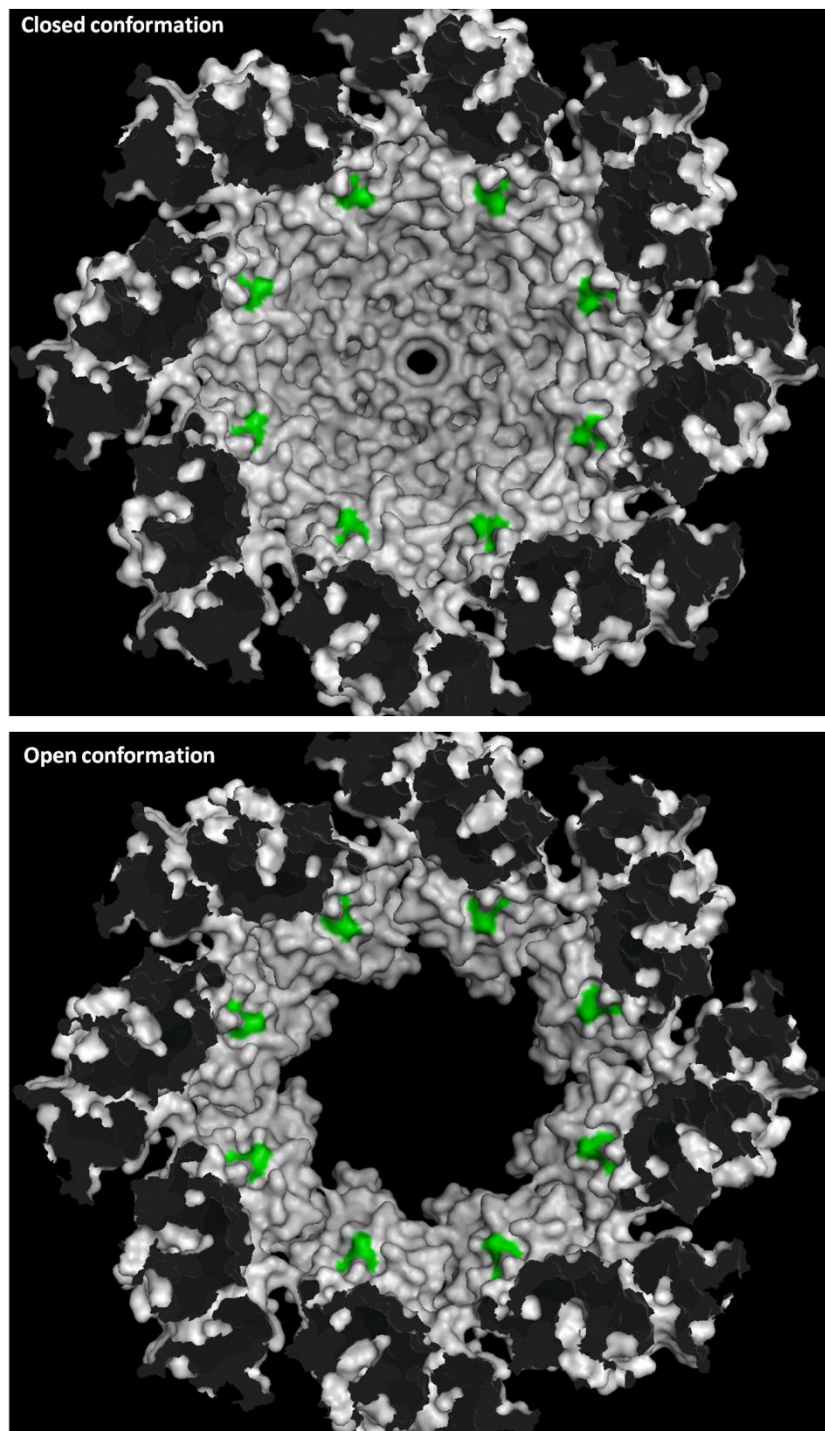
Figure 3.7

Location of K216 residues on a surface representation of MmCCT complex

The figure shows the *cis* cavity as viewed from bottom of the cross-sections of closed (above) and open (below) conformations of lidless MmCCT complex, highlighting K216 residues (green). The images were drawn in PyMOL using structures derived from PDB (ID: 3KFB and 3KFE) (Pereira *et al.*, 2010).

A
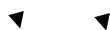
B

Figure 3.8 Conservation of K216 and M223 in 20 different archaeal species
Sequences obtained from Xbase aligned by ClustalW software (A) and distribution observed using sequence logos (B) created by WebLogo.

3.2.2 Analysis of MmCCT-M223I and MmCCT-K216E mutants in *E. coli*

This section describes the experiments conducted to test the expression and assembly of the MmCCT-M223I and MmCCT-K216E proteins in *E. coli* and further to confirm the ability to support its growth when GroEL is limiting. As discussed in the experimental design, TAB21 expression system using pET21 vector was used for all experiments.

3.2.2.1 Expression of MmCCT-M223I and K216E in TAB21

The first step towards characterization of MmCCT-M223I and K216E mutants was to test their expression in *E. coli.* Competent TAB21 cells were transformed with pET21b empty plasmid, or pET21 carrying either the *Mmcct-wt* gene, *Mmcct*-M223I, or *Mmcct*-K216E and plated on LB agar plates containing Kan, Amp and arabinose (referred to as KAA plates). Individual colonies were cultured at 37^0C in arabinose in duplicate sets. As the cultures reached an OD_{600} of 0.4, 1 mM IPTG was added to one set and the cells were grown further for four hours. All the cultures were diluted to the same starting OD (OD_{600} = 1) to have comparable densities. The protein samples were collected from these cultures and subjected to SDS gel electrophoresis.

The pET empty vector served as a negative control and pET-*Mmcct*-WT served as a positive control. Both MmCCT-M223I and K216E expressed profusely as ~60KDa monomeric subunits on SDS gel in IPTG induced conditions as seen in Figure 3.9. It should be noted however that the expression shown here is at 37^0C. The wildtype and the mutants when tested at 26^0C and 30^0C showed poorer expression (results not shown).
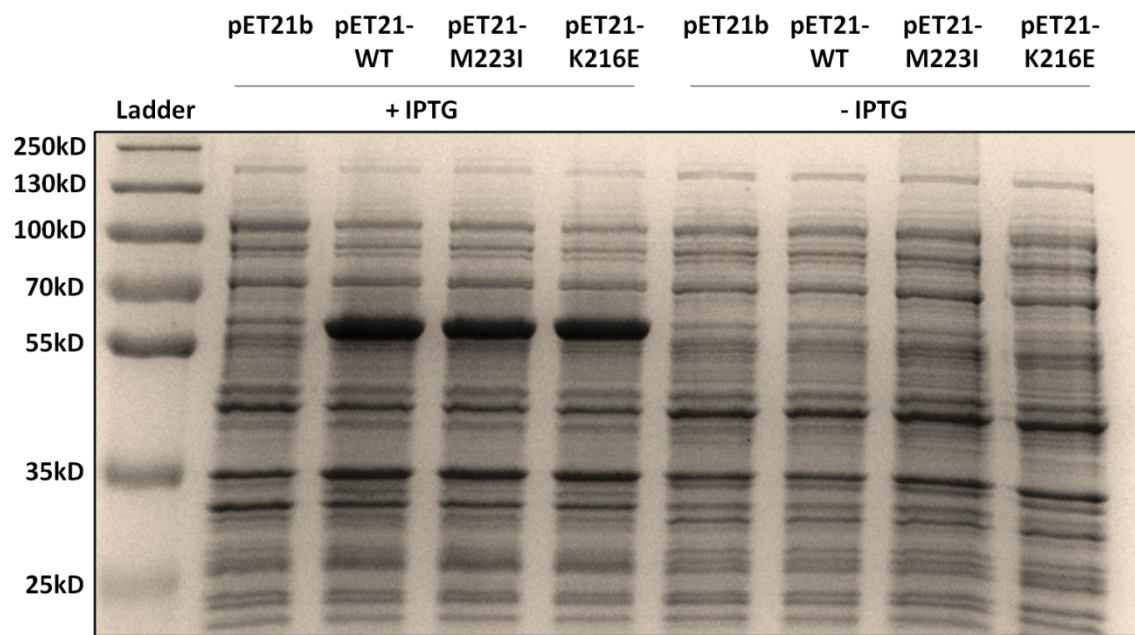
Figure 3.9

Expression of MmCCT-M223I and MmCCT-K216E in TAB21

Crude extracts from *E. coli* TAB21 cells overexpressing pET21-MmCCT-WT, pET21-MmCCT-M223I and pET21-MmCCT-K216E in IPTG induced and non-induced conditions, as analyzed by 10% SDS-PAGE and Coomassie staining.

3.2.2.2 Assembly of MmCCT-M223I and K216E in TAB21

Next, the mutants were subjected to native gel electrophoresis for testing the correct assembly into chaperonin oligomers. At every stage i.e. initial densities at sub-culturing, point of IPTG induction, cell densities for harvesting and cell lysis for protein extraction, equivalent conditions were ensured to allow comparison of soluble proteins from different samples. Initially, assembly was tested using 7.5% native-PAGE. While the MmCCT-WT and K216E mutant proteins assembled quite well at similar levels (Figure 3.10A), MmCCT-M223I showed a more variable behaviour often resulting into apparent non-assembly (Figure 3.10A, buffer with KCl). Numerous trials with varying buffer compositions and methods of cell lysis were employed to test the validity of the result. When KCl, a component added in the cell-lysis buffer for stabilization of complexes was eliminated, M223I assembled into its functional complex (Figure 3.10A, buffer without KCl), albeit in lesser proportion as compared to MmCCT-WT or K216E. However, it should be noted that assembly was observed in buffers with KCl, sometimes in equal proportions as the WT.

The assembly of the mutants was also checked by a 3-10% native gradient PAGE where both MmCCT-M223I and K216E showed assembly into their oligomers (Figure 3.10B). Consistent with the previous result, the yield for M223I protein was less than that of the WT while the levels for K216E and WT were similar. The GroEL and MmCCT-WT purified protein (provided by Dr. Andew Large) were run in parallel with the samples to determine the positions of the bands. A GroEL band can also be seen in all samples on native gels because the cultures were grown in arabinose that allows expression of *groEL* gene.
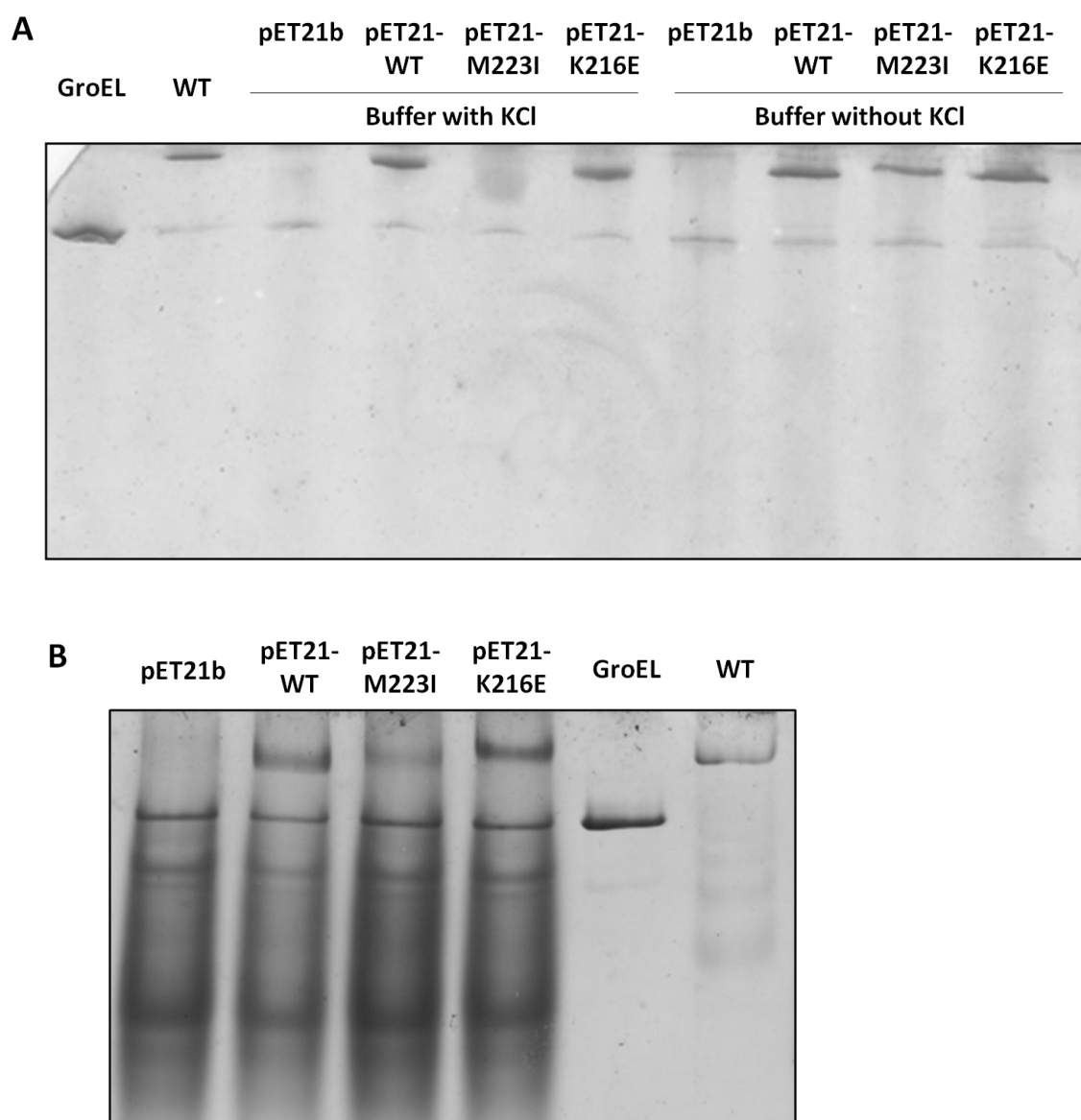
Figure 3.10

Assembly of MmCCT-M223I and MmCCT-K216E in TAB21

Soluble fractions of equal densities of *E. coli* TAB21 cells expressing pET21 empty vector, pET21-MmCCT-WT, pET21-MmCCT-M223I and pET21-MmCCT-K216E in IPTG induced conditions, prepared in non-denaturing conditions using buffer with or without KCl and analyzed by **A)** 7.5% Native PAGE. The latter were also analyzed by **B)** 3-10% Native gradient PAGE. Purified MmCCT-WT and GroEL proteins were ran as controls. Both the gels were stained with Coomassie.

3.2.2.3 Complementing ability of the mutants when *groESL* operon is switched off

To test the ability of the mutants to compensate for the loss of GroEL and GroES *in vivo,* both MmCCT-M223I and K216E were subjected to growth analysis in liquid and solid media under GroEL depleting conditions.

Liquid media analysis: Overnight grown TAB21 cultures expressing pET21-*Mmcct*-M223I or pET21-*Mmcct*-K216E along with the empty vector and wildtype controls were sub-cultured into fresh media containing glucose and the degree of growth was measured over a period of time at 37^0C. The cells were expected to grow while GroEL depletion occurs to finally reach a point where growth was no longer possible due to absence of GroEL. However, all the cultures including the TAB21 cells on their own continued to grow at normal rates (Figure 3.11A). Such an observation suggested either a carry-forward of residual GroEL or a leaky repression of T7 promoter. Indeed, when these cultures were further sub-cultured in glucose containing medium for a second cycle, absolutely no growth was observed. It was assumed hence, that the TAB21 cells grown on arabinose when sub-cultured in glucose containing medium, exhibited a normal growth pattern owing to the residual GroEL supposedly sufficient for survival.

To see what changes occur in the growth pattern of TAB21 cells that express MmCCT-WT, MmCCT-M223I and MmCCT-K216E, IPTG was added in the medium once cells reached OD_{600}=0.4 and the growth was followed by measurement of optical densities. It should be noted that addition of IPTG leads to a general growth-slowing of TAB21 cells at 37^0C. This is because of the presence of T7 RNA polymerase that competes with bacterial RNA polymerases for cell resources such as ribosomal capacity and also due to expression of Lac proteins that are known to exhibit a burden to growth on *E. coli* cells

(Malakar and Venkatesh, 2012). However this being a general phenomenon should not affect the particular differences observed between the samples. It was observed that under IPTG induced conditions, the cultures containing vector only ceased to grow after reaching an OD_{600} of approximately 0.8. While the wildtype showed a better phenotype, the M223I and K216E mutants were able to rescue growth of the cells more effectively than the wildtype (Figure 3.11B) as predicted from the phenotype of these mutants on plates.

Solid media analysis: 1 ml of the non-induced cultures were diluted to the same starting OD and harvested. The pellets were washed thrice with LB containing glucose to remove any traces of arabinose and resuspended in 1 ml of LB. Cultures were serially diluted from 10^{-1} to 10^{-7} and 10 μl of each dilution was spotted onto KAA and KAGI plates. The plates were incubated at 26^{0}C, 30^{0}C and 37^{0}C.

Consistent with the random mutagenesis results and liquid media results, both MmCCT-M223I and K216E exhibited a fairly better GroEL-complementing ability than the WT at 26^{0}C and 30^{0}C (Figure 3.12 for 30^{0}C). No growth was observed at 37^{0}C.
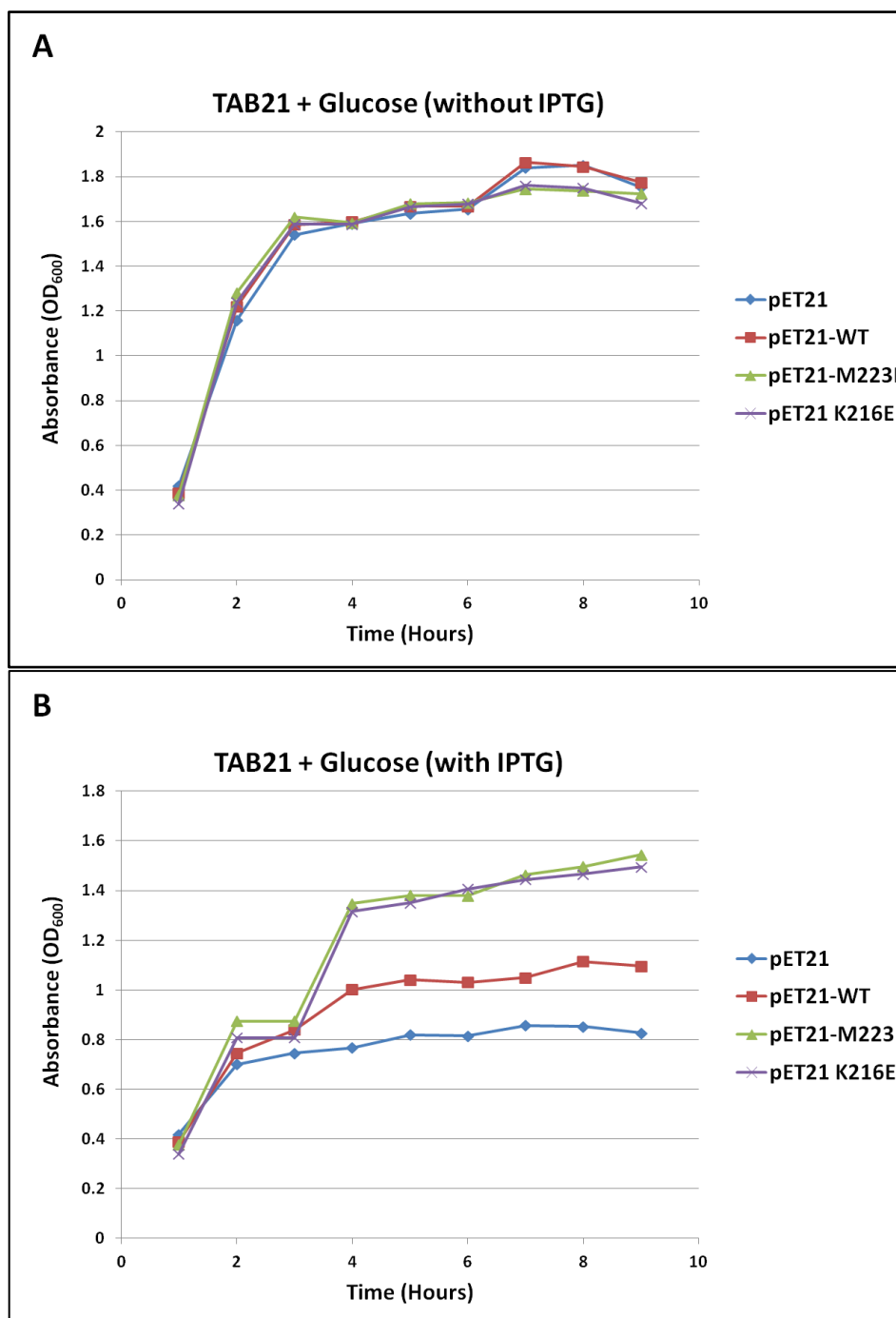
Figure 3.11

Liquid medium growth analysis of MmCCT mutants under GroEL depleting conditions

TAB21 cultures expressing pET21, pET21-*Mmcct*-WT, pET21-*Mmcct*-M223I and pET21-*Mmcct*-K216E grown in LB containing glucose with or without IPTG at 37^0C followed by measurement of optical densities over the indicated interval of time . The error bars are not included because the ODs obtained varied being dependent on the initial OD of induction. However the given pattern represents at least three independent experiments.
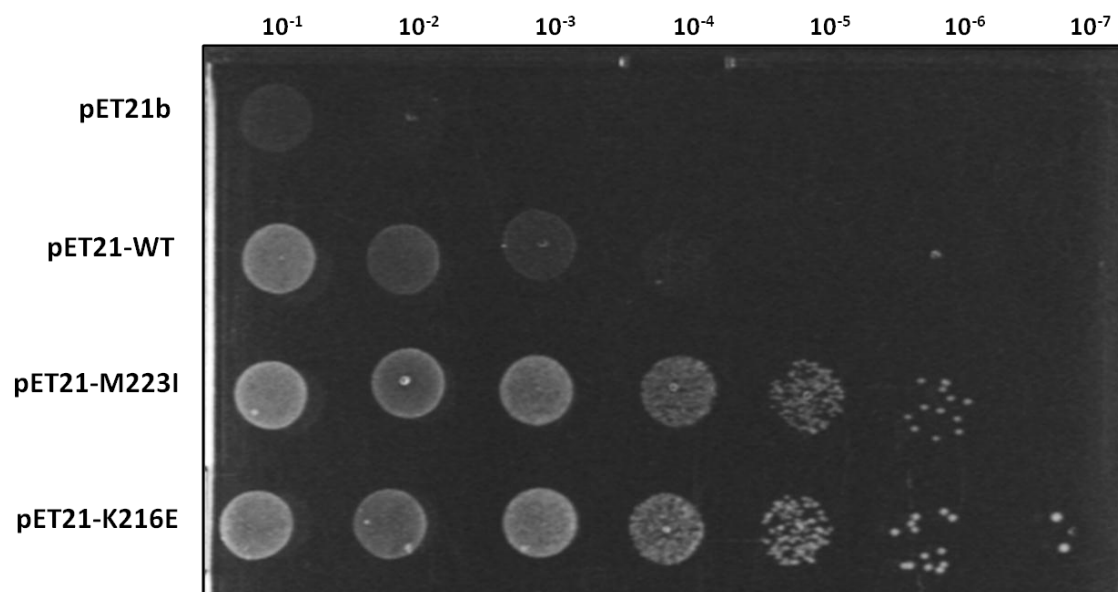
Figure 3.12

Solid medium growth analysis of mutants under GroEL depleting conditions at 30⁰C

TAB21 cells containing plasmids pET21, pET21-*Mmcct*-WT, pET21-*Mmcct*-M223I and pET21-*Mmcct*-K216E grown on 0.2% glucose and 1mM IPTG at 30^0C for five days.

3.2.3 Substitution of M223 and K216 with other amino acids

It was found that almost all amino acid substitutions at the M223 position were able to better complement for the loss of GroEL than the wild type (work carried out by Dr. Andrew Large). Similar experiment was carried out for the K216E mutant. The primers were designed such that random base pairs are inserted at K216 position generating a mixture containing different amino-acid combinations. This mixture of primers (MmCCT-K216n-F and MmCCT-K216n-R) was used to perform site directed mutagenesis in pET21-*Mmcct-wt* (Section 2.5.3, Materials and Methods). Approximately 50 colonies obtained were individually cultured, plasmid isolated and sequenced to generate a library of K216 point mutants. A total of 14 out of 19 substitutions (K216 to A,C,D,F,G,L,N,P,Q,R,S,T,W and Y) were obtained. All the mutants being in pET21 plasmids were directly subjected to complementation analysis in TAB21. Also, they were tested for their expression and assembly in TAB21 using SDS and native gel electrophoresis.

Similar to results obtained for M223I, almost all amino-acid substitutions at K216 displayed K216E-like complementation in TAB21 under *groEL* repressed conditions at 30^0C (Figure 3.13). The only exceptions were the K216R and K216W mutants that showed weaker phenotype. Accordingly, all point mutant proteins expressed and assembled in a correct manner except for K216W that showed degradation (Figure 3.14 and 3.15). These results are summarized in Table 3.1. Such results indicated that mutations at these residues imparted some sort of functional flexibility leading to improved complementation in *E. coli.*
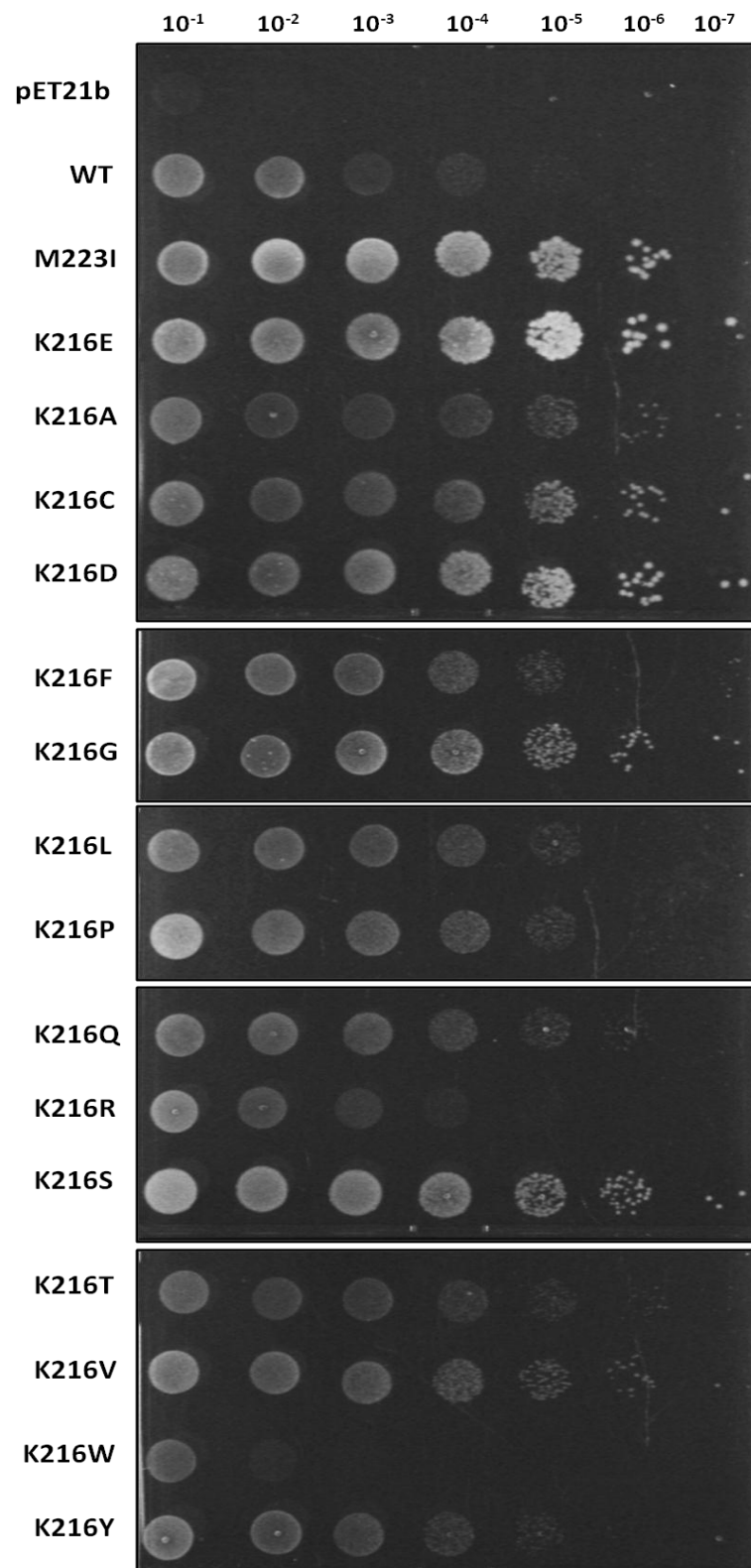
Figure 3.13

Growth of K216n mutants at 30^{0}C under GroEL depleting conditions

TAB21 cells expressing 14 different K216 mutants as indicated under pET21 vector grown on 0.2% glucose and 1 mM IPTG at 30^{0}C for five days.
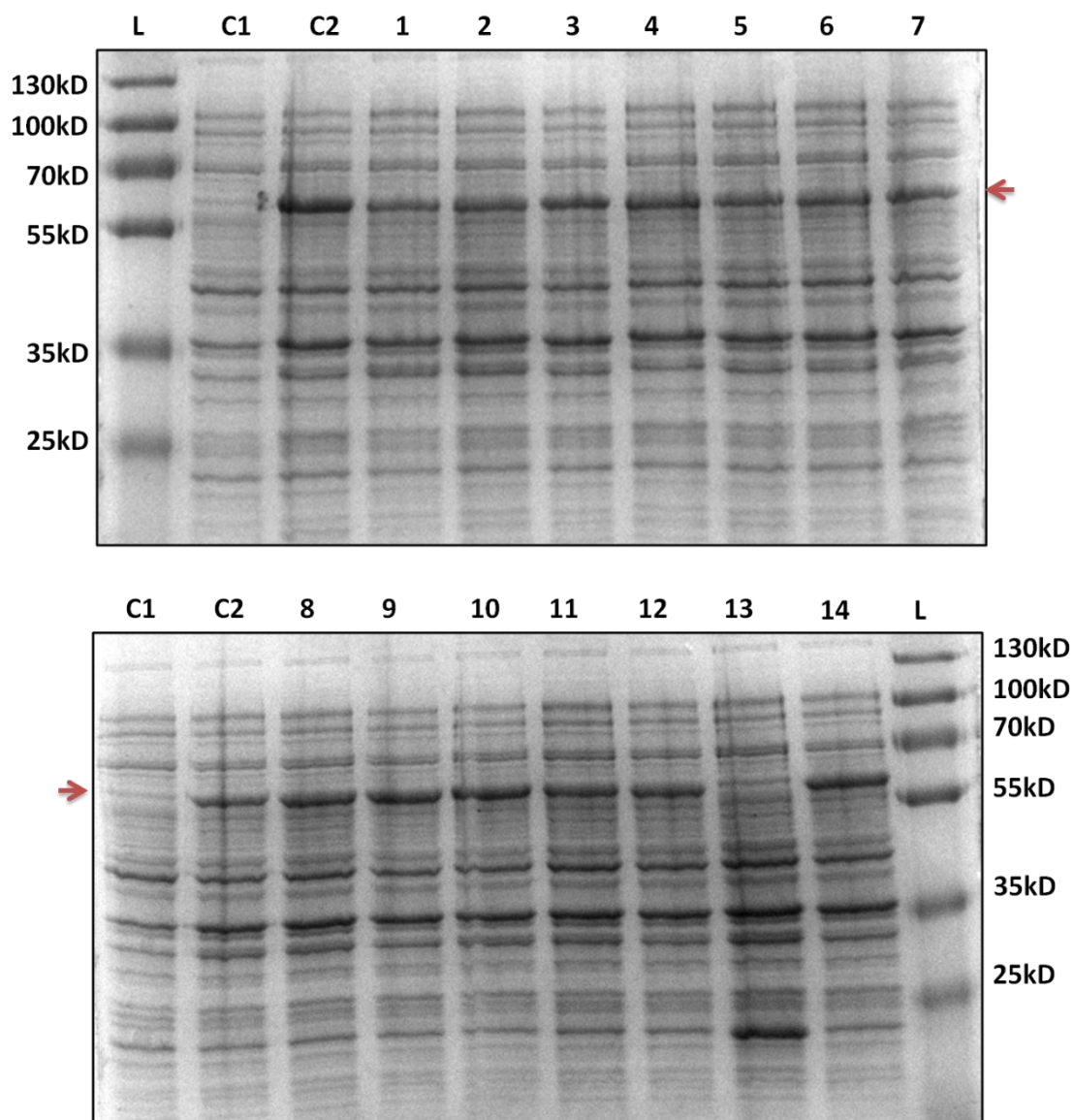
Figure 3.14

Expression of MmCCT-K216n mutants in TAB21

Crude extracts of TAB21 cells expressing 14 different MmCCT-K216n mutants i.e. K216A (1), K216C (2), K216D (3), K216F (4), K216G (5), K216L (6), K216P (7), K216Q (8), K216R (9), K216S (10), K216T (11), K216V (12), K216W (13) and K216Y (14) along with controls pET21 empty vector (C1) and pET21-MmCCT-K216E (C2) as analyzed by 12% SDS PAGE and stained by Coomassie staining. The arrows indicate the position of MmCCT bands.
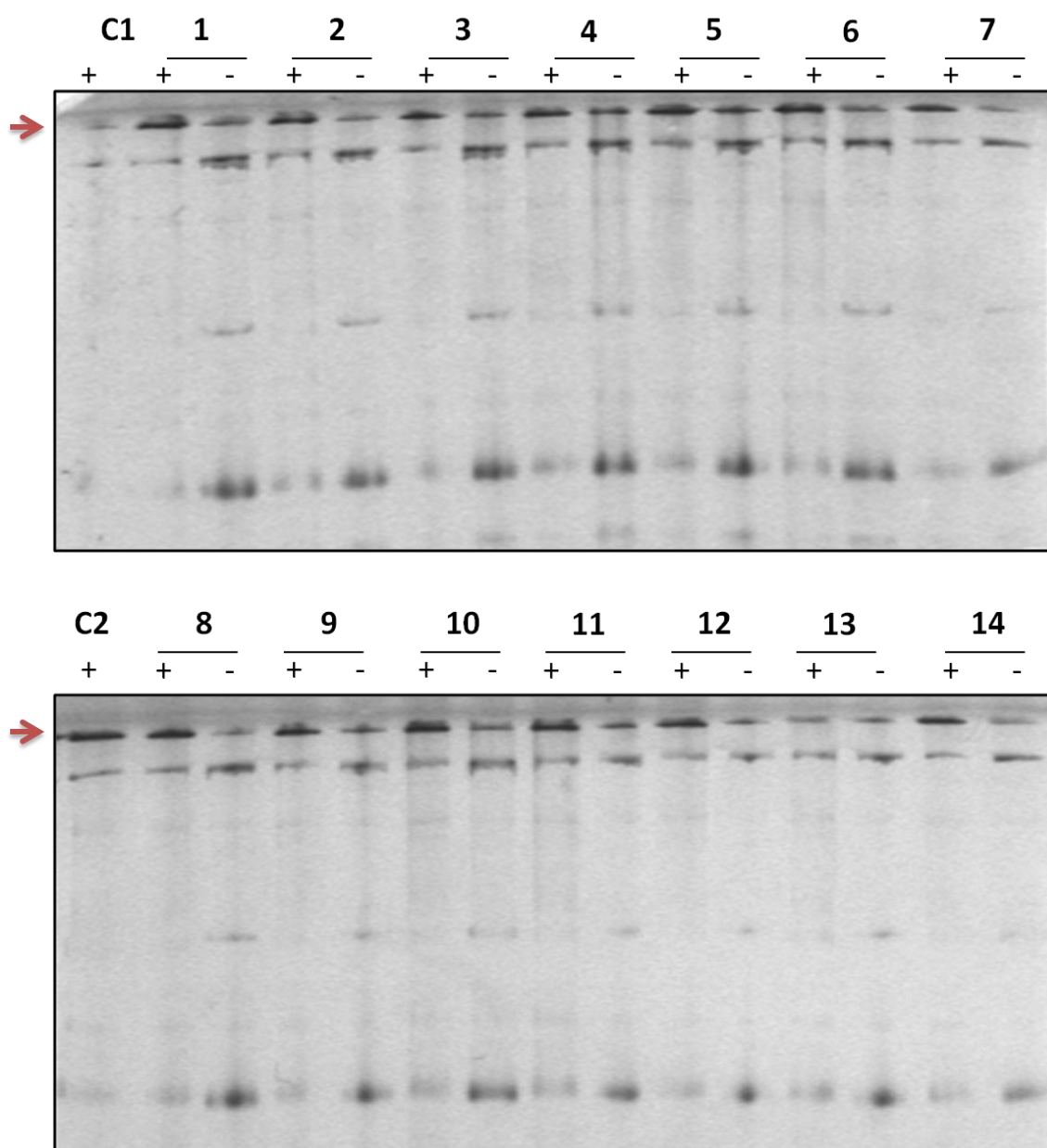
Figure 3.15

Assembly of MmCCT-K216n mutants in TAB21

Soluble fractions of induced (+) and non induced (-) TAB21 cells expressing 14 different K216 mutants i.e. K216A (1), K216C (2), K216D (3), K216F (4), K216G (5), K216L (6), K216P (7), K216Q (8), K216R (9), K216S (10), K216T (11), K216V (12), K216W (13) and K216Y (14) along with controls pET21 empty vector (C1) and pET21-*Mmcct*-K216E (C2) as analyzed by 7.5% Native PAGE and stained by Coomassie staining. The arrows indicate the position of the MmCCT bands.

Table 3.1: Summary of the functional analysis of MmCCT-M223n and K216n mutants *in vivo*

Substitutions at M223 (Work by Dr. Andrew Large)				Substitutions at K216 (This work)			
Mutation in MmCCT	Expression	Assembly	Complementation (30°C)	Mutation in MmCCT	Expression	Assembly	Complementation (30°C)
M223I	Y	Y	+++	K216A	Y	Y	++
M223F	Y	Y	+++	K216C	Y	Y	+++
M223G	Y	Y	+++	K216D	Y	Y	+++
M223E	Y	Y	+++	K216F	Y	Y	++
M223L	Y	Y	+++	K216G	Y	Y	+++
M223R	Y	Y	+++	K216L	Y	Y	++
M223S	Y	Y	+++	K216P	Y	Y	++
M223V	Y	Y	+++	K216Q	Y	Y	+++
M223W	Y	Y	+++	K216R	Y	Y	+
M223Y	Y	Y	+++	K216S	Y	Y	+++
				K216T	Y	Y	++
				K216V	Y	Y	+++
				K216W	N	N	–
				K216Y	Y	Y	++

3.2.4 Effect of an additional mutation (D386A) in MmCCT-M223I and MmCCT-K216E on the phenotype

To understand the reasoning underlying the better phenotype of the mutants MmCCT-M223I and K216E, one possibility we considered was that these mutations could cause a loss of chaperonin function, but nonetheless allow partial function that could still part replace GroEL. For example, this could be by locking the chaperonins in the open state, thereby providing more hydrophobic surface on which nascent unfolded proteins could bind, thus reducing somewhat the need for GroEL. If this is the case, we would predict that the ATPase activity of chaperonins would no longer be necessary. To test this hypothesis, a second mutation D386A that blocks the ATPase activity of the chaperonin was introduced. The D386A mutant on its own can bind to ATP but is unable to hydrolyze it (Reissmann et al., 2007), hence stopping the chaperonin from undergoing a folding cycle. In other words the protein is locked in its ATP-bound open conformation. If MmCCT-M223I and K216E are functioning better by just binding to the clients in an open conformation, the additional D386A mutation should not affect the phenotype of the mutants. If, on the other hand, the mutants are still acting as chaperonins requiring a full ATPase cycle, the additional D386A mutation should not support growth of the cells.

The D386A mutation was introduced into the *cct* gene in pET-*Mmcct-wt,* pET-*Mmcct*-M223I and pET-*MmCCT*-K216E using site-directed mutagenesis. The resultant constructs i.e. *Mmcct*-D386A, *Mmcct*-M223I-D386A and *Mmcct*-K216E-D386A were transformed into TAB21 and first checked for correct protein expression and assembly

(Figure 3.16A and B, C respectively). Subsequently the mutants were subjected to complementation analysis in TAB21. As shown in Figure 3.17C, the D386A mutant on its own only weakly supported growth as expected. However, the MmCCT-M223I-D386A and MmCCT-K216E-D386A double mutants also displayed a similar phenotype i.e. poor growth of the cells in *groEL* repressed conditions. The results suggest that blocking the ATPase activity severely affects the complementing ability of the mutants in GroEL limiting conditions. The observations strongly support the idea that the MmCCT-M223I and K216E mutants act as genuine chaperonins and undergo an ATP dependent chaperonin cycle to function.
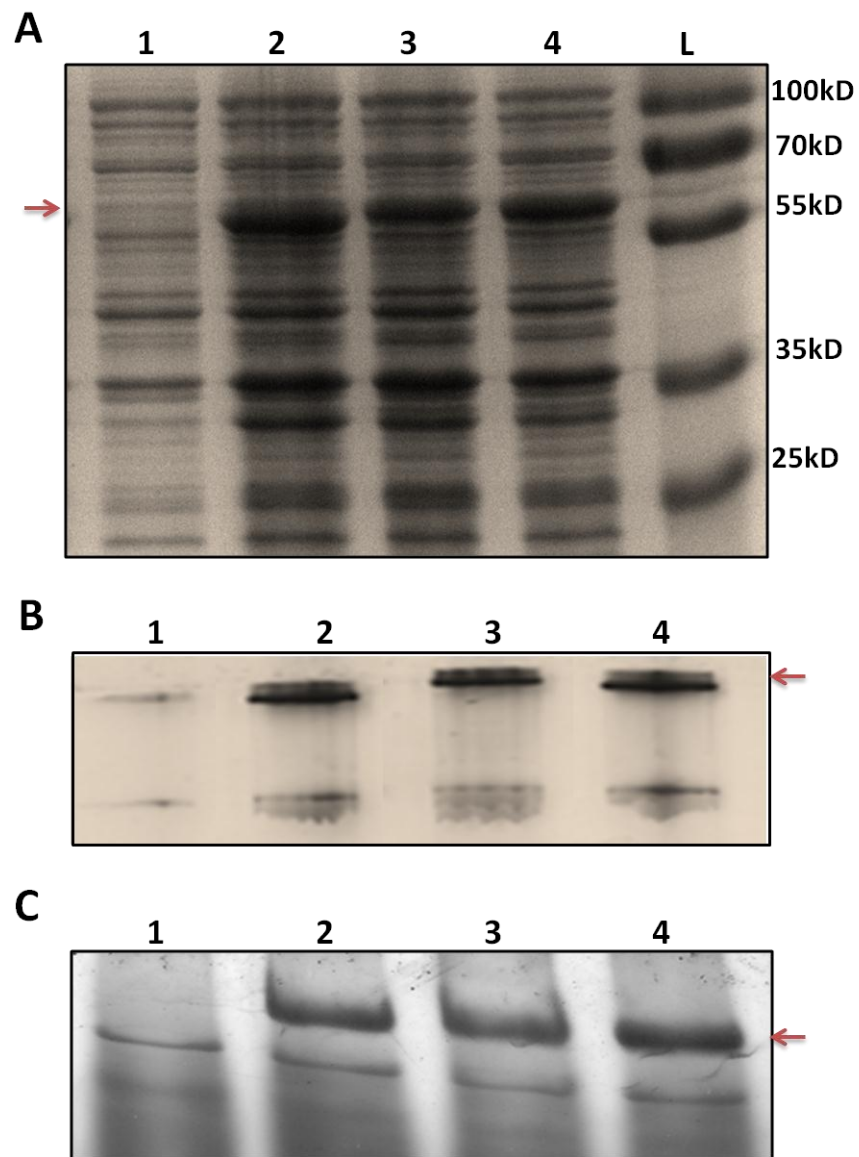
Figure 3.16

Expression and assembly of D386A mutants in TAB21

A) Crude extracts of TAB21 cells expressing pET21 empty vector (1), pET21-*Mmcct*-D386A (2), pET21- *Mmcct* -M223I-D386A (3) and pET21- *Mmcct* -K216E-D386A as analysed by 10% SDS PAGE **B and C)** Soluble fractions of the above samples as analyzed by 7.5% Native PAGE (B) and 3-10% Native gradient PAGE (C). All gels stained by Coomassie staining. The arrows indicate the position of the MmCCT bands.
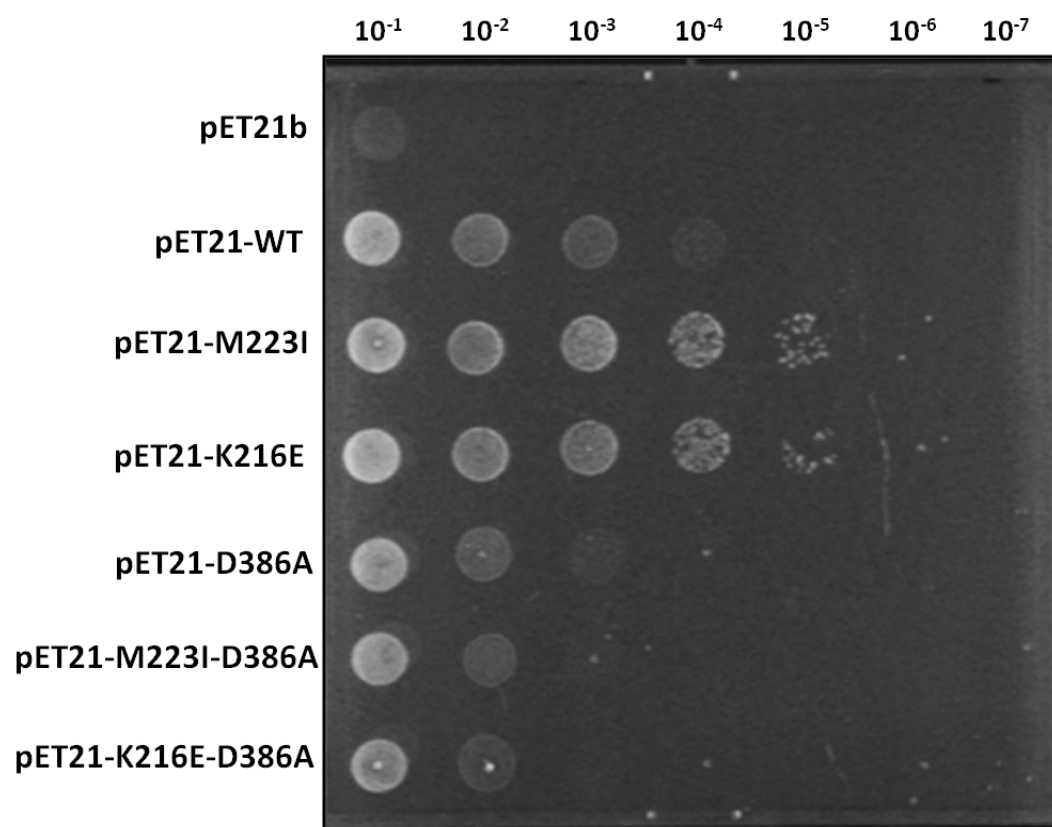
Figure 3.17

Growth of D386A mutants at 30^{0}C under GroEL depleting conditions

TAB21 cells expressing MmCCT-D386A mutants as indicated under pET21 vector grown on 0.2% glucose and 1 mM IPTG at 30^{0}C for five days.

3.2.5 Analysis of MmCCT-M223I and MmCCT-K216E in *groEL* knockout strain AI90

As noted in section 3.2.2.3, TAB21 cells could grow even after 48 hours of glucose repression. Western blotting of the proteins extracted from these cultures using anti-GroEL antibodies showed presence of residual GroEL (Dr. Andrew Large, personal communication). This suggested that glucose mediated *groEL* repression was not completely effective. It was possible hence that the growth of TAB21 cells expressing MmCCT proteins on glucose containing plates was due to residual GroEL (and not MmCCT). However the empty vector control not containing *Mmcct gene* was unable to survive under glucose repressed conditions suggesting that the residual GroEL was not sufficient for full functionality of cells and that assistance of MmCCT proteins was a requirement. The findings however did not establish as to whether or not MmCCT or its point mutants can exclusively fold all the essential client proteins of GroEL.

Therefore, to test whether MmCCT is acting as a helper-chaperonin or can completely replace GroEL function, complementation experiments were carried out in a Δ*groEL* strain of *E. coli* called as **AI90** (Ivic et al, 1996). The chromosomal copy of *groEL* in this strain is replaced by a kanamycin resistance marker cassette and the strain sustains on an external copy of *pTrc* promoter-regulated *groEL* on an expression plasmid. This expression plasmid also contains a *sacB* gene and a chloramphenicol resistance marker (Cam). The *Bacillus subtilis sacB* gene encodes an exoenzyme levansucrase (sucrose: 2,6-b-D-fructan 6-b-D-fructosyltransferasegene) that catalyzes hydrolysis of sucrose and synthesis of levans, and its expression is proved to be lethal for *E. coli* growth when sucrose is present in the medium (Gay et al., 1985; Stibitz, 1994). It can therefore be

used as a marker to select for cells that have lost the plasmid. When a plasmid expressing *Mmcct* (Amp) is introduced in AI90, a population of cells will now possess both – original plasmid containing *groEL-sacB*-Cam[R] and the new plasmid containing *Mmcct*-Amp[R]. Upon plating these cells on sucrose, the *groEL* containing plasmid is driven out and only the cells with plasmid expressing *Mmcct* are able to grow (Figure 3.18). Selection takes place against the *sacB* gene and hence the *groEL-sacB-cam*[R] plasmid is exited making the cells sucrose resistant and chloramphenicol sensitive. In such conditions, growth will occur only if MmCCT can completely replace GroEL (Figure 3.18).

Because pET expression vectors require a BL21 strain or its derivative for expression, pET21 plasmids bearing MmCCT genes could not be used in AI90 strain. Hence, *Mmcct* gene fragments (WT, M223I and K216E) from pET plasmids were cloned into HindIII and XbaI sites of pTrc99A vector under the regulation of IPTG inducible *ptrc* promoter. Competent AI90 cells were transformed with pTrc99A, pTrc-*groESL,* pTrc-*Mmcct*-wt, pTrc-*Mmcct*-M223I and pTrc-Mmcct-K216E and were plated at 26^0C, 30^0C and 37^0C on a set of plates with different components as shown in Table 3.2. The pTrc-*groEL* plasmid was used as a positive control for the experiment.
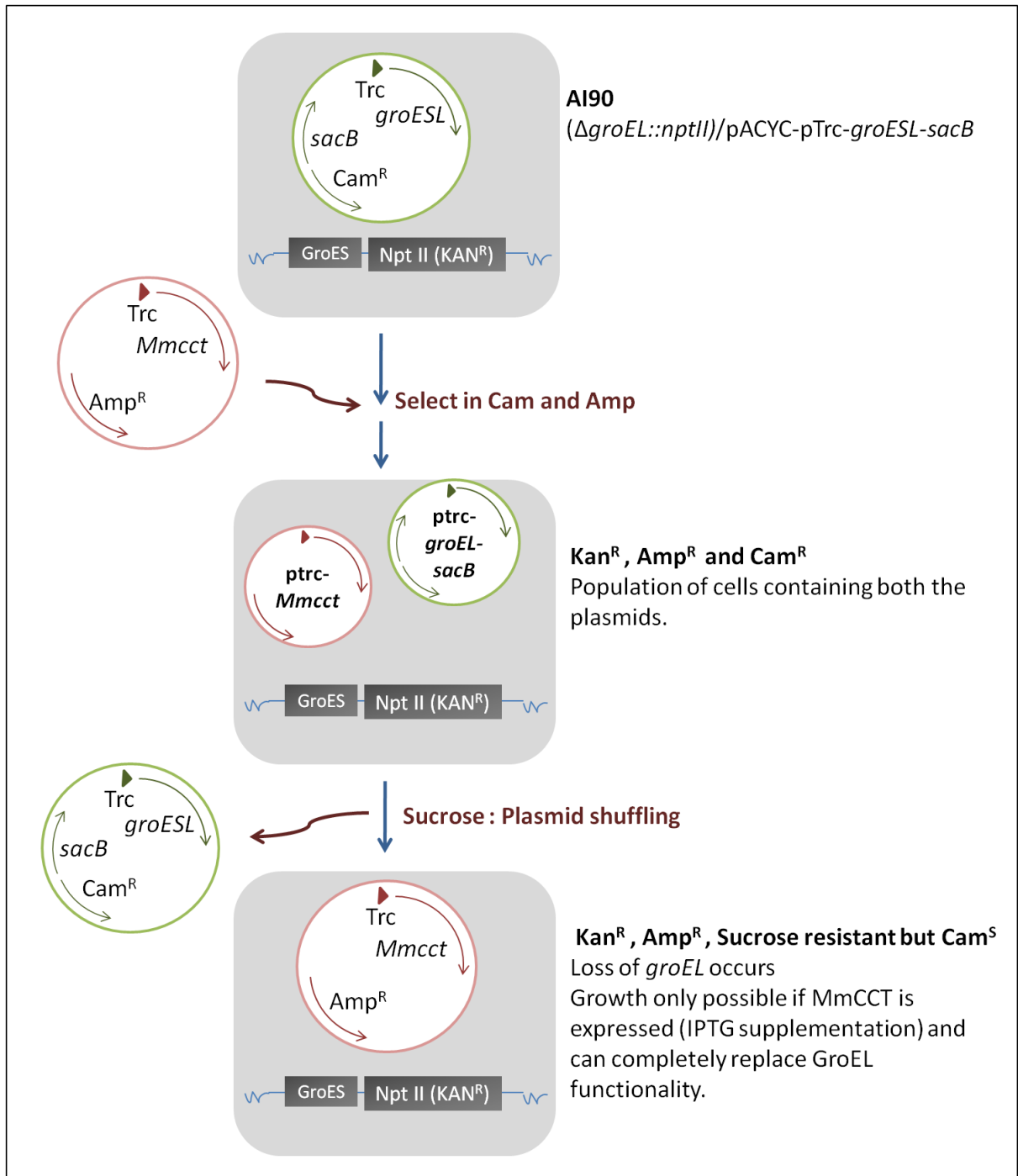
Figure 3.18
Schematic diagram explaining plasmid shuffling in AI90 strain
(Diagram adapted from Chapman *et al.*, 2006)

Table 3.2: Expected results in plasmid shuffling experiments

Components in LB agar plate	Expected observations
Kan-Cam-Amp	Growth expected for all as the two plasmids co-exist
Kan-Sucrose-Cam-Amp	In presence of sucrose, plasmid shuffling occurs. Hence, if pTrc-*sacB-cam^r-groesl* plasmid is displaced by the new plasmid, no colonies should appear because of chloramphenicol sensitivity.
Kan-Sucrose-Amp (KSA)	Growth should not occur because there is no replacement for GroEL.
Kan-Sucrose-Amp-IPTG (KSAI)	Only if MmCCT-WT or the mutants can replace the functioning of GroEL completely, growth can occur.

The positive control pTrc-*groEL* behaved as expected giving colonies on plates containing sucrose only (KSA) and sucrose + IPTG (KSAI). The pTrc99A empty vector did not give any colonies. Also, no colonies were observed for MmCCT-WT and MmCCT-M223I at any temperature suggesting that they are incapable of supporting growth in complete absence of GroEL. However, AI90 cells with MmCCT-K216E mutant showed distinct colonies after 3 days on KSAI plates at 26^0C and 30^0C. Also approximately after 5 days a few colonies were observed at 37^0C (Table 3.3).

The same experiment was carried out with selected mutants from K216n library and results were recorded as shown in Table 3.4. AI90 cells expressing MmCCT-K216D, K216Q and K216S gave similar number of colonies as K216E. MmCCT-K216A and K216L showed poorer growth however MmCCT-K216R showed no growth consistent with the results obtained for TAB21 analysis.

Table 3.3: **Analysis of MmCCT functionality in GroEL-knockout strain AI90**

	GroEL	MmCCT WT	MmCCT M223I	MmCCT K216E	
Sucrose	++	-	-	-	26°C
Sucrose and IPTG	+++	-	-	++	
Sucrose	++	-	-	-	30°C
Sucrose and IPTG	+++	-	-	++	
Sucrose	++	-	-	-	37°C
Sucrose and IPTG	+++	-	-	+	

Table 3.4: **Analysis of MmCCT-K216n functionality in GroEL-knockout strain AI90**

	MmCCT K216E	MmCCT K216A	MmCCT K216D	MmCCT K216L	MmCCT K216Q	MmCCT K216R	MmCCT K216S	
Sucrose	-	-	-	-	-	-	-	30°C
Sucrose and IPTG	++	+	++	+	++	-	++	
Sucrose	-	-	-	-	-	-	-	37°C
Sucrose and IPTG	+	-	-	+	-	-	+	

Note: The results are indicative of average number of colonies obtained for each sample from three independent experiments. The number of cells transformed, amounts of DNA used and the dilutions plated were uniformly maintained to be of equal number. (Number of colonies between 0 and 50 (+), 50 and 100 (++), 100 and more (+++))

The colonies obtained were subjected to three confirmatory experiments to ensure loss of *groEL* and expression of MmCCT-K216E:

1. Checking chloramphenicol sensitivity

To eliminate any chances of background colonies appearing (arising due to a mutation in *sacB* gene making the cells sucrose resistant), the colonies were patch plated on chloramphenicol plates. No growth was observed on chloramphenicol containing plates confirming loss of *groEL-sacB-cam^R* containing plasmid.

2. Colony PCR using GroEL internal primers

Six representative colonies supposedly expressing only MmCCT-K216E from plates grown at temperatures 26^0C, 30^0C and 37^0C were subjected to colony PCR using GroEL internal primers. DH5alpha and AI90 strains were used as positive controls. Also colonies obtained from using the pTrc-*groESL* plasmid in the above experiments served as a positive control. As shown in Figure 3.19A, the *groEL* gene band was not detected in any of the test colonies.

3. Western blotting of the protein extract from cultures of the test colonies:

The colonies were cultured in LB containing sucrose and 1 mM IPTG at respective temperatures. While the cultures at 26^0C and 30^0C reached saturated levels in approximately 36 hours, the cultures at 37^0C ceased to grow after reaching an OD_{600} of 0.3. Even when left in the incubator for seven days the cultures did not go beyond an OD_{600} of more than 0.5. Protein samples were extracted from cultures grown at 26^0C and 30^0C and were subjected to western blotting using anti-GroEL and anti-thermosome antibodies. Protein extracts from DH5α cells that produce GroEL in normal quantities

served as a positive control for anti-GroEL selection (Figure 3.19B, Upper panel, Lane 1). Purified MmCCT-WT protein (obtained from Dr. Andrew Large) served as a positive control for anti-thermosome antibodies (Lower panel, Lane 4). The cultures containing MmCCT-K216E showed complete absence of GroEL protein (Figure 3.19B, Upper panel, Lane 2 and 3). At the same time the cultures showed presence of MmCCT protein (Lower panel, Lane 2 and 3) confirming that K216E mutant can support growth of cells lacking GroEL at 26^0C and 30^0C.
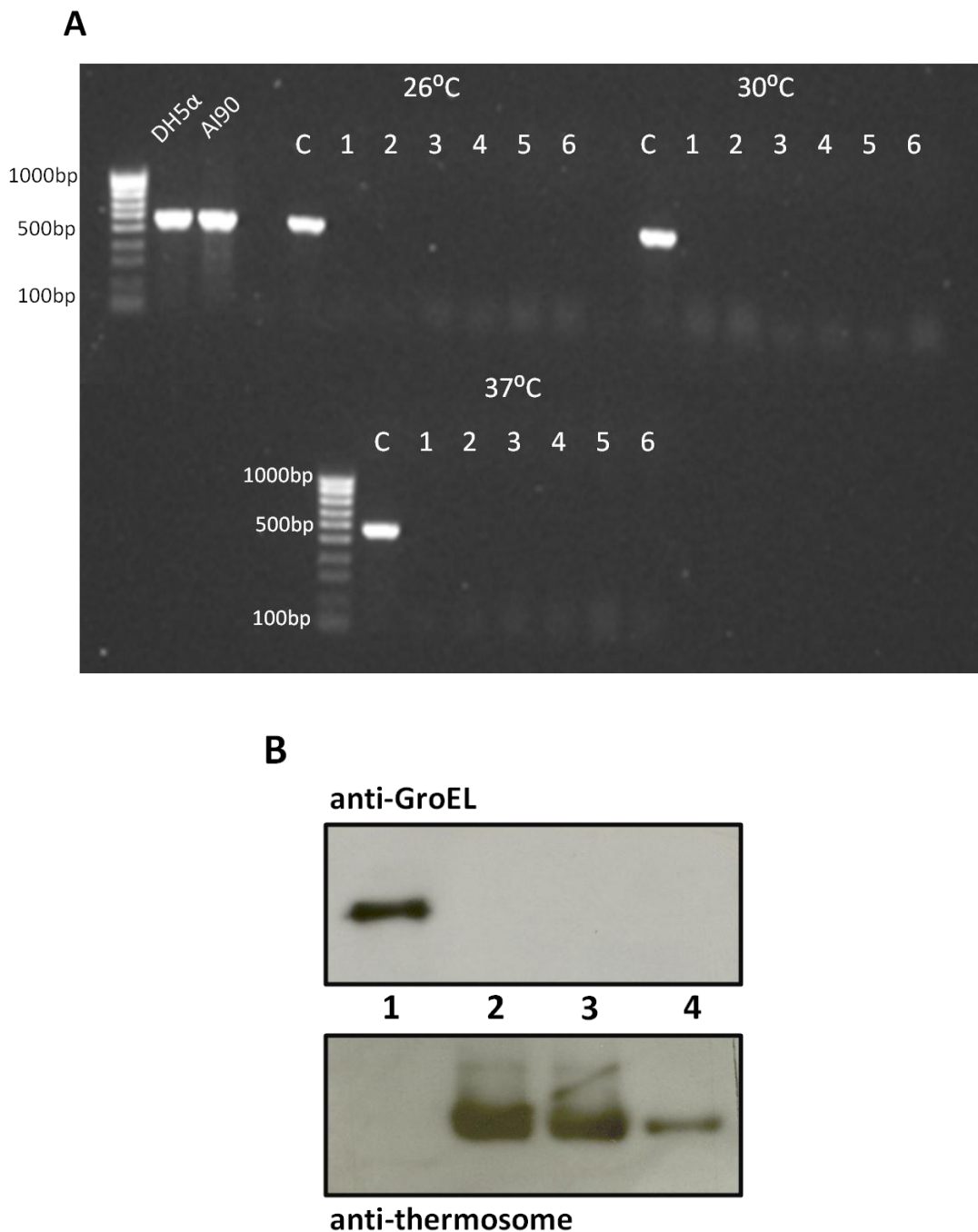
Figure 3.19

Confirmation of loss of *groEL* gene and GroEL protein from AI90 cells with MmCCT-K216E

(A) Colony PCR of AI90 cells containing MmCCT-K216E using groEL-gene internal primers. 'C' denotes positive control (AI90 colonies transformed with *ptrc-groEL*). #1 to #6 denotes six test colonies from plates grown at 25°C, 30°C and 37°C. **(B)** Western blots of soluble protein extracts from cultures using anti-GroEL (upper panel) and anti-thermosome (lower panel) antibodies. Lane 1: DH5α; Lane 2: AI90/ptrc-MmCCT-K216E (26°C); Lane 3: AI90/ptrc-MmCCT-K216E (30°C); Lane 4: MmCCT-WT purified protein (low concentration)

3.3 Discussion

Using MmCCT from archaeon *M. maripaludis* and GroEL from *E. coli* as examples, we have shown here that a Group 2 chaperonin can weakly replace the functioning of a Group 1 chaperonin at reduced temperatures. Further using random mutagenesis we established that point mutants M223I and K216E significantly improve the complementing ability by several fold and that the K216E mutant, but not M223I is able to completely replace GroEL function.

Both M223 and K216 reside in apical domain of the chaperonin where most sequence divergence is observed between Group 1 and Group 2 chaperonins (Kim *et al.*, 1994; Archibald *et al.*, 2001; Dekker *et al.*, 2011b). Because the apical domain harbours the client binding site, the simplest explanation for better functionality of MmCCT-M223I and K216E mutants could be improved client specificity. However, this is difficult to confirm as the client binding regions for Group 2 chaperonins are not explicitly defined. Previously, both hydrophobic residues in helix 10 and 11 (Feldman *et al.*, 2003; Spiess *et al.*, 2006) and hydrophilic residues lining the inner cavity (Pappenberger *et al.*, 2002; Gomez-Puertas *et al.*, 2004) have been implicated in client binding. It is interesting that M223I structurally corresponds to the F204 on GroEL, a residue involved in client binding for Group 1 chaperonins (Horwich *et al.*, 2007). In contrast, based on the solved crystal structure of MmCCT chaperonin, Pereira *et al.* put forward a hydrophobic pocket (consisting of L269, V273, A274, I276, L293, A294, Y297, L298 and I303) between helix 10 and helix 11 as the client binding region (Pereira *et al.*, 2010). Neither M223 nor K216 are in this region making the hypothesis that mutations of these residues improves client binding, less likely.

Further, we recognized that almost all amino-acid substitutions at M223 and K216 resulted in improved growth of TAB21 strain in *groEL* repressed conditions as compared to MmCCT-WT. This observation suggested a more general change in function of the mutants. One possibility we considered was that the mutants simply cause a loss of function that in turn produces a positive effect. If the loss of function is at the level of the chaperonin complex, the mutants should cease to function in an ATP dependent open-closed conformational cycle. We hypothesized that the M223I and K216E mutants locked in open conformations would bind to a large number of unfolded/misfolded proteins in the cell thus decreasing their amounts and increasing the viability of the cells (or delaying their death). However, the additional D386A mutation, that blocks the ATPase activity of MmCCT, severely affected the growth of the cells in *groEL* repressed conditions. The transition from the open to the closed conformation occurs due to ATP binding and hydrolysis. The results clearly suggest that the mutants M223I and K216E undergo a folding cycle to function. However the mechanism that enables the mutants to function in a better way than the wildtype remained unresolved.

It was observed that the TAB21 strain where the *groEL* expression should completely be repressed in presence of glucose still showed GroEL presence after 48 hours of growth in glucose containing medium. However, the amounts of GroEL present did not suffice for growth of the cells on solid medium (the strain itself does not grow on glucose containing plates; also the strain containing a pET empty vector ceases to grow on glucose containing plates). Supplementing MmCCT to such cells allows slow growth which further is improved by point mutants MmCCT-M223I and K216E. The data demonstrates that MmCCT can function as a chaperonin in *E. coli*, however it does not

prove that MmCCT can completely replace GroEL functioning. Indeed, when tested in a *groEL* knock out strain AI90, MmCCT-K216E was the only candidate to be able to support growth at 26^0C/30^0C and partially at 37^0C. While it can be concluded that MmCCT-WT and MmCCT-M223I cannot completely replace GroELS-function, the result with MmCCT-K216E is striking as it establishes that a Group 2 chaperonin, despite substantial structural differences, can replace the functioning of GroES and GroEL.

It should be noted, however, that the growth of AI90/pTrc-*Mmcct*-K216E was slower than AI90/pTrc-*groEL* (positive control)*. Moreover, despite the knowledge that the *ptrc* promoter is leaky and allows gene expression in absence of IPTG induction, MmCCT-K216E containing cells did not grow in absence of IPTG as against the positive control that showed growth in non-induced conditions. This indicates that large amounts of MmCCT-mutant protein are required for sustenance as compared to low quantities of GroEL that are generally sufficient (McLennan *et al.*, 1994; Ivic *et al.*, 1997). Furthermore, MmCCT-K216E is inefficient in replacing GroEL at 37^0C. The liquid cultures reach an OD$_{600}$ of 0.5 in 10 days. Although, it has been shown that GroEL and GroES are essential at temperatures as low as 17^0C (Fayet *et al.*, 1989). Such an observation reaffirms that the mutant MmCCT-K216E completely replaces the function of GroES and GroEL in *E. coli*. Nonetheless a number of possibilities could be considered for inefficient growth at 37^0C: **1)** The recombinant protein production of MmCCT could be more efficient at 30^0C as compared to at 37^0C. However such an explanation is less likely as MmCCT expresses at considerable levels at 37^0C in various strains tested (this work and past work in Lund group). **2)** The slower growth kinetics at 30^0C or lower can give the cells more time for MmCCT production and folding into its functional form (Assuming that MmCCT itself would be required to fold other MmCCT molecules). In

turn, the functional forms now can bring about the folding of the client proteins efficiently. At 37^0C, however, the equilibrium between the MmCCT functional molecules and the unfolded client proteins might be disturbed owing to faster and increased requirement of MmCCT protein. This could result in an increase of unfolded/misfolded molecules that cease the growth of the cells. **3)** The mechanism by which MmCCT operates could be different from GroEL. It has been hypothesized that Group 1 chaperonins are more involved in folding of clients that are aggregation-prone while Group 2 chaperonins are mainly operative on nascent chains and multi-domain proteins (Yam *et al.*, 2008; Hartl and Hayer-Hartl, 2009). It could be possible that the mechanism by which MmCCT operates is unable to efficiently rescue the aggregation prone proteins that, presumably, are higher in number at 37^0C.

A selection of K216n mutants (K216A, D, L, Q, R, and S) was also tested for their ability to sustain growth in complete absence of GroEL. Although there was not a major systematic bias observed in the biochemical properties of substituted amino acids consistent with previous analysis in TAB21, one notable result was for the mutant K216R which was unable to support growth in complete absence of GroEL. Even during TAB21 analysis the mutant K216R despite assembling into functional complexes could not improve the phenotype of the cells as compared to MmCCT-WT (Figure 3.13). Note that K216 faces inside of the chaperonin cage (Figure 3.7) and hence such a result points towards the possibility that loss of surface positive charge could facilitate folding of bacterial clients. Indeed, the wall of GroEL cavity is enriched with negatively charged residues (net surface charge is -42) in contrast to comparatively neutral distribution of MmCCT cavity (Cong *et al.*, 2010). Based on extensive mutational and biochemical analysis, the negatively charged clusters of the cage wall are strongly implicated to

facilitate folding of GroEL clients (Tang *et al.*, 2006; Chakraborty *et al.*, 2010). Moreover it has been shown that changing three negatively charged residues facing the inner cavity of GroEL to positive, resulted in strong reduction of cell growth (Tang *et al.*, 2008). This suggests the possibility of negatively charged residues having a role in efficient folding by GroEL, potentially by providing local repulsion to acidic proteins, predominantly found in *E. coli* (Horwich et al., 2009). Although the validity of these results has been challenged recently (Motojima *et al.*, 2012), it would be interesting to change the more positively charged residues facing the inner cavity of MmCCT to negative, ideally corresponding to the ones tested by Tang *et al.,* 2008 and analyze the effects on *E. coli* growth in absence of GroEL.

Nonetheless, the data clearly suggests that an archaeal chaperonin MmCCT can fold bacterial clients *in vivo* and the mutants MmCCT-M223I and K216E are more efficient in doing so. Moreover, the K216E mutant completely replaces GroEL functionality *in vivo* at 30^0C, suggesting that all of the bacterial essential clients can utilize MmCCT-K216E. However, the underlying reason for the improved function remained unclear. To answer this, the next step hence was to analyze the properties of mutants chaperonin *in vitro* the results of which are described in the next chapter.

In summary, we hereby establish a system that allows a Group 2 chaperonin to function in Group 1 host in absence of native Group 1 chaperonin. Such a system puts forward a medium for biochemical, mutational and interactome analysis of chaperonins with one-to-one comparison between the two groups that in turn could provide useful insights in understanding the divergence between the two groups. One of such applications has been studied and further described in Chapter 5.

CHAPTER 4
Biochemical characterization of MmCCT-M223I and MmCCT-K216E mutants

Note: This part of the project was carried out in collaboration with Dr. Jörg Martin's group, Max Planck Institute of Developmental biology, Tübingen, Germany. The work on mutant M223I was done by Dr. Astrid Ursinus and for the mutant K216E was done by me either during or after my visit to Max Planck Institute in Oct-Nov, 2011. The following table provides information regarding the same.

	Experiment	Responsible personnel	Figure No.
1	Protein Purification M223I	Dr. Astrid Ursinus	4.7C
2	Protein Purification K216E	Riddhi Shah	4.7B
3	Protein Purification K216E, again	Dr. Astrid Ursinus	4.7C
4	Assays	Dr. Astrid Ursinus	
	1. Protease protection assay		4.11
5	Assays	Riddhi Shah	
	1. Citrate synthase binding assay		4.8
	2. ATPase assay		4.9
	3. Rhodanese refolding assay		4.10
	4. Protease protection assay		4.11B

4.1 Background

In Chapter 3, chaperonin mutants MmCCT-M223I and MmCCT-K216E were described that partially or completely replace the function of GroEL in *E. coli* respectively at 30^0C. While the *in vivo* assays indicated that both MmCCT-M223I and MmCCT-K216E are functional chaperonins, *in vitro* data was essential to support the results and to attempt to understand the improved GroEL-complementing ability of the mutants as compared to the wildtype. In this chapter we analyse the biochemical properties of the mutants and compare it with those of the wildtype in order to understand the differences if any, that may in turn explain the better functionality of these mutants.

The additional properties gained due to the point mutations M223I and K216E could change the biochemical behaviour of chaperonin oligomers at one or more of the following stages: 1. Ability to bind and encapsulate bacterial client proteins, 2. Change in ATP turn over i.e. rate of hydrolysis of ATP, 3. Change in the ability to refold clients for various reasons for example change in the charge distribution of inner cavity, 4. Nucleotide-dependent conformational changes. All these stages are testable *in vitro* as has been described previously for wildtype MmCCT protein (Kusmierczyk and Martin, 2003b). Thus the aim of this part of the project was to purify MmCCT-M223I and MmCCT-K216E proteins and test their ability to bind and fold clients, examine the rates of ATP hydrolysis, and to check the nucleotide-dependent conformational cycle as compared to the wildtype-MmCCT. Each stage has been described in detail while discussing the results in the next section.

4.2 Results

4.2.1 Purification of MmCCT-K216E mutant

The purification of the MmCCT-K216E mutant was carried out in three steps, first two steps of anion exchange chromatography followed by a gel filtration step (Section 2.8.4, Materials and Methods). The BL21 (Gold) cells containing pET21-*Mmcct*-K216E were induced with 1 mM IPTG for 4 hours and correct expression checked using SDS-PAGE. After cell-lysis, the supernatant was loaded onto pre-equilibrated self-packed QHP anion exchange column. The elution was carried out with a linear gradient of 0-50% buffer B (30 mM Tris pH 7.5, 1 M NaCl, 5 mM MgCl$_2$). Based on the obtained chromatogram (Figure 4.1), the selected fractions were run on SDS-PAGE gel and Native-PAGE gel (Figure 4.2A and 4.2B respectively). The signal of absorbance (dark blue) and conductivity (light blue) can be seen in Figure 4.1, where the initial peak is indicative of non-specific proteins and the small peak just after that of MmCCT-K216E being eluted at approximately 30% of salt concentration.

The fractions (22-32) were pooled, concentrated and its salt concentration adjusted by equalizing conductivity to buffer A (30 mM Tris pH 7.5, 100 mM NaCl, 5 mM MgCl$_2$) using conductivity-meter. The resultant solution was subjected to the next round of anion exchange chromatography using self-packed MonoQ olumn. The elution was carried out in a similar manner as above. Because of a manual error in inserting the total volume to be used over the gradient of buffer B, the absorbance peak was observed very quickly owing to 40 – 50% of buffer B concentration in just first few fractions (Figure 4.3). Nonetheless, these fractions were run on an SDS-PAGE gel (Figure 4.4) and fraction 2 containing MmCCT was used for gel-filtration chromatography.
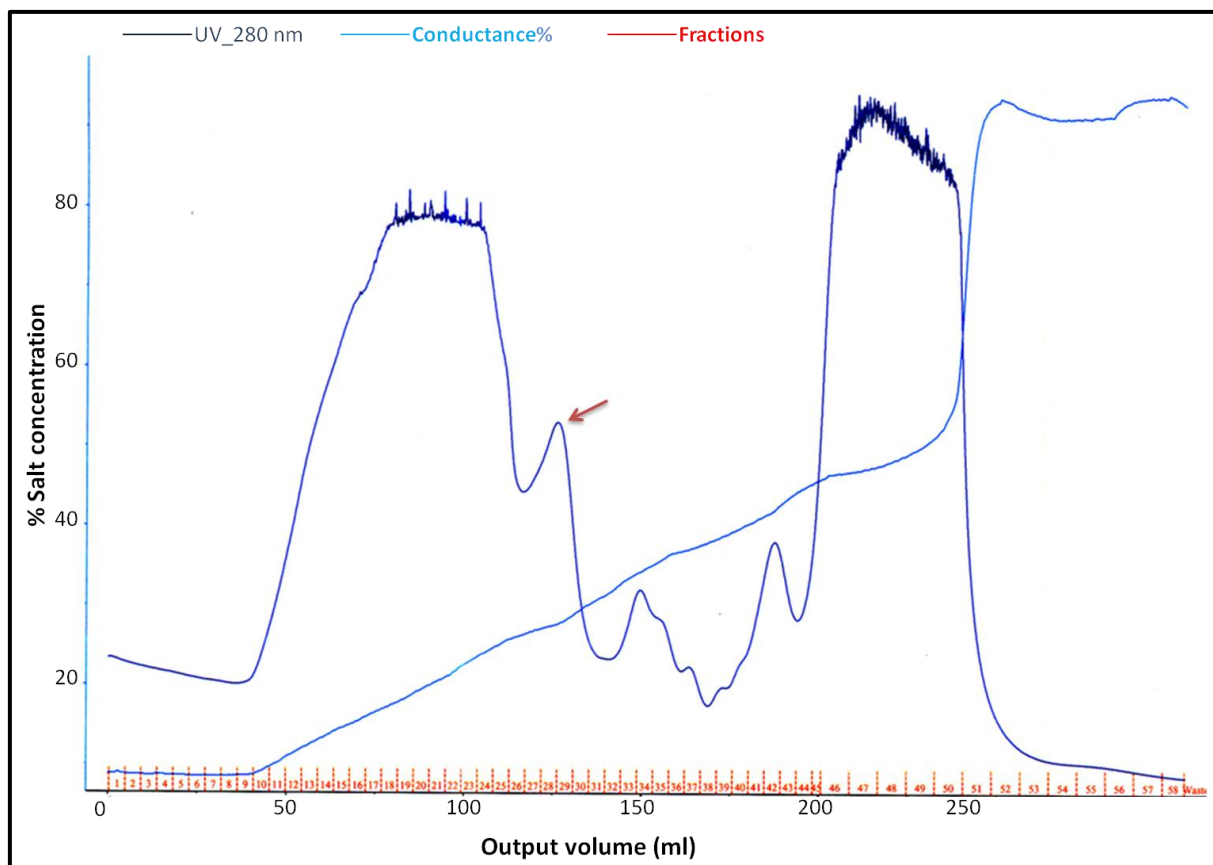
Figure 4.1

Chromatogram of QHP anion-exchange chromatography for purification of MmCCT-K216E

UV absorbance signal is seen in dark blue and the salt concentration is seen in light blue. The arrow shows the peak where MmCCT-K216E was eluted.
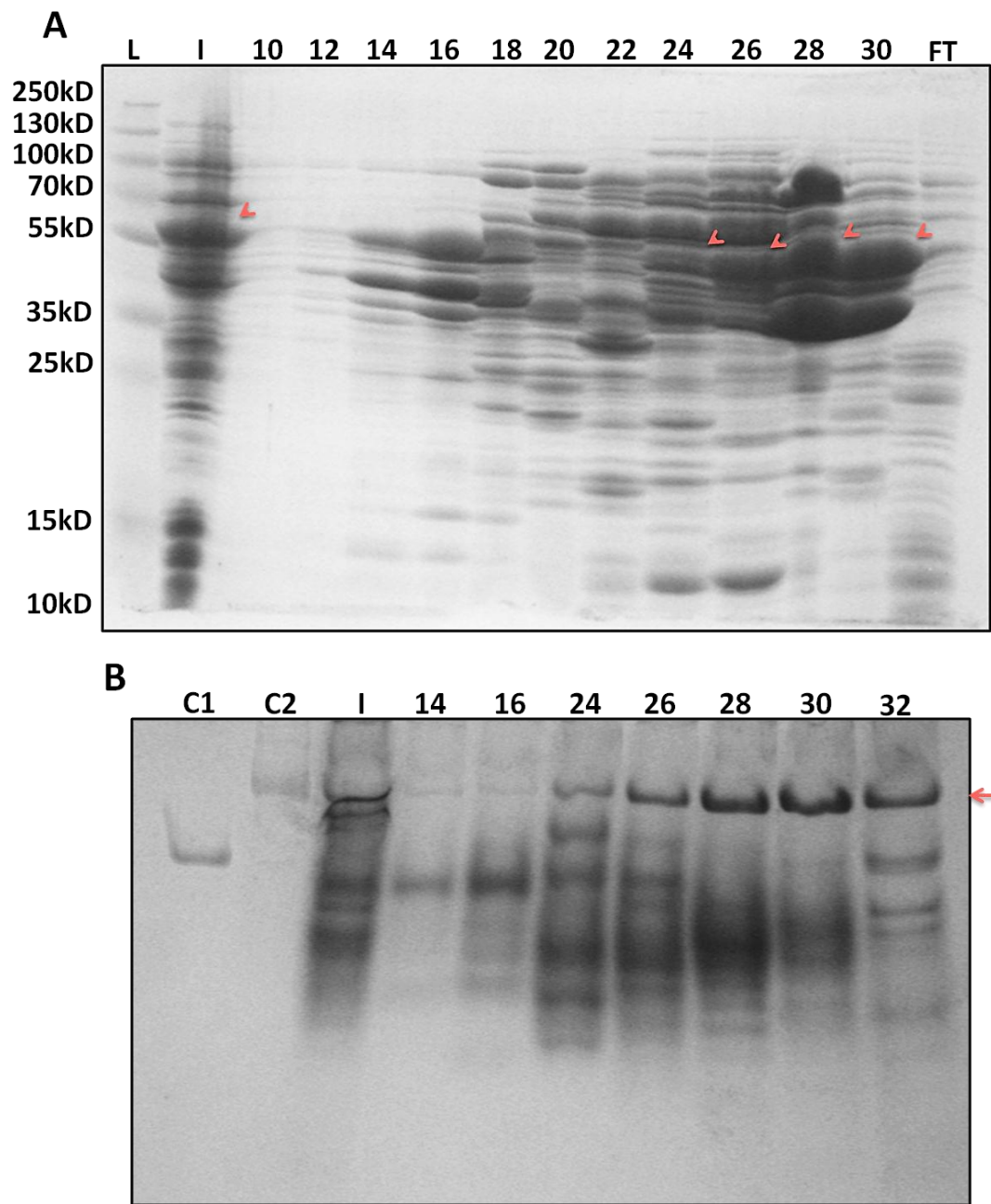
Figure 4.2

SDS-PAGE and Native-PAGE analysis of eluted fractions of MmCCT-K216E after QHP anion-exchange chromatography

(A) Alternative fractions (number 10-30) corresponding to the first two peaks in Figure 4.1 as analysed by 12% SDS-PAGE and **(B)** The fractions containing MmCCT-K216E (14, 16, 24-32) selected from SDS-PAGE gel and analysed by 3-10% gradient Native-PAGE with purified GroEL (C1) and purified MmCCT-WT (C2) as controls. 'I'indicates the initial fraction that was introduced into the column and 'FT' indicates flow-through after elution. Arrows indicate the position of MmCCT-K216E on the gels.
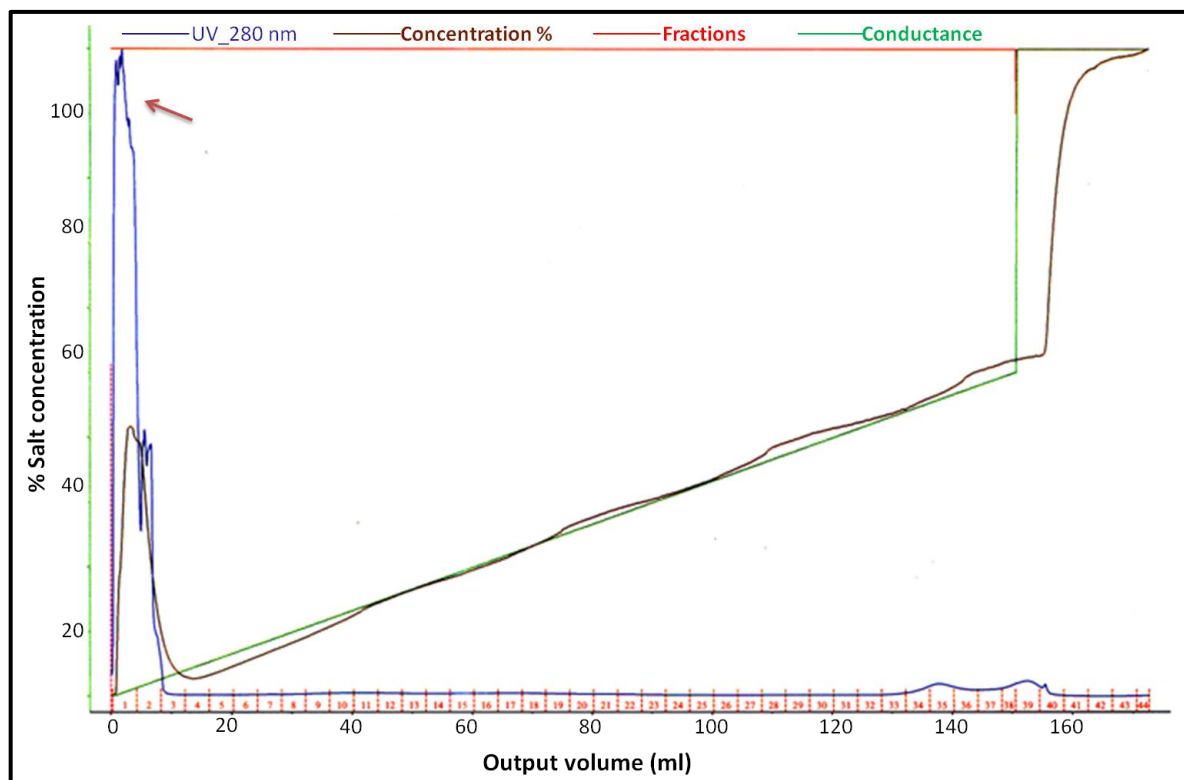
Figure 4.3
Chromatogram for MonoQ anion-exchange chromatography

UV absorbance signal is seen in dark blue and the salt concentration is seen in light orange. The arrow shows the peak where MmCCT-K216E was eluted.
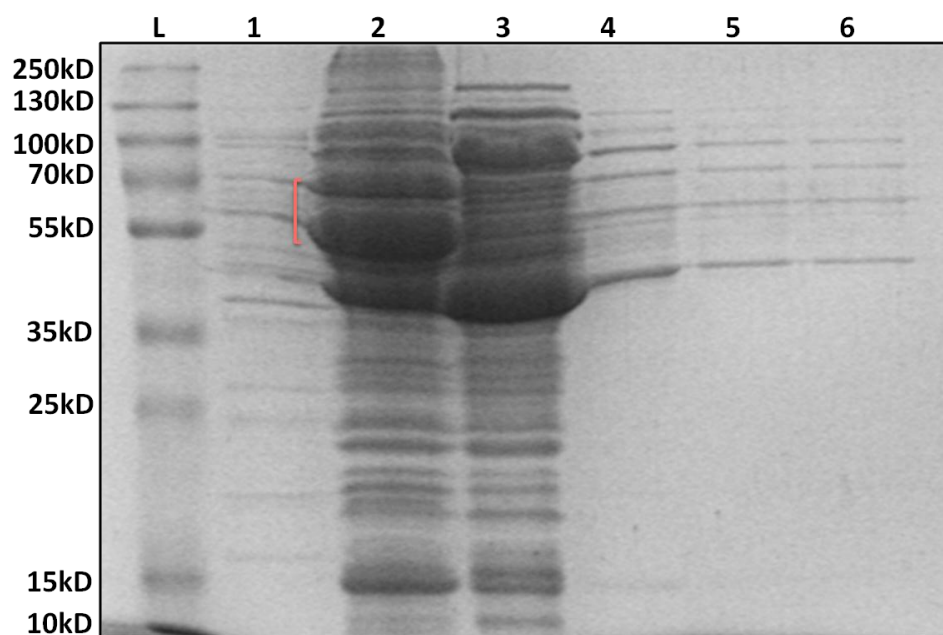
Figure 4.4

SDS-PAGE analysis of eluted fractions of MmCCT-K216E after MonoQ anion-exchange chromatography

Fractions (number 1-6) corresponding to the peak observed in Figure 4.3 as analysed by 12% SDS-PAGE and Coomassie staining. Red bracket indicates the MmCCT protein.

Gel filtration chromatography was carried out using S-300 column (320 ml bed volume) and samples were eluted with buffer C (30 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl$_2$, and 1 mM DTT). The fractions corresponding to the two peaks observed on the chromatogram (Figure 4.5) were run on SDS-PAGE gel (Figure 4.6A and B) and native-PAGE gel (Figure 4.7A). The appropriate fractions corresponding to a single peak (12 and 13) were pooled and concentrated. However when higher concentration of the purified protein was run on native-PAGE gel (Figure 4.7B), it was observed that the preparation was not satisfactorily pure especially with the presence of endogenous GroEL. Although the biochemical assays were conducted in Germany using this protein preparation, it was essential to ensure that any chaperonin activity that was observed subsequently in the assays was because of MmCCT itself and not due to contaminating GroEL. As a result the purifications were carried out again for MmCCT-K216E along with MmCCT-M223I by Dr. Astrid Ursinus following the same protocol and the samples, now satisfactorily pure (Figure 4.7C), were sent to Birmingham where the assays were developed and conducted.
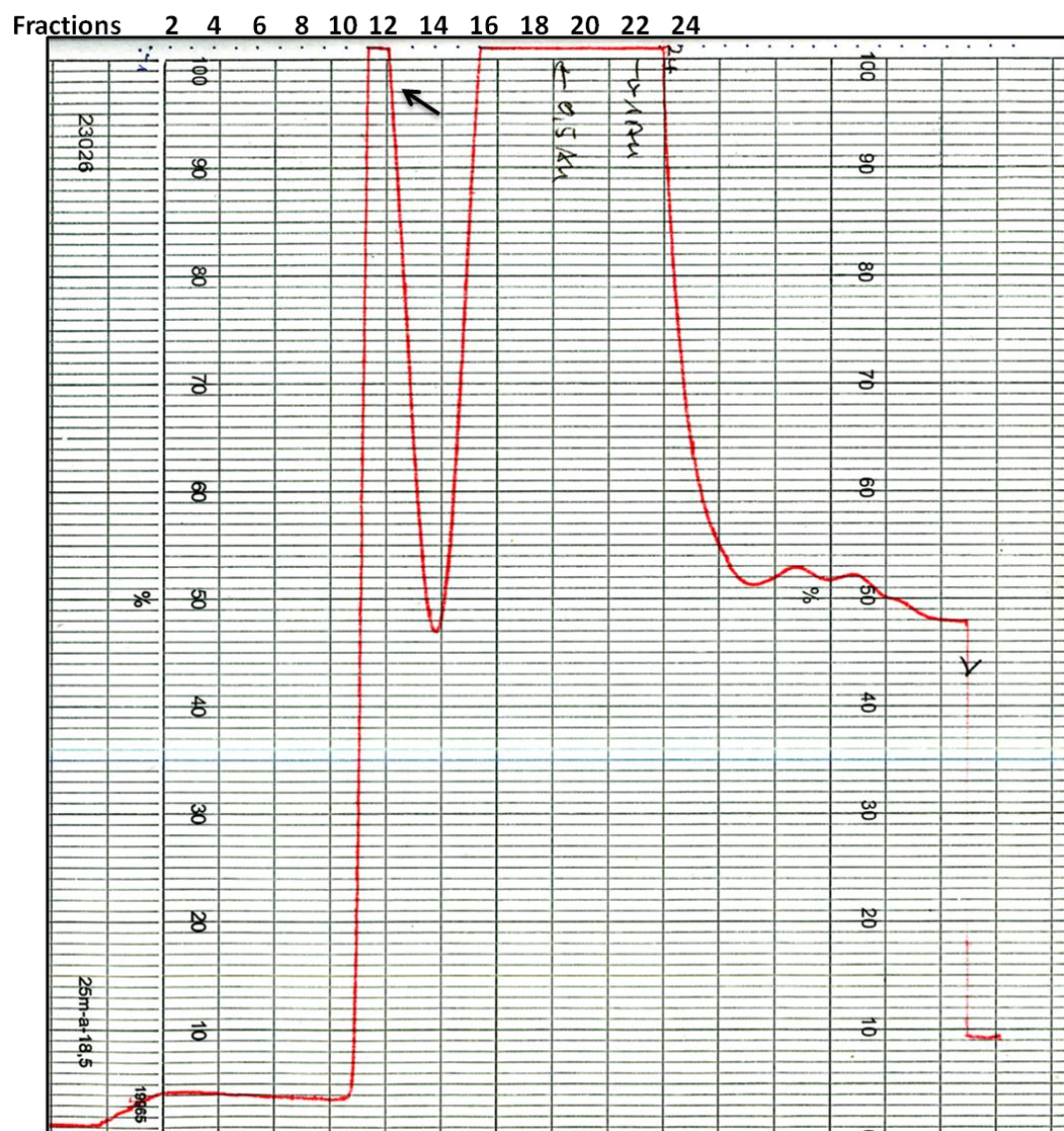
Figure 4.5

Chromatogram of S300 gel filtration analysis for the purification of MmCCT-K216E

UV absorbance signal is seen in red. The arrow shows the peak where MmCCT-K216E was eluted.
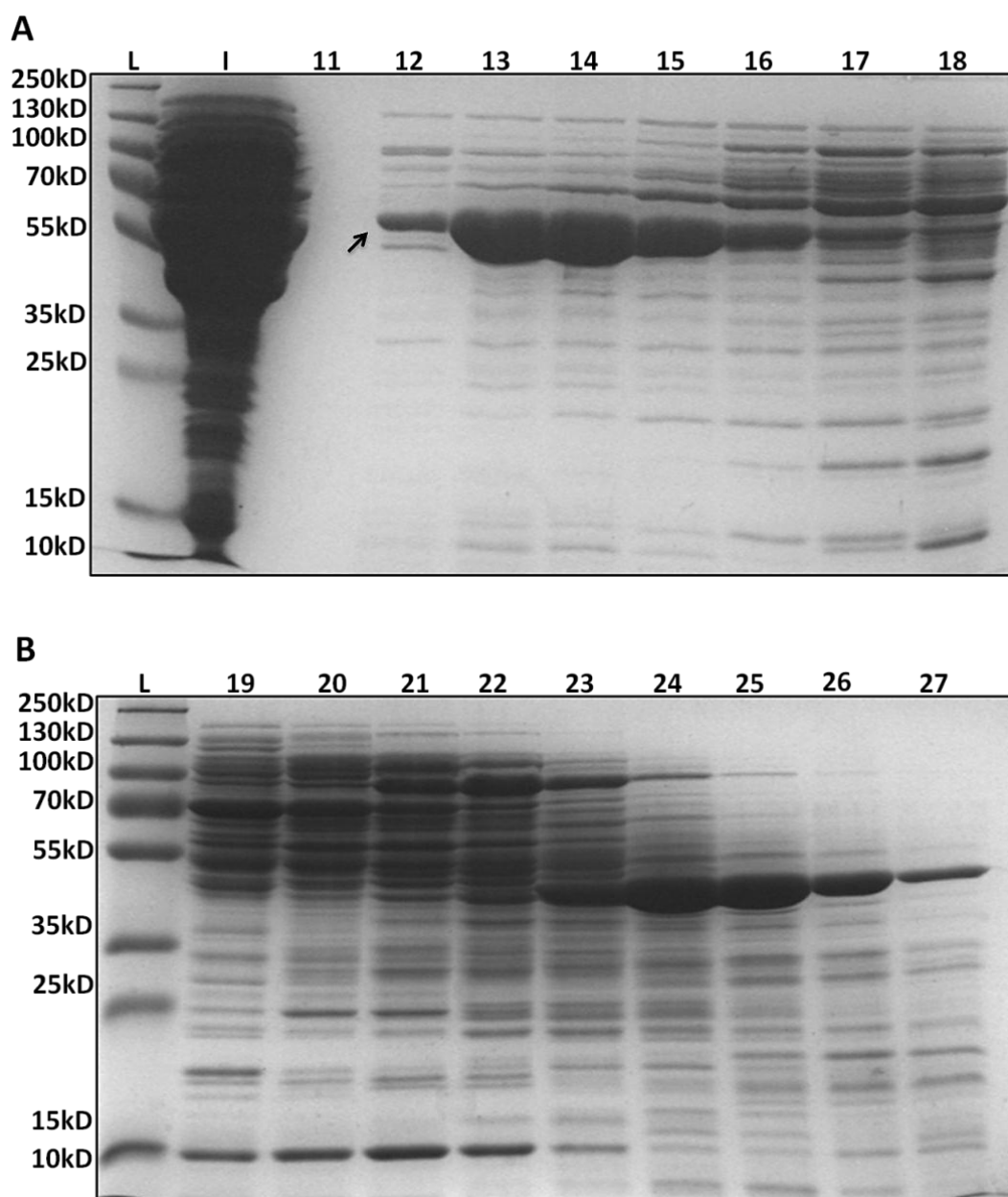
Figure 4.6

SDS-PAGE analysis of eluted fractions of MmCCT-K216E after gel filtration chromatography

Fractions (number 11-27) corresponding to the peaks observed in Figure 4.5 as analysed by 12% SDS-PAGE. 'I' indicates the initial fraction that was introduced into the column and the arrow indicates the position of MmCCT-K216E on the gel.
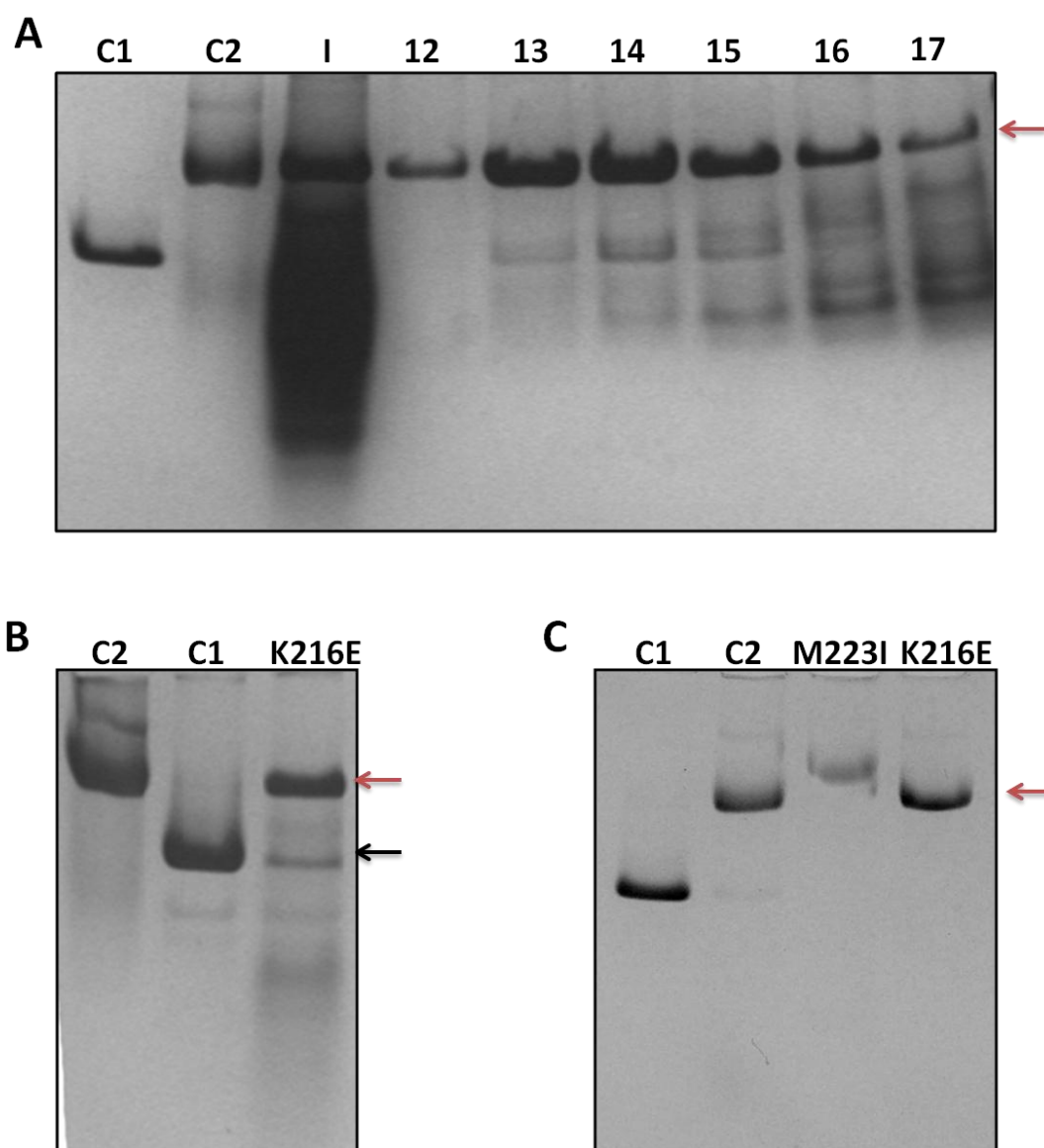
Figure 4.7

Native-PAGE analysis of eluted fractions after gel filtration chromatography and of purified protein MmCCT-K216E

(A) The fractions containing MmCCT-K216E (number 12-17) selected from SDS-PAGE gel were analysed by 3-10% gradient Native-PAGE. 'I'indicates the initial fraction that was introduced into the column.

(B) The purified protein MmCCT-K216E as analysed by Native-PAGE.

(C) Purified proteins MmCCT-M223I and MmCCT-K216E as analysed by Native-PAGE. The gel was obtained from Dr. Astrid Ursinus.

Purified GroEL (C1) and purified MmCCT-WT (C2) were used as controls to determine the position of bands on the gels. The red and black arrows indicate the position of MmCCT-K216E and GroEL respectively.

4.2.2 Client-binding ability of MmCCT-M223I and MmCCT-K216E

We first tested the ability of MmCCT mutants to bind to porcine heart citrate synthase as a non-native model substrate in an aggregation prevention assay as described previously (Kusmierczyk and Martin, 2003b). The mutants MmCCT-M223I and MmCCT-K216E were incubated with a buffer containing citrate synthase at 50^0C. Protection against aggregation of the enzyme was measured spectrophotometrically at 320 nm (Section 2.6.8.1, Materials and Methods).

Incubation of citrate synthase at 50^0C resulted in rapid increase in OD_{320} of the solution over a period of 30 minutes (Figure 4.8A). However presence of equimolar concentrations of MmCCT-WT, MmCCT-M223I or MmCCT-K216E was sufficient to suppress the aggregation (Figure 4.8A for MmCCT-K216E). The molar concentrations were calculated with respect to the 931 KDa MmCCT complex. The experiments were first carried out in a time dependent manner and then the measurements were taken after 30 minutes as an end point for the assay (Figure 4.8B). MmCCT-mutants along with controls GroEL and MmCCT-WT suppressed light scattering of citrate synthase at 50^0C indicating effective binding to the enzyme. The results indicated no obvious difference between the proteins for their client binding efficiencies by this assay.

A

Citrate synthase aggregation assay: Time curve

Absorbance (OD$_{320nm}$) vs Incubation time (mins)

Legend:
- CS
- CS+K216E

B

Citrate synthase aggregation endpoint assay

Absorbance (OD$_{320nm}$)

Categories: CS, CS + GroEL, CS + WT, CS + M223I, CS + K216E
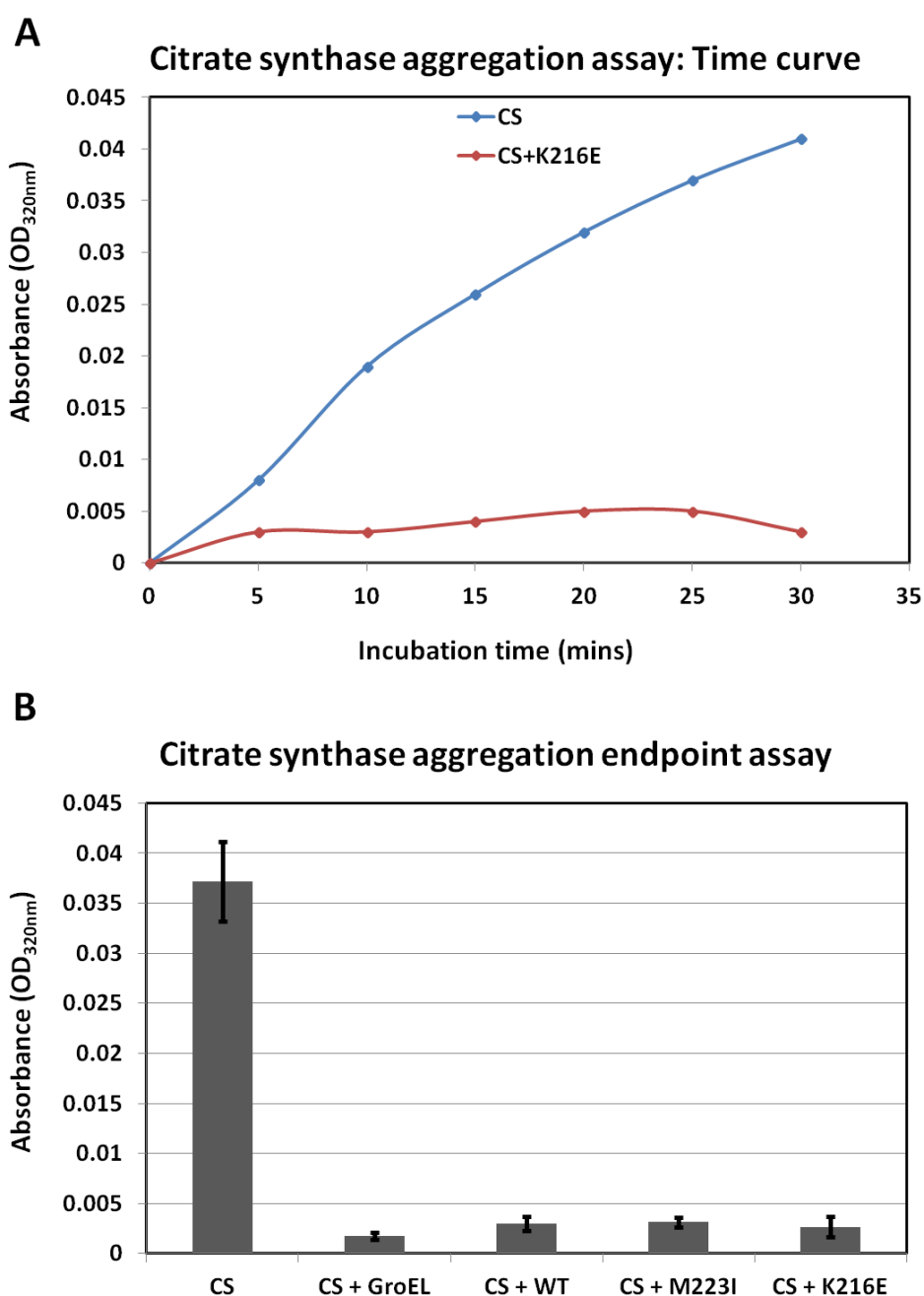
Figure 4.8

Binding ability of MmCCT mutants as assessed by citrate synthase aggregation assay

Citrate synthase (CS) in absence or presence of the various proteins as mentioned was incubated at 50°C and the increase in aggregation was measured at 320 nm over a time course of 30 minutes **(A)** or *at* 30 minutes as an end point for the assay **(B)**. The proteins and CS were used in equimolar concentrations. The data represent averages of three independent experiments.

4.2.3 ATPase activity of MmCCT-M223I and MmCCT-K216E mutants

A defining characteristic of all chaperonins is their ATP hydrolysis activity. The GroEL-complementing ability of MmCCT-M223I and K216E mutants was hindered when an additional D386A mutation was incorporated (Chapter 3) suggesting that the mutants were functional ATPases. The same was assessed *in vitro* using the malachite-green molybdate ATPase assay (Section 2.8.6.2, Materials and Methods). The protein and the appropriate controls were incubated at 37^0C with 1 mM ATP for 15 minutes in applicable buffer and the release of phosphate was measured calorimetrically at 640 nm. The amount of liberated inorganic phosphate for MmCCT increases linearly with time and an incubation for 15 minutes has been established as a standard for the assay (Kusmierczyk and Martin, 2003a and J. Martin, personal comm.). The readings obtained for spontaneous hydrolysis of ATP were deducted from the readings obtained for samples containing MmCCT proteins. The readings obtained for MmCCT-WT were set as 100% and the observations for MmCCT-M223I and MmCCT-K216E were plotted as a percentage of MmCCT as shown in Figure 4.9.

It was observed that the mutants can effectively hydrolyze ATP. The mutant MmCCT-K216E exhibited a fairly similar rate of ATP hydrolysis as compared to the wildtype whereas, comparatively, a significant drop in activity was observed for the mutant MmCCT-M223I. Similar results were consistently obtained for three independent experiments.
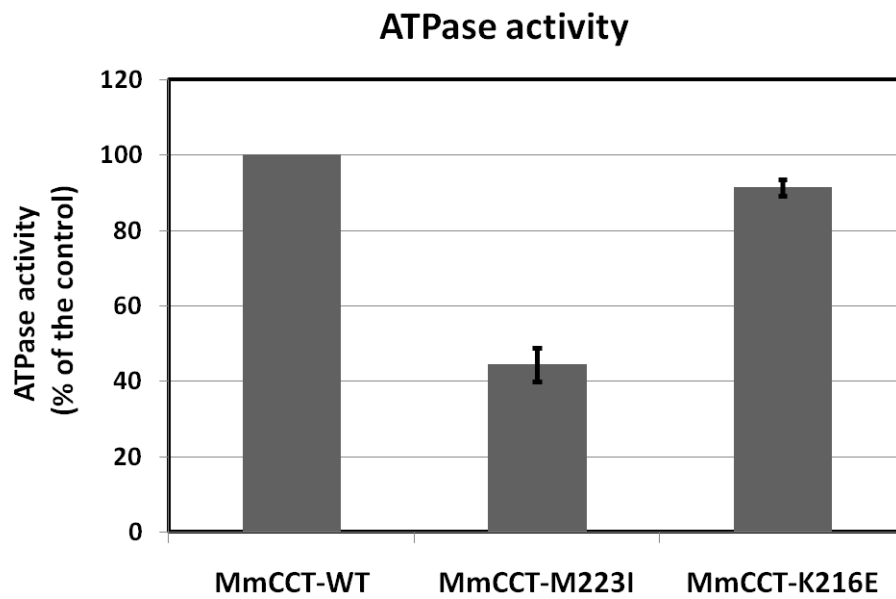
Figure 4.9

ATP hydrolysis activity of MmCCT mutants

The MmCCT proteins were incubated in appropriate buffer at 37°C and the ATPase activity was initiated by adding 1 mM ATP and continued for 15 minutes. The release of phosphate was assayed by malachite green-molybdate method and the absorbance at 640 nm was recorded. The observations for MmCCT mutants are plotted as percentage of the wildtype.

4.2.4 Client refolding capacity of MmCCT-M223I and MmCCT-K216E

The enzyme rhodanese has been widely used as a model-client to test the *in vitro* refolding ability of both GroEL (Mendoza *et al.*, 1991; Martin *et al.*, 1993) and MmCCT (Kusmierczyk and Martin, 2003b). When rhodanese is chemically denatured, it aggregates and becomes inactive. However, presence of chaperonins under appropriate conditions leads to recovery of native Rhodanese that can be assayed.

Rhodanese was denatured using 6 M guanidinium chloride (GuHCl) and incubated with the MmCCT wildtype and mutant proteins in presence of ATP as described in section 2.8.6.3 of Materials and Methods. The reaction was followed for 30 minutes and activity of refolded rhodanese was assayed by conversion of cyanide into thiocynate followed at 480 nm. As seen in Figure 4.10, approximately 60% to 70% recovery of native rhodanese was observed for MmCCT-M223I and K216E. There was not a significant difference in the values obtained for the mutants as compared to those obtained for the wildtype as observed here or reported previously (Kusmierczyk and Martin, 2003b). However MmCCT-M223I consistently showed a relatively lower refolding activity when compared to MmCCT-K216E which in turn showed activity that was almost similar to the wildtype.
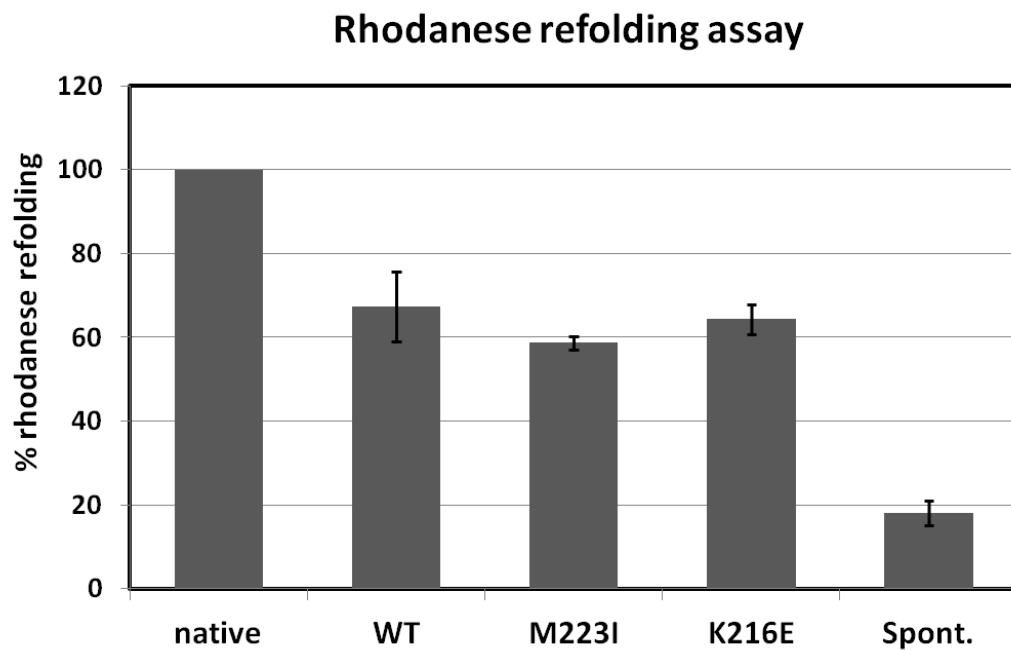
Figure 4.10

Rhodanese refolding ability of MmCCT mutants

GuHCl-denatured rhodanese was incubated in absence or presence of various proteins as mentioned in appropriate buffer containing 1 mM ATP at room temperature for 30 minutes. The observations for MmCCT proteins are plotted as a percentage activity of native rhodanese.

4.2.5 Protease sensitivity test for M223I and K216E mutants

Chaperonins exhibit different conformations at different stages of the folding cycle. To determine the asymmetric conformations of chaperonin complexes in presence of different nucleotides a simple assay based on susceptibility of digestion by a protease has been described (Langer *et al.*, 1992; Martin, 2000) and extensively used for both Group 1 and Group 2 chaperonins. In its *open* conformation, MmCCT-WT is susceptible to protease digestion because of the accessibility to the highly dynamic lid region (Reissmann *et al.*, 2007). However in the presence of ATP, the oligomer undergoes the conformational cycle and transiently attains a closed conformation. The closed state buries the sites that are susceptible to protease cleavage and hence leads to protection against protease digestion. Two considerations were made here in order to observe protease digestions effectively. Firstly, higher ATP concentration (1 mM) was used for the experiment. Higher concentration of ATP (more than 0.4 mM) has been shown to overcome the negative cooperativity between the rings resulting into a symmetrically closed complex that theoretically should be fully protected against protease treatment (Kusmierczyk and Martin, 2003b; Meyer *et al.*, 2003). Secondly, the gamma phosphate mimic AlFx was also used along with ATP. AlFx has been shown to stabilize the closed conformations leading to more efficient protease protection (Meyer *et al.*, 2003; Iizuka *et al.*, 2005). (Figure 4.11**A**, number 3).

If the point mutations M223I and K216E led to conformational changes of the complex or an alteration of time spent in the different conformations, the pattern of proteolytic digestion would be different than the wildtype. Both the mutants were subjected to Proteinase K sensitivity assay (section 2.8.6.4, Materials and Methods). Briefly, the

purified proteins were incubated with proteinase K at room temperature for 10 minutes in absence or presence of nucleotides ADP, ATP, AMP-PNP and ATP-AlF$_x$ and analysed by SDS-PAGE to check the extent of digestion.

The results showed that the mutant MmCCT-M223I was rather weakly protected against protease digestion (Figure 4.11A) suggesting that the mutant spends more time in open conformation rather than closed, consistent with its lower ATPase activity. The mutant MmCCT-K216E on the other hand was partially protected against proteolytic cleavage in presence of ATP and to a greater extent in presence of ATP-AlF$_x$ similar to the observations for wildtype protein (Figure 4.11B). Note that when the non-hydrolysable ATP analogue, AMP-PNP, was used, effective protease protection was not achieved suggesting that ATP hydrolysis is important to attain the closed conformation consistent with previously reported results (Reissmann *et al.*, 2007) (Figure 4.11B, number 4).
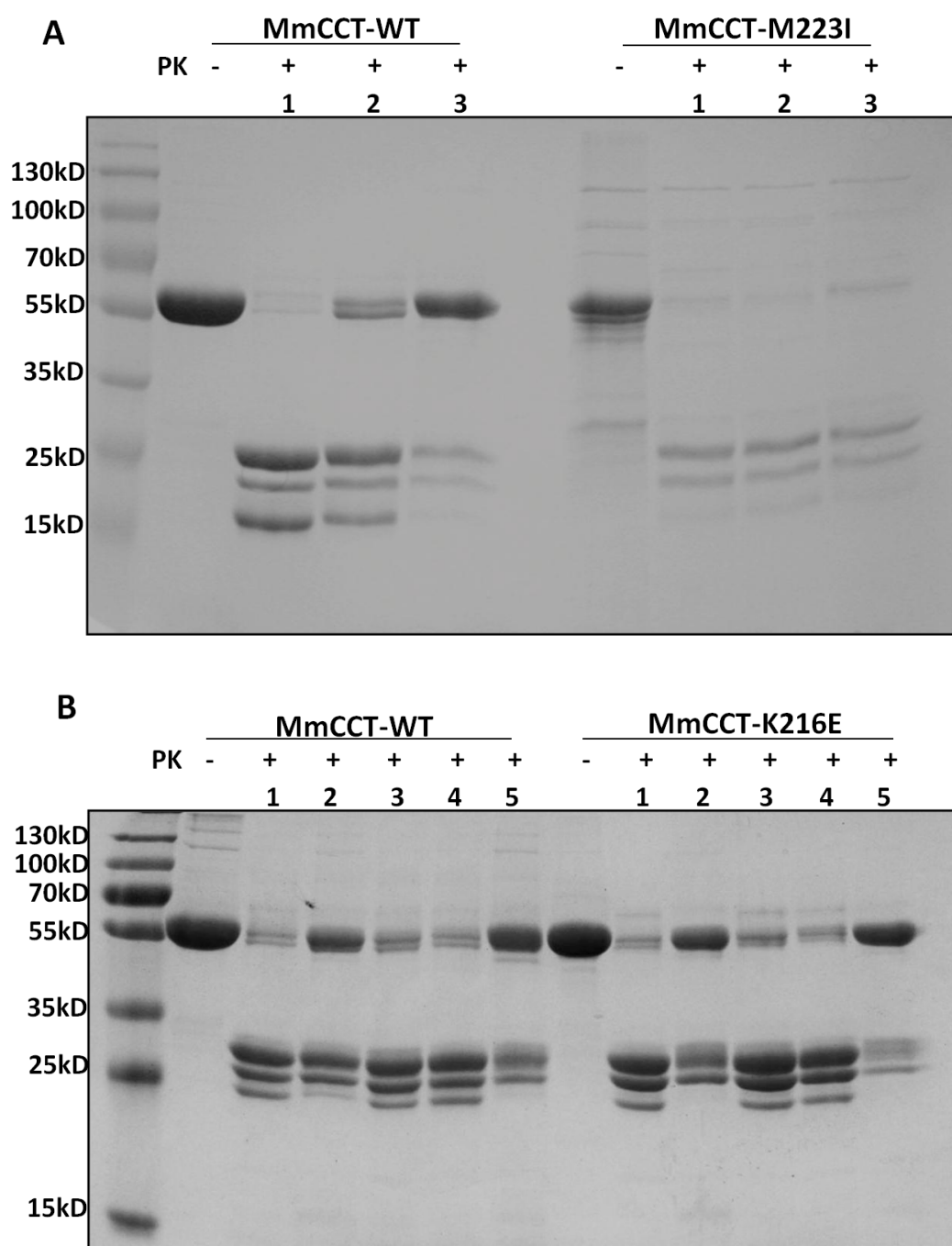
Figure 4.11

Protease sensitivity test for MmCCT-M223I and MmCCT-K216E

Protection against Proteinase K (PK) for MmCCT-WT, MmCCT-M223I and MmCCT-K216E as observed by SDS-PAGE: **(A)** Analysis of MmCCT-WT and MmCCT-M223I without (-) or with (+) PK in absence of nucleotide (1) or presence of ATP (2) and ATP-AlF$_x$ (3). **(B)** Analysis of MmCCT-WT and MmCCT-K216E without (-) or with (+) PK in absence (1) or presence of ATP (2), ADP (3), AMP-PNP (4) and ATP-AlF$_x$ (5). All gels stained by Coomassie staining.

4.3   Discussion

Group 2 chaperonin MmCCT from *Methanococcus maripaludis* and its mutants MmCCT-M223I and MmCCT-K216E rescue growth of *E. coli* when GroES and GroEL are depleting at 30^0C.   Moreover mutant MmCCT-K216E sustains *E. coli* growth in the complete absence of GroES and GroEL.   The additional properties acquired by the chaperonin complexes as a result of single point mutations M223I and K216E could be traced by investigating the biochemical properties of the mutant.   Thus experiments were conducted to purify the proteins and analyze them biochemically to understand their behaviour *in vitro.*

The purification of MmCCT was carried out with a method similar to described previously (Kusmierczyk and Martin, 2003b), which has been followed with minor modifications for purification of MmCCT in general (Reissmann *et al.*, 2007).   Two rounds of anionic exchange chromatography and a step of gel filtration, both at pH 7.5, were employed to purify MmCCT. Because the pI of MmCCT (4.97) and GroEL (4.85) are quite close to each other, anion exchange using strong anionic exchanger (quaternary resin MonoQ, in this case) has been found to be an essential step in separation of MmCCT and GroEL (Kusmierczyk and Martin, 2003b). Because of the use of a very steep salt gradient during this step, contaminating GroEL was found in the protein preparations. However subsequent purifications resulted in almost 95% pure proteins as determined by SDS-PAGE gels and Native-PAGE gels. These purified proteins were then subjected to biochemical assays.

The biochemical profile for MmCCT-M223I differed from that of wildtype and MmCCT-K216E. While the mutant was able to bind the client citrate synthase efficiently, the

rates of ATP hydrolysis and rhodanese-refolding were lower than corresponding rates of MmCCT-WT and MmCCT-K216E. Moreover MmCCT-M223I was not efficiently protected against protease digestion in presence of ATP or ATP-AlF$_x$. Using D386A mutants of MmCCT, it has been shown that ATP hydrolysis is absolutely required for attaining the protease-resistant closed conformation (Reissmann *et al.*, 2007). In this regard, it can be considered that the mutation M223I has a negative effect on ATP hydrolysis and hence the cavity-closure, particularly when lower ATPase activity was consistently observed for the mutant. Alternatively, it has also been reported that susceptibility to protease digestion is in general due to the lid region and that Δlid-MmCCT variants are protected even in absence of any nucleotides (Reissmann *et al.*, 2007). Because M223 does not lie in this region, it can be speculated that protease sensitivity results due to some secondary conformational changes or steric hindrances that in turn blocks the movement of lid region. The close proximity of M223 residue to the lid region (Figure 3.4) might be instrumental for such changes. Indeed structural studies like electron microscopy or crystallography are required to understand in detail the implications of this mutation on the oligomeric rearrangement and assembly.

In either case, it can be proposed that the mutant MmCCT-M223I resists the formation of closed conformation and tends to lock the chaperonin complex in its open state that is able bind to clients but not fold them. Such an action however is not fully achieved and the chaperonin does undergo albeit a slow ATPase cycle to bring about a certain degree of productive folding as demonstrated for rhodanese. This model is consistent with the initially described hypothesis (Chapter 3) that suggests that the improved GroEL-complementing ability of the MmCCT-M223I could be partly because of a loss of

chaperonin function and that the better phenotype results due to slower aggregation of the misfolded proteins that are transiently stabilized by binding to the mutants. In addition, folding of at least some clients (even with slow ATP turn-over) depend on MmCCT-M223I as completely obstructing the ATPase activity by an additional D386A mutation leads to cessation of growth (Chapter 3). The results thus suggest that both active folding of clients and slow aggregation kinetics of the misfolded proteins could probably contribute together, for improved GroEL-complementing ability of MmCCT-M223I. Future experiments should elucidate the findings better, for example by examining the growth effects when a non-native toxic protein known to aggregate in *E. coli* is over-expressed along with the mutant M223I.

The biochemical properties of MmCCT-K216E on the other hand were almost identical to that of wildtype. The mutant was able to bind the client, effectively hydrolyze ATP and refold rhodanese with efficiency comparable to that of the wildtype MmCCT. Also the mutant showed nucleotide dependent open and closed conformation cycle based on protease sensitivity assay. Because MmCCT-K216E was able to completely replace GroEL functionality, it was expected to behave differently in fact in a better way than the wildtype during *in vitro* analysis. However no significant differences were observed suggesting a similar reaction cycle and mechanism of folding clients.

Such results point towards the possibility that the improved functioning of MmCCT-K216E could be client-specific instead of better functioning of the chaperonin complex itself. The additional properties acquired by point mutant MmCCT-K216E could be a structural adjustment for either accommodating more bacterial clients than the wildtype or to increase the efficacy of folding of particular bacterial clients. Such a

proposition is in line with the previous results where mutation of K216 to any amino acid but a positively charged residue leads to better rescue of *E. coli* in GroEL and GroES limiting conditions. Elimination of positive charge at this residue thus perhaps allows more efficient interaction or folding of bacterial clients. The validity of these conclusions indeed requires further experimentation, an attempt for which has been made in the next chapter by elucidating interacting clients.

Of note, we also considered certain broader issues such as stability of the MmCCT mutant proteins and their total amount in the cell to be contributing towards their improved function. Careful experiments were done to compare the protein levels of the mutants to the wildtype with appropriate controls. The results showed no significant difference in their expression (Chapter 3 and results not shown). Also the melting temperatures (T_m) of both the mutants (63^0C) were found to be similar to that of the wildtype (65^0C) as deduced from circular dichroism melting curves (J. Martin, pers. comm.). This suggests that the mutants are as thermally stable as the wildtype protein and rules out the issues described above for better function in *E. coli.* The results however do not exclude more general effects arising due to mutations like cross-talks with other chaperone systems and/or other regulatory events which although are unlikely to play a role, would require a more elaborate scheme of experiments in a case where a direct answer to reason the better function of the mutants is not derived.

Chapter 5
Towards investigation of MmCCT clients and client binding site in *E. coli.*

Note: The results discussed in sections 5.3.2.1, 5.3.2.2 and 5.3.3.3 were done by undergraduate students Kate Wilkinson, Wing Chun, Kiren Baines, and Lynette Jude supervised by myself. The following table provides information regarding the same.

	Experiment	Responsible personnel	Figure No.
1	Generating V273n mutants library	Kate Wilkinson	-
2	Generating Y297n mutants library	Wing Chun	-
3	Expression and assembly of V273n mutants	Kiran Bains	5.16
4	Complementation of V273n mutants	Kiran Bains	5.17
5	Complementation of Y297n mutants	Lynette Jude	5.18

5.1 Background

In chapter 3, an *in vivo* system was described wherein MmCCT was shown to be active as a chaperonin in *E. coli* under GroEL depletion conditions. Such a system generated a platform for investigating properties of Group 2 chaperonin, MmCCT, in a Group 1 host. Using this system, in this chapter, we describe our preliminary attempts to elucidate the MmCCT dependent clients in *E. coli* (Section 5.2) and dissect its client binding site (Section 5.3) by means of proteomic and mutational analysis.

5.2   Elucidation of MmCCT-clients in *E. coli*

5.2.1 Introduction

An important aspect in the field of chaperonins is identification of the proteins that partially or stringently require a particular chaperonin to fold, re-fold or maintain its native structures. Several genome and proteome-wide studies have been conducted to identify GroEL-interactors (Houry *et al.*, 1999; Kerner *et al.*, 2005; Fujiwara *et al.*, 2010), yeast and mammalian TRiC-interactors (Thulasiraman *et al.*, 1999; Dekker *et al.*, 2008; Yam *et al.*, 2008) and interactors of archaeal chaperonins from *M. mazei* (Hirtreiter *et al.*, 2009).

Our finding that the chaperonin MmCCT has GroEL-like function in *E. coli* offered a unique opportunity to identify bacterial proteins that interact with and are potentially folded by an archaeal chaperonin. We initiated this study on account of two reasons: Firstly, as discussed in Chapter 4, we hypothesized that the better GroEL-complementing ability of the point mutants MmCCT-M223I and MmCCT-K216E was because of their ability to interact with and perhaps fold more clients than the wildtype. This hypothesis could be tested by comparing the interactors of MmCCT with those of the mutants. The differences if obtained might provide reasons for the enhanced GroEL-complementing ability of the mutants as compared to the wildtype. Secondly, the study is interesting on its own right as it will enable isolation of proteins from a Group 1 host, *E. coli,* that are possibly recognized and folded by a Group 2 chaperonin, MmCCT. Comparison of such proteins with the already known GroEL-interactors can identify overlapping and novel sets of potential MmCCT-interactors. We suggest that this initiative provides a platform for comparative analysis that can help understand the

differences between interactors and in turn can reflect on evolutionary co-relation between chaperonins and clients.

A combination of diverse techniques have been reported in the literature to isolate the chaperonin interactors for example pulse chase assays followed by two dimensional chromatography of the immunoprecipitated chaperonins (Houry *et al.*, 1999; Yam *et al.*, 2008), immunoprecipitation technique using specific antibodies (Kerner *et al.*, 2005; Hirtreiter *et al.*, 2009) or usage of internal tagging methods and subsequent pull-downs (Dekker *et al.,* 2008) among others. The details of these studies are outlined in Table 5.1. Initially we had undertaken two different approaches to identify potential MmCCT interactors in *E. coli.* Firstly, we tried utilizing an additional point mutant of MmCCT that contains a G to C substitution at the highly conserved glycine 61 residue. The equivalent mutation (G65C) in the archaeal chaperonin from the *Thermococcus strain* KS-1 has been shown to yield a 'trap variant' of this chaperonin. The variant binds to and supposedly encapsulates the clients - citrate synthase and green fluorescent protein (GFP) with significantly higher affinity than that of the wildtype, but restricts their release and hence productive folding, even in the presence of ATP. This refolding arrest activity has been proposed to occur due to inhibition of the ATP dependent conformational changes possibly by steric hindrance of side-chain of the residue position (Iizuka et al., 2001). We speculated that if the soluble fractions of cells expressing MmCCT proteins containing the G61C mutation are run on a Native-PAGE gel, the 'trapped' interactors should remain associated with the chaperonin proteins and become identifiable by MS analysis. While a significant number of proteins were isolated with this method, a high background possibly due to slower migration of certain protein on native-PAGE gels was

consistently observed, questioning the obtained results. It was concluded that a more pure protein preparation was essential.

Hence we further attempted to isolate MmCCT proteins by C- or N- terminal His-tagging method. However, these experiments were again not successful and resulted in conformational inconsistencies of the chaperonin complexes. Such observations were attributed to the poor accessibility of cavity-buried termini and subsequent falling apart of the complexes during metal affinity chromatography as had been previously observed (Dr. Andrew Large, former post-doc). We speculated that using an internal tag could overcome these difficulties and allow identification of potential MmCCT interactors. Thus, we here describe a comprehensive internal-tagging approach undertaken for our analysis as previously described by Dekker *et al.,* 2008. The method utilizes calmodulin binding protein (cbp) affinity tag incorporated at a specific solvent exposed site of one of the subunits yeast CCT *(cct3)*. The tag enables isolation of CCT complexes using a calmodulin resin along with the bound proteins that can be further analysed by mass spectrometry (Pappenberger *et al.*, 2006).

Table 5.1: A summary of client identification studies for various chaperonins.

Chaperonin	Study by	Description of the approach
GroEL	Houry *et al.,* 1999	1. Pulse chase analysis of live *E. coli* cells followed by immunoprecipitation, 2D PAGE and MS analysis
	Kerner *et al.,* 2005	1. Biochemical assay analysis of individual proteins detected by Houry *et. al.* 2. MS analysis of isolated client containing GroEL-GroES complexes by immobilized metal-affinity chromatography using a C-terminal His-tagged GroES
	Chapman *et al.,* 2006	1. Identification of clients from inclusion bodies after depletion of cellular GroEL levels
	Fujiwara *et al.,* 2010	1. Proteome wide MS analysis of GroE depleted cells and identification by comparing abundance ratios. 2. Individual characterization based on solubility in GroE depleted cells of proteins detected by above method and by Kerner *et al.* 3. Identifying metabolites by direct assays in GroE depleted *E. coli* cells
Mammalian TRiC	Thulasiraman *et al.,* 1999	1. Pulse chase analysis of cells with newly translated proteins followed by immunoprecipitation and 2D gel analysis.
	Yam *et al.,* 2008	1. Pulse chase assays of newly translated proteins, immunoprecipitation, 2D PAGE, MS analysis 2. Genomic screens in a cell free mammalian translation system
Yeast CCT	Dekker *et. al.,* 2008	1. Pull-downs of yeast-CCT and the bound interactors using internal tagging method 2. Testing genetic interaction spectrum of yeast-CCT using a temperature sensitive allele by synthetic gene array screen.

5.2.2 Results

To comprehensively identify MmCCT-interactors we employed a recently described approach used for identification of yeast CCT clients by Dekker *et al.* 2008. They have utilized an internal tagging method where Histidine (His), streptavidin (strep) and calmodulin binding protein (cbp) affinity tags were incorporated in tandem at a specific solvent exposed site of one of the subunits yeast CCT *(cct3)*. The CCT complexes were then isolated using a calmodulin resin along with the bound proteins. These bound proteins were further purified by sucrose gradients and analysed by mass spectrometry to identify yeast CCT clients. They showed that addition of ATP in the final wash step removed these proteins and hence suggested that the isolated proteins were *bona fide* CCT clients. The site for tag insertion (between proline 374 and lysine 375 on yeast *cct3* subunit) was finalized after several attempts of terminal or internal tagging at different positions on yeast *cct* gene using different tags (Pappenberger *et al.*, 2006). Because of successful results obtained with the chosen tags and the tag insertion site for yeast CCT (Dekker *et al.*, 2008), the same combination of tags and equivalent site on MmCCT was employed for this study as described below.

5.2.2.1 Insertion of CBP-tag in MmCCT

In the case of yeast CCT the tag was inserted internally between the residues P374 and K375 of yeast *cct gene*. These residues reside on the outer surface of the chaperonin complex and hence are more accessible in comparison to the N and C termini that lie inside of the complex. The corresponding residues for MmCCT, P362 and K363 were found to be positioned in a similar manner (Figure 5.1). A 56 amino-acid long tag containing his- strep-cbp sequences in tandem was inserted in *Mmcct* gene by overlap

extension PCR (Figure 5.2 and section 2.8.5.1, Materials and Methods). Although only cbp tag was required for downstream calmodulin based pull-down analysis, the combination of tags described by Pappenberger *et al.* was retained for this analysis. This is because, as stated before, such a combination was finalized after several trials taking into account the stability and *in vivo/in vitro* function of the differentially tagged yeast CCT proteins (Pappenberger *et al.*, 2006). It should be noted however that MmCCT is a homooligomeric chaperonin and hence tagging would occur in all 16 subunits of the complex as against on a single subunit for yeast CCT. Thus, the tagged MmCCT protein was first checked for its expression and solubility in *E. coli* as described in next section. For the ease of understanding in the document, a protein that is not tagged will be referred to as untagged and not as the wildtype protein.

5.2.2.2 Expression and solubility of cbp-tagged MmCCT protein

The construct *Mmcct-wt-tag* was cloned into pET21 vector and tested for its expression and solubility in *E. coli* TAB21 strain in a similar manner as described in chapter 3. MmCCT-WT-tag protein migrated as a 64KDa monomer on SDS-PAGE gels (Figure 5.3A) however the expression levels were lower than that of the wildtype. When tested for its solubility, the tagged protein was not detected in soluble fractions of cells as analysed by SDS-PAGE and western blotting (Figure 5.3B and C). In addition the tagged protein was not able to restore the GroEL complementing ability of the wildtype MmCCT protein in TAB21 (Figure 5.3D).

A


```
                              ↓↓
Tacidophilum_alpha    RKIGDDRMTFVMGCKNPKAVSILIRGGTDHVVSEVERALNDAIRVVAITKEDGKFLWGGG 409
Mmaripaludis_mmcct    RKISGDSMIFVEECKHPKAVTMLIRGTTEHVIEEVARAVDDAVGVVGCTIEDGRIVSGGG 405
Mmusculus_ccty        KKIGDEYFTFITDCKDPKACTILLRGASKEILSEVERNLQDAMQVCRNVLLDPQLVPGGG 412
Scerevisiae_cct3      EMIGDEYFSFLDNCKEPKACTIMLRGGSKDILNEIDRNLQDAMAVARNVMLSPSLSPGGG 417
                      . *..: : *:  **.*** ::::** :..::.*: * ::**: *    .  .  :  ***
```

B
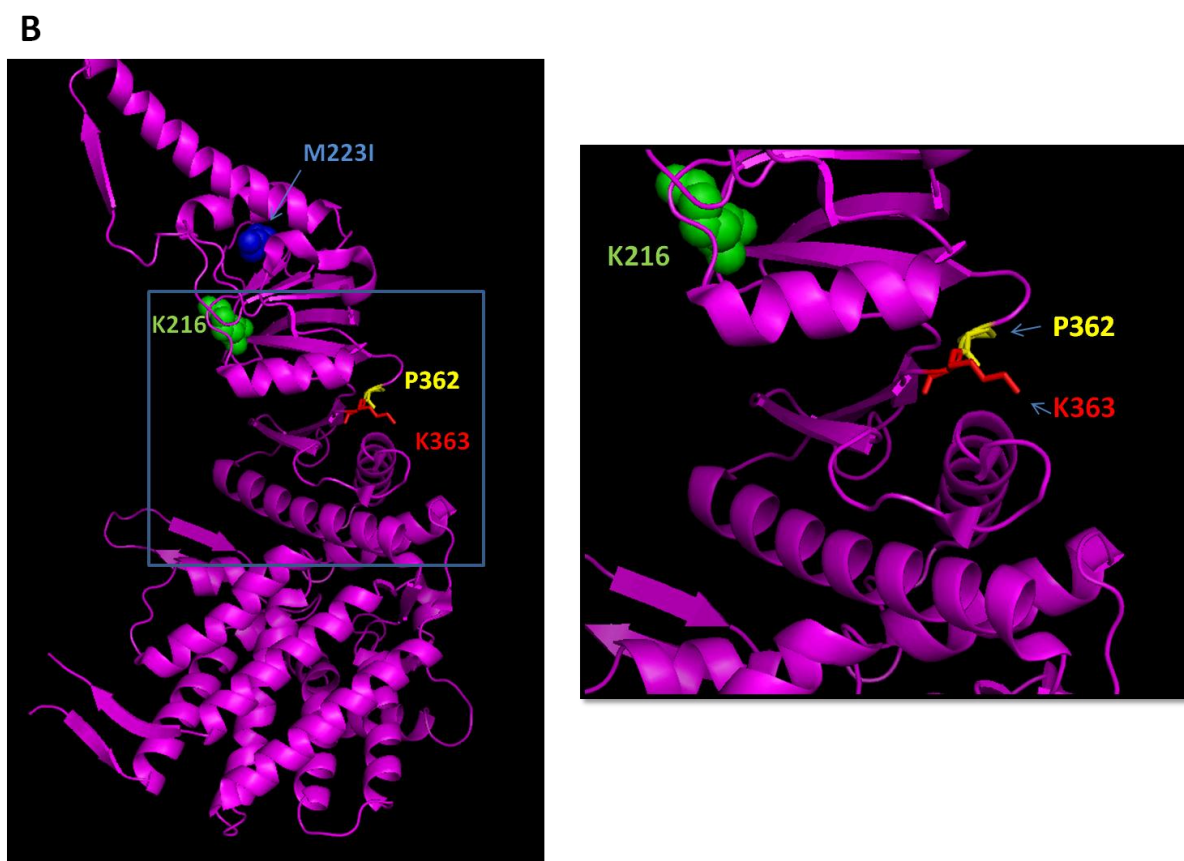
Figure 5.1

Choice and location of tag insertion site for MmCCT

(A) Alignments of selected sequences of Group 2 chaperonins from organisms as mentioned. **(B)** MmCCT subunit showing P362 and K363 residues highlighted as yellow and red respectively. The subunits were coloured in PYMOL from structures obtained from Protein Data Bank (PDB ID: 3KFK) submitted by Pereira *et al.* 2010.

A

Primer 1: Mmcctgene-for

P362/K363

Mmcct gene

Primer 2: Mmcctgene-rev

5' *Mmcct* region   5' tag sequence   tag overlap region   3' tag sequence   3' *Mmcct* region

Primer 3
CBP- A-rev primer

Primer 4
CBP-B-for primer

B

PCR 1

Primer 1

Primer 3

Product 1

PCR 2

Primer 4

Primer 2

Product 2

Overlap extension PCR

Product 1

Product 2

Primer 1

P362   K363

his-strep-cbp-tag
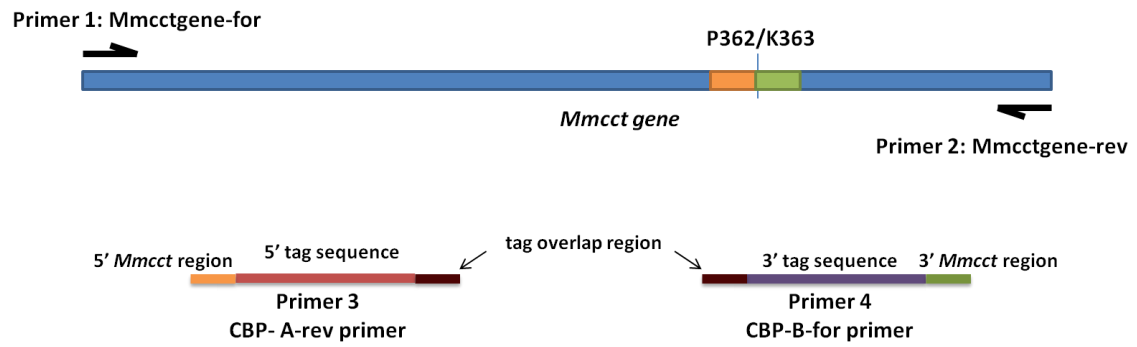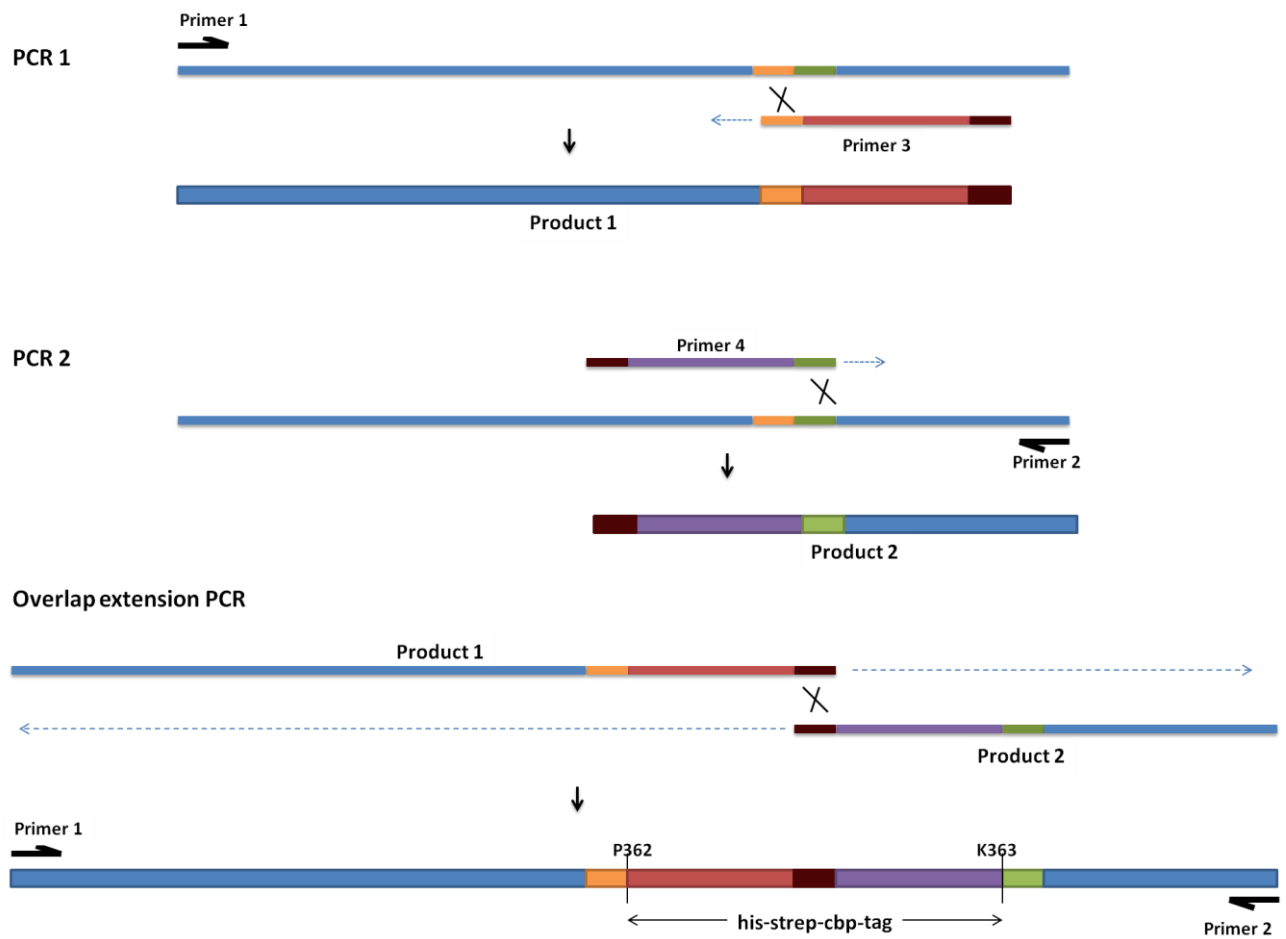
Primer 2

Figure 5.2

Schematic illustrating the overlap extension PCR method used to generate MmCCT-tag.

(A) The diagram shows the sites of the four primers required for overlap extension PCR. The calmodulin binding protein (cbp) tag is contained within primers 3 and 4. Both these primers have a tag overlapping region (coloured red), in addition to the 5' and 3' regions respectively (coloured orange and green) that are complementary to 21 bp upstream and downstream of the tag insertion site P363|K363 in *Mmcct* gene. Primers 1 and 2 contain NdeI and BamHI restriction sites respectively.

(B) In the first PCR step, primers 1 and 3 are used to amplify the first part of the *Mmcct* gene. The product now has an adjoining tag region with a 3' patch that can overlap with the product of second PCR step that is similarly obtained by amplifying *Mmcct* gene with primers 2 and 4. In the final PCR step, the two products are mixed together, denatured, annealed, and the 3' tails are extended by DNA polymerase. This generates a small amount of chimeric product containing the whole tag sequence. The full length product was amplified using primers 1 and 4. This product was then cloned into the NdeI/BamHI site of the pET21b expression vector.
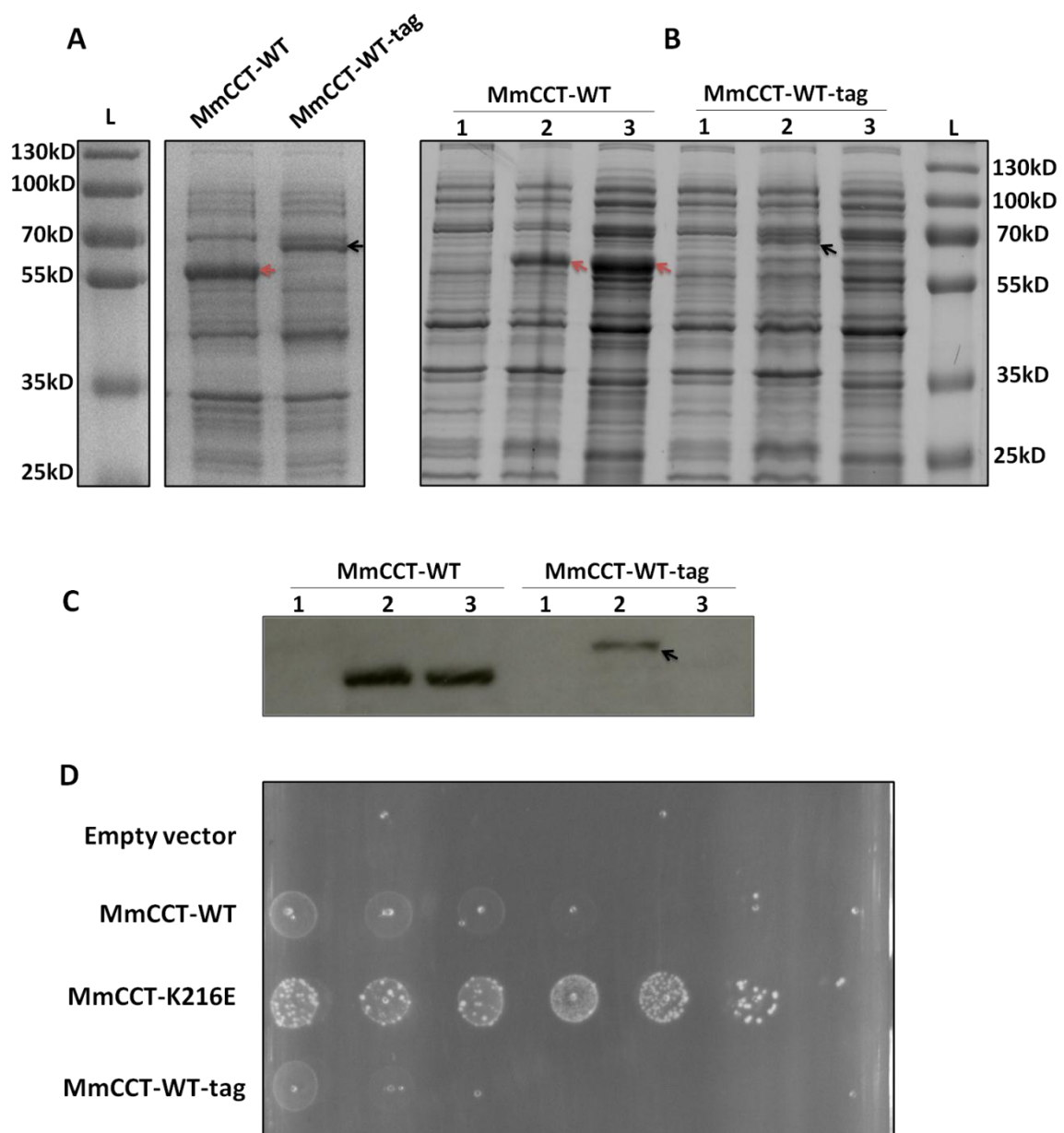
Figure 5.3

Expression, solubility and complementation analysis of MmCCT-WT-tag protein in *E. coli*

(A) Crude extracts of *E. coli* TAB21 cells expressing pET21-MmCCT-WT and pET21-MmCCT-WT-tag as analyzed by 12% SDS PAGE and Coomassie staining.

(B) and **(C)** Non-induced crude extracts (1), IPTG-induced crude extracts (2) and IPTG induced soluble extracts (3) of *E. coli* TAB21 cells expressing pET21-MmCCT-WT and pET21-MmCCT-WT-tag as analyzed by 12% SDS PAGE **(B)** and western blotting using anti-thermosome antibody **(C).** Black and red arrows show tagged and untagged MmCCT proteins respectively.

(D) TAB21 cells expressing MmCCT proteins as indicated grown on 0.2% glucose and 1 mM IPTG at 30°C for five days.

5.2.2.3 Co-expression of tagged and untagged MmCCT proteins

The insolubility of tagged MmCCT protein was thought either be to be due to lack of formation of functional complexes or instability of already assembled complexes because of the presence of the tag on all 16 subunits. Thus an attempt was made to co-express the tagged and untagged MmCCT wildtype proteins together such that heterooligomers with lesser number of tagged subunits per complex would result. Such heteroligomers could therefore be more stable and hence functional in *E. coli*.

Because both tagged and untagged *Mmcct* constructs were in the same vector pET21, it was first necessary to subclone one of the genes into a compatible vector for co-expression. We chose a lower copy number plasmid, pACYC184, to express the tagged MmCCT protein so that complexes with fewer tagged subunits could be obtained owing to its relatively lower expression. The *Mmcct-tag* fragment was cloned into pACYC184 by disrupting its tetracycline marker cassette (Figure 5.4). The resulting construct pACYC-*Mmcct-tag* was transformed together with pET21-*Mmcct-wt* into *E. coli* TAB21 cells and tested for its expression and solubility. While IPTG was added to the medium for pET21 based expression of the untagged wildtype protein, the expression of *Mmcct-tag* relied on the tetracycline promoter upstream of the gene on the pACYC plasmid. Selection was based on chloramphenicol marker for pACYC184 and the standard ampicillin marker for pET guided expression of the untagged wildtype.

As shown in Figure 5.5A, both tagged and untagged MmCCT proteins were resolved as monomeric subunits on SDS-PAGE gels when expressed together. However, only the untagged monomers were detected in the soluble fractions of the cells (Figure 5.5B and C). This could either mean that the complexes containing only the wildtype untagged

monomers are soluble or that complexes containing relatively lower numbers of tagged monomers per complex are soluble but undetectable because of their smaller number. It should be noted that the proportion of the tagged subunits was lower than that of the untagged in whole cell lysates (Figure 5.5A lane 5). The ratio of tagged:untagged monomeric subunits was reproducibly obtained between 1:3 to 1:4 when analysed by densitometry. If it is assumed that the assembly of the chaperonin complexes is random inside the cells, the approximate relative distribution of complexes containing 1 to 16 tagged subunits could be calculated using the formula $((p)^k)((1-p)^{(n-k)})(n!/(k!(n-k)!))$ where k are untagged subunits and n-k represent tagged subunits, and where the relative probabilities of k and n-k are p and 1-p respectively. With the stoichiometry 1:3 the distribution obtained can be represented as shown in Figure 5.6. Indeed these figures are based on the approximation that the binary distribution assumes an infinite pool of both sorts of subunits, but as subunits assemble, the number of remaining subunits goes down. Nonetheless, the data suggested that at a given time ~ 80% of the total chaperonin population contained less than 5 tagged subunits. If this was the case it could have been possible that the complexes containing fewer tags were stable and soluble however because of their lower numbers were not detected during gel analysis. Based on this assumption, the analysis was carried forward and pull-downs were performed.
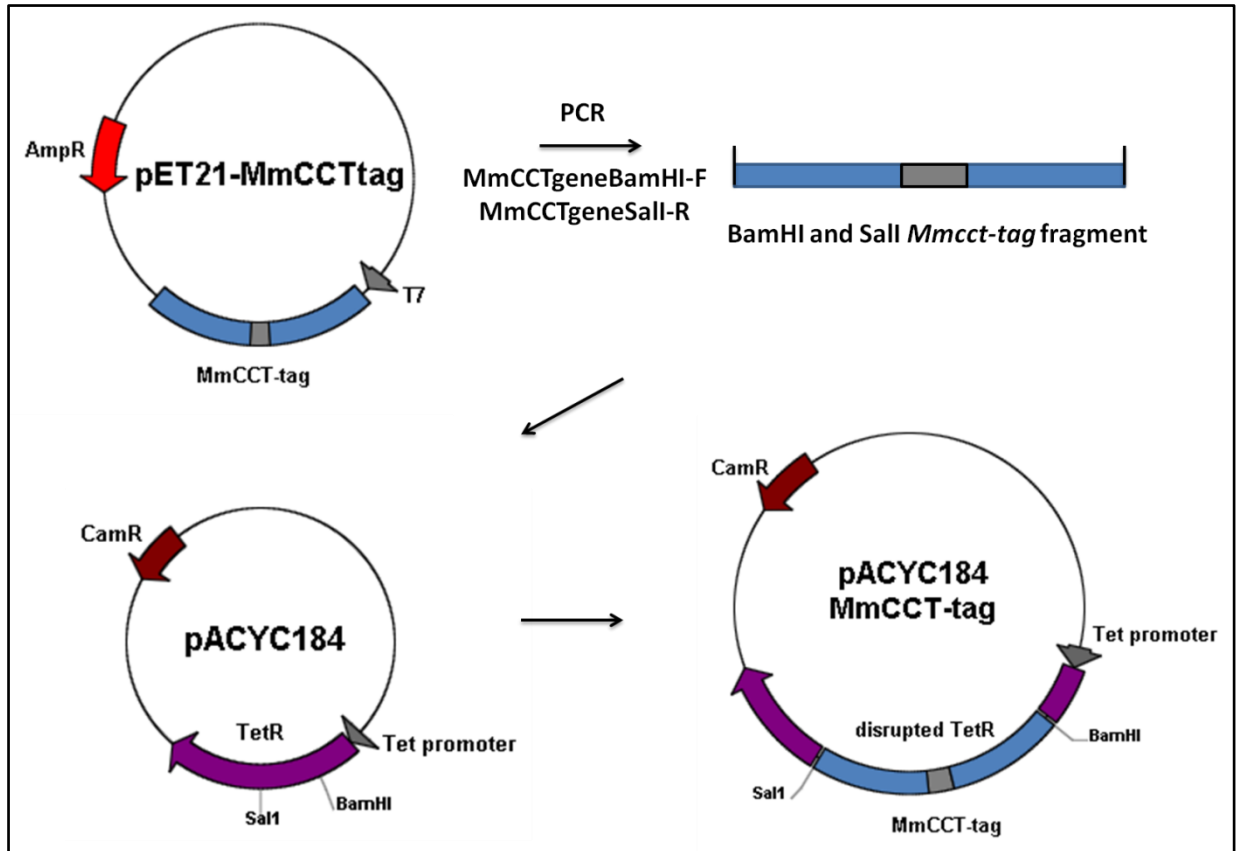
Figure 5.4
Schematic diagram to explain construction of pACYC-MmCCT-tag.
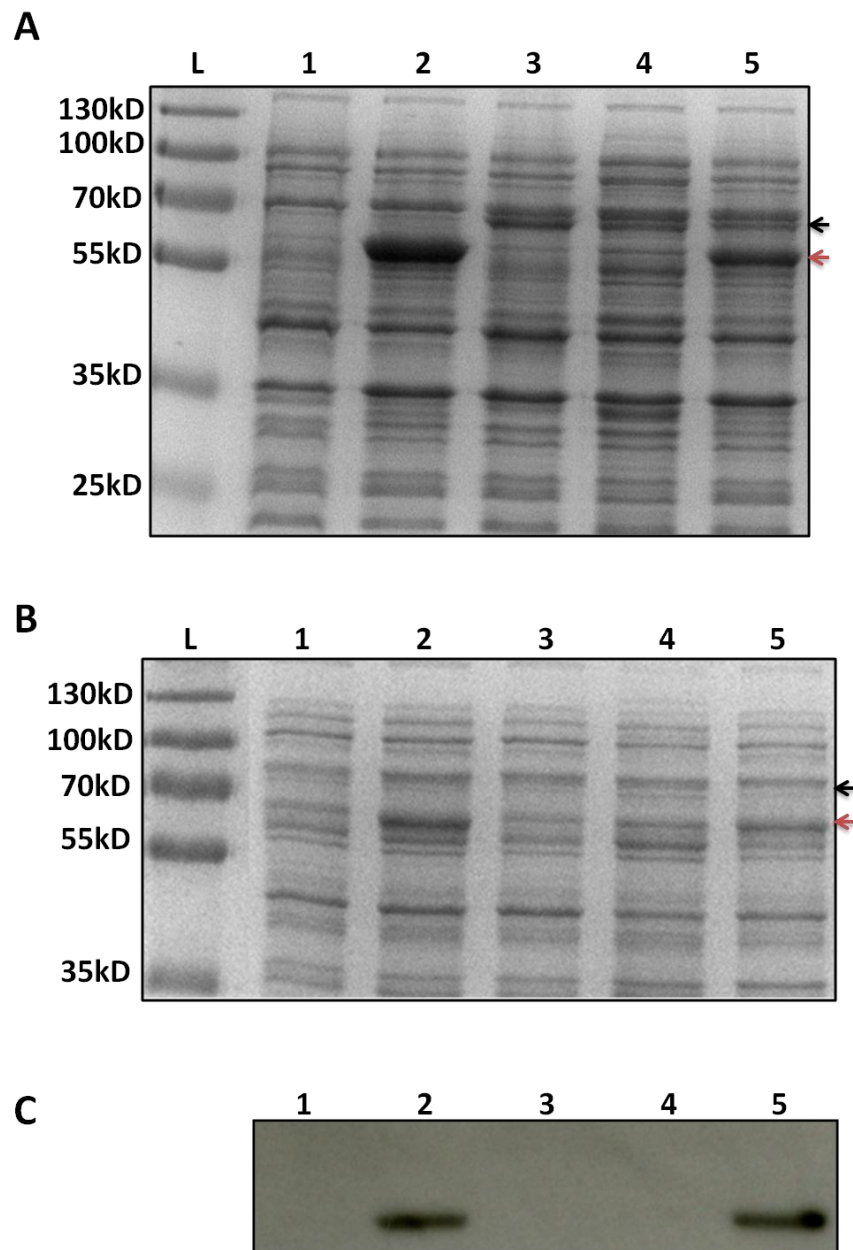
Figure 5.5

Expression and solubility of coexpressed MmCCT-WT and MmCCT-WT-tag proteins in *E.coli*

Whole cell crude lysates **(A)** and soluble extracts **(B)** of *E. coli* TAB21 cells expressing MmCCT proteins as analyzed by 12% SDS PAGE and Coomassie staining. Soluble extracts also analysed by western blotting using anti-thermosome antibodies **(C)**.

(1) TAB21/pET21 empty vector
(2) TAB21/pET21-*Mmcct*-WT
(3) TAB21/ pET21-*Mmcct* -tag
(4) TAB21/ pACYC184-*Mmcct*-tag
(5) TAB21/pET21-*Mmcct* -WT+ pACYC184-*Mmcct* -tag
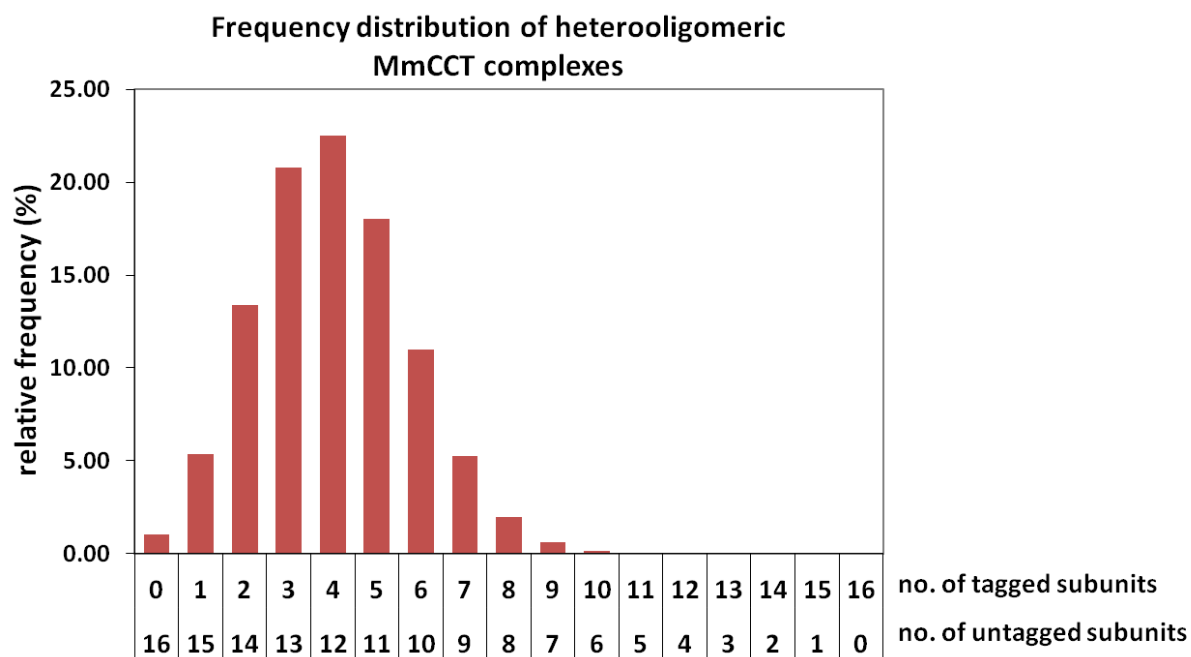
Figure 5.6

Frequency distribution of MmCCT complexes containing tagged and untagged subunits.

The numbers are based on the approximation that the binary distribution assumes an infinite pool of both sorts of subunits.

5.2.2.4 Pull-downs using calmodulin-affinity resin

The TAB21 cells co-expressing MmCCT wildtype tagged and untagged subunits were subjected to pull-down analysis using calmodulin resin beads as described in section 2.8.5.2, Materials and Methods. Briefly, the cells were lysed and soluble fraction was incubated with the resin for binding of tagged complexes. The unbound and non-specific material was removed by several washes with wash buffer and protein was eluted using EDTA. Any proteins bound to MmCCT should also be eluted along with it and become detectable in downstream analysis.

The eluted fractions showed the presence of both tagged and untagged subunits of MmCCT when resolved by SDS-PAGE gel (Figure 5.7) suggesting that the method had worked and that at least some tagged complexes were soluble. When analysed using densitometry, the ratio of tagged to untagged subunits was reproducibly obtained between 1:4 and 1:5. A lower co-expression ratio was consistent with our earlier estimates and therefore indicated a lower number of tagged subunits per complex as hypothesized earlier. The reason for detection of soluble tagged MmCCT monomers in the eluate but not in the soluble fractions during the previous experiments could be due to a smaller population of tagged monomers. To test whether the same was happening for complexes with all tagged subunits, the above experiment was repeated with TAB21 cells expressing only pACYC184-*Mmcct*-WT-tag. However the complexes with all tagged subunits failed to isolate during the analysis (Figure 5.8, Lane 4b), verifying the previous results and confirming that heteromers with lesser number of tagged subunits per complex were required to form the oligomeric complexes. The proteins isolated with MmCCT tagged protein were visually observed on the gels stained with silver staining (Figure 5.9, Lane 3).
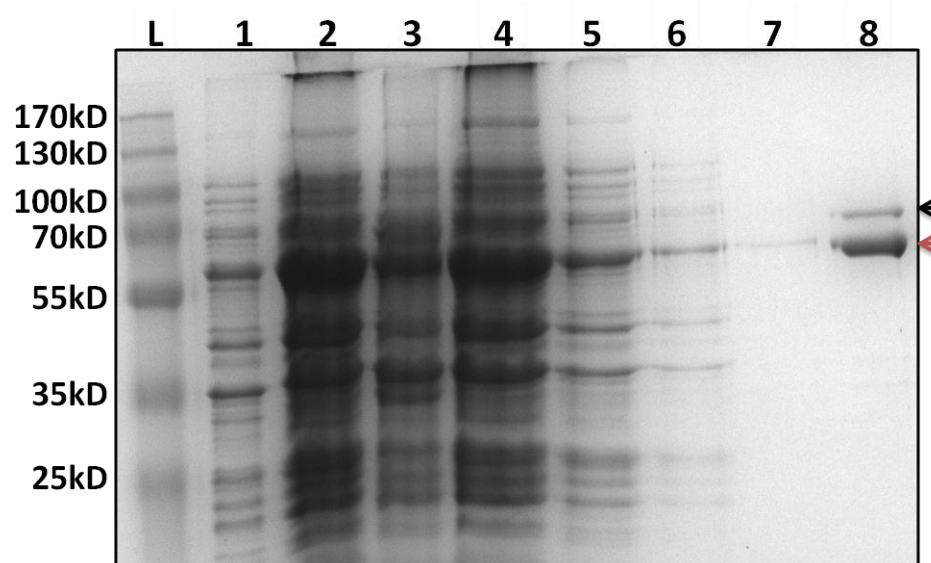
Figure 5.7

Calmodulin-based pull-downs of cbp tagged MmCCT hetero-oligomers analysed by SDS-PAGE and Coomassie staining.

SDS-PAGE analysis of fractions obtained at different stages (1 – 8) during the calmodulin based pull-down analysis of TAB21 cells expressing both pET21-*Mmcct*-WT and pACYC184-*Mmcct*-WT-tag. The gel was stained by Coomassie staining. Black and red arrows indicate the tagged and untagged MmCCT monomers respectively.

(1)   Crude extracts (OD_{600}=1), 4 hours after IPTG induction
(2)   Soluble fraction after lysis
(3)   Insoluble fraction after lysis
(4)   Unbound material
(5)   Wash 1
(6)   Wash 2
(7)   Wash 3
(8)   Elution
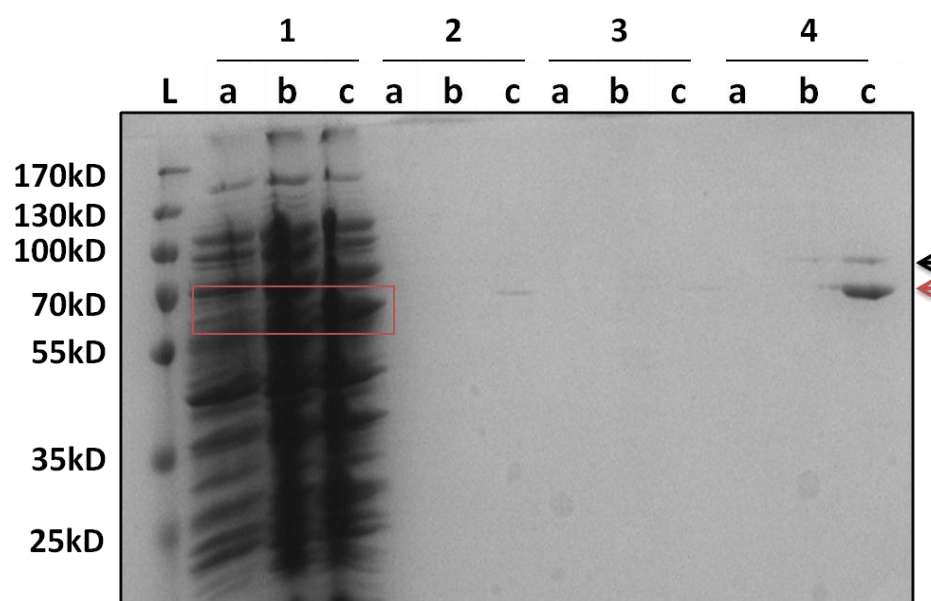
Figure 5.8

Calmodulin-based pull-downs of cbp tagged MmCCT homo- and hetero-oligomers analysed by SDS-PAGE and Coomassie staining.

SDS-PAGE analysis of fractions obtained at different stages (1 – 4) during the calmodulin based pull-down analysis of TAB21 cells expressing MmCCT proteins as indicated below (a-c). The gel was stained by Coomassie staining. Black and red arrows indicate the tagged and untagged MmCCT monomers respectively. (1) Soluble fractions after lysis; (2) Wash 2; (3) Wash 2; (4) Eluates.

- **(a)** TAB21/pET21b + pACYC184 empty vectors
- **(b)** TAB21/pACYC184-*Mmcct*-WT-tag
- **(c)** TAB21/pET21-*Mmcct*-WT+ pACYC184-*Mmcct*-WT-tag
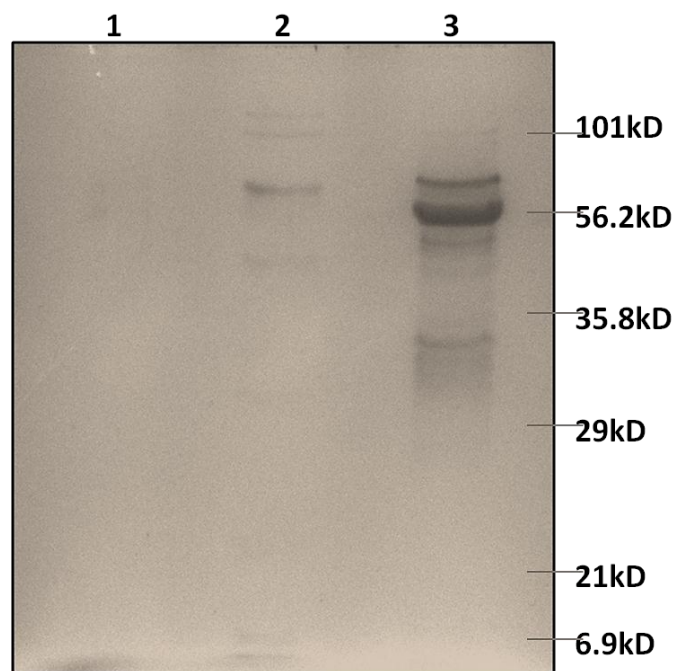
Figure 5.9
Potential MmCCT- interactors detected by SDS-PAGE and silver staining

Protein eluates of pull-down analysis resolved by SDS-PAGE and silver staining.

(1) TAB21 + pET21b + pACYC184 empty vectors

(2) TAB21 + pET21-*Mmcct*-WT

(3) TAB21 + pET21- *Mmcct*-WT + pACYC184- *Mmcct*-WT-tag

5.2.2.5 MS analysis of MmCCT-interactors

The next step was to identify the potential MmCCT-interactors using MS analysis. Due to a limited amount of time this analysis was only done once and currently the work is ongoing in the laboratory to test reproducibility of the results described below.

The gel-lane expressing the potential bound interactors (Figure 5.9, Lane 3) was sliced into three vertical pieces as shown in Figure 2.4, Materials and Methods. Care was taken to avoid the gel-area expressing the MmCCT tag and untagged proteins because it has been noted that presence of higher amounts of a particular protein during the MS analysis reduces the chances of picking up the proteins that are relatively present in lower concentrations. As a control, three slices at similar positions to the sample were cut out from the gel-lane that resolved pull-downs of cells expressing untagged MmCCT-WT protein (Figure 5.9, Lane 2). Such a control served for two purposes: first to test if untagged MmCCT protein has any affinity for the resin and second to determine whether there are any other calmodulin binding proteins, that if present would give rise to false positives. All the gel slices were individually subjected to in-gel trypsination and LTQ-Orbitrap MS analysis and the peptides were identified by MASCOT search engine using the BL21 (DE3) database.

The protein hits obtained from all three gel samples of tagged MmCCT protein as a result of the analysis were combined and arranged score-wise. The overall analysis revealed a total of 96 potential MmCCT-interacting proteins (Table 5.2). As seen, MmCCT-tag and MmCCT-untagged subunits were the first scorers of the list. Even though the samples were made particularly leaving the bands for MmCCT, the high scores of these proteins suggested a potential carry forward in the sample gel-slices possibly due to a certain

extent of degradation. Out of these 96 proteins, 16 proteins overlapped with the proteins obtained from analysis of untagged MmCCT control (Table 5.2, 81-96). Additionally 60 other proteins were detected in the control sample and again MmCCT-WT protein was one of the top scoring candidates. These signals possibly represent proteins with an affinity for calmodulin resin. However also note that ~50% of the 77 control protein signals consisted of ribosomal proteins. Because ribosomal proteins are among the highly abundant proteins, the results also suggest that the proteins present in higher amounts (and hence MmCCT) are not completely removed during the washes and can be eluted as background contamination. For simplicity of this analysis the non-overlapping control proteins have not been considered.

Out of the obtained protein signals, 30 proteins were found to be previously reported as GroEL-interactors (Table 5.3) (Kerner *et al.*, 2005; Fujiwara *et al.*, 2010). Using the data by Kerner *et. al.*, the probability for a protein to be a GroEL client out of total 2400 soluble proteins of *E. coli* can be calculated as 0.103 (247 GroEL-interactors). Taking this value into consideration and assuming that the selection is totally random, the cumulative probability for selection of 30 or more GroEL-interactors out of 96 was solved as 1.68E-08. This suggested that the obtained data was statistically significant.

GroEL-interactors have been divided into four groups (Azia *et al.*, 2012): Class I that can fold independently of GroEL but still are found to be associated with GroEL, Class II that partially depend on GroEL for folding, and the obligate GroEL clients belonging to either Class III- described by Kerner *et. al.* (Kerner *et al.*, 2005) or Class IV described by Fujiwara *et. al.*, 2010. Out of the 30 GroEL clients obtained in this analysis, 7 belonged to Class I (highlighted as blue), 10 belonged to Class II (highlighted as red), 3 to Class III-

(highlighted as green) and 10 to Class IV (highlighted as purple). Almost all proteins were well below the 60kDa cut-off.

The remaining signals (approximately 70 proteins) could represent a novel set of proteins that probably interact with MmCCT but not GroEL. However note that 67% of these proteins belong to the group of 1100 most abundant proteins in *E. coli* (Ishihama et al., 2008) (Table 5.2, A1-A5 where A1 represents first 20%, A2 next 20% and so on). It is possible thus that several low abundant GroEL-interactors that are recognized by MmCCT have escaped this analysis. Moreover, the study does not establish whether the potential MmCCT-interacting proteins are *bona fide* clients. Testing the ATP dependent release of the obtained interactors would hence be the next immediate step for a clearer analysis. Nonetheless the results described here demonstrate that the internal tagging method has been successful and can be utilized for identification of MmCCT interactors. Once the reproducibility of the results is tested, future experiments will be targeted at identifying potential interactors of different MmCCT mutants.

Table 5.2: MmCCT-interactors in *E. coli* determined using tagging approach

Note1 : Highlighted cells represent four groups of GroEL clients as described by Kerner *et al.* and Fujiwara *et al.* Blue – Group I ; Red – Group II; Green – Group III- and Purple – Group IV.

Note2: Abundance by the emPAI scores for 1100 *E. coli* proteins by Ishihama *et. al.* 2008 (A1 belongs to first 20% of abundant proteins, A2 belongs to next 20% and so on).

No	Accession No.	Description	Gene name	Score	Coverage	# AAs	MW [kDa]	calc. pI	Essen tiality	Abund ance
1	MmCCT_tag	Mmcct_tag		345.3	67.11	599	64.2	5.97		
2	MmCCT	WILDTYPE_Mmcct		341.9	73.85	543	58.2	5.06		
3	253775633	F0F1 ATP synthase subunit beta	atpB	35.23	21.74	460	50.3	5.01		A2
4	253775592	Transcription termination factor Rho	rho	34.04	22.91	419	47.0	7.25		
5	253774547	Efflux transporter RND family, MFP subunit (Multidrug efflux transporter)	acrA	16.73	12.59	397	42.2	7.99		
6	253774856	Dihydrolipoamide dehydrogenase	lpD	16.61	11.81	474	50.7	6.15		
7	253772597	Exodeoxyribonuclease VII large subunit	xseA	15.46	11.84	456	51.8	9.98		
8	253775631	F0F1 ATP synthase subunit alpha	atpA	15.02	12.67	513	55.2	6.13		A2
9	253772127	Outer membrane channel protein	tolC	13.96	6.69	493	53.7	5.66		
10	253774875	Cell division protein ftsz	ftsZ	12.51	13.58	383	40.3	4.78	E	A3
11	253774876	Cell division protein ftsa	ftsA	11.9	8.33	420	45.3	6.24	E	
12	253772794	NADH dehydrogenase I subunit F	nuof	10.52	10.56	445	49.3	6.86		A4
13	253774042	Outer membrane protein F	ompF	9.64	6.63	362	39.3	4.96		A3
14	253774355	(dimethylallyl)adenosine trna methylthiotransferase	miaB	8.68	5.91	474	53.6	5.34		
15	253775430	ATP-dependent protease ATP-binding subunit hslu	hslU	7.63	8.35	443	49.6	5.35		A3
16	253775201	Ftsh protease regulator hflk	hflK	6.16	6.44	419	45.5	6.43		A5
17	253773284	Glutamate dehydrogenase	gdhA	6.08	8.05	447	48.6	6.4		A5
18	253774310	Type II citrate synthase	gltA	6	5.15	427	48.0	6.68		A2
19	253771741	Glucose-1-phosphate adenylyltransferase	glgC	5.75	4.64	431	48.7	6.14		

20	253773266	Glyceraldehyde-3-phosphate dehydrogenase	gapA	5.33	7.85	331	35.5	7.11	E	A1
21	253772553	Serine hydroxymethyltransferase	glyA	5.05	4.32	417	45.3	6.48	E	A1
22	253774981	DNA repair protein rada	radA	5.01	6.96	460	49.4	7.24		
23	253775140	Peptidase pmba	pmbA	4.7	4.89	450	48.3	5.6		A5
24	253775449	Two-component sensor protein	cpxA	4.49	3.94	457	51.6			
25	253773798	Gtp-dependent nucleic acid-binding protein engd	ychF	3.96	6.61	363	39.6			
26	253774289	Translocation protein tolb	tolB	3.95	3.26	430	45.9			
27	253773411	Electron transport complex protein rnfg	rnxG	3.93	5.83	206	21.9	7.53		
28	253773420	Adenosine deaminase	add	3.73	3.3	333	36.4	5.6		A5
29	253774575	Trigger factor	tig	3.72	5.09	432	48.2	4.88		A1
30	253773263	Mlta-interacting mipa family protein	mipA	3.68	5.24	248	27.8	5.85		A4
31	253774279	Lac repressor	lacI	3.48	5.51	363	38.9	6.89		
32	253775352	Aspartate kinase III	lysC	3.47	2.67	449	48.5	5.11		A5
33	253773868	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	ndh	3.24	3.23	434	47.3	8.85		
34	253772448	Alanyl-trna synthetase	alaS	3.23	1.37	876	96.0	5.92	E	A2
35	253772354	Hypothetical protein ECBD_0934	ygdH	3.15	2.42	454	50.9	6.48		A5
36	253772574	Cysteine desulfurase	iscS	3.06	2.97	404	45.1	6.37	E	A2
37	253774964	Threonine synthase	thrC	3.04	3.04	428	47.1	5.4		A3
38	253771552	L-lactate dehydrogenase	lldD	2.98	2.78	396	42.7	6.81		
39	253774804	30S ribosomal protein S2	rpsB	2.88	6.64	241	26.7	7.14		A1
40	253775469	Formate dehydrogenase subunit beta	fdnH	2.87	4.33	300	33.1	5.33		A5
41	253772392	Sulfate adenylyltransferase subunit 2	cysD	2.83	3.64	302	35.2	8.02		
42	253773044	Lipopolysaccharide biosynthesis protein	wzzE	2.81	3.37	326	36.3	5.34		
43	253773768	Hypothetical protein ECBD_2388	rssA	2.78	3.5	314	34.3	6.73		
44	253771825	FKBP-type peptidyl-prolyl cis-trans isomerase	slyD	2.76	4.59	196	20.8	5.05	E	A1
45	253772976	D-tagatose-bisphosphate aldolase, class II, non-catalytic subunit	gatZ	2.72	2.62	420	47.0	5.78		A3
46	253774769	Hypothetical protein ECBD_3415	yafD	2.67	5.02	259	29.1	9.41		
47	253775630	F0F1 ATP synthase subunit delta	atpH	2.66	6.21	177	19.3	5.02		A2

48	253773850	Putrescine/spermidine ABC transporter atpase protein	potA	2.65	3.17	378	43.0	5.3		A5
49	253772462	Glycine betaine transporter ATP-binding subunit	proV	2.6	2.75	400	44.1	5.57		A3
50	253774993	Patatin (predicted esterase)	yjjU	2.52	3.08	357	39.8	8.48		
51	253775034	5-methylcytosine-specific restriction enzyme subunit mcrb	mcrB	2.51	2.15	465	53.7	5.86		
52	253772387	Phosphoadenosine phosphosulfate reductase	cysH	2.48	4.1	244	28.0	5.91		
53	253773947	Outer membrane N-deacetylase	pgaB	2.46	1.19	672	77.4	6.19		
54	253775242	Lysyl-trna synthetase	lysS	2.45	2.18	505	57.8	5.24		A5
55	253774573	ATP-dependent protease ATP-binding subunit clpx	clpX	2.43	2.12	424	46.3	5.35		A4
56	253774088	ATP-dependent Clp protease ATP-binding subunit	clpA	2.43	1.32	758	84.2	6.32		A5
57	253772822	Undecaprenyl phosphate 4-deoxy-4-formamido-L-arabinose transferase	arnC	2.38	3.73	322	36.3	7.3		
58	253772446	Recombinase A	recA	2.32	3.12	353	37.9	5.19		A4
59	253772053	Hypothetical protein ECBD_0630 (Predicted role as L-cysteine desulfidase)	yhaM	2.3	2.06	436	45.3	5.53		
60	253774829	Poly(A) polymerase I	pcnB	2.29	4.03	472	54.6	9.66		A5
61	253774775	DL-methionine transporter substrate-binding subunit	metQ	2.24	4.06	271	29.4	5.29		A3
62	253773882	3-oxoacyl-(acyl carrier protein) synthase II	fabF	2.19	3.39	413	43.0	6.09		A2
63	253775603	Branched-chain amino acid aminotransferase	ilvE	2.19	3.56	309	34.1	5.82		A4
64	253775229	Chaperonin GroEL	groL	2.17	1.82	548	57.3	4.94	E	A1
65	253772304	Sugar transporter	araE	2.08	1.91	472	51.7	9.26		A3
66	253772112	Bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase	rfaE	2.07	2.52	477	51.0	5.41		A5
67	253772329	N-acetylmuramoyl-L-alanine amidase	amiA	2.06	2.64	417	45.6	9.58		
68	253772791	NADH dehydrogenase subunit B	nuoB	1.98	3.18	220	25.0	5.74		A3
69	253775399	Soluble pyridine nucleotide transhydrogenase	sthA	1.91	1.72	466	51.5	6.57		A5
70	253772496	Signal recognition particle protein	ffh	1.89	2.21	453	49.8	9.51	E	A5
71	253773042	6-phosphogluconate dehydrogenase	gnd	1.89	1.92	468	51.5	5.07		A1
72	253773003	Deoxycytidine triphosphate deaminase	dcd	1.89	4.66	193	21.2	5.91		A5
73	253775198	Adenylosuccinate synthetase	purA	1.8	1.85	432	47.3	5.49		A1

74	253773813	D-amino acid dehydrogenase small subunit	dadA	0	3.01	432	47.6	6.62	E	
75	253774382	D-alanyl-D-alanine carboxypeptidase fraction A	dacA	0	2.73	403	44.4	8.28	E	A4
76	253774880	N- acetylglucosaminyltransferase murg	murG	0	2.25	355	37.8	9.73		A5
77	253771935	Malate dehydrogenase	mdh	0	2.24	312	32.3	5.77		A2
78	253773756	Oligopeptide transporter ATP-binding component	oppD	0	4.45	337	37.1	6.1	E	
79	253771537	ADP-heptose:LPS heptosyl transferase I	rfaC	0	3.68	326	36.2	8.79	E	
80	253771835	Elongation factor Tu	tuf	31.38	32.49	394	43.3	5.45		A1
81	253772760	Acetyl-coa carboxylase subunit beta	accD	11.01	12.5	304	33.3	7.68		A3
82	253774014	Outer membrane protein A	ompA	10.76	13.58	346	37.2	6.42		A1
83	253773728	Transcriptional regulator cysb	cysB	10.01	12.35	324	36.1	7.42		A5
84	253772660	Ethanolamine ammonia-lyase	eutB	9.95	9.27	453	49.4	4.93		A5
85	253771881	DNA-directed RNA polymerase subunit alpha	rpoA	8.41	10.33	329	36.5	5.06		A1
86	253774788	Acetyl-coa carboxylase carboxyltransferase subunit alpha	accA	6.09	8.46	319	35.3	6.04	E	A3
87	253771920	Rod shape-determining protein mreb	mreB	5.48	7.2	347	36.9	5.26	E	A2
88	253774954	Chaperone protein dnaj	dnaJ	3.09	3.19	376	41.0	7.66		A3
89	253771856	50S ribosomal protein L3	rplC	2.82	4.78	209	22.2	9.91	E	A1
90	253774102	NAD-dependent epimerase/dehydratase	ybjS	2.72	2.87	349	39.6	8.63	E	
91	253771862	30S ribosomal protein S3	rpsC	2.72	5.15	233	26.0	10.2		A1
92	253771859	50S ribosomal protein L2	rplB	2.59	4.03	273	29.8	10.9	E	A1
93	253772512	Protein disaggregation chaperone	clpB	2.52	1.4	857	95.5	5.52		A2
94	253771880	30S ribosomal protein S4	rpsE	2.5	5.34	206	23.5	10	E	A1
95	253775387	50S ribosomal protein L1	rplA	2.26	5.56	234	24.7	9.64		A1
96	253774410	Alkyl hydroperoxide reductase subunit C	ahpC	2.04	7.49	187	20.7	5.17		A1

Table 5.3: Distribution of GroEL-clients among the potential MmCCT interactors

Total MmCCT clients detected	120						
Total GroEL client proteins out of 120	30	Abundant? (Ishihama *et.* al.)		Essential?		Overlap with control?	
Group 1	7	First 25%	1	-		1	ompA
		25%-50%	3	2	gapA glyA	-	
		50%-75%	0	-		1	ahpC
		75%-end	3	-		-	
Group 2	10	First 25%	5	-		1	rpoA
		25%-50%	4	2	accA, mreB	2	accA, mreB
		50%-75%	1	-		-	
		75%-end	0	-		-	
Group 3	3	First 25%	0	-		-	
		25%-50%	1	-		1	dnaJ
		50%-75%	2	-		-	
		75%-end	0	-		-	
Group 4	10	First 25%	1	-		1	eutB
		25%-50%	5	-		-	
		50%-75%	2	2	ybjS, fabF	1	ybjS
		75%-end	2	1	dadA	-	
Essential proteins	23/120						
Total essential proteins amongst 30 GroEL clients	7/30						
Note: ybjS, FabF, dadA not described essential by Fujiwara *et al.* but by Kerner *et al.*							

5.3   Preliminary investigation of MmCCT client-binding site

5.3.1 Background

Notable structural, biophysical and mutagenic studies have converged to identify the patch of hydrophobic residues in helices H and I of the GroEL apical domain as the client binding region in Group 1 chaperonins (Fenton *et al.*, 1994; Buckle *et al.*, 1997; Chen and Sigler, 1999; Tanaka and Fersht, 1999).   And accordingly, a number of studies have recognized GroEL-client interactions to be primarily hydrophobic in nature (Hayer-Hartl *et al.*, 1994; Lin *et al.*, 1995; Wang and Boisvert, 2003; Chaudhuri and Gupta, 2005; Li *et al.*, 2009).

For Group 2 chaperonins however, the client binding site has not been explicitly defined and has remained a point of debate. Electron microscopic studies of actin and tubulin bound to TRiC chaperonin have identified subunit-specific binding of these clients that occur mainly via polar and electrostatic interactions (Llorca et al., 1999; Llorca et al., 2000; Llorca et al., 2001). These observations were supported by biochemical and mutagenic studies that recognized polar surface regions on actin and tubulin to be involved in binding to CCT (Hynes and Willison, 2000; Ritco-Vonsovici and Willison, 2000; McCormack et al., 2001). Indeed the solved crystal structure of mouse CCTγ apical domain and corresponding phylogenetic analysis confirmed these results and the study further proposed a conserved patch of polar and charged residues facing the inside of the cavity, as the potential site for client recognition and interaction for TRiC (Pappenberger et al., 2002; Gomez-Puertas et al., 2004). The recently solved crystal structure of yeast CCT have further suggested these 'signature' residues that are distributed differently in all eight subunits to be potentially involved in client-binding

(Dekker *et al.*, 2011a). However, mutational analysis of these potential client binding residues was shown to have no effect on binding of TRiC to its obligate client, von Hippel-Lindau (VHL) tumor suppressor protein (Feldman *et al.*, 2003; Spiess *et al.*, 2006). Instead, based on further mutational and cryo-EM reconstruction studies, hydrophobic residues in helices 10 and 11 were proposed to be involved in VHL-binding to TRiC (Spiess *et al.*, 2006). These observations are consistent with several other findings such as genetic screens where deletion of hydrophobic sequences in actin was shown to disturb its interaction with CCT (Rommelaere *et al.*, 1999), proteolytic analysis that recognized hydrophobic-rich interdomain segment in tubulin as the TRiC-interaction site (Dobrzynski *et al.*, 1996) and stability studies where for example TRiC has been shown to preferentially recognize hydrophobic β-strands and stabilize reconstituted WD-40 proteins (Kubota *et al.*, 2006). While a clear consensus remains to be established, all these studies suggest a non-generalized and subunit-specific mechanism of client interaction for eukaryotic chaperonins.

For archaeal chaperonins on the other hand, previously, solved X-ray structures of archaeal chaperonins combined with the electron microscopy and NMR studies suggested two patches of hydrophobic residues, one occluding the cleft of the helical protrusion region and other at the base of the apical domain in a region corresponding to Group 1 client binding site, to be involved in client binding (Klumpp *et al.*, 1997; Schoehn *et al.*, 2000b; Heller *et al.*, 2004). However, the helical protrusion region was found to be dispensable for client binding (Iizuka *et al.*, 2004; Reissmann *et al.*, 2007). Albeit, detailed client binding studies to determine the nature of interactions have not been done for archaeal chaperonins so far, it has been shown that the ability of *S. solfataricus* chaperonin to prevent aggregation of clients, lysozyme and yeast α-

glucosidase, is largely governed by hydrophobic interactions (Guagliardi *et al.*, 1995). Evidently, on the basis of the recently solved crystal structure of *M. maripaludis* chaperonin (MmCCT) in an open conformation, an interface between helices 10 and 11 consisting of 9 hydrophobic residues has been put forward as the potential client binding site (Pereira *et al.*, 2010) (Figure 5.13). However precise genetic or mutational studies to complement these structural propositions have not yet been done.

The experimental system described in Chapter 3 provided an *in vivo* platform to examine properties of archaeal chaperonins such as analysis of client binding region in this case. The system allows genetic dissection of the archaeal chaperonin MmCCT and the resulting changes can be studied on the basis of growth of bacterial cells. Also of interesting note is that the results would be produced from the totality of the cell including at least all possible essential clients rather than a specific case.

Thus an attempt has been made to provide a genetic basis to the proposed client binding site for MmCCT (Pereira *et al.*, 2010) by mutating the residues and studying the effects of the resultant mutants on the growth of *E. coli* in absence of GroEL and GroES. To further substantiate the results, a preliminary experiment has been initiated where the mutants with limited ability to restore GroEL function, presumably because of restricted client binding function are further subjected to calmodulin-pull down analysis to test whether or not (and if yes, to what extent) proteins isolate with the potential client-site mutants.
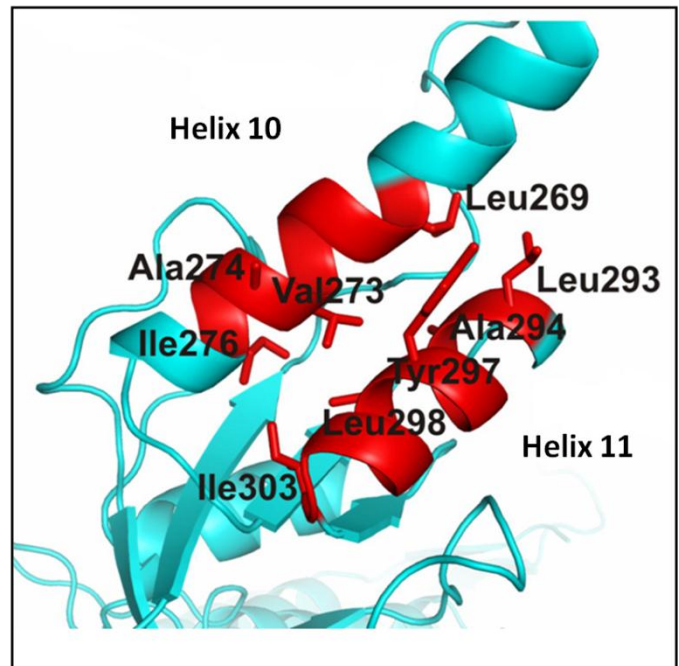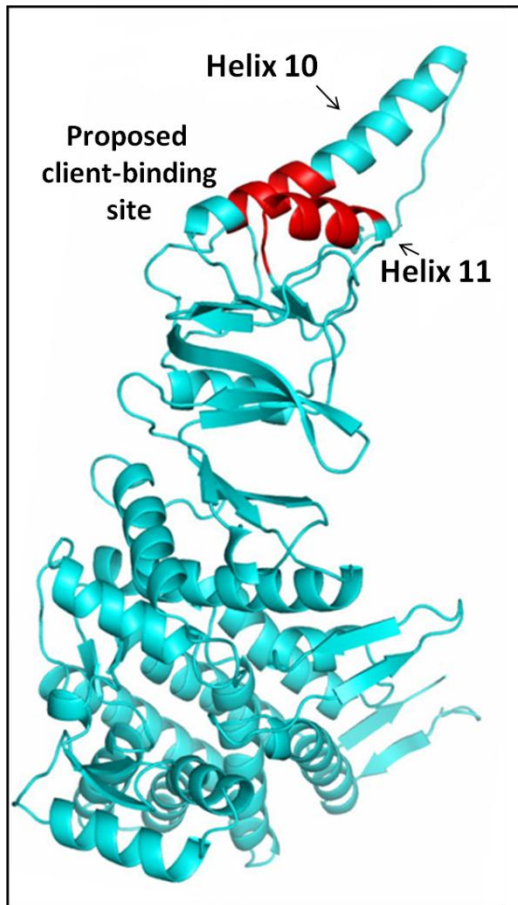
Figure 5.10

Proposed client binding region by Pereira *et al.*, 2010

The potential client binding site for MmCCT is shown as a red patch in the apical domain of MmCCT monomer on the left diagram. On the right are shown the nine hydrophobic residues proposed to be involved in client binding.

Note: Picture taken from Pereira *et al.,* 2010.

5.3.2 Results

5.3.2.1 Selection of residues for mutagenesis

The 9 hydrophobic residues i.e. L269, V273, A274, I276, L293, A294, Y297, L298, I303 encompassing the potential client binding site for MmCCT (Pereira *et al.*, 2010) are highly conserved across all archaeal chaperonins (Figure 5.11A and B). Out of these, V273, A274, L293 and Y297 residues are solvent-exposed and have been proposed to support direct interactions with the clients (Pereira *et al.*, 2010). Thus for this analysis two residues, V273 from helix 10 and Y297 from helix 11, on the basis of their particularly high conservation among the archaeal chaperonins were selected for mutagenesis (Figure 5.11 and 5.12).

5.3.2.2 Mutagenesis of residues V273 and Y297

Site-directed mutagenesis using two sets of primers that contain randomised bases at the positions corresponding to V273 and Y297 respectively was employed to generate a series of point mutants with different substitutions at V273 and Y297 (method as described in section 3.2.3, Chapter 3). The pET21 plasmid containing *Mmcct* gene that already bears the K216E mutation was used as a template for these experiments. Because MmCCT-K216E, as compared to the wildtype, is more efficient in rescuing the growth of the strain TAB21 in GroEL and GroES depleting conditions, it was speculated that a clearer comparative phenotypic analysis of the mutants would result with the additional K216E background mutation. As a consequence a more robust system is generated that allows better resolution of growth-based results of the potential client-site mutants. The mutagenesis experiments yielded a library of following mutants: MmCCT-V273A, D, E, G, I, N, S, T, W, Y and MmCCT-Y297A, C, E, F, G, N, R, S, T and W.

Figure 5.11A

Conservation of proposed client binding residues across 20 archaeal species

The arrows indicate the residues of the potential client binding region as suggested by Pereira *et al.*, 2010. Conservation scoring is performed by PRALINE (Simossis and Heringa, 2005). The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position. The colour assignments are:

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

Figure 5.14B

Conservation of proposed client binding residues across 20 archaeal species

The arrows indicate the residues of the potential client binding region as suggested by Pereira *et al.*, 2010. The sequence logos were generated by WebLogo server. The size of the amino acid letter in the sequence logo indicates the degree of conservation on a logarithmic scale. The colour assignements are: blue - hydrophilic amino acids (RKDENQ), green - neutral aminoacids (SGHTAP) and black - hydrophobic amino acids (YVMCLFIW).
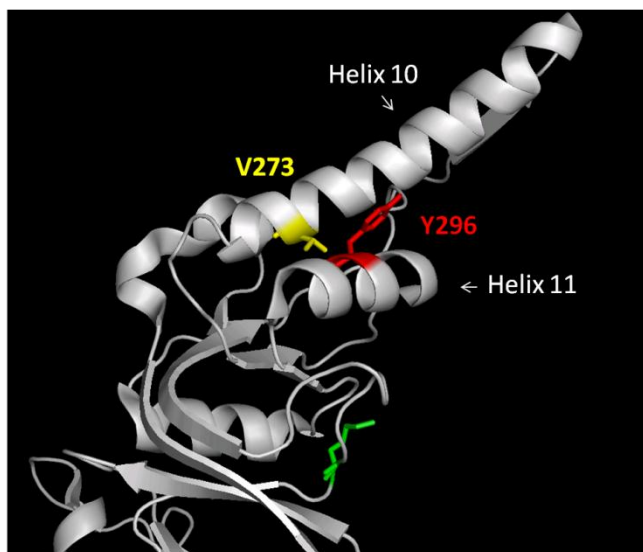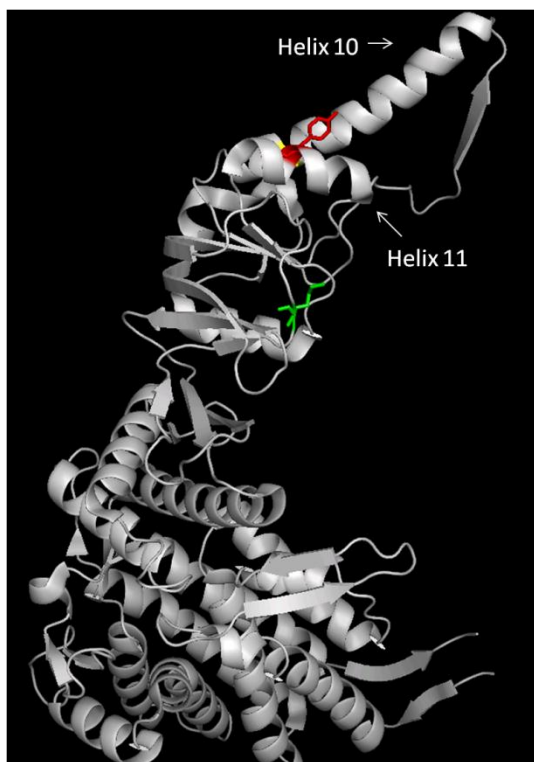
Figure 5.12

Location of residues V273 and Y297 on MmCCT

The figure shows MmCCT subunit highlighting residues V273 and Y297 in yellow and red respectively. The subunits were coloured in PYMOL from structures obtained from Protein Data Bank (PDB ID: 3KFK) submitted by Pereira *et al.*, 2010. K216 residue is highlighted in green.

5.3.2.3 GroEL-complementing ability of the potential client-site mutants

The obtained mutants were first checked for their correct expression and assembly in *E. coli* TAB21 strain by SDS-PAGE and native-PAGE analysis respectively. Figure 5.13 shows the expression and assembly of five of these mutants (MmCCT-K216E-V273A, D, L, S and W), and essentially represents the results obtained for rest of the tested mutants, the observations for which are recorded in Table 5.4.

All the mutants were then subjected to complementation analysis in TAB21. Out of the ten MmCCT-K216E-V273n point mutants, the proteins with mutations - V273A and V273I preserved the improved GroEL-complementing ability of MmCCT-K216E (Figure 5.14 and Table 5.4). Indeed, from a function-conservation point of view these are among the most preferred amino acid substitutions mostly due to the similarity in their amino acid properties (Betts and Russell, 2003). Along the same lines thus, MmCCT-K216E-Y297F and MmCCT-K216E-Y297W out of the ten Y297 point mutants were found to be able to restore the characteristic improved phenotype (Figure 5.15 and Table 5.4). The amino acids phenylalanine and tryptophan share similar properties with tyrosine such as having aromatic side-chains, partial hydrophobicity etc., and hence the corresponding substitutions are less likely to alter the chaperonin function. Indeed, a constitutive replacement of tyrosine to phenylalanine at the position equivalent to Y297 is observed in several archaeal species (Figure 5.11).

On the other hand, substituting V273 or Y297 to relatively distinct hydrophilic residues (D, E, N or R) or neutral residues (S and G) resulted in a significantly weaker phenotype of the cells (Figure 5.14 and 5.15, Table 5.4). This combined with the earlier observations indicated a requirement of hydrophobic residues at the positions V273 and

Y297 in MmCCT. Intriguingly however, substitutions of V273 to hydrophobic but aromatic residues like tryptophan and tyrosine severely affected the phenotype of the cells and conversely, the same effect was observed when Y297 was replaced by aliphatic residues like alanine and cysteine (Figure 5.14 and 5.15, Table 5.4). Such results probably suggest that in addition to a certain degree of hydrophobicity, other factors such as presence or absence of an aromatic ring, reactivity and conformational flexibility of side-chains among others, may also contribute to the functional significance of V273 and Y297 residues. Such an implication is further supported by the threonine mutants of V273 and Y297 that in contrast to other polar or neutral substitutions (like to serine or glycine) improved the phenotype (Figure 5.14 and 5.15). This observation particularly holds for the V273T mutants as the cells expressing MmCCT-K216E-V273T showed no apparent difference in growth pattern when compared to corresponding MmCCT-K216E control (Figure 5.14).

In sum, the results established that substituting V273 and Y297 to dissimilar residues has a rather negative effect on MmCCT function in *E. coli*. The next obvious question hence was to ask whether V273 or Y297-mutants actually impede client recognition or binding. We hypothesized that if the reduced function of such mutants was because of a perturbed client-binding ability, there should be a decline (or absence) in the number of proteins that interact with them. As a preliminary experiment thus, two of the potential client-site mutants were subjected to calmodulin based pull down assays to test the effects on the interacting proteins that in absence of the client-site mutations should isolate with MmCCT.
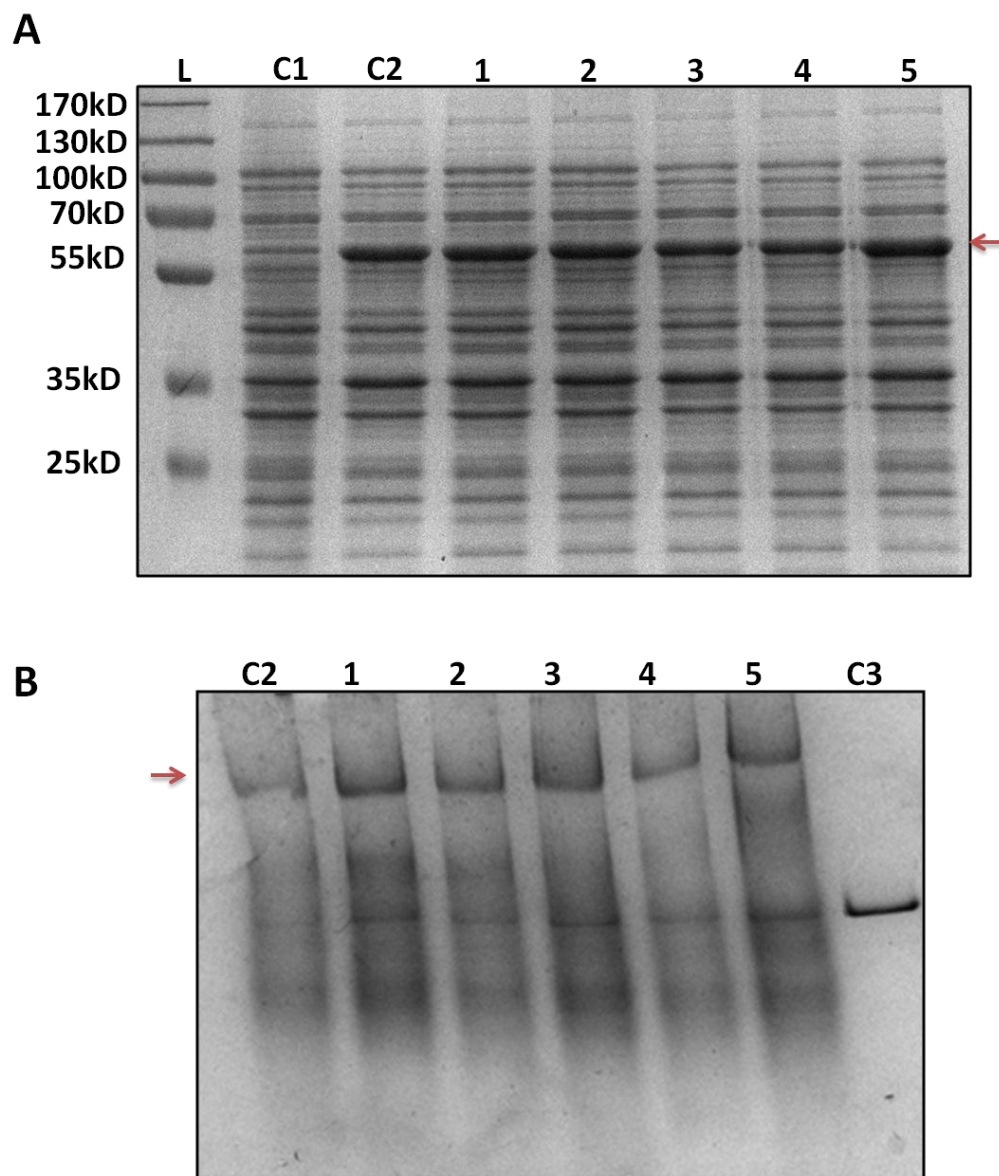
Figure 5.13

Expression and assembly of potential client-site mutants in TAB21

A) Crude extracts of TAB21 cells expressing 5 different MmCCT-K216E-V273n mutants namely V273A (1), V273D (2), V273L (3), V273S (4) and V273W (5) under pET21 vector along with controls pET21 empty vector (C1) and pET21-*Mmcct*-K216E (C2) as analysed by 10% SDS-PAGE.

B) Soluble fractions of the above samples as analyzed by 3-10% gradient native-PAGE.

The control C3 represents purified GroEL protein.

Both the gels are stained by Coomassie staining. The arrows indicate the position of the MmCCT bands.

Figure 5.14

Solid medium growth analysis of MmCCT-K216E-V273n mutants in TAB21 at 30^{0}C under GroEL depleting conditions

Dilutions (10^{-1} to 10^{-7}) of TAB21 cells expressing the MmCCT mutant proteins as indicated, grown on agar plates containing 0.2% glucose and 1 mM IPTG at 30°C for five days.

Work done by undergraduate students: Kiren Bains, Wingyun Cheung, Kate Wilkinson
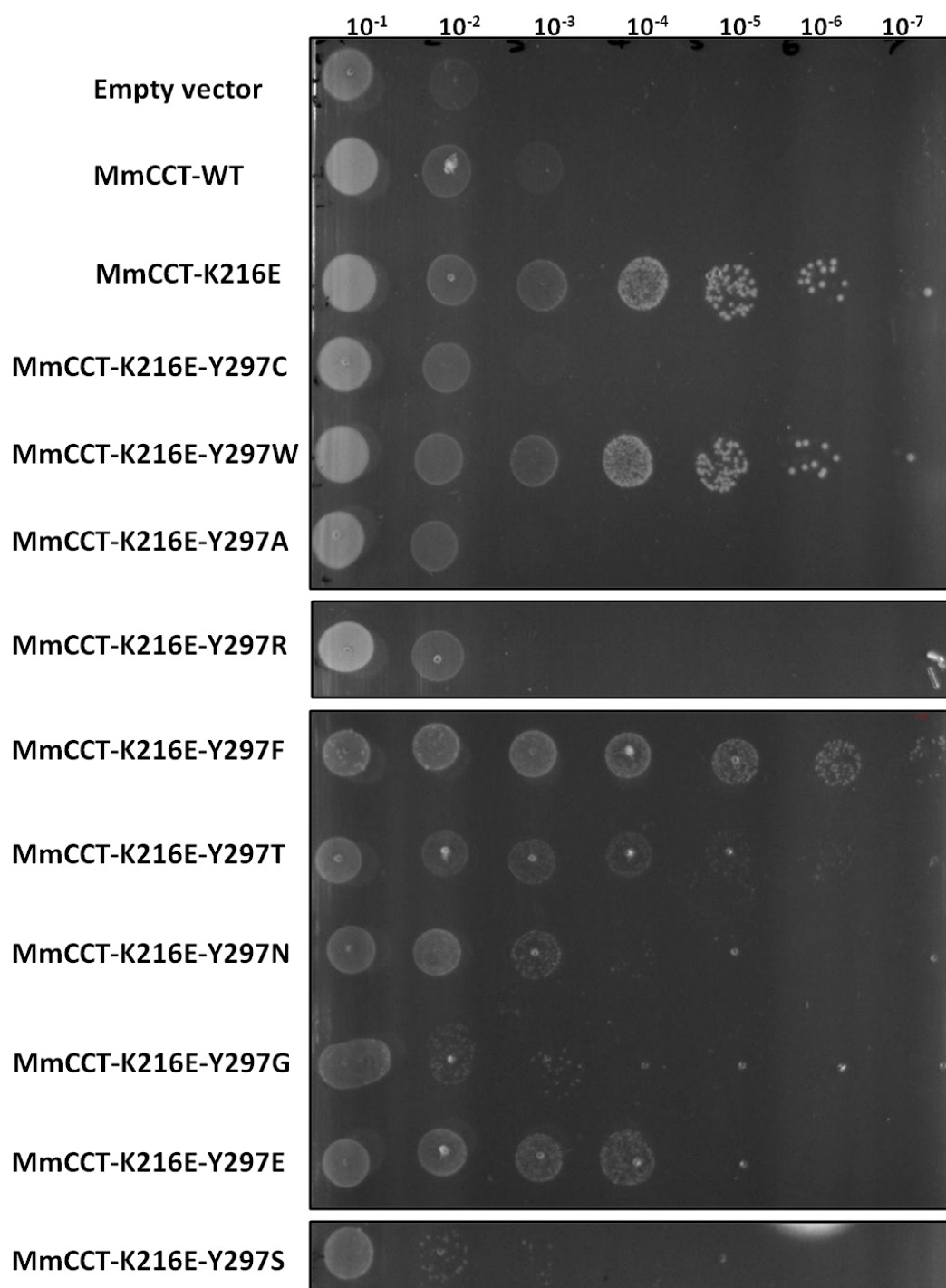
Figure 5.15

Solid medium growth analysis of MmCCT-K216E-Y297n mutants in TAB21 at 30^{0}C under GroEL depleting conditions

Dilutions (10^{-1} to 10^{-7}) of TAB21 cells expressing the MmCCT mutant proteins as indicated, grown on agar plates containing 0.2% glucose and 1 mM IPTG at 30°C for five days.

Work done by undergraduate students: Lynette Jude, Wingyun Cheung, Kate Wilkinson

Table 5.4: *In vivo* functional analysis of V273 and Y297 mutants in *E. coli*

Substitutions at V273				Substitutions at Y297			
Mutation in MmCCT	Expression	Assembly	Complementation (30°C)	Mutation in MmCCT	Expression	Assembly	Complementation (30°C)
V273D	Yes	Yes	+	Y297E	Yes	Yes	++
V273E	Yes	Yes	+	Y297N	Yes	Yes	++
V273N	Yes	Yes	+	Y297R	Yes	Not checked	+
V273S	Yes	Yes	++	Y296S	Yes	Yes	+
V273G	Yes	Not checked	+	Y297G	Yes	Not checked	+
V273T	Yes	Yes	+++	Y297T	Yes	Yes	++
V273A	Yes	Yes	+++	Y297A	Yes	Yes	+
V273I	Yes	Yes	+++	Y297C	Yes	Not checked	+
V273W	Yes	Yes	+	Y297F	Yes	Yes	+++
V273Y	Yes	Not checked	+	Y297W	Yes	Yes	+++

Note:

1. All mutants have a background K216E mutation.
2. Colour assignments are based the increasing hydrophobicity index of the substituted amino acids (Monera et al., 1995):
   Blue – hydrophilic substitutions, green – neutral substitutions, black – hydrophobic substitutions
3. Symbol assignments:
   +++   Growth similar to MmCCT-K216E mutant
   ++   Growth similar to MmCCT-WT
   +   Growth poorer than MmCCT-WT

5.3.2.4 Pull-down assays of potential client site mutants

The mutants MmCCT-K216E-V273W and MmCCT-K216E-Y296A were selected for the calmodulin based pull down analysis on account of two reasons: 1) The cells expressing these mutants showed a severely impaired phenotype and 2) The results obtained from these substitutions could highlight the importance of properties such as aliphaticity or aromaticity at V273 and Y297 respectively, in addition to their hydrophobic character. As the mutant proteins were not tagged, the experiments were conducted by co-transforming the TAB21 cells with pET21 vector containing the mutated *Mmcct* genes and pACYC184 containing tagged *Mmcct*-wt. Because the number of tagged subunits per a stable oligomeric complex was calculated to be less than five at any given time (Figure 5.9), it was anticipated that any resultant effect on bound proteins would be because of the client binding mutations.

Figure 5.16 shows the eluates from these pull-down experiments as analysed by SDS-PAGE and silver staining. Differences in the profiles of the proteins which co-elute with MmCCT can be seen between the wildtype and K216E controls (lanes 3, 5) and the potential client site mutants (lanes 4, 5). However the resolution of the gel is not sufficiently clear to determine whether this difference is significant and it is clear that many proteins still co-elute with the versions of MmCCT which carry the potential client binding site mutants. Additional experiments are needed to determine whether or not the apparent difference is significant, which if so would support the identification of the client binding site. It should be noted that the complexes in this experiment will be a mix of wild-type and mutated sub-units, so an obvious early experiment would be to introduce the V273W and V297A mutations into the tagged sub-units.
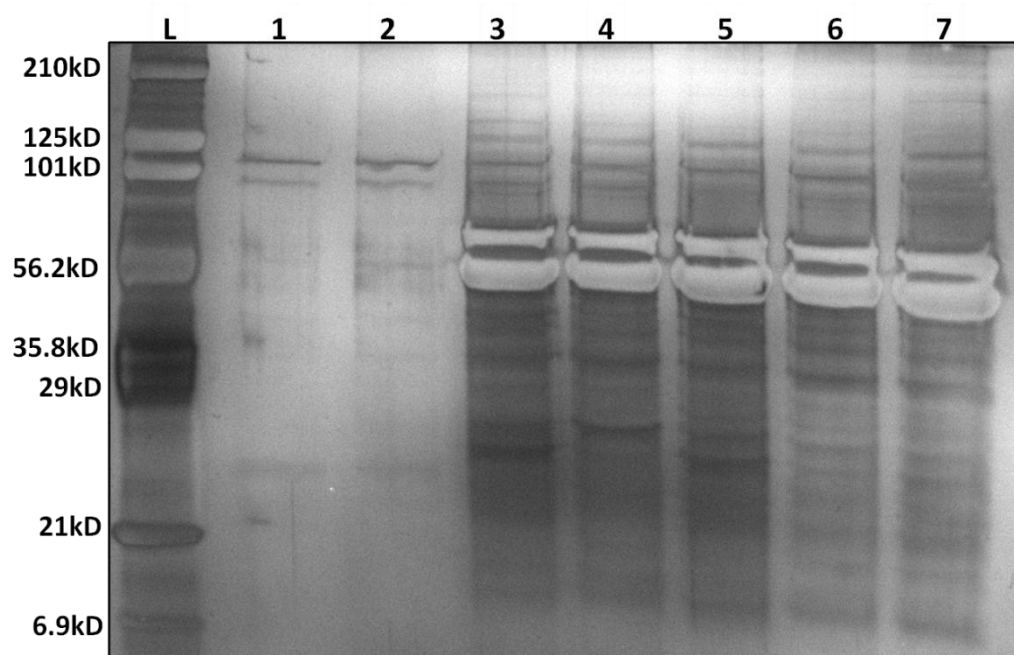
Figure 5.16

Calmodulin based pull down analysis of potential MmCCT client-site mutants

The figure shows protein eluates from the pull-down analysis as analysed by 10% SDS-PAGE and silver staining.

(1)   TAB21/ pET21b + pACYC184 empty vectors

(2)   TAB21/ pET21-*Mmcct*-WT

(3)   TAB21/ pET21-Mmcct-WT+ pACYC184-*Mmcct*-WT-tag

(4)   TAB21/ pET21-Mmcct-M223I + pACYC184-*Mmcct*-WT-tag

(5)   TAB21/ pET21-Mmcct-K216E + pACYC184-*Mmcct*-WT-tag

(6)   TAB21/ pET21-Mmcct-K216E-V273W + pACYC184-*Mmcct*-WT-tag

(7)   TAB21/ pET21-Mmcct-K216E-Y297A + pACYC184-*Mmcct*-WT-tag

5.3.3 Discussion

The client binding sites for Group 2 chaperonins have not been clearly defined despite the various structural propositions for both eukaryotic and archaeal chaperonins. However based on recently solved crystal structure of MmCCT in its open conformation, a pocket like structure formed by nine highly conserved hydrophobic residues at the interface of helices 10 and 11, which structurally correspond to client binding region between helices H and I of GroEL, has been put forward as the potential client binding site for MmCCT (Pereira *et al.*, 2010). Using the experimental system described in Chapter 3, we initiated a mutagenic analysis of two of these residues, V273 and Y297, and tested the effects of the resultant MmCCT-mutants on growth of *E. coli* TAB21 cells in GroEL depleting conditions.

Our findings suggest that mutating residues V273 and Y297 indeed affects the chaperonin function of MmCCT as seen by altered GroEL-complementing capacity of most of the mutants. The impaired cell growth due to hydrophilic and certain neutral substitutions probably highlights the importance of hydrophobicity at V273 and Y297 positions. However, severe growth reduction due to aromatic-hydrophobic substitutions at V273 and in contrast, the enhanced growth effects due to a polar threonine-substitution indicates that the functional contribution of the residue V273 might not be restricted solely to its hydrophobic property. Similarly, the growth impairment observed due to hydrophobic-aliphatic amino acid substitutions at Y297 more obviously suggests that in addition to hydrophobicity, the aromatic nature of this residue is possibly critical. The equivalent Y301 residue in *T. acidophilum* has been shown to undergo severe positional changes with the movement of highly dyanamic

helical protrusion, suggesting that this residue probably due to its aromatic stacking interactions might also be playing a stabilizing role by causing compensatory changes in conformations of surrounding residues (Klumpp *et al.*, 1997). Moreover, improved growth because of polar theronine substitutions at both V273 and Y297 suggest that a certain degree of functional redundancy is allowed at these residues. While it is possible that other hydrophobic residues might still permit the supposed hydrophobic client interactions, the overall results question to a certain extent the hypothesis that these residues govern client recognition and binding solely on the basis of their hydrophobic nature (Spiess *et al.*, 2006; Pereira *et al.*, 2010).

It should be noted however that the experiments do not rule out any secondary effects that could arise due to additional mutations such as altered ATP hydrolysis activity, conformational restrictions, oligomeric stability among others. Such changes, although unlikely to occur because of the mutations being in apical domain helices 10 and 11, could lead to a functional compromise or enhancement that in turn affects the growth of the cells. While purification of the mutants and direct client binding assays (e.g. citrate synthase aggregation assay) will be required to establish whether or not these residues support client binding, we attempted a pull down assay as a preliminary step to test if proteins isolate with the potential client-site mutants. The apparent decrease in the number of pulled-down proteins for the mutants MmCCT-V273W and MmCCT-Y297A as compared to the wildtype controls possibly indicate a negative effect on client binding. However the result cannot be ascertained because a significant amount of proteins still seem to co-elute with the mutants. Because it was a heterologous expression experiment, it is possible that the proteins that isolate with the client-site mutants are

because of their supposedly unhampered binding to native valine or tyrosine present on the tagged subunits. The next series of experiments thus with all tagged subunits containing mutations should more precisely answer this question. This assay will not only show a direct demonstration of the clients that isolate or not with the mutants but also will be helpful in answering questions such as whether or not all subunits are involved in binding.

Indeed a more detailed mutagenic analysis of all nine residues and in addition, of other polar and charged residues in archaeal chaperonins that for example coincide with the proposed client binding region of mouse CCTγ or yeast CCT (Pappenberger *et al.*, 2002; Dekker *et al.*, 2011a) or are found at the structurally equivalent positions of GroEL client site residues (Klumpp *et al.*, 1997; Horwich *et al.*, 2007), will be required to fully delineate the basis of client binding interactions in archaeal chaperonins. Nonetheless, our findings clearly indicate an essential role of residues V273 and Y297 in MmCCT function.

CHAPTER 6
Analysis of additional archaeal chaperonins for their ability to replace GroEL

6.1 Background

It was established in Chapter 4 and Chapter 5 that the archaeal chaperonin from the species *Methanococcus maripaludis,* MmCCT, and its single point mutants MmCCT-M223I and MmCCT-K216E complemented for the loss of GroEL in *E. coli* at reduced temperatures (Chapter 3). We next wished to ask whether this was a generalized case for all archaeal chaperonins or a specialized functioning confined only to MmCCT.

Our initial hypothesis was that the archaeal chaperonins should not be able to complement for the loss of GroEL in *E. coli* owing to approximately 2.8 billion years of gap in their divergence (Battistuzzi *et al.*, 2004) and the structural and mechanistic differences that prevail between the groups (Chapter 1). However, the case of MmCCT clearly suggested that a functional overlap between the two groups is possible, more so by simple genetic manipulations. It should be considered however that *M. maripaludis* is a mesophile with a growth optimum at 37^0C (Jones *et al.,* 1983), similar to *E. coli*. Also, *M. maripaludis* contains only one copy of *cct* gene and hence the chaperonin MmCCT exhibits a GroEL like homologous subunit arrangement, with eight fold symmetry (Kusmierczyk and Martin, 2003b; Large and Lund, 2009). Archaeal chaperonins from almost all other genera posses more than one chaperonin gene (Large and Lund, 2009). It is possible therefore that the ability of MmCCT to function in *E. coli* is because like GroEL, it is a homo-oligomeric complex and that the same might not hold true for other divergent archaeal chaperonins. Indeed it has been speculated that the evolution of Group 1 and Group 2 chaperonins has occurred in parallel to the proteome evolution to attain client-specific mechanisms of working (Kubota *et al.,* 1994; Archibald *et al.,* 2001; Fares and Wolfe, 2003; Dekker *et al.,* 2011b).

However note that there are stark differences between the evolutionary courses of eukaryotic and archaeal chaperonins. In contrast to the eukaryotic chaperonin genes that have been shown to have diverged very early during the evolution of eukaryotes (Archibald *et al.*, 2000; Archibald and Roger, 2002a; Fares and Wolfe, 2003), the gene duplication events of archaeal chaperonins are combined with recurrent gene conversions and gene loss within individual archaeal lineages. Such observations suggest a role of neutral fixation in hetero-oligomerism of archaeal chaperonins over specific functionalization of the diverged subunits (Archibald *et al.*, 1999; Archibald and Roger, 2002b; Ruano-Rubio and Fares, 2007). Indeed while it has been shown that all eight subunits of eukaryotic chaperonins are essential in yeast (Stoldt *et al.*, 1996), subunits of archaeal chaperonin *H. volcanii* have been found to be individually dispensable (Kapatai *et al.*, 2006).

Moreover, functional homo-oligomeric assembly of paralogs of several archaeal chaperonins have been successfully shown to occur in *E. coli.* For example, expression of α-only and β-only chaperonins from *T. acidophilum* and *Thermococcus* strain KS-1 have been shown to form homo-oligomeric complexes in *E. coli* with optimal ATPase activities at particular temperatures that can effectively protect various proteins against thermal aggregation and also promote *in vitro* refolding (Waldmann *et al.*, 1995b; Yoshida *et al.*, 1997; Hirai *et al.*, 2008). Similar observations have been also noted for α and γ subunits of *M. mazei* (Klunker et al., 2003) and the two subunits of *A. pernix K1* (Lee et al., 2013). Although, there also have been certain cases like for the chaperonins of *Sulfolobus sp.* S7 (Yoshida *et al.*, 1998) and *P. occultum* (Minuth et al., 1998), where despite being assembled into functionally active homo-oligomers in *E. coli*, the chaperonin complexes

are thermally unstable. Nonetheless, all the above examples suggest that the hetero-oligomeric nature of archaeal chaperonins might not limit the cross-over of function between archaeal and bacterial chaperonins that has been demonstrated by MmCCT. It was intriguing hence to find out whether or not other archaeal chaperonins could replicate the results obtained of MmCCT.

Consequently, the aim of the project was to test the ability of chaperonins from distinct archaeal species to replace GroEL function in *E. coli*. Taking into account the above discussed points, this part of the project had following objectives: 1) To select chaperonins from distinct archaeal species and test their GroEL-complementing ability in *E. coli*. 2) To construct M223I and K216E mutants of these archaeal chaperonins and test them for the same.

The archaeal species to be tested were selected on the basis of one or more of the following criteria: 1) Genetic tractability 2) Previous successful examples of expression in *E. coli* and 3) Availability of the cloned chaperonin gene. The chaperonins from following species were chosen for the experiments.

Methanosarcina mazei: *M. mazei* is an anaerobic mesophilic methanogen containing three different thermosome genes annotated as α, β, and γ. The α and γ chaperonins have been shown to express and assemble independently in *E. coli* (Klunker et al., 2003). The peculiarity of this species is that it contains both Group 1 and Group 2 chaperonins; along with the three thermosome genes, a Group 1 chaperonin (Mma-groESL) gene probably acquired by lateral gene transfer is also found (Klunker *et al.*, 2003; Figueiredo *et al.*, 2004). As discussed in Chapter 1 (section 1.2.4), in addition to an overlapping set of protein interactors, Mma-GroESL and the thermosomes have been shown to

exclusively associate to various archaeal proteins. Also of interest was that a fraction of the sole thermosome interactors were of prokaryotic origin, out of which some were found to be homologues of GroEL clients, thought to be acquired by horizontal gene transfer (Hirtreiter et al., 2009). It was interesting hence to test whether these thermosome proteins could complement for GroEL function.

Thermoplasma acidophilum: *T. acidophilum* is a thermoacidophile containing two thermosome genes α and β and as noted above, both the genes have been shown to express and assemble into functionally active homo-oligomers in *E. coli* with comparable biochemical properties as to the corresponding hetero-oligomers (Waldmann *et al.*, 1995b; Bigotti and Clarke, 2005; Hirai *et al.*, 2008). However in its native form, the oligomer shows a strict α-β repeat (Ditzel *et al.*, 1998). We chose to express both the subunits together in *E. coli* to test their GroEL-complementing ability.

Haloferax volcanii: This species is an extreme halophile and contains three chaperonin genes annotated as *cct1, cct2* and *cct3* (Kuo *et al.*, 1997; Large *et al.*, 2002). Interestingly, all three *cct* genes have been shown to form homo-oligomeric complexes in *H. volcanii* and both *cct1* and *cct2* genes have been reported to be able to sustain growth of *H. volcanii* independently (Kapatai et al., 2006). Because *cct1* gene was shown to be more efficient in supporting growth than *cct2, cct1* gene was chosen for analysis.

All three species selected were from a different family than the *Methanococcus* and of a different nature i.e. a methanogen, a thermophile and a halophile. Moreover, while single *cct* genes were individually tested for *M. mazei and H. volcanii*, both α and β *cct* genes were expressed together for *T. acidophilum*. Such distinctive features were employed to facilitate generalization of the results to the archaeal Group 2 chaperonins.

The characteristics of the archaeal chaperonins chosen and their nomenclature used henceforth are given in Table 6.1. The standard TAB21 expression system using pET vectors was employed for the analysis.

Table 6.1: The archaeal species chosen for GroEL-complementation analysis and their characteristics.

Archaeal species	Number of CCT subunits in native host	Nomenclature	Subunits chosen for analysis	Source and References
Methanosarcina mazei	Three (α, β, and γ)	MmaCCT-α MmaCCT-β MmaCCT-γ	MmaCCT-α and MmaCCT-γ (MmaCCT-β unable to assemble in *E. coli*)	F.U. Hartl group, Max Plank Institute, Germany. (Klunker *et al.*, 2003; Figueiredo *et al.*, 2004)
Haloferax volcanii	Three (1, 2 and 3)	HvCCT-1 HvCCT-2 HvCCT-3	HvCCT-1	Laboratory colleague (Kapatai *et al.*, 2006; Large *et al.*, 2007)
Thermoplasma acidophilum	Two (α and β)	TaCCT_α TaCCT_β	Both subunits in tandem as TaCCT_αβ	A. R. Clarke group, University of Bristol. (Bigotti and Clarke, 2005)

6.2 Results

6.2.1 Cloning of chaperonin genes in a pET expression plasmid

The constructs containing *cct* genes from *M. mazei* (*Mmacct-α and Mmacct-γ*), *H. volcanii* (*Hvcct-1*) and *T. acidophilum* (*Tacct-αβ*) were obtained in pET22, PrV and pET27 vector backbones respectively. The first step hence was to clone them into appropriate pET expression vectors. Because the TAB21 strain contained a kanamycin resistance gene (Kan^R) in itself, pET expression vectors bearing selectable markers that are different from Kanamycin were essential. From the available wide range of pET expression plasmids, pET21, pET22, pET31 and pET32 containing ampicillin resistant (Amp^R) genes and a T7lac promoter qualified for the given criteria (pET system manual, Novagen).

The *M. mazei cct* gene constructs were already in pET22 plasmid and hence proved directly acceptable for further analysis. For *Hvcct-1*, the gene sequence was excised from the parent PrV plasmid using unique restriction sites NdeI and EcoRI and introduced into the pET21 expression plasmid using a cloning vector pOK12 as shown in Figure 6.1B. The pET21 plasmid was chosen because of the availability in the laboratory. The pET27 plasmid containing *T. acidophilum* chaperonin genes had Kan^R as a selectable marker. Hence using restriction sites NheI and HpaI unique to pET27 and pET21 plasmids, a fragment swap was performed to yield pET21- *Tacct-αβ* (Figure 6.1). The resulting four constructs namely pET22- *Mmacct-α,* pET22- *Mmacct-γ,* pET21-*Hvcct-1* and pET21-*Tacct-αβ* were tested for their ability to express, assemble and complement for loss of GroEL in *E. coli* TAB21 strain as discussed below.
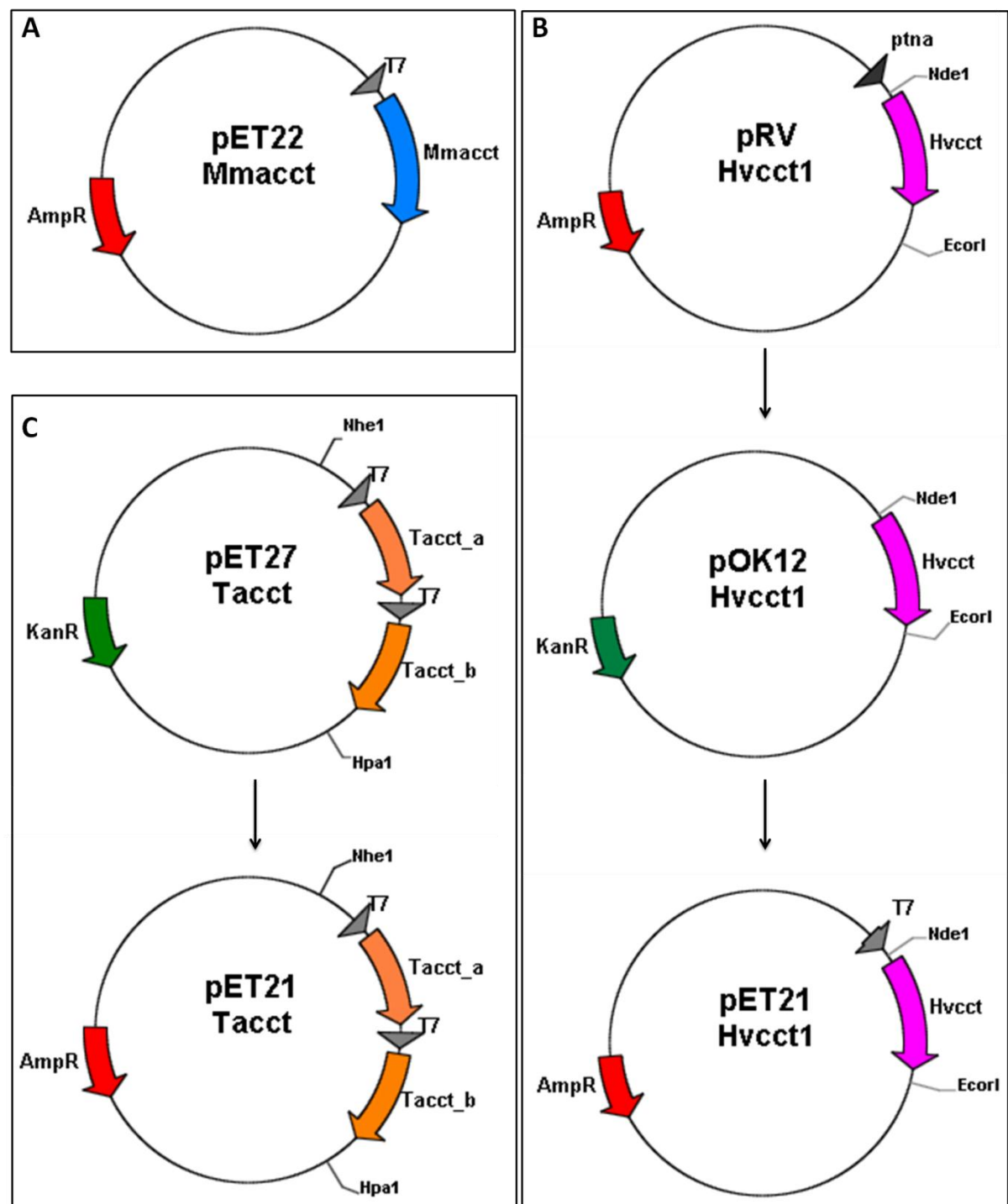
Figure 6.1
Schematic diagram showing cloning of archaeal chaperonin genes into pET expression plasmids.

6.2.2 Characterization of MmaCCTs, HvCCT-1 and TaCCT-αβ in *E. coli*

6.2.2.1   Protein expression by SDS-PAGE

The pET constructs of all archaeal chaperonins were transformed into TAB21 and grown in IPTG-induced and non-induced conditions. The protein samples were collected and first tested for expression of the chaperonins by SDS-PAGE. The MmaCCT-α and MmaCCT-γ proteins were highly overexpressed under IPTG induction as shown in Figure 6.2. The expression of MmaCCT-α was consistently observed to be higher than that of the MmaCCT-γ. The HvCCT-1 expressed well in IPTG induced state albeit in an anomalously slowly running position (Figure 6.3). The discrepancy between the molecular mass deduced theoretically from the protein sequence of HvCCT-1 (~60kDa) and the one observed on the gel has been routinely observed and attributed to the excess of negatively charged acidic amino acids (calculated Pi = 6.2) that affects the mobility of the protein on SDS-polyacrylamide gels (Large *et al.*, 2002; Kapatai *et al.*, 2006). The expression of TaCCT-αβ under IPTG induced conditions was lower than MmaCCTs or HvCCT-1. However, a time dependent analysis confirmed correct expression of TaCCT-αβ (Figure 6.4). Unexpectedly, protein samples from non-induced cultures also showed over-expression of a protein. The over-expressed protein in the non-induced sample ran at slightly lower position on the gel as compared to the over-expressed protein in the induced samples (Figure 6.3). However when checked for assembly by native-PAGE as described in the next section, both the samples showed assembly of chaperonin complex of similar molecular masses suggesting that the over expressed protein in the non-induced samples could be truncated TaCCT or one of its individual subunits.
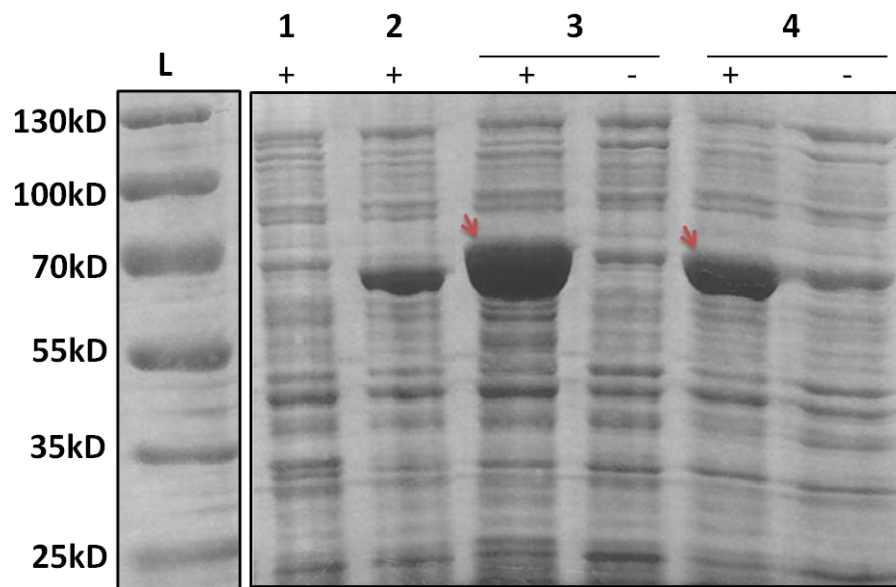
Figure 6.2

Expression of *M. mazei* chaperonin alpha and gamma subunits in *E. coli.*

Crude extracts from *E. coli* TAB21 cells overexpressing CCT proteins in IPTG (1 mM) induced (+) and non-induced (-) conditions as analyzed by 10% SDS-PAGE and Coomassie staining. Arrowheads point to the positions of the overexpressed proteins.

(1)   TAB21/pET21 empty vector
(2)   TAB21/pET21-*Mmcct*-WT
(3)   TAB21/pET22-*Mmacct*- α-WT
(4)   TAB21/pET22-*Mmacct*- γ-WT

Figure 6.3

Expression of *H. volcanii and T. acidophilum* chaperonins in *E. coli*.

Crude extracts from *E. coli* TAB21 cells overexpressing CCT proteins in IPTG (1 mM) induced (+) and non-induced (-) conditions as analyzed by 10% SDS-PAGE and Coomassie staining. Arrowheads point to the positions of the overexpressed proteins.

(1)   TAB21/pET21 empty vector

(2)   TAB21/pET21-*Mmcct*-WT

(3)   TAB21/pET21-*Hvcct*- 1

(4)   TAB21/pET21- *Tacct*-α_β
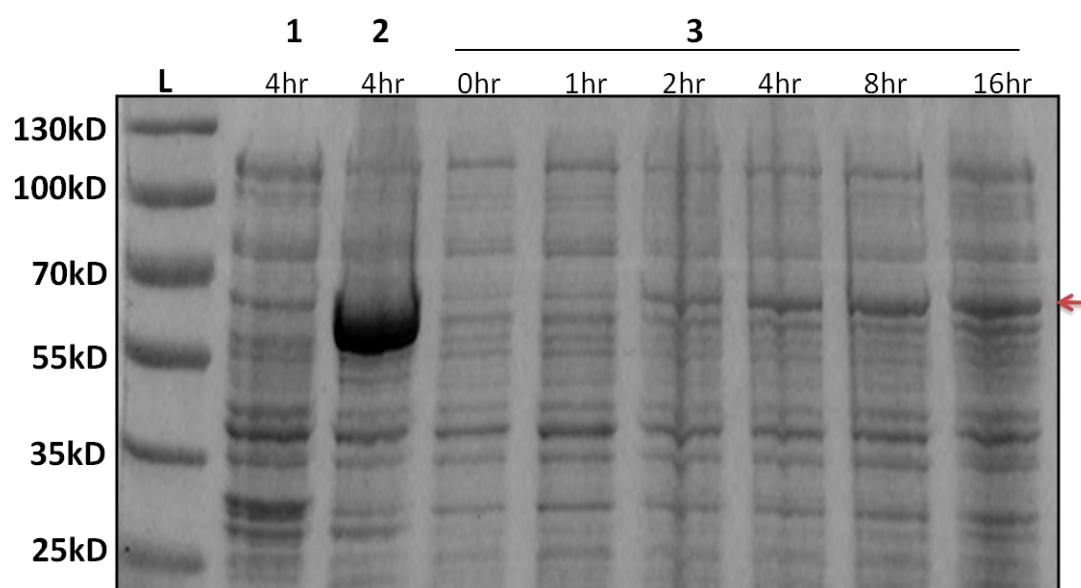
Figure 6.4

Expression of *T. acidophilum* chaperonins TaCCT-α_β in a time-dependent manner.

Crude extracts from *E. coli* TAB21 cells overexpressing CCT proteins prepared at different time intervals after IPTG induction as indicated, analyzed by 10% SDS-PAGE and Coomassie staining. Arrowhead points to the position of the overexpressed proteins.

    (1)   TAB21/pET21 empty vector

    (2)   TAB21/pET21-*Mmcct*-WT

    (3)   TAB21/pET21- *Tacct*-α_β

6.2.2.2 Native-PAGE analysis:

The protein samples once checked for their expression were then subjected to native gel electrophoresis to test their correct assembly. MmaCCT-α and MmaCCT-γ assembled into their complexes as demonstrated by discrete bands on native-PAGE gel (Figure 6.5). Protein bands were not observed for HvCCT-1 (Figure 6.6, No.3), suggesting either failure in oligomeric assembly or formation of unstable complexes that fell apart rapidly. Several variations in lysis protocols, pH and buffers were adopted to ensure that the results were not limited by technical errors. Ultimately it was concluded that HvCCT-1 did not form functional oligomers in *E. coli*. TaCCT-αβ demonstrated weak assembly under IPTG induced conditions (Figure 6.6, No.4). However, a rather stronger band with equivalent molecular mass was consistently observed under non-induced conditions (Figure 6.6, No.4). The result suggested that the over-expressed protein seen in non-induced conditions during SDS-PAGE analysis was probably that of TaCCT. The possible reasons for higher expression of protein in the non-induced cultures could be attributed to how the construct has been made and will be discussed at length later.
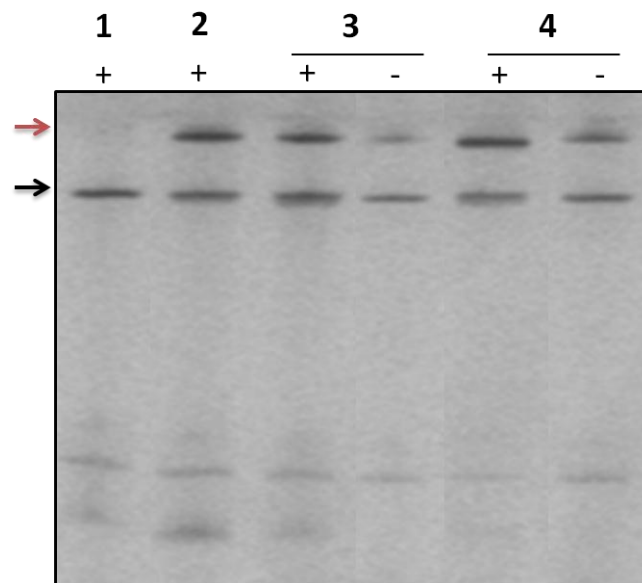
Figure 6.5

Assembly of *M. mazei* chaperonin alpha and gamma subunits in *E. coli.*

Soluble fractions of *E. coli* TAB21 cells expressing CCT proteins in IPTG induced (+) and non-induced (-) conditions as analyzed by 7.5% Native PAGE and Coomassie staining. Red and black arrows show CCT and GroEL proteins respectively.

    (1)   TAB21/pET21 empty vector

    (2)   TAB21/pET21-*Mmcct*-WT

    (3)   TAB21/pET22-*Mmacct*- α-WT

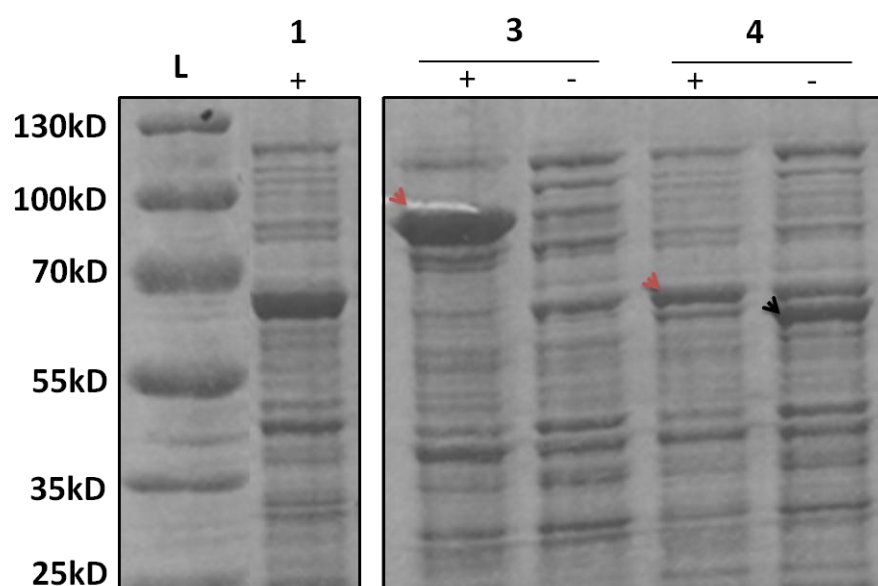   (4)   TAB21/pET22-*Mmacct*- γ-WT
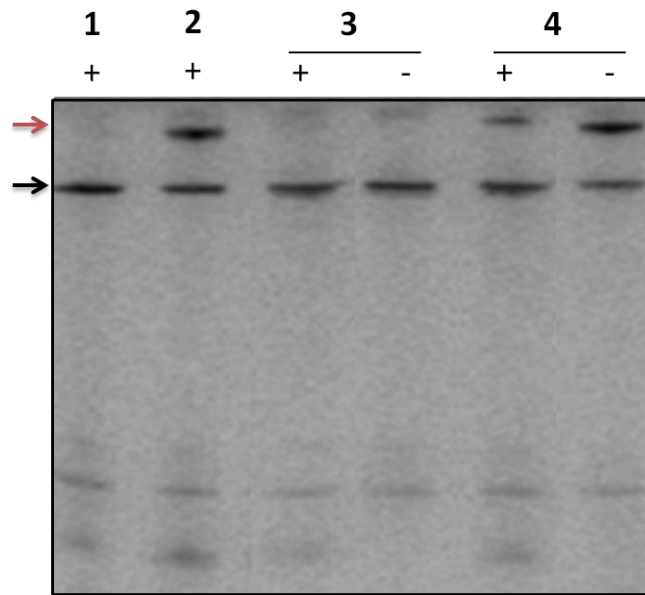
Figure 6.6

Assembly of *H. volcanii and T. acidophilum* chaperonins in *E. coli*.

Soluble fractions of *E. coli* TAB21 cells expressing CCT proteins in IPTG induced (+) and non-induced (-) conditions as analyzed by 7.5% Native PAGE and Coomassie staining. Red and black arrows show CCT and GroEL proteins respectively.

    (1)   TAB21/pET21 empty vector
    (2)   TAB21/pET21-*Mmcct*-WT
    (3)   TAB21/pET21-*Hvcct*- 1
    (4)   TAB21/pET21- *Tacct*-αβ

6.2.2.3   Complementation analysis:

All the archaeal chaperonins were then investigated for their ability to rescue growth of *E. coli* in absence of the GroEL and GroES using TAB21 strain. Because Hv-CCT1 did not assemble, it was not expected to complement. However, it was included in the analysis as it could serve as a negative control. The complementation assay was done in the same way as discussed in Chapter 3. None of the archaeal chaperonins tested could rescue growth of *E. coli* in presence of glucose at 30^0C (Figure 6.7). The experiments were repeated at a lower temperature (26^0C) and higher temperatures (37^0C and 42^0C) and with longer incubation times, however no change in the results was seen. The positive controls MmCCT-WT and MmCCT-M223I worked as expected each time, confirming right experimental conditions. All strains grew well in presence of arabinose, where native GroEL functioning is restored. Complementation experiments were repeated several times with the same results.

These findings suggested that none of chaperonin proteins tested were able to replace the GroE function in *E. coli*. However, we speculated that if the same mutations that improved the function of MmCCT in GroE depleting *E. coli* cells were introduced in these archaeal chaperonins, a similar effect could arise.

A

TAB21/pET22-*Mmacct*-α

TAB21/pET22-*Mmacct*-β

TAB21/pET22-*Mmacct*-γ

TAB21/pET21-*Tacct*-α_β

TAB21/pET21- *Hvcct*-1

Empty vector

TAB21/pET21-*Mmcct*-M223I

TAB21/*Mmcct*-WT

B

TAB21/pET22-*Mmacct*-α

TAB21/pET22-*Mmacct*-β

TAB21/pET22-*Mmacct*-γ

TAB21/pET21-*Tacct*-α_β

TAB21/pET21- *Hvcct*-1

Empty vector

TAB21/pET21-*Mmcct*-M223I

TAB21/*Mmcct*-WT
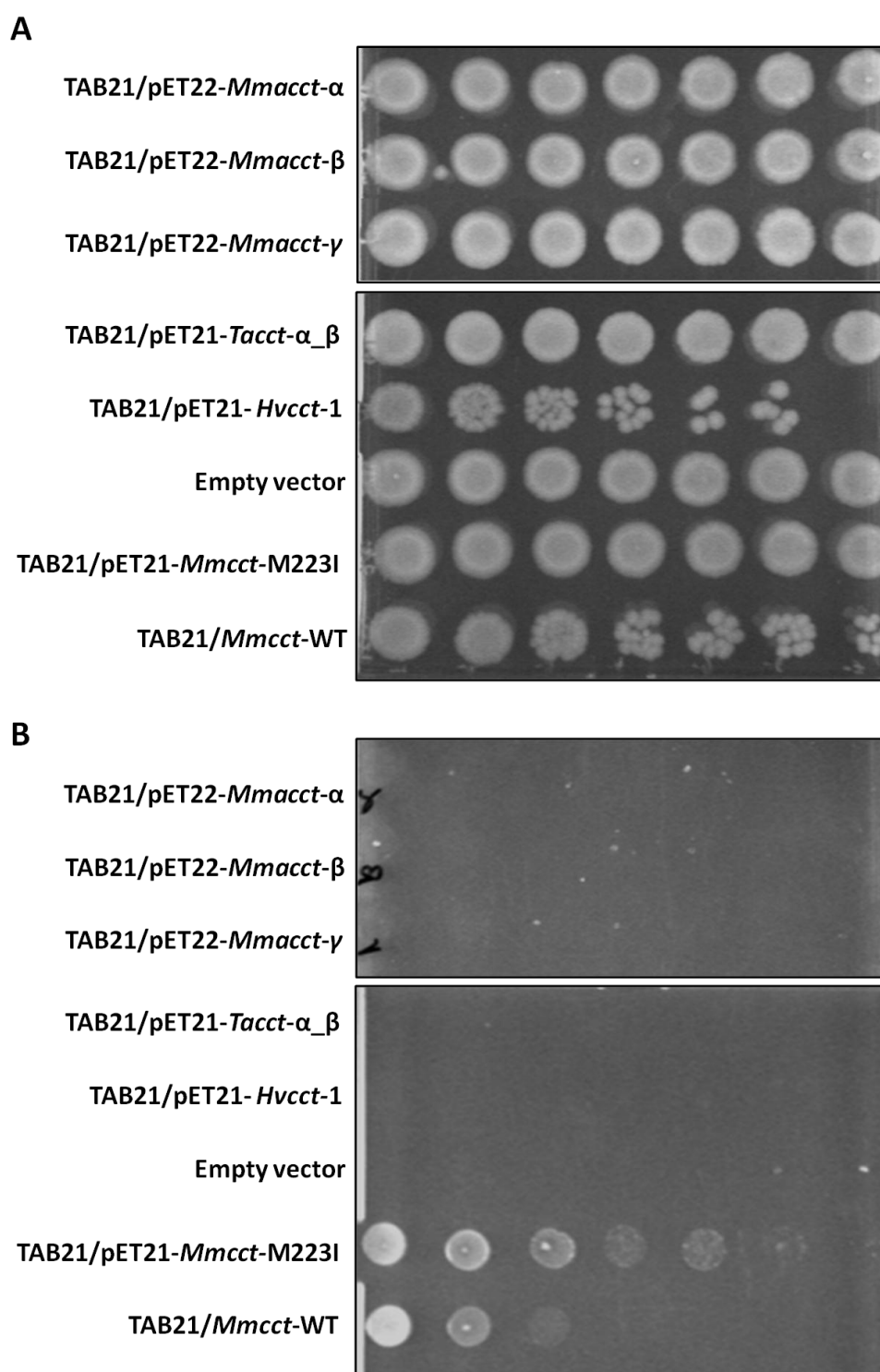
Figure 6.7

Solid medium growth analysis of *M. mazei, H. volcanii* and *T. acidophilum* chaperonins

Dilutions (10^{-1} to 10^{-7}) of TAB21 cells expressing the CCT proteins as indicated, grown on LB agar plates containing arabinose (*groE* induction) at 37°C **(A)** and on 0.2% glucose (*groE* repression) and 1mM IPTG (*cct* induction) at 30°C **(B)**.

6.2.3 Analysis of MmaCCT-α-M223I and MmaCCT-α-K216E mutants.

Single point mutations, M223I and K216E in MmCCT significantly increased its ability to complement for GroEL and GroES in *E. coli* (Chapter 3). These two residues are highly conserved across all archaeal chaperonins as shown in Chapter 3 (Figure 3.8). Because *M. mazei* is a mesophile as *M. maripaludis,* it can be speculated that chances of MmaCCTs to replicate MmCCT results would be higher than the other archaeal chaperonins. Indeed, a sequence alignment of MmCCT with MmaCCTs, TaCCTs and HvCCTs showed maximum similarity between MmCCT and MmaCCT-α (~30% identity and ~50% similarity) (Table 6.2). Thus, MmaCCT-α was selected for the initial analysis. If the results obtained were positive for MmaCCT-α chaperonin, the experiments could be extended for the other archaeal chaperonins.

Using site-directed mutagenesis, equivalent mutations were introduced at the conserved M226 and K219 positions of MmaCCT-α (Figure 6.8). The resultant MmaCCT-α-M226I and MmaCCT-α-K219E constructs were subjected to SDS-PAGE, Native PAGE and complementation analysis in the similar manner as the discussed in the previous sections.

Table 6.2: Sequence similarity scores of archaeal chaperonins (ClustalW)

No.	Name	Length	Name	Length	Score
1	MmCCT	1632	MmaCCT-alpha	1656	61.0
2	MmCCT	1632	MmaCCT-beta	1704	53.0
3	MmCCT	1632	MmaCCT-gamma	1629	59.0
4	MmCCT	1632	HvCCT-1	2033	56.0
5	MmCCT	1632	TaCCT-alpha	1656	57.0
6	MmCCT	1632	TaCCT-beta	1659	56.0

Figure 6.8

Sequence alignments of residues 212-262 of MmCCT with equivalent regions of MmaCCTs.

6.2.3.1 Protein expression of MmaCCT-α mutants by SDS and Native PAGE

To determine whether the mutants expressed and assembled into their oligomers, lysates of TAB21 cells transformed with pET22-*Mmacct-α-M226I* and pET22-*Mmacct-α-K219E* were analyzed by SDS-PAGE and Native-PAGE. Both MmaCCT-α-M226I and MmaCCT-α-K219E mutants were highly over-expressed similar to the wildtype protein expression (Figure 6.9). Moreover both the mutants showed clear oligomeric assembly as demonstrated by native-PAGE analysis (Figure 6.10). The mutants hence were further subjected to standard GroEL-complementation analysis.

6.2.3.2 Complementation analysis

The TAB21 cells expressing MmaCCT-α-M226I and MmaCCT-α-K219E were tested for their ability to retain growth in *groE* repressed conditions at 30°C. As seen in Figure 6.11B, both the mutants were inefficient in rescuing growth of *E. coli* when grown on glucose, similar to the results obtained for MmaCCT-α wildtype protein. Again, experiments were repeated with varying parameters however the same results were obtained reproducibly. The results showed that these mutations did not measurably increase the ability of MmaCCT-α-M226I and MmaCCT-α-K219E to function. This is therefore a distinct difference to the results obtained with MmCCT. However the results do not rule out the possibility that co-expression of different *M. mazei* CCT subunits could be required for serving all potential clients and hence supporting growth. Even if the additional mutations were functionally co-operative, lack of MmaCCT-β or MmaCCT-γ subunit could have been responsible for misfolding of a group of particular clients and hence cessation of growth. Attempts were hence made to clone *M. mazei* CCT subunits together or in different combinations under a *pET* expression vector however these

were not successful. The reasons were predominantly related to 30% – 40% identity shared by the alpha and the gamma subunit that led to recombination events and hence problems in cloning. Because the beta subunit has been shown to be insoluble in *E. coli* anyway (Klunker *et al.*, 2003), the experiments were not continued.

Subsequently, a random mutagenesis approach was also attempted wherein vectors carrying the *M. mazei* chaperonin genes were passaged using XL-1 Red mutator strain and the resultant random mutants were tested for their function in *E. coli*. The GroEL-complementation experiments did result into two clones expressing randomly mutated MmaCCT-α, that supported *E. coli* growth at 30^0C. The phenotype of the cells was indeed carried with the plasmid ruling out strain-specific changes but no mutations were found in the *Mmacct- α* gene or the T7 promoter. The results thus probably indicated changes in the plasmid backbone and the resultant secondary effects such as increase in the copy number or increased expression of the proteins that in turn would allow growth of *E. coli* to a certain extent. While these observations were interesting in their own right, they did suggest that isolation of functional variants of MmaCCT-α that allow GroEL complementation was rather difficult highlighting the inefficiency of this archaeal chaperonin to replace GroEL.
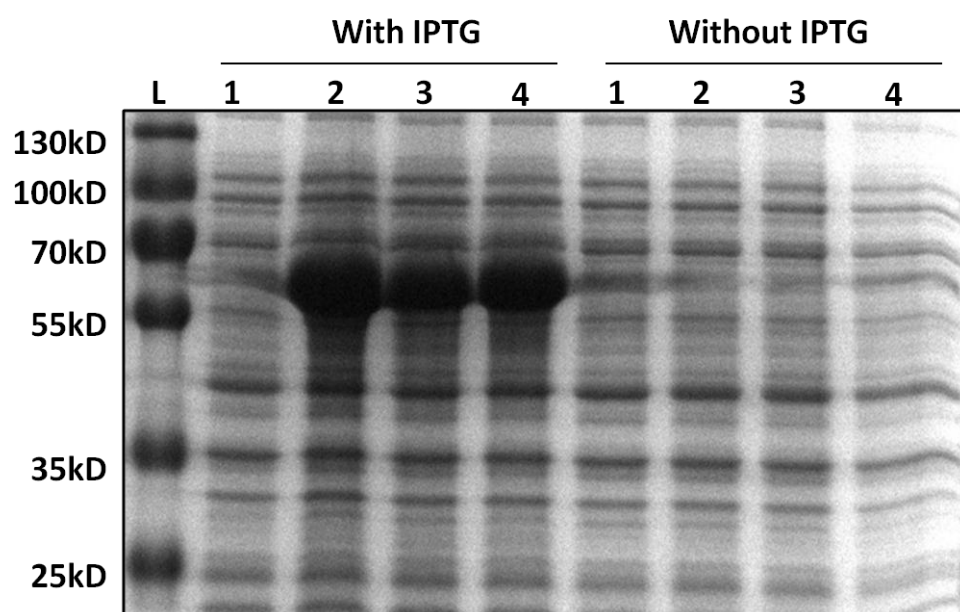
Figure 6.9

Expression of M226I and K219E mutants of *M. mazei* chaperonin alpha subunit in *E. coli.*

Crude extracts from *E. coli* TAB21 cells overexpressing CCT proteins in IPTG induced and non-induced conditions as analyzed by 10% SDS-PAGE and Coomassie staining.

(1) TAB21/pET21 empty vector

(2) TAB21/pET22-*Mmacct*- α-WT

(3) TAB21/pET22-*Mmacct*- α-M226I

(4) TAB21/pET22-*MMacct*- α-K219E
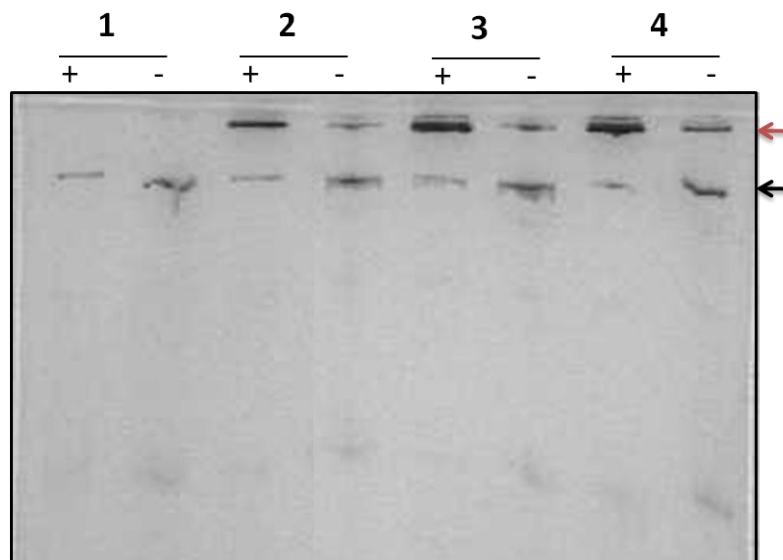
Figure 6.10

Assembly of M226I and K219E mutants of *M. mazei* chaperonin alpha subunit in *E. coli*.

Soluble fractions of *E. coli* TAB21 cells expressing CCT proteins in IPTG induced (+) and non-induced (-) conditions as analyzed by 7.5% Native PAGE and Coomassie staining. Red and black arrows show CCT and GroEL proteins respectively.

- (1) TAB21/pET21 empty vector
- (2) TAB21/pET22-*Mmacct*- α-WT
- (3) TAB21/pET22-*Mmacct*- α-M226I
- (4) TAB21/pET22-*MMacct*- α-K219E
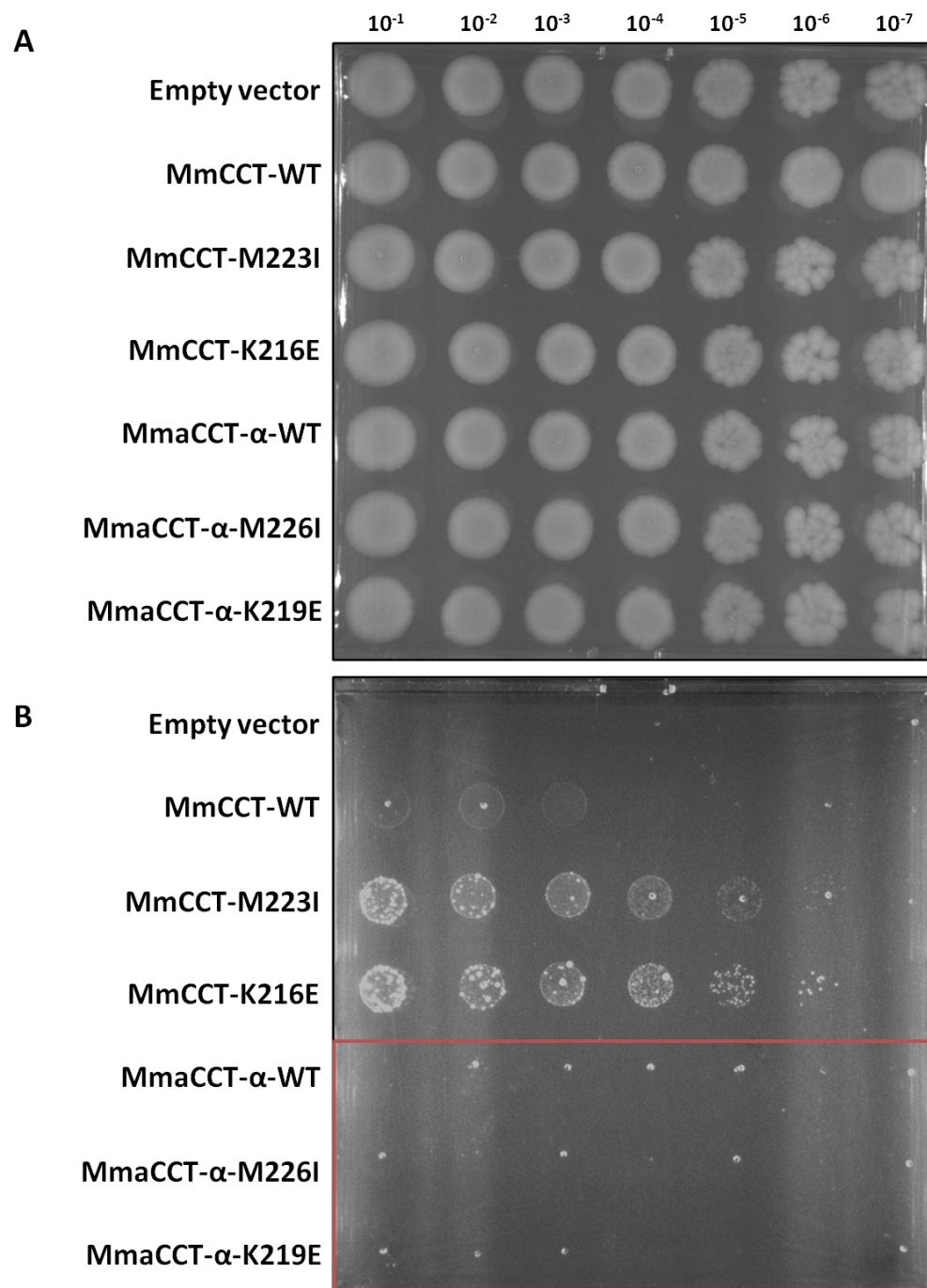
Figure 6.11

Solid medium growth analysis of *M. mazei* CCT-α mutants, M226I and K219E in TAB21.

Dilutions (10^{-1} to 10^{-7}) of TAB21 cells expressing the MmCCT proteins as indicated, grown on LB agar plates containing arabinose (*groE* induction) at 37°C **(A)** and on 0.2% glucose (*groE* repression) and 1mM IPTG (*cct* induction) at 30°C **(B)**.

6.3 Discussion

We show here that archaeal chaperonins from species other than *M. maripaludis* are unable to complement for the loss of GroEL in *E. coli.* Chaperonins from three different species, *M. mazei, T. acidophilum* and *H. volcanii,* were selected and subjected to the analysis.

The MmaCCTs and HvCCT-1 proteins expressed well in *E. coli* however only MmaCCTs assembled into their functional complexes. HvCCT-1 subunits did not form functional oligomers possibly because of the requirement of higher salt concentrations. This was not very surprising because it has been observed rather generally for many proteins from *H. volcanii* and other typical halophilic enzymes from extremely halophilic archaea that despite being soluble in *E. coli*, such proteins remain inactive and unstable. The recovery of these proteins is usually carried out from inclusion bodies, followed by denaturation and refolding in hypersaline solutions (Rao and Argos, 1981; Connaris *et al.*, 1998; Allers, 2010).

Expression of *T. acidophilum* CCTs was relatively weaker under IPTG induced conditions however this was expected. It had been observed that when TaCCT-α and TaCCT-β subunits were expressed in tandem in *E. coli*, the expression of alpha subunit dominated despite the two genes being driven by separate (T7) promoters (Bigotti and Clarke, 2005). The gene construct used for this study was hence modified in a way that a lac promoter and a spacer region were inserted upstream of *Tacct-α* and *Tacct-β* genes respectively. Such an alteration led to a stringent control on the over-production of TaCCT-α and enhanced expression of TaCCT-β, together ensuring formation of heterooligomers (Bigotti and Clarke, 2005). The fact that the expression was lower in

induced samples only confirmed that the major population was of heterooligomeric TaCCT-αβ complexes. However it should be noted that an unexpected protein band of molecular weight slightly lower than 60KDa was found in non-induced samples. The Native-PAGE gels confirmed correct bands of assembled proteins yet again the expression was stronger in non-induced conditions. Because these bands corresponded to the correct size of a chaperonin complex (when compared to positive control, MmCCT), it was assumed that the unexpected band observed during SDS-PAGE was that of TaCCT. It has been shown that TaCCT-β subunit when expressed in *E. coli* on its own results in a slightly lower molecular weight protein on SDS-PAGE as compared to TaCCT-α or TaCCT-αβ complex (Waldmann *et al.*, 1995b). It can be speculated thus that the expression of TaCCT-β occurs in absence of IPTG induction, possibly due to a strong but not a stringent T7 promoter. This is consistent with the protein expression observed in non-induced conditions for MmaCCTs. The reason for expression of only TaCCT-β and not TaCCT-α probably relies on the fact that the construct had been genetically manipulated to strictly control expression of TaCCT-α subunit and enhance the expression of TaCCT-β subunit, as mentioned earlier (Bigotti and Clarke, 2005). Because the subject of interest was to study the extent to which heterooligomeric TaCCT-αβ could complement for the loss of GroEL in *E. coli,* the events occurring in non-induced conditions were not studied further and all experiments were carried out in presence of IPTG.

The final complementation analysis showed that none of the chaperonins tested could rescue the growth of *E. coli* cells under *groE* repressed conditions at any temperatures. Furthermore, the supplementary mutations - M226I and K219E in the case of *M. mazei*

chaperonin also did not change the outcome. The results are indeed consistent with the hypothesis of functional divergence between the two groups of chaperonins. The case of MmCCT thus becomes an exception here where native GroEL-clients, vastly diverged from archaeal proteins are potentially served by a Group 2 chaperonin and its point mutations. It can be recalled that growth of the cells expressing MmCCT in GroE depleting conditions (TAB21) or in absence of GroE (Ai90) was very slow at 37^{0}C emphasizing a certain degree of insufficiency of this chaperonin to replace GroE that perhaps exaggerates with the case of other archaeal chaperonins. The underlying reasons for the GroEL-complementing ability of MmCCT have been difficult to decipher. However these results suggest there could exist at a certain level similarity between properties of GroEL- and MmCCT- clients, which probably reduces with the divergent clients of other extremophilic archaeal chaperonins. Such a speculation is consistent with the homo-oligomeric nature of MmCCT, where lineage-specific gene duplications characteristic to almost all other archaeal species have not occurred (Archibald *et al.*, 1999; Archibald and Roger, 2002b).

However we also suspect that the native heteromeric property of the other archaeal chaperonins tested here might not be an exhaustively limiting factor for their inability to replace GroE function. This is because firstly, there are several examples where every single subunit of heteromeric archaeal chaperonins has been individually shown to form functional homo-oligomers in *E. coli* (Yoshida *et al.*, 1997; Klunker *et al.*, 2003; Hirai *et al.*, 2008; Lee *et al.*, 2013) as against only two out of eight subunits for human TRiC (Sergeeva *et al.*, 2013). Secondly, subunits of various archaeal chaperonins have been shown to have a certain degree of functional redundancy. This is demonstrated by for

example the ability of homo-oligomeric CCT1 or CCT2 to substitute for the loss of two other native chaperonins in *H. volcanii* (Kapatai *et al.*, 2006) or the temperature-dependent interchangeable functions of two chaperonins from *T. kodakarensis,* CpkA and CpkB (Fujiwara *et al.*, 2008). All these studies suggest that in contrast to the distinct roles of each subunit for TRiC (Spiess *et al.*, 2004; Dekker *et al.*, 2011a), an overlap of function in addition to a certain degree of species-specific diversification exists for duplicated subunits of heterologous archaeal chaperonins, as has also been proposed by evolutionary studies (Archibald and Roger, 2002b; Ruano-Rubio and Fares, 2007).

We speculate therefore that instead of very client-specific limitations, general conditional changes that might have occurred for species-specific properties like thermo or salt tolerance might be responsible for their inability of the tested archaeal chaperonins to rescue growth of GroE depleting *E. coli* cells. For example HvCCT1 did not assemble possibly because of low salt concentrations. The result thus does not rule out the possibility that the native chaperonin if fully assembled could bind and fold GroEL clients. Accordingly for TaCCT_αβ, it is possible that the chaperonin adapted to work at higher temperatures behaves differently at a lower temperature. Indeed the optimal temperature for the ATP hydrolysis activity of the alpha homo-oligomer of TaCCT has been shown to be at 60^0C (Hirai *et al.*, 2008). It is intriguing as to why the mesophilic *M. mazei* chaperonins were not able replicate the function of MmCCT. While it is possible that this is because all three chaperonin subunits have evolved to meet the diverse chaperoning requirements also indicated by acquisition of Group 1 chaperonin by lateral gene transfer (Figueiredo *et al.*, 2004), we suggest that the complementation experiments conducted here are not completely exhaustive. Further mutagenesis

experiments could lead to identification of genetic variants of some of these archaeal chaperonins that would allow them to replace GroEL function, as has been the case for M223I and K216E mutants of MmCCT.

Collectively, the data suggests that for archaeal chaperonins to substitute the functioning of GroE is rather difficult because they have evolved for working on a diverse set of proteins. Although they assemble and presumably become functional in *E. coli*, not all bacterial client proteins are acted upon by archaeal chaperonins suggesting a client specific mechanism of chaperonins. However, the case of MmCCT clearly suggests that a functional cross-over is possible. Hence, in sum it can be speculated that although the archaeal chaperonins cannot easily replace GroEL function in bacteria, small structural changes can improve their complementing ability suggesting a similar mechanism of action of both groups but with difference that is almost certainly related to client specificity.

CHAPTER 7
Towards complementation analysis of Group 2 chaperonin mutants in archaea

7.1 Background

The results described in Chapter 3 and 4, established that mutants MmCCT-M223I and MmCCT-K216E of *M. maripaludis* chaperonin substantially improved the growth phenotype of the *groELS*-depleted strain of *E. coli* when compared to the phenotype of the strain expressing wildtype MmCCT. However, the precise nature of these mutations and their effects on chaperonin function was uncertain. It was speculated that the changes due the point mutations could either be client-specific, for example a more negatively charged cavity for GroEL-clients; or they could be chaperonin-specific that would improve the chaperonin function in general irrespective of the clients. However, the mutants did not show extensive biochemical disparity when compared to the wildtype MmCCT (Chapter 4), suggesting a more client-specific change. If this was the case, the same point mutations - M223I and K216E - could affect the chaperonin function for the native archaeal clients. This can be tested by introducing the same mutations back into the native hosts and deleting the endogenous genes.

This chapter describes the initial experiments that it is hoped will further culminate in testing the effects of the chaperonin mutations M223I and K216E in archaeal species. The chapter first describes preliminary cloning work to express the mutants in the parent species, *M. maripaludis,* for planned experiments in collaboration with Dr. James Chong, University of York. The later part describes attempts to establish a chaperonin knock-out system in the archaeal species *H. volcanii* that would allow growth-based analysis to study the mutants in *H. volcanii* chaperonin HvCCT-1. The potential of both the approaches are described in final discussion section. The experiments with *H. volcanii* were done prior to the work for *M. maripaludis* expression because of an

existing set up in the laboratory; however the later has been described first to maintain the flow of the story.

7.2 Experimental approach to complementation analysis in *M. maripaludis*

An efficient and markerless mutagenesis protocol has been developed in recent years for genetic manipulation in *M. maripaludis* (Moore and Leigh, 2005). The cloning required for the same has been done so that this method can be used for the analysis. The method utilizes the principle of homologous recombination and a hypoxanthine phosphoribosyltransferase (Hpt) based negative selection. The Hpt encoding *MMP0145* gene renders the species *M. maripaludis* susceptible to nucleotide base analogue 8-azahypoxanthine (8-aza-2,6-diaminopurine). For markerless mutagenesis, an *hpt* knock-out strain of *M. maripaludis* (MM900) and a non-replicating vector (pCRPrtNeo) have been constructed (Moore and Leigh, 2005). The pCRPrtNeo vector contains the *hpt* gene for negative selection, a Neomycin marker (Neo^R) for positive selection and a cloning site where the gene of interest, carrying the desired mutation/s, is inserted (Figure 7.1). Here, gene of interest being *Mmcct* with either M223I or K216E mutation. For point mutagenesis, only a selected part of the gene (usually 1000 – 1200 bases) with the site of mutation in the middle is cloned into the vector (Figure 7.1). For this, using primers, XbaI-MmCCTinternal-For and MmCCTinternal-NotI-Rev, 1.2 Kb long sequences from *Mmcct* WT gene or *Mmcct* gene containing mutations M223I and K216E in the middle were first amplified (Figure 7.2A, Lanes 1,2 and 3) and cloned into XbaI-NotI cut pCRPrtNeo vector (Figure 7.2A, Lane 4). The final constructs were checked for their correct inserts by screening digestions using XbaI and NotI (Figure 7.2B).

When the cells of MM900 strain are transformed with pCRPrtNeo carrying the gene of interest, the vector will integrate into the genome with a single recombination step between the homologous regions of the gene on the vector and on the chromosomal DNA, rendering the cells neomycin resistant (Figure 7.3). If these cells are then grown in absence of the antibiotic, a second recombination is a possible event where the vector can recombine out of the chromosome and can either return into the wildtype genotype or incorporate the mutated sequence that was introduced with pCRPrtNeo (Figure 7.3). In either case the cells become resistant to base analogue 8-azahypoxanthine with the loss of *hpt* gene. Single colonies are then screened by PCR to identify the mutated strain. The method has been schematically shown in Figure 7.3.

The constructs containing the gene sequences of MmCCT flanking the mutation M223I or K216E in pCRPrtNeo have been made (as shown in Figure 7.1 and 7.2) and sent to Dr. James Chong's group, University of York where the experiments are currently being conducted. These constructs will be used to introduce the mutations M223I and K216E in the chromosomal DNA of *M. maripaludis* by the markerless mutagenesis method as described above. There are two possible outcomes to the experiment:

1) Positive clones containing the mutations are identified. This would suggest that the mutated CCT is still able to show some function in *M maripaludis*. The extent to which the additional mutations affect the phenotype of the cells would be tested by competition experiments wherein the growth of both, the wildtype and the mutant strain, will be simultaneously monitored by solid and liquid media analysis.

2) Clones containing the mutations are not identified. In this case, it would become necessary to test whether the cessation of growth is because of the additional mutations or because of any inconsistencies in the recombination process. A different approach

with straight selection would be required to determine clearly whether or not the mutations are detrimental to the strain. In such a scenario, a method would be adopted where effects of *Mmcct* mutants will be studied in an *Mmcct* knock-out strain sustaining on a copy of *Mmcct* gene on a self-replicating vector. This method will be fairly similar to the plasmid shuffling method discussed in Chapter 3, and will rely on displacement of wildtype plasmid with the externally introduced plasmid containing mutant *Mmcct.* Direct growth-based results would yield a clear picture as to whether or not the mutants *Mmcct*-M223I or K216E allow growth of native host *M. maripaludis.*

While the above experiments are being pursued at the University of York, an attempt was also made to establish a similar system in the species *H. volcanii* in the laboratory.
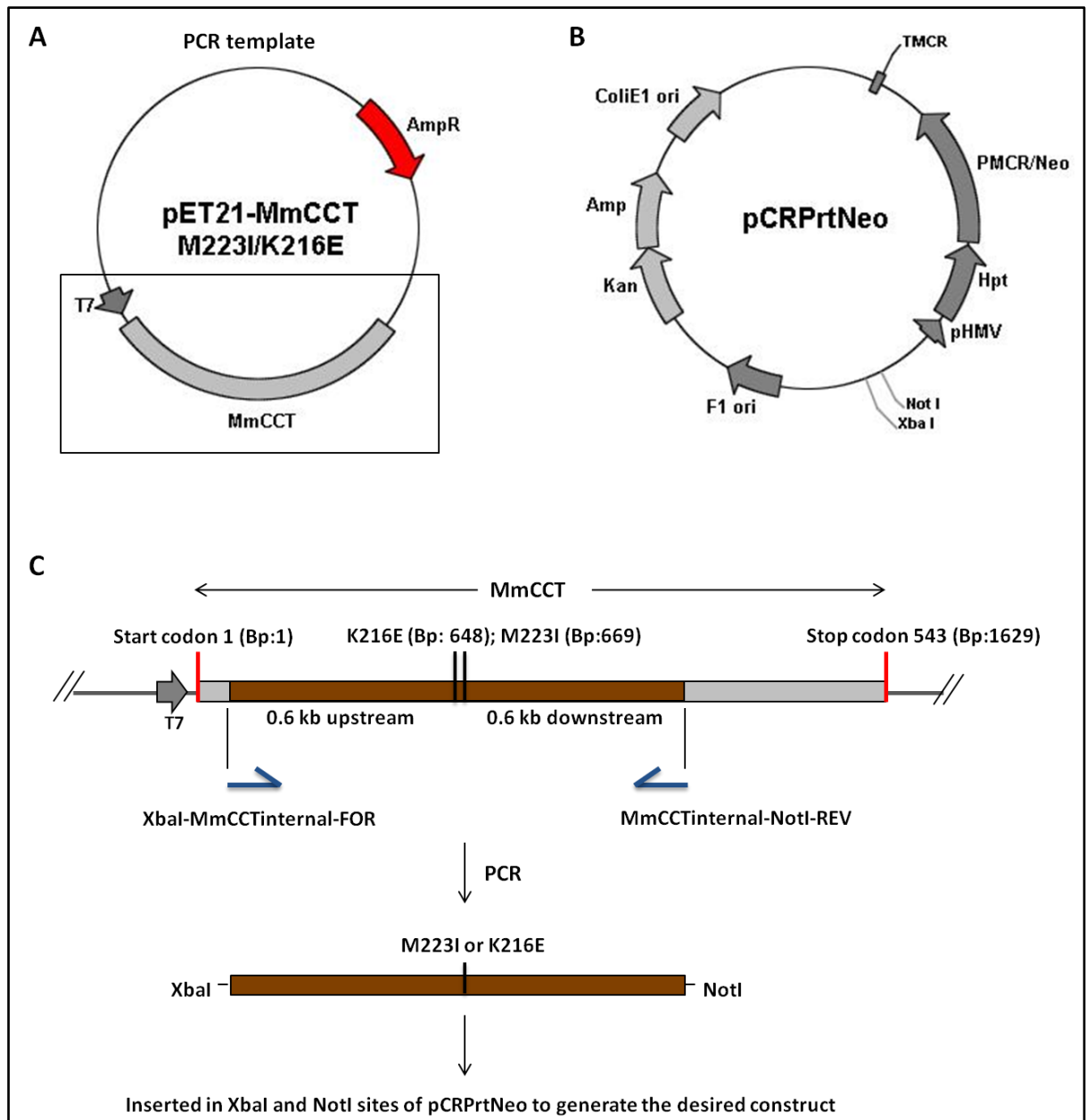
Figure 7.1

Schematic diagram to explain construction of pCRPrtNeo containing *Mmcct* gene sequence for markerless mutagenesis in *M. maripaludis.*

A. Vector pET21 containing MmCCT gene was used as the DNA template for the cloning.

B. Vector pCRPrtNeo as the recipient for markerless mutagenesis in *M. maripaludis*

C. Cloning strategy

(Plasmid maps drawn using *Plasm* software.)
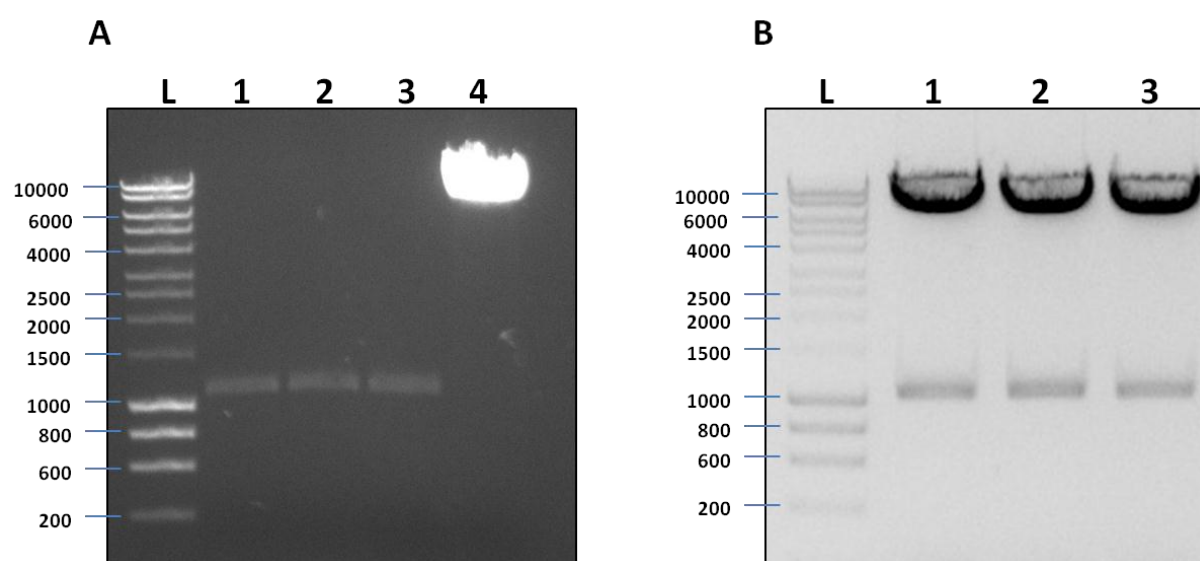
Figure 7.2

Agarose gel electrophoresis showing the construction of pCRPrtNeo *Mmcct* gene constructs for markerless mutagenesis in *M. maripaludis*

(A) Agarose gel analysis of amplified *Mmcct* wt gene (Lane 1), *Mmcct*-M223I (Lane 2) and *Mmcct*-K216E (Lane 3) using primers, XbaI-MmCCTinternal-For and MmCCTinternal-NotI-Rev. Lane 4 shows XbaI-NotI cut pCRPrtNeo vector (~6.5 Kb)

(B) Agarose gel analysis showing XbaI-NotI digests to confirm correct 1.2 Kb *Mmcct* wt (Lane 1), *Mmcct*-M223I (Lane 2) and *Mmcct*-K216E (Lane 3) inserts into pCRPrtNeo.
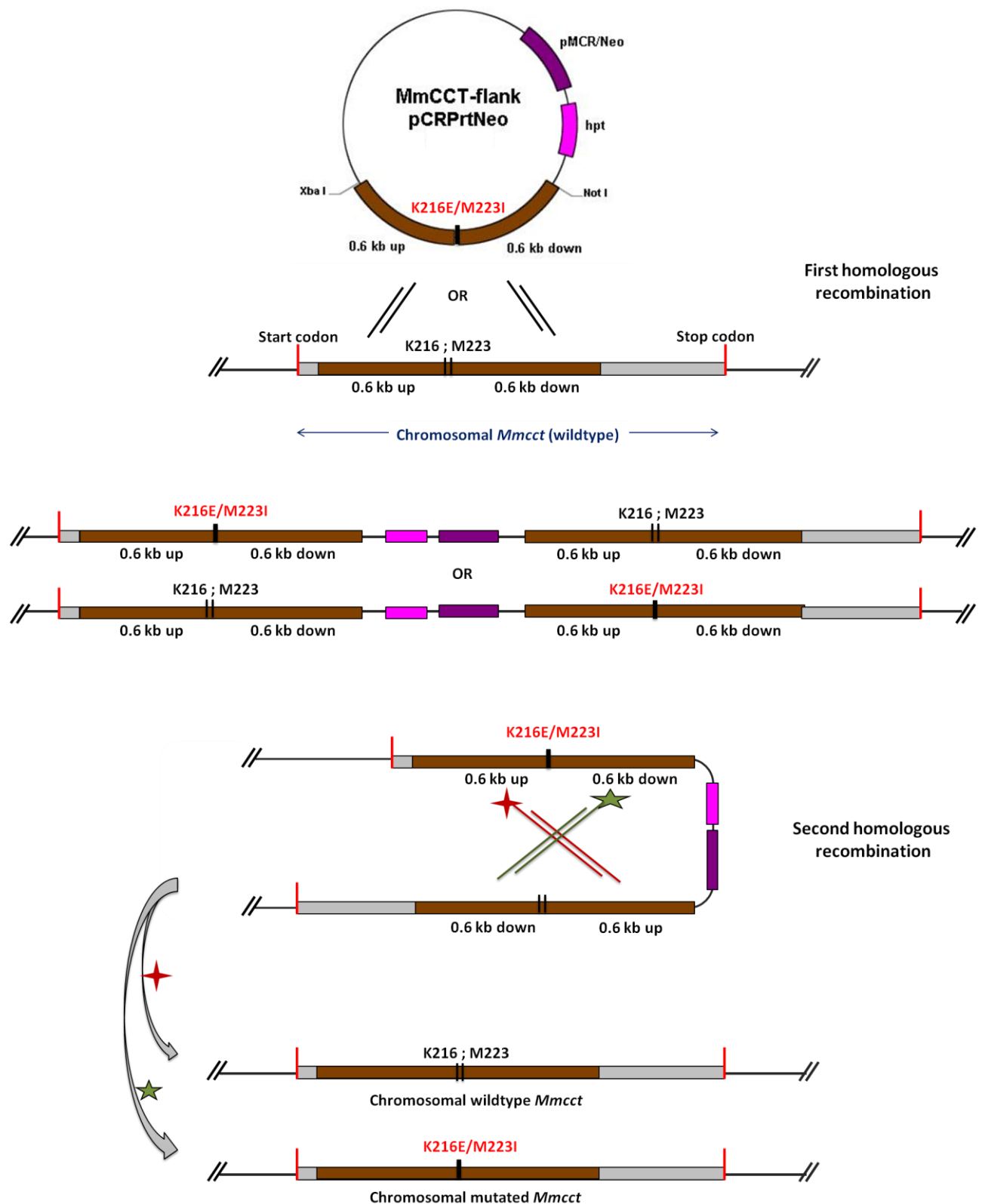
Figure 7.3

Schematic diagram to explain the technique of markerless mutagenesis in *M. maripaludis.*

7.3 Experimental approach to complementation analysis in *H. volcanii*

H. volcanii, an obligate halophilic archeaon originally isolated from Dead Sea, is an emerging model for *in vivo* and *in vitro* archaeal studies and has been routinely used in our laboratory in the past (Large *et al.*, 2002; Kapatai *et al.*, 2006; Large *et al.*, 2007). Archaeal genetics has benefited immensely since the discovery of this organism as it offers advantages of simple cultivation techniques (Dyall-Smith *et al.*, 2006), shuttle vectors (Lam and Doolittle, 1989), selectable markers (Allers *et al.*, 2004), efficient gene knock-out system (Bitan-Banin *et al.*, 2003) and a whole annotated genome sequence (Hartman *et al.*, 2010). Moreover, a tryptophan induced *tna* promoter has been specially developed for *H. volcanii* in our laboratory (Large et al., 2007). This promoter (which normally drives the gene encoding the enzyme tryptophanase) can be strongly induced in the presence of tryptophan and be tightly switched-off in its absence.

H. volcanii possesses three chaperonin genes referred to as *cct1, cct2* and *cct3* (Kuo *et al.*, 1997; Large *et al.*, 2002). It is established that while the functioning of these genes is essential for survival, either *cct1* or *cct2* alone can sustain growth of the organism in respective double knock-outs, with the complementation efficiency of *cct1* being better than *cct*2 (Kapatai *et al.*, 2006). Taking this into account we hypothesized that a mutational analysis of an array of *cct1* gene mutants targeting specific regions, that have some anticipated role in the total working of chaperonin, (for example *Hvcct1-*MtoI and *Hvcct1-*KtoE) could be conducted that in turn will delineate the structure-function co-relation of this chaperonin and give insights into the functional properties of Group 2 chaperonins in general. The first step towards such analysis in *H. volcanii* was development of a gene expression system that will specifically allow phenotypic analysis

as a net effect of a mutation in *cct1* gene without masking the effects by the chromosomal *cct* genes. A *Haloferax* strain lacking all three *cct* genes, when introduced with a mutant *cct1* gene, will grow only if the mutant is functional. However, the strain lacking all three genes will perish because at least one out of *cct1* or *cct2* is essential for growth. Consequently, the aim of this part of the project was to construct a triple knock-out *Haloferax* strain that contains an externally introduced *cct2* gene under an inducible promoter which could be turned off when an external *cct1* mutant is introduced, allowing its analysis when all other *cct* genes are non-functional (Figure 7.4).

The system hence generated would permit the analysis of conditional lethal mutants in essential gene *cct1* (when the *cct2* gene is switched off) by simple manipulation of growth conditions. As a matter of fact, instead of the *cct2* gene, *cct1* gene can also be incorporated under the inducible promoter to construct the desired system. However, as the mutations to be introduced later in the system are also in the *cct1* gene, homologous recombination between the wild type and the mutant is a possible event, known to occur, that could replace the mutant gene to the wild type. To avoid any such occurrences *cct2* gene was chosen to create the triple knock-out strain.
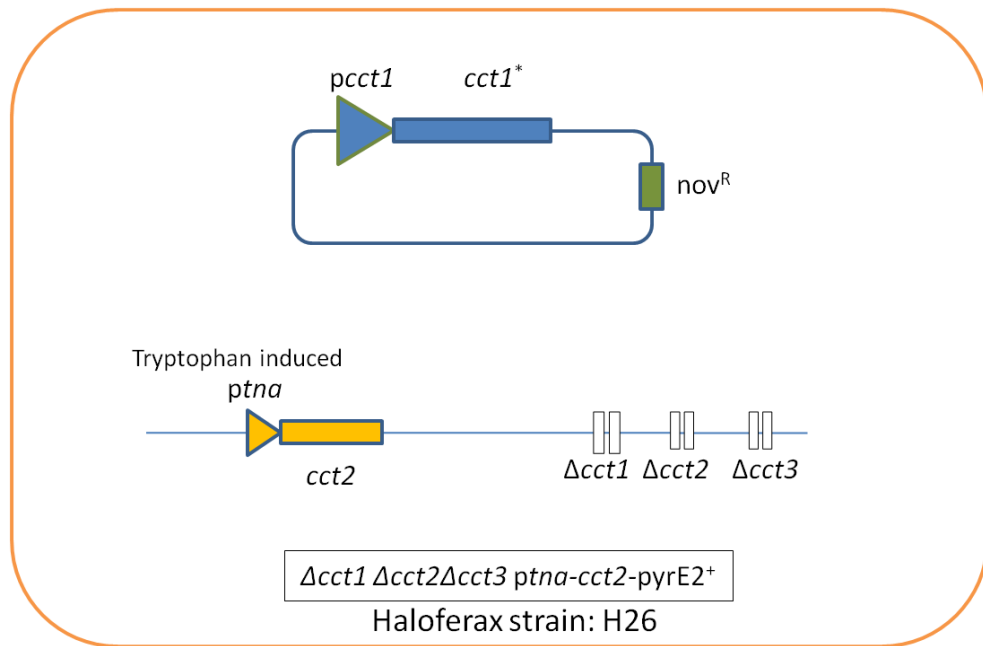
Figure 7.4

Cartoon representation of a triple knock-out strain of *Haloferax volcanii*

The diagram represents a triple knock-out strain of *H. volcanii* that can sustain on an externally introduced *cct2* gene under a tryptophan-inducible p*tna* promoter. The *cct2* gene could then be turned off when an external *cct1* mutant *(cct1*)* is introduced. The system altogether will allow analysis of mutant *cct1* genes when all other native *cct* genes are non-functional.

7.3.1   Experimental design

For construction of the desired gene expression system, the strain used is *Haloferax volcanii* **H26** that lacks the *pyrE2* gene essential for uracil biogenesis. The vector used at every stage is **p131** containing ampicillin resistance cassette and a *pyrE2* gene under its own promoter. This plasmid does not contain an archaeal origin of replication and hence is an integrative, non-replicative plasmid. The use of H26 strain and p131 enable the application of a *pyrE2* based gene knock-out method as discussed in the following section (Bitan-Banin et al., 2003).

1. **Construction of a single knock-out strain lacking *cct3* gene - H26:Δ*cct3*.**

This step has already been accomplished previously in the laboratory using the same gene knock-out method as described below.

2. **Construction of a double knock-out strain - H26:Δ*cct2*Δ*cct3***

As described for markerless mutagenesis in previous section, the gene knock-out method utilizes the principle of homologous recombination and a *pyrE2* based selection. In *H. volcanii* the *pyrE2* gene codes for the physiological enzyme orotate phosphoribosyl transferase. This is a key enzyme in uracil biosynthesis and a deletion of *pyrE2* gene results in inability to synthesize uracil. Hence the H26 strain (Δ*pyrE2)* of *H. volcanii*, used for the study, is a uracil auxotroph. Firstly, a non replicative plasmid (p131) containing *pyrE2* gene and a cloned chromosomal DNA fragment containing flanking sequences of the target gene (here *cct2* gene) is transformed into the Δ*pyrE2* strain. The transformants are then plated onto medium lacking uracil and the uracil prototrophs are selected. The uracil prototrophs are a result of integration of the plasmid and hence the *pyrE2* gene into the chromosome (Figure 7.5). Concomitantly, a homologous event

between the flanking sequences on the plasmid and those around the native target gene is expected. The cells are then grown in liquid cultures to allow integrative recombination events, between the plasmid-inserted flanking sequences and the chromosomal flanking sequences, that could result either into reconstitution of the wild type gene (containing *pyrE2*) or a deletion of the target gene (loss of *pyrE2*) (Figure 7.5). The cells are then plated onto selective medium again. This time the selection is based on the fact that the cells that can synthesize *de novo* uracil are sensitive to the toxic uracil analogoue 5-fluoroorotic acid (5-FOA). The cells that still contain the integrated plasmid, and hence *pyrE2,* would metabolize 5-FOA to produce 5-fluorouracil and die. While the cells that have lost the integrated plasmid and hence *pyrE2,* will survive but will need uracil in their growth medium (Figure 7.5) (Bitan-Banin et al., 2003; Sato et al., 2003). Employing the above method, deletion of *cct2 from* the strain H26:Δ*cct3* requires following steps:

A) Construction of p131-5'3'flank*cct2*, plasmid p131 containing flanking sequences of *cct2* gene.

B) Transforming p131-5'3'flank*cct2* in H26:Δ*cct3* and selecting the popped in plasmid containing strain on uracil lacking medium (POP IN)

C) Allowing integral recombination by growing the transformants and selecting the popped out strain on 5-FOA and uracil containing medium (POP OUT)

D) Screening of selected colonies by colony-PCR using primers specific to *cct2* gene

E) Confirmation of the positively screened colonies by western blotting using anti-thermosome antibodies that will contain CCT1 protein but not CCT2 and CCT3 proteins.
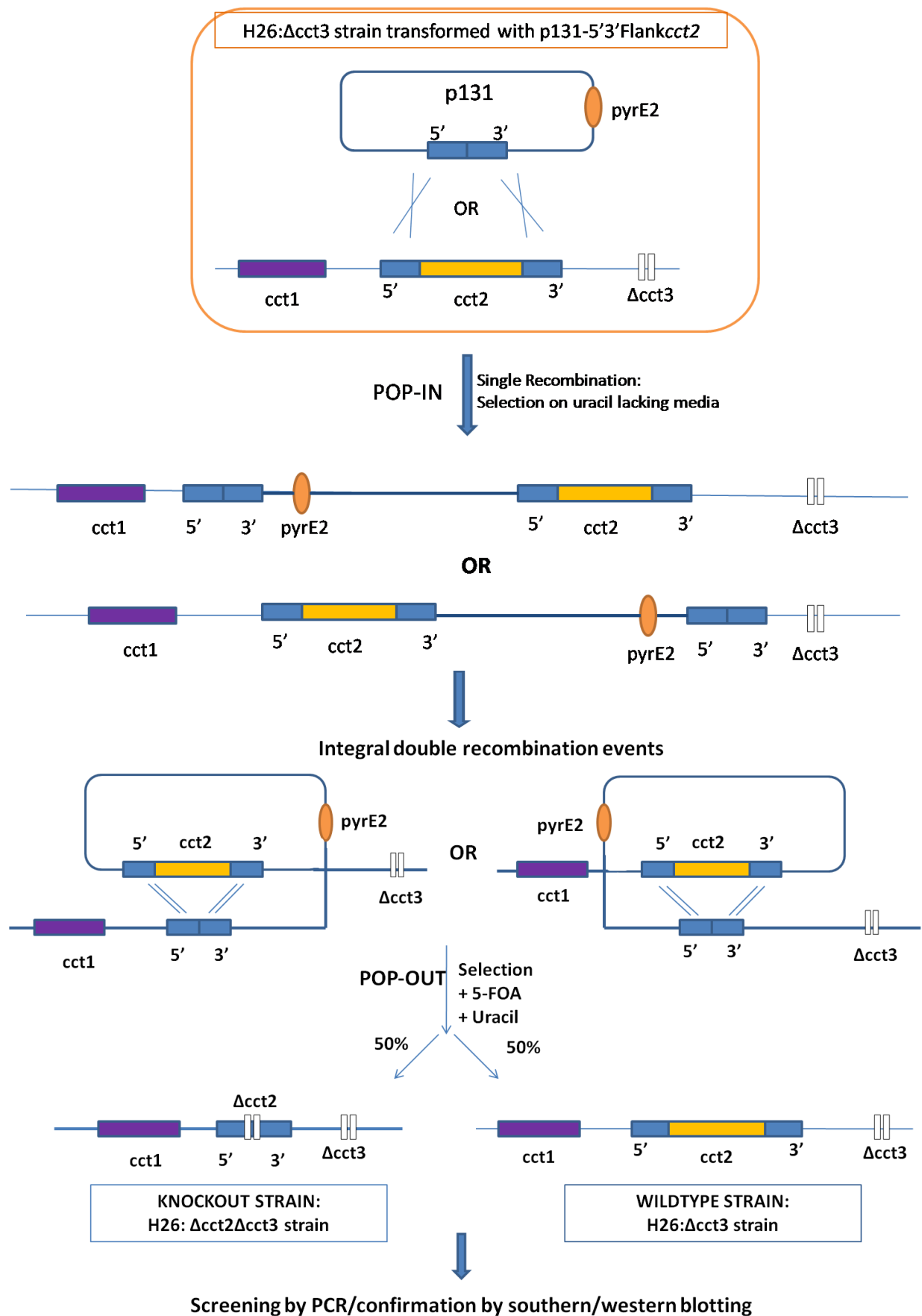
Figure 7.5:

Schematic diagram to demonstrate the *pyrE2* based gene knock-out system for the disruption of *cct2* gene in the strain H26:Δ*cct3* strain

Flanking sequences (blue) just upstream and downstream of the *cct2* gene (yellow) are cloned into the p131 vector to give rise p131-5'-3'flank*cct2*.This plasmid when transformed in H26:Δ*cct3* strain, integrates into the chromosome by homologous single recombination events (single recombination event because circular plasmid) between the flanking sequences of the chromosomal *cct2* gene and the plasmid (POP-IN). The strain containing popped in plasmid now also contain *pyrE2* gene and hence selected on medium lacking uracil. The transformed colonies, further grown on medium lacking uracil may now undergo integral double recombination events i.e simultaneous recombination at 3' and 5' flanking region, which can result in either the reconstitution of the wild type *cct2* or knocked out *cct2* gene (POP-OUT). The popped-out colonies are selected on plates containing 5-FOA and screened by colony PCR to isolate knocked out *cct2* strains.

3. Construction of a triple knock-out strain containing *cct2* gene under an inducible promoter *ptna* - H26:Δ*cct1*Δ*cct2*Δ*cct3* p*tna-cct2-pyrE2*⁺

The triple knock-out strain will be created by swapping the chromosomal *cct1* gene in H26:Δ*cct2*Δ*cct3* with an external *cct2* gene under the *ptna* promoter using the same gene knock-out method, but with a double homologous recombination event (Figure 7.6). Firstly, a plasmid vector will be constructed containing the *cct2* gene (p*tna* promoter) and *pyrE2* gene (under its own promoter), both placed between the *cct1* flanking sequences. The plasmid will be linearised using specific enzymes prior to the transformation. When the double knock strain will be transformed with this construct, a double recombination event will take place between the flanking sequences of *cct1* gene on the vector and on the chromosome. The process will yield a triple knock-out strain with *cct1* gene replaced by an external *cct2* gene and *pyrE2* gene, that can be selected on uracil lacking medium (Figure7.6). This strain would now allow study of phenotypic effects of externally introduced wildtype or mutant *cct1* gene in absence of the native CCTs. The construction of the above system requires following steps:

A) Construction of the plasmid p131 containing upstream and downstream flanking sequences of *cct1* gene sandwiching the *cct2* gene and the *pyrE2* genes.

B) Transformations of H26:Δ*cct2*Δ*cct3* with linearised p131-5'*cct1* flank-p*tna-cct2-pyrE2*-3'*cct1* flank and selection of the transformants on uracil lacking but tryptophan containing medium.

C) Screening of the transformants by PCR using primers specific to *cct1* gene and confirmation by western blotting.
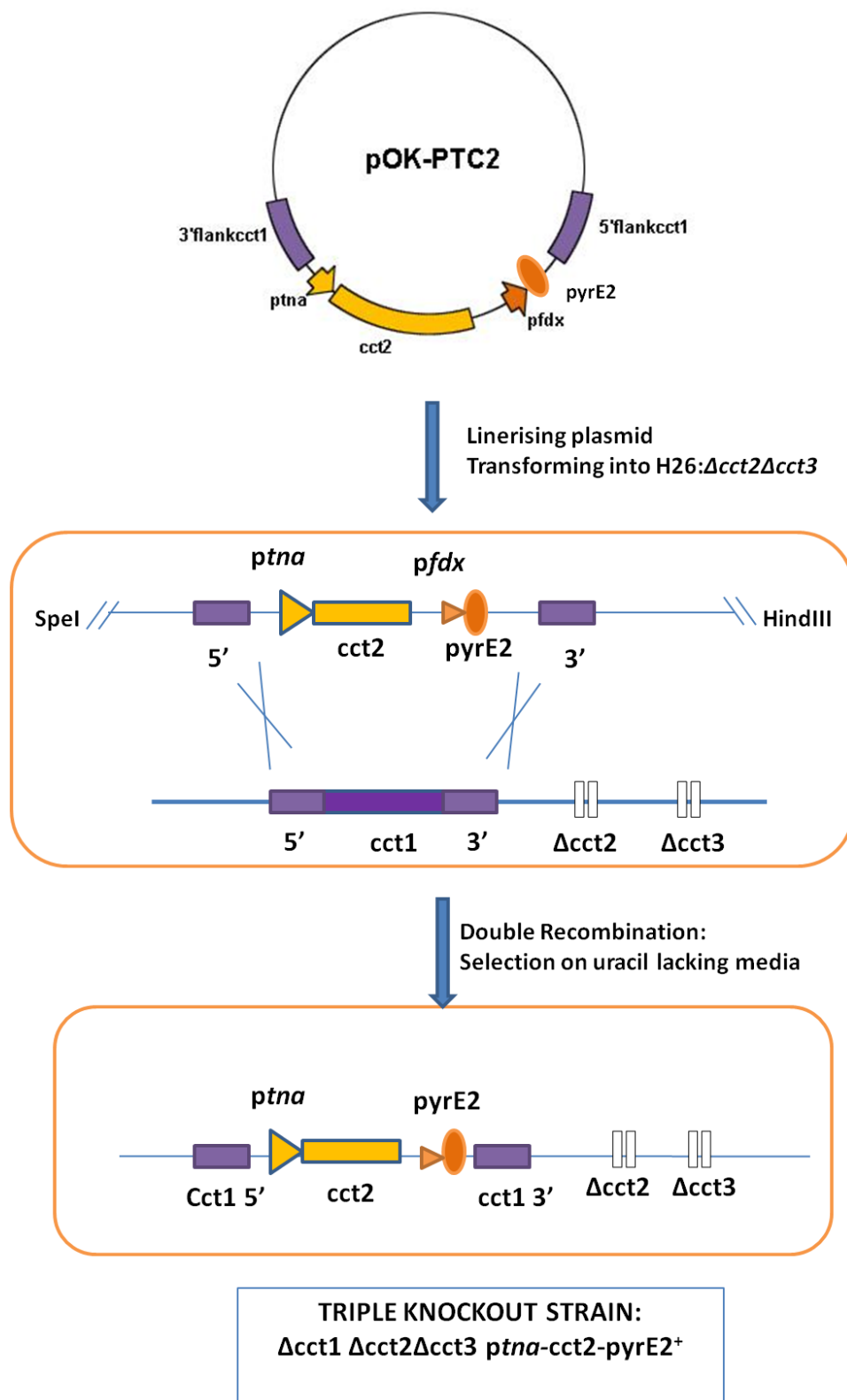
Figure 7.6

Schematic diagram demonstrating construction of a triple knock-out strain containing externally introduced cct2gene under tryptophan induced ptna promoter

The plasmid p131-5'cct1 flank-ptna-*cct2-pyrE2-3'cct1* flank will be transformed in the H26:Δ*cct2*Δ*cct3* strain. As the plasmid will be linearised prior to transformations using specific restriction enzymes, a double recombination event will take place between the flanking sequences of *cct1* gene on plasmid and the chromosome (violet regions), replacing the chromosomal *cct1* gene by *cct2* under p*tna* and *pyrE2* gene under its own ferridoxin (p*fdx*) promoter. The transformants will be selected on uracil lacking medium, screened and confirmed, to finally give the desired strain i.e. H26:Δ*cct1*Δ*cct2*Δ*cct3* p*tna-cct2-pyrE2+*

7.3.2 Results

7.3.2.1 Western blot to check *cct3* deletion in H26:Δ*cct3* strain

H26:Δ*cct3* strain had already been created in the laboratory (Georgia Kapatai, former PhD student). The strain was confirmed for its uracil auxotrophy and further that it lacks CCT3 protein expression by western blotting using anti-thermosome antibodies. The controls, DS70 strain and H26 strain that express of all three *cct* genes constitutively show a clear band of CCT3 which is absent in H26: Δ*cct3* strain (Figure 7.7).

7.3.2.2 Construction of a double knock-out strain - H26:Δ*cct2*Δ*cct3*

I. Construction on p131-5'3'Flank*cct2* plasmid:

The first step in construction of the double knock-out was to generate a non-replicative plasmid containing flanking sequences of *cct2* gene. DNA sequences upstream and downstream of the *cct2* gene were amplified individually from H26 chromosomal DNA using primers (5'flankingcct2-FOR and REV, 3'flankingcct2-FOR and REV) containing specific restriction sites. The two fragments were called **5'Flank*cct2*** - upstream of the *cct2* gene containing Ecor1/BamH1 sites (474bp) and **3'Flank*cct2*** - downstream of the *cct2* gene containing BamH1/Xba1 restriction sites (524bp) (Figure 7.8). The fragments were individually cloned into the plasmid vector p131 in between Ecor1 and Xba1 sites as shown schematically in Figure 7.9 to give final plasmid construct p131-3'5'flank*cct2*.
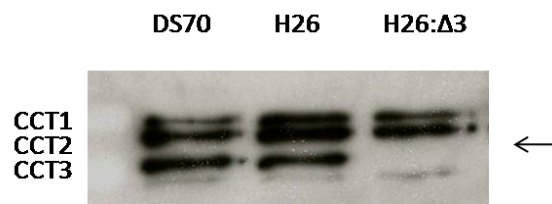
Figure 7.7

Western blot showing lack of CCT3 expression in H26:Δ*cct3* strain.

Figure shows the expression pattern of *cct* genes in strain DS70, H26 and H26:Δ*cct3* strains. As seen, the H26:Δ*cct3* strain lacks CCT3 protein as compared to the controls.
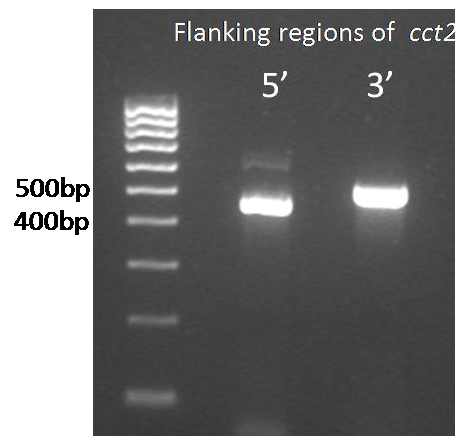
Figure 7.8

Agarose gel electrophoresis of the PCR amplified 5'-flanking and 3'-flanking regions of the *cct2* gene

Lane1: Hyper Ladder

Lane2: cct2 5' flanking sequence (474bp)

Lane3: cct2 3' flanking sequence (524bp)
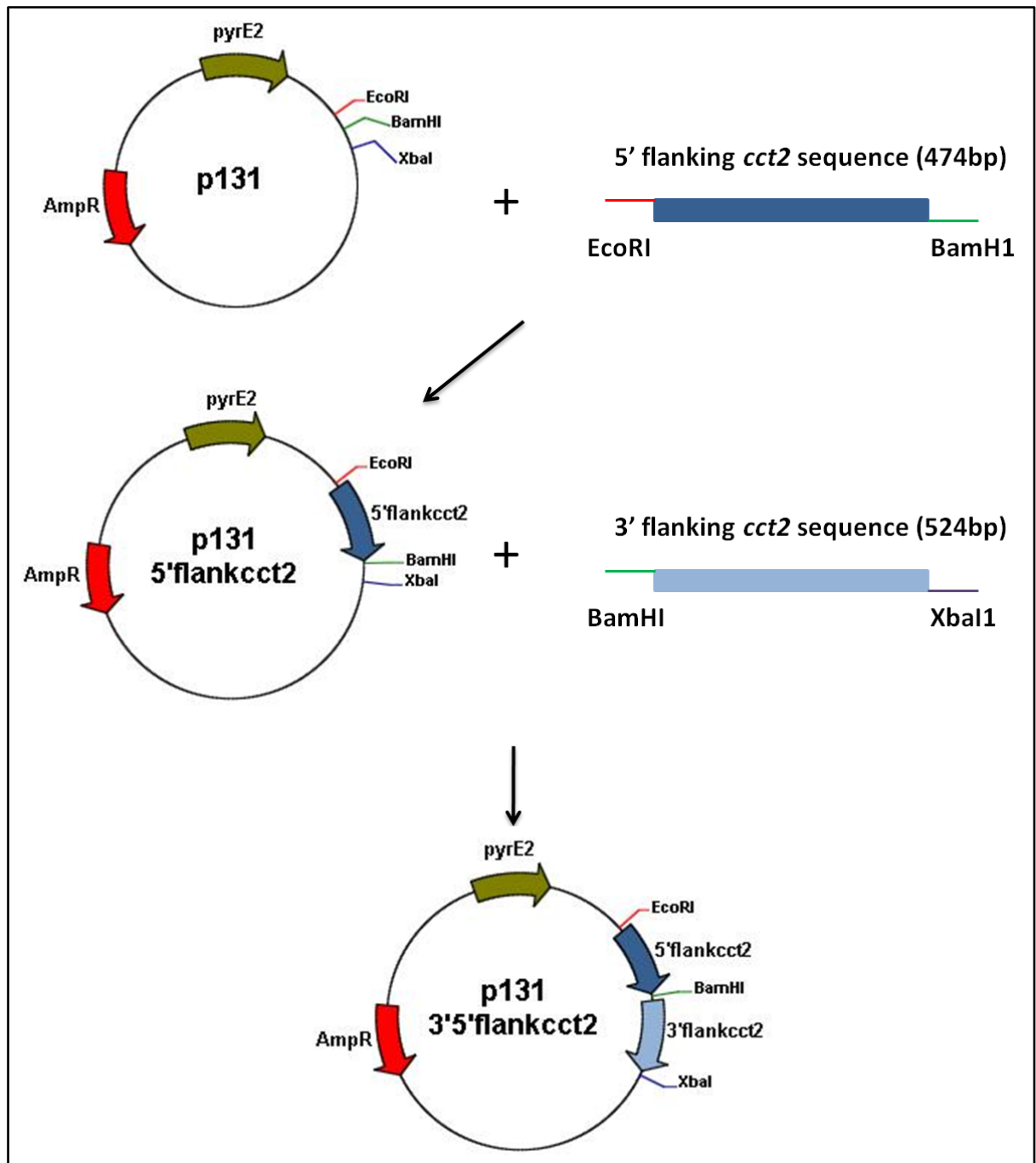
Figure 7.9
Schematic diagram to explain the construction of p131- 5'3'flank*cct2* cloning vector.

II. Transformations of the p131-5'3'flank*cct2* into H26:Δ*cct3*

(POP-IN and POP-OUT selection)

The plasmid p131-5'3'flank*cct2* was passaged through JM110 *E.coli* strain *(dam⁻dcm⁻)* to demethylate the DNA prior to transformations into H26:Δ*cct3* strain. The transformants were selected for uracil prototrophy on minimal medium plates. Colonies arising from these transformations resulted from integration of the plasmid into the chromosome (POP-IN). A couple of individual colonies were grown in liquid cultures allowing a second recombination that could result into either reconstitution of the native *cct2* gene or the desired knock-out strain (POP-OUT). To select for the loss of *pyrE2* gene, dilutions of these cultures were grown on MGM medium (rich media including uracil) containing 50μg/ml 5-FOA. Wild type strain DS70 that contains *pyrE2* gene was used as a negative control. As expected, while no colonies came up for DS70 strain due to toxicity, several colonies came up on the test plates. These colonies were subjected to colony PCR using *cct2* gene internal primers. However, all the colonies (~200 colonies screened) were found to contain the band for *cct2* gene indicating that reconstitution to original was preferred over popping out of the gene.

Because of the availability of yet another strain in the laboratory, H26:Δcct2$_{trun}$Δ*cct3*, which is a partial double knock-out (Figure 7.10), it was decided to use this as the starting strain instead of H26:Δ*cct3*. The H26:Δ*cct2*$_{trun}$Δ*cct3*, previously created (Georgia Kapatai, PhD thesis) for the same experiment, contains a truncated version of *cct2* gene and hence should lessen the chances of *cct2* gene reconstitution. Using the same method as described above, the H26:Δ*cct2*$_{trun}$Δ*cct3* was transformed with plasmid p131-5'3'flank*cct*, popped-in colonies were selected on minimal media and further

grown in liquid cultures. The popped-out colonies were selected on MGM plates containing 5-FOA. Similar to results obtained for previous experiment, several colonies came up on 5-FOA containing plates confirming loss of *pyrE2* gene. The obtained colonies were screened by PCR using *cct2* gene internal primers (Figure 7.11A). Out of ~50 colonies tested, few candidates showed faint bands or no bands at all (For e.g. colony 7, d, h, i and k in the figure). These colonies were selected, grown again on selective media and subjected to confirmatory colony PCR reactions using 2 sets of primers. Set 1 comprised of a 5'flank-forward primer 474 bp upstream of the 5' flank region of *cct2* and a 3'flank-reverse primer 524bp downstream of 3' flank region of *cct2* gene (Red arrows, Figure 7.11B). Set 2 were the same *cct2* gene internal primers (*cct2* gene FOR and REV) at the start and end of the gene (Green arrows, Figure 7.11B). The *cct2* knock-out should give a ~1000 bp band with set 1 primers and no band with the set 2 while the H26:Δ*cct2*$_{trun}$Δ*cct3* colonies, used as a control, should give ~2400 bp band with set 1 and ~1400 with set 2 primers (Figure 7.11B).

As seen in Figure 7.11C, the potential *cct2* knockout candidates – clones i and k, show a 1000bp band with primer set 1 and no band with primer set 2. The parent strain that contains the truncated version of *cct2* gene gives a band of ~2000bp with primer set 1 which is a bit lower than expected (~2300bp). However, it gives expected 1300bp product with primer set 2. The results suggest that clones 'i' and 'k' are successful products of the experiment and do not possess *cct2* gene. It should be noted that western blotting cannot confirm the absence of CCT2 protein expression in the knockout candidates because the parent strain itself contains truncated version of *cct2* gene that is unable to produce CCT2 protein, thus leaving no difference between the two strains.
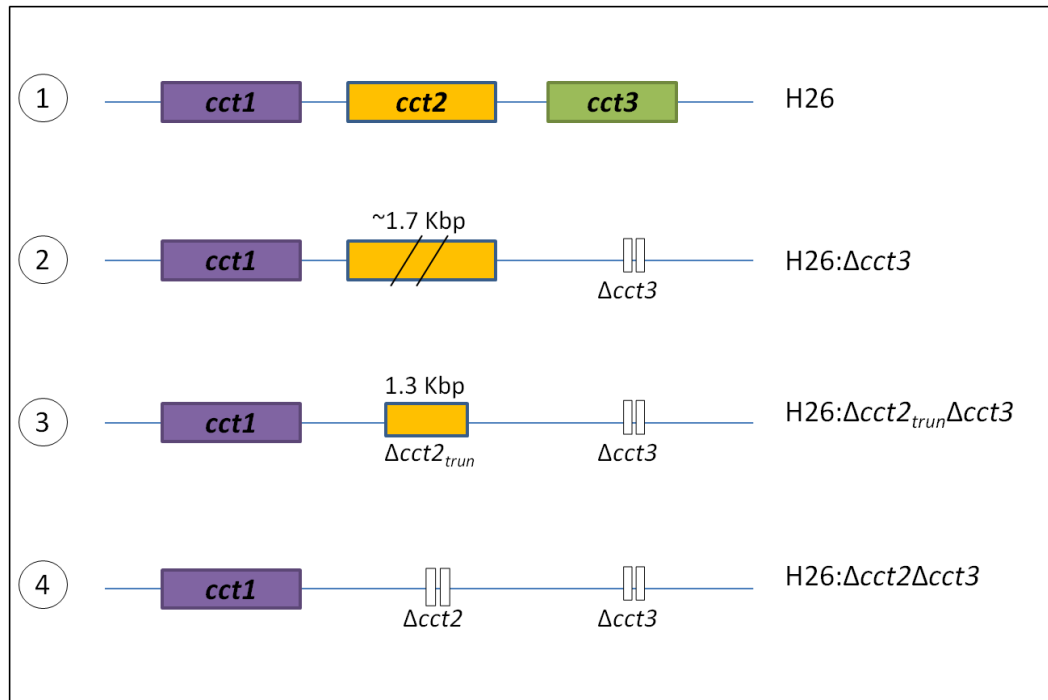
Figure 7.10

Cartoon representations of the genotypes of the three cct genes in different strains

The H26:Δ*cct2*_{trun}Δ*cct3* has a truncated version of *cct2* gene where a fragment of 400bp is deleted from the middle of the gene.
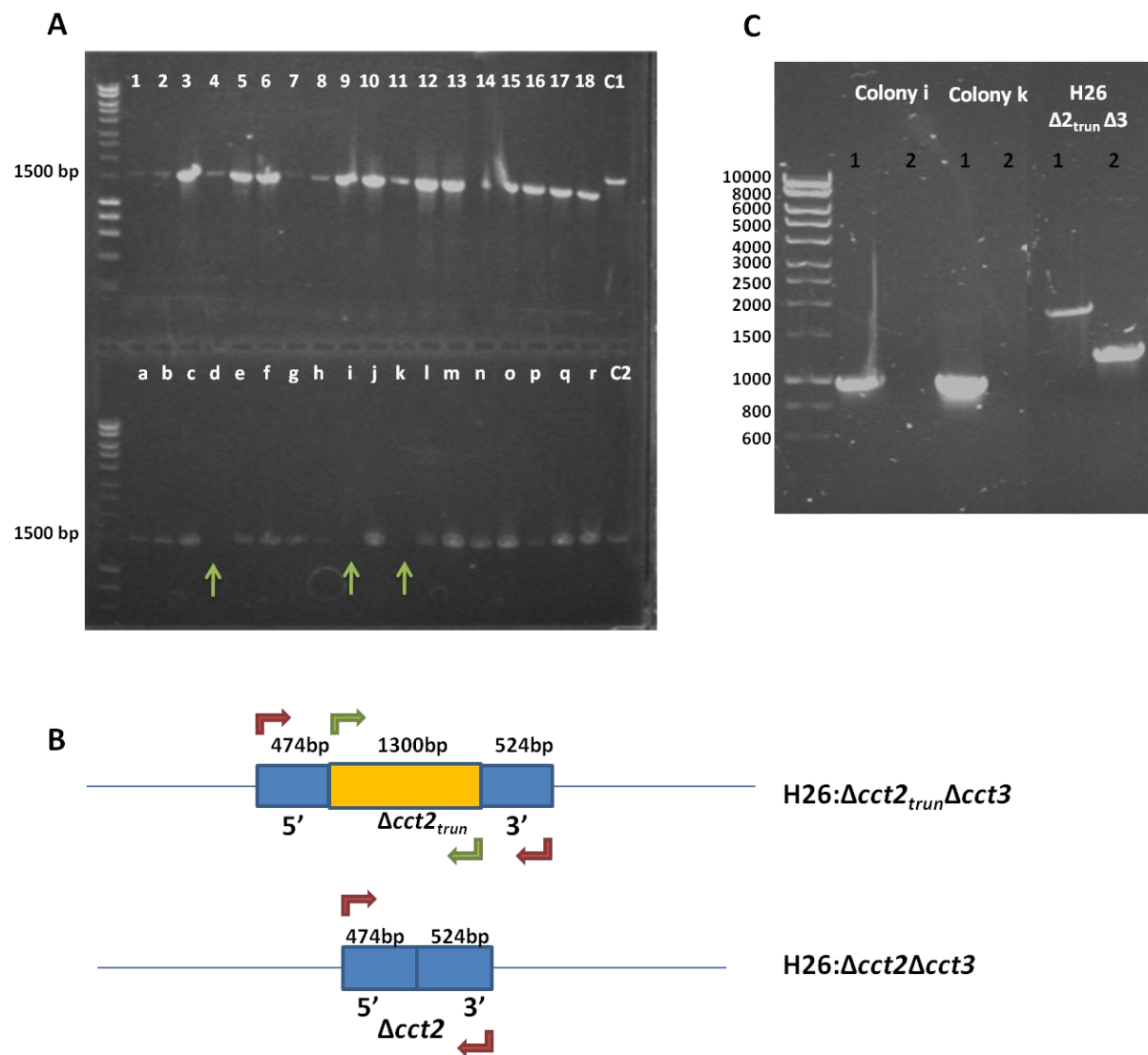
Figure 7.11

Screening and confirmation *of* cct2 knockout candidates by colony PCR

A) Agarose gel electrophoresis of the colony PCR products obtained using internal *cct2* gene primers. The test colonies are numbered 1 to 18 in upper lane and *a* to *r* in the lower lane. C1 and C2 correspond to the controls H26:Δcct3 and H26:Δcct2$_{trun}$Δcct3.

B) Schematic diagram showing the expected sizes of DNA fragments in the parent strain and *cct2* knock-out. The red and green arrows represent primer set 1 (5'flankcct2-FOR/3'flankcct2-REV) and 2 (*cct2*gene-FOR/*cct2*gene-REV) respectively.

C) Agarose gel electrophoresis of confirmatory colony PCR reactions of two potential knock-out colonies i and k, along with the parent strain, using both sets of primers labelled as 1 and 2 respectively.

7.3.2.3   Construction of the triple knock-out - H26:Δ*cct1*Δ*cct2*Δ*cct3*: p*tna-cct2*

The first step in making the triple knock-out strain was to construct a plasmid vector containing *cct2* gene (under *ptna* promoter) and *pyrE2* gene inside the flanking sequences of *cct1* gene. Owing to previous attempts of making a triple knock-out strain, such a plasmid was already made in the laboratory (Dr. Andrew Large, former post-doc). This plasmid was called pOK-5'C1-PTC2-3'C1 (**p**yrE2 – p**t**na – **c**ct**2** in cct1 locus under pOK12 background) (Figure 7.6). The plasmid was linearized using HindIII restriction enzyme prior to transformations in the knock-out strain. Two double knock-out strains, H26:Δ*cct2*Δ*cct3*i and H26:Δ*cct2*Δ*cct3*k, obtained from the previous step were transformed with the linearised plasmid pOK12-PTC2. A no-DNA negative control and a positive control, a vector p354 were also incorporated. The p354 vector contains *pyrE2* gene and an archaeal origin of replication which allows the strain to grow in the minimal medium. This control ensured robustness of the transformation procedure. The transformed cultures were grown on a set of plates containing 1) minimal media 2) minimal media + 5mM tryptophan 3) minimal media + 5-FOA (Figure 7.12). Selection takes place on the basis of the ability of the strain to grow in absence of uracil in the medium because of the acquired *pyrE2* gene (Figure 7.6). Also, considering that the triple knock-out strain would only have p*tna*-promoter driven *cct2* gene for sustenance, tryptophan would be required in the medium for expression of *cct2* gene. Plates containing 5-FOA were used as a negative control.

With the initial experiments, growth was observed on minimal media plates for all the transformants including the no-DNA negative control. A possible carry forward from the rich regeneration medium during the transformation was thought to be an issue. Hence the experiments were repeated with regeneration carried out in minimal medium.

While the results for the positive and the negative controls were as expected, the strain transformed with pOK-5'C1-PTC2-3'C1 grew on minimal medium plates without tryptophan (Figure 7.12). Although, colonies did appear on tryptophan containing plates, such a result was considered ambiguous due to presence of almost the same number of colonies on respective minimal media plates. The experiment was repeated several times ensuring that there was no tryptophan contamination in minimal media via any source. However the outcome was similar to as described above.

These results were rather unanticipated because the p*tna* is a tightly regulated promoter and would require tryptophan in the media for *cct2* gene expression and hence growth. It was speculated thus that a single recombination event between either of the flanking sequences of *cct1* gene instead of a double recombination event could have taken place where despite the integration of *pyrE2*, intact *cct1* gene would still be present, allowing growth on minimal medium. This suggested that the plasmid pOK-5'C1-PTC2-3'C1 was perhaps not linearised prior to the transformation despite being digested using the HindIII enzyme. Attempts were made using another single-cutter enzyme, SpeI for linearising the plasmid however it resulted into unexpected band patterns. The results suggested either an error in the construct pOK-PTC2 possibly by accidental disruption of the restriction sites during its handling or less likely, certain random recombination events in the cells that would inhibit the process. If former was the case either the whole plasmid used had to be sequenced or a new plasmid had to be constructed. However because of the amount of time that the process required and because the experimental system set up at the University of York should be able to test the mutants MmCCT-M223I and MmCCT-K216E in its host species *M. maripaludis*, the project with *H. volcanii* was not carried further.
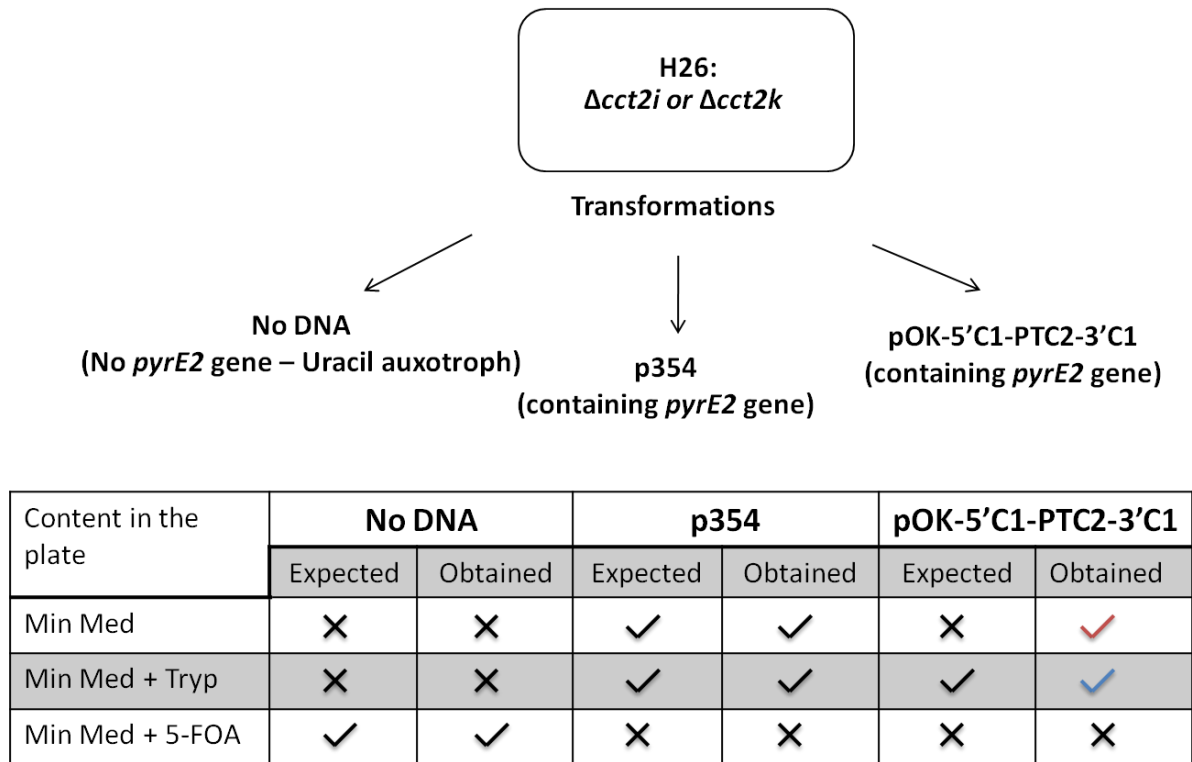
Figure 7.12

The expected and obtained results of transformations in double knock-out strains.

7.4 Discussion

The overall aim of this part of the project was to test the effects of mutants MmCCT-M223I and MmCCT-K216E in archaeal species. As described in section 7.1.1, a markerless mutagenesis method will be used to introduce these mutations in the native strain *M. maripaludis*, the required plasmid constructs for which have already been made and sent to Dr. James Chong's laboratory where further experiments will be pursued. This study will enable us to investigate the effects of mutants MmCCT-M223I and MmCCT-K216E in their native hosts. The questions that would be answered are (a) are these mutants functional in *M. maripaludis* i.e. is growth observed when the wildtype gene is replaced with the mutant gene; (b) if growth is not observed, does it imply that expression of the mutant genes is actually deleterious to cells i.e. do these mutants actually reduce organismal function?; (c) if growth is observed, then do these mutants function as well as wildtype or in a different manner?

Because the residues, M223 and K216, are highly conserved among all archaeal chaperonins and are located in the most diverged apical domain (Chapter 3), it is likely that the corresponding mutations M223I and K216E would affect the chaperonin function in the native host. Based on its relatively lower ATPase activity and the tendency to spend more time in the open conformation (Chapter 4), it can be speculated that the chaperoning function of mutant MmCCT-M223I may possibly be compromised in the native host. However it is also possible that such an effect is not very extensive because the change from methionine to isoleucine is rather less drastic and that the mutant does not completely replace GroEL function (Chapter 3). It is intriguing thus to know how the mutant MmCCT-K216E with a satisfactory biochemical profile and its

ability to restore complete GroEL function would behave in its native host, particularly when the substitution is from a positively charged residue to a negatively charged residue in a cavity that is dominated by positive charges (Cong *et al.*, 2010). We speculate that both the mutants would confer a rather negative effect on the chaperonin function in the native host and that the effects of MmCCT-K216E mutant would be more extensive that MmCCT-M223I.

The second part of the project was to develop a triple knock-out strain that would allow analysis of conditional lethal *cct1* gene mutants in *H. volcanii*. As described, attempts to construct a double knock-out strain from H26:Δ*cct3* were initially unsuccessful. A preference for reconstitution to wild-type strain rather than a *cct2* knock-out strain was observed despite the fact that theoretically, the frequency of both the events should be the same. Such a bias could be because of the tendency of the cells to oppose the changes due to loss of the *cct2* gene (Kapatai *et al.*, 2006) thus increasing the chances of cells with intact *cct2* gene to be selected on 5-FOA plates. The use of H26:Δ*cct2*$_{trun}$Δ*cct3* presumably decreased such selection competition between the cells and hence resulted into successful construction of the double knock-out strain. The approximate frequency of the potential *cct2* knock-out colonies was 1 in 25, which was rather low when compared to certain other gene knock-out studies in *H. volcanii* using the same method (Bitan-Banin *et al.*, 2003; Altman-Price and Mevarech, 2009). However, the observed lower frequency was consistent with the results obtained for similar *cct* knock-out experiments done in the lab (P. Gowarinathan, PhD thesis), highlighting the difficulty in knocking-out of the chaperonin genes from *H. volcanii*. Indeed the trials to create a triple knock-out strain did not reach to completion owing to certain unexplained errors

in the construct or the recombination events. Nevertheless, the results remain open-ended with a perfectly delineated protocol to achieve a triple knock-out strain in future.

In essence, both the studies described here, once completed, would generate a platform to test a range of chaperonin mutants in their native hosts. The experiments with *H. volcanii* in particular, if successful, would establish a first ever *in vivo* system for direct genetic dissection of the properties of an archaeal chaperonin. Not only the system would permit extensive mutagenic studies of archaeal chaperonins but will also be able to validate their structural and biochemical conclusions derived only from *in vitro* studies. Equally important, a comparative analysis would become possible wherein the effects of the same chaperonin mutation can be simultaneously studied in both, their native hosts and *E. coli* (using the TAB21 or Ai90 system). Such parallel experiments would facilitate identification of point-mutants that behave differently in archaeal and bacterial hosts. Subsequent phylogenetic and biochemical analysis of such mutants could reflect on the basis of structural differences between the archaeal and bacterial chaperonins. And such structural-functional correlations in turn can give useful insights in understanding the divergence between the two groups of chaperonins.

References

Allers, T. (2010) Overexpression and purification of halophilic proteins in Haloferax volcanii. *Bioeng Bugs* **1**:(288-90).

Allers, T. & Mevarech, M. (2005) Archaeal genetics - the third way. *Nat Rev Genet* **6**:(58-73).

Allers, T., Ngo, H. P., Mevarech, M. & Lloyd, R. G. (2004) Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes. *Appl Environ Microbiol* **70**:(943-53).

Altieri, D. C. (2013) Hsp90 regulation of mitochondrial protein folding: from organelle integrity to cellular homeostasis. *Cell Mol Life Sci* **70**:(2463-72).

Altman-Price, N. & Mevarech, M. (2009) Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii. *J Bacteriol* **191**:(1610-7).

Amann, E., Ochs, B. & Abel, K. J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene* **69**:(301-15).

Andra, S., Frey, G., Nitsch, M., Baumeister, W. & Stetter, K. O. (1996) Purification and structural characterization of the thermosome from the hyperthermophilic archaeum Methanopyrus kandleri. *FEBS Lett* **379**:(127-31).

Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**:(223-30).

Anfinsen, C. B., Haber, E., Sela, M. & White, F. H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**:(1309-14).

Archibald, J. M., Blouin, C. & Doolittle, W. F. (2001) Gene duplication and the evolution of group II chaperonins: implications for structure and function. *J Struct Biol* **135**:(157-69).

Archibald, J. M., Logsdon, J. M. & Doolittle, W. F. (1999) Recurrent paralogy in the evolution of archaeal chaperonins. *Curr Biol* **9**:(1053-6).

Archibald, J. M., Logsdon, J. M., Jr. & Doolittle, W. F. (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol Biol Evol* **17**:(1456-66).

Archibald, J. M. & Roger, A. J. (2002a) Gene conversion and the evolution of euryarchaeal chaperonins: a maximum likelihood-based method for detecting conflicting phylogenetic signals. *J Mol Evol* **55**:(232-45).

Archibald, J. M. & Roger, A. J. (2002b) Gene duplication and gene conversion shape the evolution of archaeal chaperonins. *J Mol Biol* **316**:(1041-50).

Azia, A., Unger, R. & Horovitz, A. (2012) What distinguishes GroEL substrates from other Escherichia coli proteins? *FEBS J* **279**:(543-50).

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**:(905-20).

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**:(9188-93).

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

Battistuzzi, F. U., Feijao, A. & Hedges, S. B. (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* **4**:(44).

Bell, S. D. & Jackson, S. P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**:(208-13).

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

Bigotti, M. G. & Clarke, A. R. (2005) Cooperativity in the thermosome. *J Mol Biol* **348**:(13-26).

Bitan-Banin, G., Ortenberg, R. & Mevarech, M. (2003) Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene. *J Bacteriol* **185**:(772-8).

Boisvert, D. C., Wang, J., Otwinowski, Z., Horwich, A. L. & Sigler, P. B. (1996) The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat Struct Biol* **3**:(170-7).

Booth, C. R., Meyer, A. S., Cong, Y., Topf, M., Sali, A., Ludtke, S. J., Chiu, W. & Frydman, J. (2008) Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nat Struct Mol Biol* **15**:(746-53).

Braig, K., Adams, P. D. & Brunger, A. T. (1995) Conformational variability in the refined structure of the chaperonin GroEL at 2.8 A resolution. *Nat Struct Biol* **2**:(1083-94).

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* **371**:(578-86).

Braig, K., Simon, M., Furuya, F., Hainfeld, J. F. & Horwich, A. L. (1993) A polypeptide bound by the chaperonin groEL is localized within a central cavity. *Proc Natl Acad Sci U S A* **90**:(3978-82).

Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U. & Hayer-Hartl, M. (2001) Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**:(223-33).

Bryngelson, J. D. & Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**:(7524-8).

Buckle, A. M., Zahn, R. & Fersht, A. R. (1997) A structural model for GroEL-polypeptide recognition. *Proc Natl Acad Sci U S A* **94**:(3571-5).

Bui, E. T., Bradley, P. J. & Johnson, P. J. (1996) A common evolutionary origin for mitochondria and hydrogenosomes. *Proc Natl Acad Sci U S A* **93**:(9651-6).

Bukau, B. & Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**:(351-66).

Camasses, A., Bogdanova, A., Shevchenko, A. & Zachariae, W. (2003) The CCT chaperonin promotes activation of the anaphase-promoting complex through the generation of functional Cdc20. *Mol Cell* **12**:(87-100).

Chakraborty, K., Chatila, M., Sinha, J., Shi, Q., Poschner, B. C., Sikor, M., Jiang, G., Lamb, D. C., Hartl, F. U. & Hayer-Hartl, M. (2010) Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**:(112-22).

Chaudhuri, T. K. & Gupta, P. (2005) Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. *Cell Stress Chaperones* **10**:(24-36).

Cheetham, M. E. & Caplan, A. J. (1998) Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones* **3**:(28-36).

Chen, L. & Sigler, P. B. (1999) The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* **99**:(757-68).

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R. & Saibil, H. R. (1994) Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature* **371**:(261-4).

Clare, D. K., Stagg, S., Quispe, J., Farr, G. W., Horwich, A. L. & Saibil, H. R. (2008) Multiple states of a nucleotide-bound group 2 chaperonin. *Structure* **16**:(528-34).

Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. & Doolittle, W. F. (1989) Transformation methods for halophilic archaebacteria. *Can J Microbiol* **35**:(148-52).

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

Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., Van Der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003) An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi. *Mol Microbiol* **47**:(1495-512).

Cong, Y., Baker, M. L., Jakana, J., Woolford, D., Miller, E. J., Reissmann, S., Kumar, R. N., Redding-Johanson, A. M., Batth, T. S., Mukhopadhyay, A., Ludtke, S. J., Frydman, J. & Chiu, W. (2010) 4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc Natl Acad Sci U S A* **107**:(4967-72).

Connaris, H., West, S. M., Hough, D. W. & Danson, M. J. (1998) Cloning and overexpression in Escherichia coli of the gene encoding citrate synthase from the hyperthermophilic Archaeon Sulfolobus solfataricus. *Extremophiles* **2**:(61-6).

Creighton, T. E. (1974) Intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J Mol Biol* **87**:(579-602).

Creighton, T. E. (1983) An empirical approach to protein conformation stability and flexibility. *Biopolymers* **22**:(49-58).

Dekker, C., Roe, S. M., Mccormack, E. A., Beuron, F., Pearl, L. H. & Willison, K. R. (2011a) The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. *EMBO J* **30**:(3078-90).

Dekker, C., Stirling, P. C., Mccormack, E. A., Filmore, H., Paul, A., Brost, R. L., Costanzo, M., Boone, C., Leroux, M. R. & Willison, K. R. (2008) The interaction network of the chaperonin CCT. *EMBO J* **27**:(1827-39).

Dekker, C., Willison, K. R. & Taylor, W. R. (2011b) On the evolutionary origin of the chaperonins. *Proteins* **79**:(1172-92).

Dill, K. A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry* **24**:(1501-9).

Dill, K. A. & Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* **4**:(10-9).

Ditzel, L., Lowe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R. & Steinbacher, S. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**:(125-38).

Dobrzynski, J. K., Sternlicht, M. L., Farr, G. W. & Sternlicht, H. (1996) Newly-synthesized beta-tubulin demonstrates domain-specific interactions with the cytosolic chaperonin. *Biochemistry* **35**:(15870-82).

Dobson, C. M. & Karplus, M. (1999) The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* **9**:(92-101).

Douglas, N. R., Reissmann, S., Zhang, J., Chen, B., Jakana, J., Kumar, R., Chiu, W. & Frydman, J. (2011) Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell* **144**:(240-52).

Dyall-Smith, M., Oren, A., Jiang, H. & Dong, H. (2006) Culture-dependent study of microbial diversity of Lake Chaka. *Appl Environ Microbiol* **72**:(7427; author reply 7427).

Ellis, J. (1987) Proteins as molecular chaperones. *Nature* **328**:(378-9).

Ellis, R. J. (1990) Molecular chaperones: the plant connection. *Science* **250**:(954-9).

Ellis, R. J. (2006) Protein folding: inside the cage. *Nature* **442**:(360-2).

Ellis, R. J. (2007) Protein misassembly: macromolecular crowding and molecular chaperones. *Adv Exp Med Biol* **594**:(1-13).

Ewalt, K. L., Hendrick, J. P., Houry, W. A. & Hartl, F. U. (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90**:(491-500).

Fares, M. A. & Wolfe, K. H. (2003) Positive selection and subfunctionalization of duplicated CCT chaperonin subunits. *Mol Biol Evol* **20**:(1588-97).

Farr, G. W., Fenton, W. A., Chaudhuri, T. K., Clare, D. K., Saibil, H. R. & Horwich, A. L. (2003) Folding with and without encapsulation by cis- and trans-only GroEL-GroES complexes. *EMBO J* **22**:(3220-30).

Farr, G. W., Fenton, W. A. & Horwich, A. L. (2007) Perturbed ATPase activity and not "close confinement" of substrate in the cis cavity affects rates of folding by tail-multiplied GroEL. *Proc Natl Acad Sci U S A* **104**:(5342-7).

Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. (1989) The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. *J Bacteriol* **171**:(1379-85).

Feldman, D. E., Spiess, C., Howard, D. E. & Frydman, J. (2003) Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding. *Mol Cell* **12**:(1213-24).

Feldman, D. E., Thulasiraman, V., Ferreyra, R. G. & Frydman, J. (1999) Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. *Mol Cell* **4**:(1051-61).

Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994) Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**:(614-9).

Figueiredo, L., Klunker, D., Ang, D., Naylor, D. J., Kerner, M. J., Georgopoulos, C., Hartl, F. U. & Hayer-Hartl, M. (2004) Functional characterization of an archaeal GroEL/GroES chaperonin system: significance of substrate encapsulation. *J Biol Chem* **279**:(1090-9).

Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P. & Hartl, F. U. (1992) Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J* **11**:(4767-78).

Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. & Taguchi, H. (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J* **29**:(1552-64).

Fujiwara, S., Aki, R., Yoshida, M., Higashibata, H., Imanaka, T. & Fukuda, W. (2008) Expression profiles and physiological roles of two types of molecular chaperonins from the hyperthermophilic archaeon Thermococcus kodakarensis. *Appl Environ Microbiol* **74**:(7306-12).

Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H. & Cowan, N. J. (1992) A cytoplasmic chaperonin that catalyzes beta-actin folding. *Cell* **69**:(1043-50).

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**:(918-21).

Gomez-Puertas, P., Martin-Benito, J., Carrascosa, J. L., Willison, K. R. & Valpuesta, J. M. (2004) The substrate recognition mechanisms in chaperonins. *J Mol Recognit* **17**:(85-94).

Grabowski, B. & Kelman, Z. (2001) Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol* **183**:(5459-64).

Grayling, R. A., Sandman, K. & Reeve, J. N. (1996) Histones and chromatin structure in hyperthermophilic Archaea. *FEMS Microbiol Rev* **18**:(203-13).

Guagliardi, A., Cerchia, L. & Rossi, M. (1995) Prevention of in vitro protein thermal aggregation by the Sulfolobus solfataricus chaperonin. Evidence for nonequivalent binding surfaces on the chaperonin molecule. *J Biol Chem* **270**:(28126-32).

Guenther, M. G., Yu, J., Kao, G. D., Yen, T. J. & Lazar, M. A. (2002) Assembly of the SMRT-histone deacetylase 3 repression complex requires the TCP-1 ring complex. *Genes Dev* **16**:(3130-5).

Haas, I. G. & Wabl, M. (1983) Immunoglobulin heavy chain binding protein. *Nature* **306**:(387-9).

Haber, E. & Anfinsen, C. B. (1962) Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J Biol Chem* **237**:(1839-44).

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**:(557-80).

Hartl, F. U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**:(571-9).

Hartl, F. U. & Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**:(1852-8).

Hartl, F. U. & Hayer-Hartl, M. (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**:(574-81).

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F.,

Allers, T. & Eisen, J. A. (2010) The complete genome sequence of Haloferax volcanii DS2, a model archaeon. *PLoS One* **5**:(e9605).

Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* **12**:(842-6).

Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E. & Hartl, F. U. (1994) Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. *EMBO J* **13**:(3192-202).

Heller, M., John, M., Coles, M., Bosch, G., Baumeister, W. & Kessler, H. (2004) NMR studies on the substrate-binding domains of the thermosome: structural plasticity in the protrusion region. *J Mol Biol* **336**:(717-29).

Hemmingsen, S. M., Woolford, C., Van Der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W. & Ellis, R. J. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**:(330-4).

Hendrix, R. W. (1979) Purification and properties of groE, a host protein involved in bacteriophage assembly. *J Mol Biol* **129**:(375-92).

Hickey, A. J., Conway De Macario, E. & Macario, A. J. (2002) Transcription in the archaea: basal factors, regulation, and stress-gene expression. *Crit Rev Biochem Mol Biol* **37**:(537-99).

Hirai, H., Noi, K., Hongo, K., Mizobata, T. & Kawata, Y. (2008) Functional characterization of the recombinant group II chaperonin alpha from Thermoplasma acidophilum. *J Biochem* **143**:(505-15).

Hirtreiter, A. M., Calloni, G., Forner, F., Scheibe, B., Puype, M., Vandekerckhove, J., Mann, M., Hartl, F. U. & Hayer-Hartl, M. (2009) Differential substrate specificity of group I and group II chaperonins in the archaeon Methanosarcina mazei. *Mol Microbiol* **74**:(1152-68).

Hodson, S., Marshall, J. J. & Burston, S. G. (2012) Mapping the road to recovery: the ClpB/Hsp104 molecular chaperone. *J Struct Biol* **179**:(161-71).

Hoffmann, A., Bukau, B. & Kramer, G. (2010) Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta* **1803**:(650-61).

Horovitz, A. & Willison, K. R. (2005) Allosteric regulation of chaperonins. *Curr Opin Struct Biol* **15**:(646-51).

Horwich, A. L., Apetri, A. C. & Fenton, W. A. (2009) The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS Lett* **583**:(2654-62).

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

Horwich, A. L., Fenton, W. A., Chapman, E. & Farr, G. W. (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol* **23**:(115-45).

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F. U. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**:(147-54).

Huber, H., Hohn, M. J., Stetter, K. O. & Rachel, R. (2003) The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res Microbiol* **154**:(165-71).

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

Huo, Y., Hu, Z., Zhang, K., Wang, L., Zhai, Y., Zhou, Q., Lander, G., Zhu, J., He, Y., Pang, X., Xu, W., Bartlam, M., Dong, Z. & Sun, F. (2010) Crystal structure of group II chaperonin in the open state. *Structure* **18**:(1270-9).

Hynes, G. M. & Willison, K. R. (2000) Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of beta-actin. *J Biol Chem* **275**:(18985-94).

Iizuka, R., So, S., Inobe, T., Yoshida, T., Zako, T., Kuwajima, K. & Yohda, M. (2004) Role of the helical protrusion in the conformational change and molecular chaperone activity of the archaeal group II chaperonin. *J Biol Chem* **279**:(18834-9).

Iizuka, R., Yoshida, T., Ishii, N., Zako, T., Takahashi, K., Maki, K., Inobe, T., Kuwajima, K. & Yohda, M. (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**:(40375-83).

Iizuka, R., Yoshida, T., Maruyama, T., Shomura, Y., Miki, K. & Yohda, M. (2001) Glycine at the 65th position plays an essential role in ATP-dependent protein folding by Archael group II chaperonin. *Biochem Biophys Res Commun* **289**:(1118-24).

Ikai, A. & Tanford, C. (1971) Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* **230**:(100-2).

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J. & Frishman, D. (2008) Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **9**:(102).

Ivic, A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) Deletion of Escherichia coli groEL is complemented by a Rhizobium leguminosarum groEL homologue at 37 degrees C but not at 43 degrees C. *Gene* **194**:(1-8).

Jahn, T. R. & Radford, S. E. (2005) The Yin and Yang of protein folding. *FEBS J* **272**:(5962-70).

Jakob, U., Gaestel, M., Engel, K. & Buchner, J. (1993) Small heat shock proteins are molecular chaperones. *J Biol Chem* **268**:(1517-20).

Jewett, A. I. & Shea, J. E. (2008) Do chaperonins boost protein yields by accelerating folding or preventing aggregation? *Biophys J* **94**:(2987-93).

Jewett, A. I. & Shea, J. E. (2010) Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell Mol Life Sci* **67**:(255-76).

Kafri, G., Willison, K. R. & Horovitz, A. (2001) Nested allosteric interactions in the cytoplasmic chaperonin containing TCP-1. *Protein Sci* **10**:(445-9).

Kapatai, G., Large, A., Benesch, J. L., Robinson, C. V., Carrascosa, J. L., Valpuesta, J. M., Gowrinathan, P. & Lund, P. A. (2006) All three chaperonin genes in the archaeon Haloferax volcanii are individually dispensable. *Mol Microbiol* **61**:(1583-97).

Kawata, Y., Kawagoe, M., Hongo, K., Miyazaki, T., Higurashi, T., Mizobata, T. & Nagai, J. (1999) Functional communications between the apical and equatorial domains of GroEL through the intermediate domain. *Biochemistry* **38**:(15731-40).

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M. & Hartl, F. U. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122**:(209-20).

Kim, P. S. & Baldwin, R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **51**:(459-89).

Kim, S., Willison, K. R. & Horwich, A. L. (1994) Cystosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem Sci* **19**:(543-8).

Klumpp, M., Baumeister, W. & Essen, L. O. (1997) Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* **91**:(263-70).

Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D. J., Pfeifer, G., Muller, V., Deppenmeier, U., Gottschalk, G., Hartl, F. U. & Hayer-Hartl, M. (2003) Coexistence of group I and group II chaperonins in the archaeon Methanosarcina mazei. *J Biol Chem* **278**:(33256-67).

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

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

Kubota, S., Kubota, H. & Nagata, K. (2006) Cytosolic chaperonin protects folding intermediates of Gbeta from aggregation by recognizing hydrophobic beta-strands. *Proc Natl Acad Sci U S A* **103**:(8360-5).

Kuo, Y. P., Thompson, D. K., St Jean, A., Charlebois, R. L. & Daniels, C. J. (1997) Characterization of two heat shock genes from Haloferax volcanii: a model system for transcription regulation in the Archaea. *J Bacteriol* **179**:(6318-24).

Kusmierczyk, A. R. & Martin, J. (2003a) Nested cooperativity and salt dependence of the ATPase activity of the archaeal chaperonin Mm-cpn. *FEBS Lett* **547**:(201-4).

Kusmierczyk, A. R. & Martin, J. (2003b) Nucleotide-dependent protein folding in the type II chaperonin from the mesophilic archaeon Methanococcus maripaludis. *Biochem J* **371**:(669-73).

Lam, W. L. & Doolittle, W. F. (1989) Shuttle vectors for the archaebacterium Halobacterium volcanii. *Proc Natl Acad Sci U S A* **86**:(5478-82).

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. (1993) Characterization of a functionally important mobile domain of GroES. *Nature* **364**:(255-8).

Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F. U. (1992) Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J* **11**:(4757-65).

Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J. & Lund, P. A. (2007) Characterization of a tightly controlled promoter of the halophilic archaeon Haloferax volcanii and its use in the analysis of the essential cct1 gene. *Mol Microbiol* **66**:(1092-106).

Large, A. T., Kovacs, E. & Lund, P. A. (2002) Properties of the chaperonin complex from the halophilic archaeon Haloferax volcanii. *FEBS Lett* **532**:(309-12).

Large, A. T. & Lund, P. A. (2009) Archaeal chaperonins. *Front Biosci* **14**:(1304-24).

Laskey, R. A., Honda, B. M., Mills, A. D. & Finch, J. T. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275**:(416-20).

Lee, J. W., Kim, S. W., Kim, J. H., Jeon, S. J., Kwon, H. J., Kim, B. W. & Nam, S. W. (2013) Functional characterization of the alpha- and beta-subunits of a group II chaperonin from Aeropyrum pernix K1. *J Microbiol Biotechnol* **23**:(818-25).

Leroux, M. R. & Hartl, F. U. (2000) Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr Biol* **10**:(R260-4).

Li, Y., Gao, X. & Chen, L. (2009) GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side. *J Biol Chem* **284**:(4324-31).

Lin, Z., Madan, D. & Rye, H. S. (2008) GroEL stimulates protein folding through forced unfolding. *Nat Struct Mol Biol* **15**:(303-11).

Lin, Z., Schwartz, F. P. & Eisenstein, E. (1995) The hydrophobic nature of GroEL-substrate binding. *J Biol Chem* **270**:(1011-4).

Liou, A. K. & Willison, K. R. (1997) Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J* **16**:(4311-6).

Llorca, O., Martin-Benito, J., Gomez-Puertas, P., Ritco-Vonsovici, M., Willison, K. R., Carrascosa, J. L. & Valpuesta, J. M. (2001) Analysis of the interaction between the eukaryotic chaperonin CCT and its substrates actin and tubulin. *J Struct Biol* **135**:(205-18).

Llorca, O., Martin-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G. M., Willison, K. R., Carrascosa, J. L. & Valpuesta, J. M. (2000) Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J* **19**:(5971-9).

Llorca, O., Mccormack, E. A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. L., Willison, K. R., Fernandez, J. J. & Valpuesta, J. M. (1999) Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature* **402**:(693-6).

Lund, P. (2011) Insights into chaperonin function from studies on archaeal thermosomes. *Biochem Soc Trans* **39**:(94-8).

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

Maier, T., Ferbitz, L., Deuerling, E. & Ban, N. (2005) A cradle for new proteins: trigger factor at the ribosome. *Curr Opin Struct Biol* **15**:(204-12).

Malakar, P. & Venkatesh, K. V. (2012) Effect of substrate and IPTG concentrations on the burden to growth of Escherichia coli on glycerol due to the expression of Lac proteins. *Appl Microbiol Biotechnol* **93**:(2543-9).

Martin-Benito, J., Boskovic, J., Gomez-Puertas, P., Carrascosa, J. L., Simons, C. T., Lewis, S. A., Bartolini, F., Cowan, N. J. & Valpuesta, J. M. (2002) Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J* **21**:(6377-86).

Martin-Benito, J., Grantham, J., Boskovic, J., Brackley, K. I., Carrascosa, J. L., Willison, K. R. & Valpuesta, J. M. (2007) The inter-ring arrangement of the cytosolic chaperonin CCT. *EMBO Rep* **8**:(252-7).

Martin, J. (2000) GroEL/GroES interaction assayed by protease protection. *Methods Mol Biol* **140**:(71-4).

Martin, J., Mayhew, M., Langer, T. & Hartl, F. U. (1993) The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature* **366**:(228-33).

Maurizi, M. R. & Xia, D. (2004) Protein binding and disruption by Clp/Hsp100 chaperones. *Structure* **12**:(175-83).

Mayer, M. P. & Bukau, B. (2005) Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* **62**:(670-84).

Mccormack, E. A., Altschuler, G. M., Dekker, C., Filmore, H. & Willison, K. R. (2009) Yeast phosducin-like protein 2 acts as a stimulatory co-factor for the folding of actin by the chaperonin CCT via a ternary complex. *J Mol Biol* **391**:(192-206).

Mccormack, E. A., Llorca, O., Carrascosa, J. L., Valpuesta, J. M. & Willison, K. R. (2001) Point mutations in a hinge linking the small and large domains of beta-actin result in trapped folding intermediates bound to cytosolic chaperonin CCT. *J Struct Biol* **135**:(198-204).

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

Mclennan, N. F., Mcateer, S. & Masters, M. (1994) The tail of a chaperonin: the C-terminal region of Escherichia coli GroEL protein. *Mol Microbiol* **14**:(309-21).

Meimaridou, E., Gooljar, S. B. & Chapple, J. P. (2009) From hatching to dispatching: the multiple cellular roles of the Hsp70 molecular chaperone machinery. *J Mol Endocrinol* **42**:(1-9).

Mendoza, J. A., Lorimer, G. H. & Horowitz, P. M. (1991) Intermediates in the chaperonin-assisted refolding of rhodanese are trapped at low temperature and show a small stoichiometry. *J Biol Chem* **266**:(16973-6).

Meyer, A. S., Gillespie, J. R., Walther, D., Millet, I. S., Doniach, S. & Frydman, J. (2003) Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **113**:(369-81).

Minuth, T., Frey, G., Lindner, P., Rachel, R., Stetter, K. O. & Jaenicke, R. (1998) Recombinant homo- and hetero-oligomers of an ultrastable chaperonin from the archaeon Pyrodictium occultum show chaperone activity in vitro. *Eur J Biochem* **258**:(837-45).

Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. & Hodges, R. S. (1995) Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci* **1**:(319-29).

Moore, B. C. & Leigh, J. A. (2005) Markerless mutagenesis in Methanococcus maripaludis demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J Bacteriol* **187**:(972-9).

Motojima, F., Motojima-Miyazaki, Y. & Yoshida, M. (2012) Revisiting the contribution of negative charges on the chaperonin cage wall to the acceleration of protein folding. *Proc Natl Acad Sci U S A* **109**:(15740-5).

Motojima, F. & Yoshida, M. (2010) Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. *EMBO J* **29**:(4008-19).

Nielsen, K. L. & Cowan, N. J. (1998) A single ring is sufficient for productive chaperonin-mediated folding in vivo. *Mol Cell* **2**:(93-9).

Nitsch, M., Klumpp, M., Lupas, A. & Baumeister, W. (1997) The thermosome: alternating alpha and beta-subunits within the chaperonin of the archaeon Thermoplasma acidophilum. *J Mol Biol* **267**:(142-9).

Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H. & Allers, T. (2007) Genetic and physical mapping of DNA replication origins in Haloferax volcanii. *PLoS Genet* **3**:(e77).

O'brien, E. P., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. (2012) Trigger factor slows co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J Am Chem Soc* **134**:(10920-32).

Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M. & Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **8**:(2759-69).

Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Socci, N. D. (2000) The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv Protein Chem* **53**:(87-152).

Onuchic, J. N. & Wolynes, P. G. (2004) Theory of protein folding. *Curr Opin Struct Biol* **14**:(70-5).

Pappenberger, G., Mccormack, E. A. & Willison, K. R. (2006) Quantitative actin folding reactions using yeast CCT purified via an internal tag in the CCT3/gamma subunit. *J Mol Biol* **360**:(484-96).

Pappenberger, G., Wilsher, J. A., Roe, S. M., Counsell, D. J., Willison, K. R. & Pearl, L. H. (2002) Crystal structure of the CCTgamma apical domain: implications for substrate binding to the eukaryotic cytosolic chaperonin. *J Mol Biol* **318**:(1367-79).

Pearl, L. H. & Prodromou, C. (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem* **75**:(271-94).

Pelham, H. R. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46**:(959-61).

Pereira, J. H., Ralston, C. Y., Douglas, N. R., Kumar, R., Lopez, T., Mcandrew, R. P., Knee, K. M., King, J. A., Frydman, J. & Adams, P. D. (2012) Mechanism of nucleotide sensing in group II chaperonins. *EMBO J* **31**:(731-40).

Pereira, J. H., Ralston, C. Y., Douglas, N. R., Meyer, D., Knee, K. M., Goulet, D. R., King, J. A., Frydman, J. & Adams, P. D. (2010) Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. *J Biol Chem* **285**:(27958-66).

Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**:(3551-67).

Qiu, X. B., Shao, Y. M., Miao, S. & Wang, L. (2006) The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol Life Sci* **63**:(2560-70).

Ranson, N. A., Clare, D. K., Farr, G. W., Houldershaw, D., Horwich, A. L. & Saibil, H. R. (2006) Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nat Struct Mol Biol* **13**:(147-52).

Rao, J. K. & Argos, P. (1981) Structural stability of halophilic proteins. *Biochemistry* **20**:(6536-43).

Reading, D. S., Hallberg, R. L. & Myers, A. M. (1989) Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature* **337**:(655-9).

Reissmann, S., Parnot, C., Booth, C. R., Chiu, W. & Frydman, J. (2007) Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat Struct Mol Biol* **14**:(432-40).

Richter, K. & Buchner, J. (2006) hsp90: twist and fold. *Cell* **127**:(251-3).

Ritco-Vonsovici, M. & Willison, K. R. (2000) Defining the eukaryotic cytosolic chaperonin-binding sites in human tubulins. *J Mol Biol* **304**:(81-98).

Rivenzon-Segal, D., Wolf, S. G., Shimon, L., Willison, K. R. & Horovitz, A. (2005) Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. *Nat Struct Mol Biol* **12**:(233-7).

Rommelaere, H., De Neve, M., Melki, R., Vandekerckhove, J. & Ampe, C. (1999) The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* **38**:(3246-57).

Rommelaere, H., Van Troys, M., Gao, Y., Melki, R., Cowan, N. J., Vandekerckhove, J. & Ampe, C. (1993) Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven related subunits. *Proc Natl Acad Sci U S A* **90**:(11975-9).

Rospert, S., Dubaquie, Y. & Gautschi, M. (2002) Nascent-polypeptide-associated complex. *Cell Mol Life Sci* **59**:(1632-9).

Rothman, J. E. (1989) Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. *Cell* **59**:(591-601).

Ruano-Rubio, V. & Fares, M. A. (2007) Testing the neutral fixation of hetero-oligomerism in the archaeal chaperonin CCT. *Mol Biol Evol* **24**:(1384-96).

Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. & Horwich, A. L. (1999) GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97**:(325-38).

Saibil, H. R., Fenton, W. A., Clare, D. K. & Horwich, A. L. (2013) Structure and allostery of the chaperonin GroEL. *J Mol Biol* **425**:(1476-87).

Salminen, A., Ojala, J., Kaarniranta, K., Hiltunen, M. & Soininen, H. (2011) Hsp90 regulates tau pathology through co-chaperone complexes in Alzheimer's disease. *Prog Neurobiol* **93**:(99-110).

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**:(5463-7).

Santagata, S., Bhattacharyya, D., Wang, F. H., Singha, N., Hodtsev, A. & Spanopoulou, E. (1999) Molecular cloning and characterization of a mouse homolog of bacterial ClpX, a novel mammalian class II member of the Hsp100/Clp chaperone family. *J Biol Chem* **274**:(16311-9).

Sato, T., Fukui, T., Atomi, H. & Imanaka, T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *J Bacteriol* **185**:(210-20).

Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. (1996) HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci* **21**:(289-96).

Schoehn, G., Hayes, M., Cliff, M., Clarke, A. R. & Saibil, H. R. (2000a) Domain rotations between open, closed and bullet-shaped forms of the thermosome, an archaeal chaperonin. *J Mol Biol* **301**:(323-32).

Schoehn, G., Quaite-Randall, E., Jimenez, J. L., Joachimiak, A. & Saibil, H. R. (2000b) Three conformations of an archaeal chaperonin, TF55 from Sulfolobus shibatae. *J Mol Biol* **296**:(813-9).

Sergeeva, O. A., Chen, B., Haase-Pettingell, C., Ludtke, S. J., Chiu, W. & King, J. A. (2013) Human CCT4 and CCT5 chaperonin subunits expressed in Escherichia coli form biologically active homo-oligomers. *J Biol Chem* **288**:(17734-44).

Sharma, S., Chakraborty, K., Muller, B. K., Astola, N., Tang, Y. C., Lamb, D. C., Hayer-Hartl, M. & Hartl, F. U. (2008) Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* **133**:(142-53).

Shomura, Y., Yoshida, T., Iizuka, R., Maruyama, T., Yohda, M. & Miki, K. (2004) Crystal structures of the group II chaperonin from Thermococcus strain KS-1: steric hindrance by the substituted amino acid, and inter-subunit rearrangement between two crystal forms. *J Mol Biol* **335**:(1265-78).

Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999) Chaperonin function: folding by forced unfolding. *Science* **284**:(822-5).

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

Silver, L. M., White, M. & Artzt, K. (1980) Evidence for unequal crossing over within the mouse T/t complex. *Proc Natl Acad Sci U S A* **77**:(6077-80).

Simossis, V. A. & Heringa, J. (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33**:(W289-94).

Spiess, C., Meyer, A. S., Reissmann, S. & Frydman, J. (2004) Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol* **14**:(598-604).

Spiess, C., Miller, E. J., Mcclellan, A. J. & Frydman, J. (2006) Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. *Mol Cell* **24**:(25-37).

Stibitz, S. (1994) Use of conditionally counterselectable suicide vectors for allelic exchange. *Methods Enzymol* **235**:(458-65).

Stirling, P. C., Srayko, M., Takhar, K. S., Pozniakovsky, A., Hyman, A. A. & Leroux, M. R. (2007) Functional interaction between phosducin-like protein 2 and cytosolic chaperonin is essential for cytoskeletal protein function and cell cycle progression. *Mol Biol Cell* **18**:(2336-45).

Stoldt, V., Rademacher, F., Kehren, V., Ernst, J. F., Pearce, D. A. & Sherman, F. (1996) Review: the Cct eukaryotic chaperonin subunits of Saccharomyces cerevisiae and other yeasts. *Yeast* **12**:(523-9).

Stuart, S. F., Leatherbarrow, R. J. & Willison, K. R. (2011) A two-step mechanism for the folding of actin by the yeast cytosolic chaperonin. *J Biol Chem* **286**:(178-84).

Sun, Z., Scott, D. J. & Lund, P. A. (2003) Isolation and characterisation of mutants of GroEL that are fully functional as single rings. *J Mol Biol* **332**:(715-28).

Tanaka, N. & Fersht, A. R. (1999) Identification of substrate binding site of GroEL minichaperone in solution. *J Mol Biol* **292**:(173-80).

Tang, Y. C., Chang, H. C., Chakraborty, K., Hartl, F. U. & Hayer-Hartl, M. (2008) Essential role of the chaperonin folding compartment in vivo. *EMBO J* **27**:(1458-68).

Tang, Y. C., Chang, H. C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U. & Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**:(903-14).

Thirumalai, D. & Lorimer, G. H. (2001) Chaperonin-mediated protein folding. *Annu Rev Biophys Biomol Struct* **30**:(245-69).

Thulasiraman, V., Yang, C. F. & Frydman, J. (1999) In vivo newly translated polypeptides are sequestered in a protected folding environment. *EMBO J* **18**:(85-95).

Tian, G., Vainberg, I. E., Tap, W. D., Lewis, S. A. & Cowan, N. J. (1995) Specificity in chaperonin-mediated protein folding. *Nature* **375**:(250-3).

Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996) Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc Natl Acad Sci U S A* **93**:(4030-5).

Trent, J. D., Nimmesgern, E., Wall, J. S., Hartl, F. U. & Horwich, A. L. (1991) A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**:(490-3).

Tsong, T. Y., Baldwin, R. L. & Elson, E. L. (1971) The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc Natl Acad Sci U S A* **68**:(2712-5).

Tyagi, N. K., Fenton, W. A. & Horwich, A. L. (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**:(20264-9).

Ursic, D., Sedbrook, J. C., Himmel, K. L. & Culbertson, M. R. (1994) The essential yeast Tcp1 protein affects actin and microtubules. *Mol Biol Cell* **5**:(1065-80).

Vainberg, I. E., Lewis, S. A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H. L. & Cowan, N. J. (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**:(863-73).

Van Montfort, R., Slingsby, C. & Vierling, E. (2001) Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv Protein Chem* **59**:(105-56).

Vieira, J. & Messing, J. (1991) New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene* **100**:(189-94).

Vinh, D. B. & Drubin, D. G. (1994) A yeast TCP-1-like protein is required for actin function in vivo. *Proc Natl Acad Sci U S A* **91**:(9116-20).

Waldmann, T., Lupas, A., Kellermann, J., Peters, J. & Baumeister, W. (1995a) Primary structure of the thermosome from Thermoplasma acidophilum. *Biol Chem Hoppe Seyler* **376**:(119-26).

Waldmann, T., Nitsch, M., Klumpp, M. & Baumeister, W. (1995b) Expression of an archaeal chaperonin in E. coli: formation of homo- (alpha, beta) and hetero-oligomeric (alpha+beta) thermosome complexes. *FEBS Lett* **376**:(67-73).

Walter, S. (2002) Structure and function of the GroE chaperone. *Cell Mol Life Sci* **59**:(1589-97).

Wang, J. & Boisvert, D. C. (2003) Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)14 at 2.0A resolution. *J Mol Biol* **327**:(843-55).

Wegrzyn, R. D. & Deuerling, E. (2005) Molecular guardians for newborn proteins: ribosome-associated chaperones and their role in protein folding. *Cell Mol Life Sci* **62**:(2727-38).

Wendoloski, D., Ferrer, C. & Dyall-Smith, M. L. (2001) A new simvastatin (mevinolin)-resistance marker from Haloarcula hispanica and a new Haloferax volcanii strain cured of plasmid pHV2. *Microbiology* **147**:(959-64).

Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* **70**:(697-701).

Willardson, B. M. & Howlett, A. C. (2007) Function of phosducin-like proteins in G protein signaling and chaperone-assisted protein folding. *Cell Signal* **19**:(2417-27).

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**:(4576-9).

Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995) Navigating the folding routes. *Science* **267**:(1619-20).

Won, K. A., Schumacher, R. J., Farr, G. W., Horwich, A. L. & Reed, S. I. (1998) Maturation of human cyclin E requires the function of eukaryotic chaperonin CCT. *Mol Cell Biol* **18**:(7584-9).

Xu, Z., Horwich, A. L. & Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**:(741-50).

Xu, Z. & Sigler, P. B. (1998) GroEL/GroES: structure and function of a two-stroke folding machine. *J Struct Biol* **124**:(129-41).

Yaffe, M. B., Farr, G. W., Miklos, D., Horwich, A. L., Sternlicht, M. L. & Sternlicht, H. (1992) TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature* **358**:(245-8).

Yam, A. Y., Xia, Y., Lin, H. T., Burlingame, A., Gerstein, M. & Frydman, J. (2008) Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat Struct Mol Biol* **15**:(1255-62).

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:(103-19).

Yifrach, O. & Horovitz, A. (1995) Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry* **34**:(5303-8).

Yifrach, O. & Horovitz, A. (1996) Allosteric control by ATP of non-folded protein binding to GroEL. *J Mol Biol* **255**:(356-61).

Yoshida, T., Yohda, M., Iida, T., Maruyama, T., Taguchi, H., Yazaki, K., Ohta, T., Odaka, M., Endo, I. & Kagawa, Y. (1997) Structural and functional characterization of homo-oligomeric complexes of alpha and beta chaperonin subunits from the hyperthermophilic archaeum Thermococcus strain KS-1. *J Mol Biol* **273**:(635-45).

Yoshida, T., Yohda, M., Suzuki, M., Yazaki, K., Miura, K. & Endo, I. (1998) Characterization of homo-oligomeric complexes of alpha and beta chaperonin subunits from the acidothermophilic archaeon, Sulfolobus sp. strain 7. *Biochem Biophys Res Commun* **242**:(640-7).

Zhang, J., Baker, M. L., Schroder, G. F., Douglas, N. R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C. J., Levitt, M., Ludtke, S. J., Frydman, J. & Chiu, W. (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463**:(379-83).

Zhang, J., Ma, B., Dimaio, F., Douglas, N. R., Joachimiak, L. A., Baker, D., Frydman, J., Levitt, M. & Chiu, W. (2011) Cryo-EM structure of a group II chaperonin in the prehydrolysis ATP-bound state leading to lid closure. *Structure* **19**:(633-9).

(256 references)