Structural and Functional Studies of
Mycobacterium tuberculosis Chaperonin60s

Thesis Submitted in Partial Fulfillment for the Degree of

Doctor of Philosophy

to

The Department of Biochemistry
School of Life Sciences, University of Hyderabad

by

Rohini Qamra

Centre for DNA Fingerprinting and Diagnostics

ECIL Road, Nacharam, Hyderabad 500 076

2004

Registration number: 01LBPH06

University of Hyderabad
School of Life Sciences
Department of Biochemistry
Hyderabad 500 046. India

Declaration

The research work embodied in this thesis entitled, "Structural and Functional Studies of *Mycobacterium tuberculosis* Chaperonin60s", has been carried out by me at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad, under the guidance of Dr. J. Gowrishankar. I hereby declare that this work is original and has not been submitted in part or full for any other degree or diploma of any other university.

Rohini Qamra

University of Hyderabad
School of Life Sciences
Department of Biochemistry
Hyderabad 500 046. India

Certificate

This is to certify that this thesis entitled, "Structural and Functional Studies of *Mycobacterium tuberculosis* Chaperonin60s", submitted by Ms. Rohini Qamra for the degree of Doctor of Philosophy to the University of Hyderabad is based on the work carried out by her at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad. This work is original and has not been submitted for any diploma or degree of any other university or institution.

Dr. J. Gowrishankar
Thesis supervisor
CDFD, Hyderabad

Prof. C. K. Mitra
Head, Department of Biochemistry
University of Hyderabad

Prof. A. S. Raghavendra
Dean, School of Life Sciences
University of Hyderabad

Table of Contents

Chapter 2

Cloning, Overexpression and Purification of *Mycobacterium tuberculosis* Chaperonin60s

Chapter 3

Crystallization, Data Collection and Structure Solution of Chaperonin60.2

Chapter 4

Mycobacterium tuberculosis GroEL homologues unusually exist as dimers and retain the ability to suppress aggregation of substrate proteins

ACKNOWLEDGEMENTS

Rohini Qamra

ABBREVIATIONS

Measurements

A	Angstrom
bp	basepairs
°C	degree Centigrade
hrs	hours
kb	kilobases
kDa	kiloDalton
K	Kelvin
ul	microlitre
mg	milligram
ml	millilitre
mM	milliMolar
min	minute
μg/ml	microgram per millilitre
mg/ml	milligram per millilitre
nm	nanometer
O.D.	Optical Density
rmsd	Root mean square deviation

Proteins

Cpn10	Chaperonin-10
Cpn60	Chaperonin-60
GroES	*Escherichia coli* Chaperonin-10
GroEL	*Escherichia coli* Chaperonin-60

Microorganisms

Binomial Latin or single letter abbreviated generic nomenclature was used throughout the text

Chemicals

ADP	Adenosine 5'-diphosphate
ATP	Adenosine 5'-triphosphate
EDTA	Ethylene diamine tetraacetic acid
HEPES	(N-[2-hydroxyethylpiperazine- N' [2-ethanesulfonic acid])
IPTG	isopropyl -D-thio-galactoside
LDAO	N, N-dimethyl dodecylamine N-e
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
SDS	Sodium dodecyl sulphate
TCA	Tri-chloro acetic acid
Tris	Tris (hydroxymethyl) aminomethane

Others

C-terminal	Carboxyl-terminus
DNA	Deoxyribonucleic acid
HSP	Heat shock protein
N-terminal	Amino-terminus
PAGE	Polyacrylamide gel electrophoresis

Amino-acids

Standard single or three letter codes

Nucleotides

Standard single letter codes

PREFACE

Cells from virtually all organisms respond to a variety of stresses by rapid synthesis of a highly conserved set of polypeptides, termed as heat shock proteins. These proteins have been appropriately referred to as chaperones as they serve to prevent protein misfolding and aggregation in the crowded environment of the cell, thus promoting efficient folding of many newly synthesized proteins. During stress, however, heat shock proteins stabilize and refold proteins that have been partially denatured and guide the irreversibly denatured proteins to degradation.

A pathogen is dependent on the synthesis of its heat shock proteins for survival inside its host. However, elevated levels of pathogen heat shock proteins in the host cells lead to their rapid degradation by the host. The pathogen-derived determinants, when presented to the host immune system, promote recognition of these infected cells and hence their elimination. Heat shock proteins thus form a major group of antigens recognized by the immune system in defense against infectious agents. In fact heat shock proteins have been shown to be responsible for induction of strong humoral and cellular immune responses in a variety of infections.

The heat shock protein-60 class of stress proteins, also known as the Chaperonin60s, has been shown to possess the dual function as chaperones and as antigens. The Chaperonin60 from *Escherichia coli,* the best characterized molecular chaperone, exists as a large oligomeric assembly of two heptameric rings and is aided in its chaperoning role by a co-chaperonin, Chaperonin10. Interaction between the two proteins, and communication between the two Chaperonin60 rings is necessary for the folding of a variety of polypeptides in an ATP-dependent manner.

Mycobacterium tuberculosis has been shown to possess two copies of Chaperonin60s, both of which have been shown to be potent cytokine stimulators. The role of these proteins as chaperones however remains to be characterized. This thesis takes an account of the biochemical, biophysical and structural studies of the *M. tuberculosis* Chaperonin60s, undertaken to understand the dual role of these proteins. Chapter 1 reviews the current understanding of the role of heat shock proteins as molecular chaperones and as antigens. The importance of the quaternary state of the chaperone and the protein folding mechanism by the *E. coli* Chaperonin60 is discussed in detail.

Chapter 2 gives a detailed account of the experimental procedures adopted in the cloning and over-expression of *M. tuberculosis* Chaperonin60s to obtain high levels of purified protein, which was utilized for the various biochemical assays, and production of diffraction quality crystals. The chapter also details the different procedures adopted to promote solubilization of Chaperonin60.1.

Solution of the crystal structure of Chaperonin60.2 was attempted by a number of different methods available. Chapter 3 gives a complete and detailed description of the procedures carried out to produce suitable heavy atom derivatives and generation of Seleno-methionyl Chaperonin60.2. The several attempts made to structure determination by molecular replacement and the difficulties confronted are discussed.

The oligomeric assembly of the Chaperonin60 as tetradecamers has been shown to be essential for the functioning of these proteins as molecular chaperones in an ATP-dependent manner. The oligomeric state of the *M. tuberculosis* Chaperonin60s was established by a variety of techniques. The Chaperonin60s showed unusual existence as homogenous dimers, yet retaining the ability to act as chaperones. Interestingly, the chaperoning ability was observed to be independent of ATP. Reasons for the dimeric state of

Chaperonin60.1 could be inferred by sequence analysis and further established by site directed mutagenesis of the *E. coli* Chaperonin60. Chapter 4 presents the biochemical characterization of the *M. tuberculosis* Chaperonin60s and their unusual behaviour.

Chapter 5 describes the crystal structure of Chaperonin60.2 elucidated at a resolution of 3.2A. The structure gives an explanation for the dimeric state and the loss of ATPase activity observed in this protein. The structure reveals an unusual open quaternary association of this protein. Detailed examination of the surface properties gives us insights into the functioning of Chaperonin60.2 as a chaperone. The protein surface properties also suggest the basis of its long studied role as a strong antigen.

The unusual behaviour of *M. tuberculosis* Cpn60s suggested that the proteins might have lost their role as canonical chaperonins. While *in vitro,* the proteins were less efficient in promoting folding of substrate proteins when compared with *E. coli* GroEL, attempts were made to determine their potential as molecular chaperones *in vivo.* Appendix A describes the complementation studies carried out in an *E. coli groEL* mutant strain with the *M. tuberculosis cpn60s.*

Apart from the experimental work described above on Chaperonin60s, sequences of β-barrel proteins were also analyzed. This work though not related to the biochemical and structural characterization of Chaperonin60s, showed interesting folding determinants of small β-barrels. In this study extensive structural and sequence comparisons were carried out on proteins belonging to the four stranded β-barrel family of proteins. Analysis showed a high conservation of hydrophobic residues at the core of the protein barrel. Correlation analysis suggested the importance of the total core volume for maintenance of the protein fold. Moreover, conservation of a glycyl-aspartyl

dipeptide at a type II P-turn across the different protein families was identified. Molecular dynamics simulations exhibited the importance of this dipeptide in maintaining the integrity of the P-barrel fold. The work described in Appendix B has been published in **Qamra** *et al.,* (2002). *Protein Engg.* **15,** 967-977.

Chapter 1

Introduction and Review of Literature

1.1   INTRODUCTION

Heat shock proteins are among the most highly conserved protein families in all forms of life. Although referred to as heat shock proteins, most of these proteins are expressed at significant levels in all cells under normal growth conditions and are essential for cellular growth at all physiologically relevant temperatures. Heat shock proteins perform important functions in the folding of proteins, their translocation across different compartments within a cell, as well as in the assembly of protein complexes. Heat shock proteins have thus been termed as molecular chaperones.

Increased synthesis of heat shock proteins occurs in prokaryotic and eukaryotic cells when they are exposed to stress conditions such as hypoxia, nutrient deprivation, oxygen radicals, metabolic disruption, viral infection, phagocytosis and transformation (Young and Elliot, 1989; Morimoto and Milarski, 1990; Young, 1990). As much as 15% of prokaryotic cellular protein mass is constituted by heat shock proteins under conditions of stress. The increase in the cellular content of heat shock proteins enables the cells to protect themselves from the various lethal assaults. The protection that the heat shock proteins offer to cells is primarily through their properties as molecular chaperones, whereby they interfere with the uncontrolled protein unfolding that occurs under conditions of stress. Heat shock proteins thus play a major role in cell protection following different stressful stimuli.

Another form of stress occurs upon invasion of a foreign host by a pathogen. Alterations in the cellular environment result in an increase in the cellular content of heat shock proteins within the pathogen. The host reacts to the assault by pathogens by rapidly degrading the foreign heat shock proteins. These heat shock protein-derived determinants therefore form a major group of antigens inducing strong humoral and cellular immune responses (Ziigel and Kaufmann, 1999).

Heat shock proteins thus play a dual role in the cells, as antigens in a wide spectrum of infections and as molecular chaperones.

1.2 HEAT SHOCK PROTEINS AND IMMUNITY

1.2.1 Significance of Heat Shock Proteins in Immunity

Heat shock proteins (Hsps) play a variety of important roles in immunity. For many pathogenic species, Hsps represent prominent antigens in the humoral and cellular immune response. In addition, these proteins also play an important role in presentation of antigenic peptides to the immune system. Although Hsps are widely distributed in nature, and are highly homologous among different species, the extent of their immunogenicity is different. This suggests that the immunological properties of Hsps are dependent on their sequence and structure. The different roles that Hsps play in immunity are discussed below.

1.2.1.1 *Role as Antigens*

The vertebrate immune system encounters an enormous variety of pathogens. Pathogen infection results in dramatic alterations in the living conditions of both the host cell and the microbe. With these changing conditions, induction of synthesis of Hsps within the pathogen has often been observed. Increased pathogen Hsp levels in cells lead to rapid degradation of the foreign proteins by the host processing machinery. Since they are so abundantly expressed, Hsps swamp the immune system with these epitopes. Pathogen-derived determinants are then efficiently presented by host cells and promote recognition of infected cells by the immune system (Kaufmann, 1991; Ziigel and Kaufmann, 1999).

The abundance of Hsps, especially under conditions of stress, is one of the major factors contributing to their role as antigens. Moreover, conservation of Hsps across different life forms also appears to contribute to the antigenicity of these proteins. A high conservation results in the presence of cross-reactive epitopes on different Hsps. For the host, frequent interaction with microbes results in the generation of an immunological memory for these cross-reactive determinants

(Kaufmann and Schoel, 1994). As a result, the immune system of an infected individual is already prepared to react quickly to subsequent infections. Furthermore, an immune response to the conserved epitopes of Hsps shared by different microbes may prevent colonization of the host by microbial pathogens. Thus, due to their wide distribution in nature, and high homology among different species, Hsps represent important immunogenic components of different pathogens.

1.2.1.2 *Role as Antigen Presenters*

The Hsps mediate yet other powerful reactions that engage the immune system. Many Hsps have been shown to be associated *in vivo* with a large repertoire of cellular peptides, including antigenic peptides generated by degradation of the proteins within a cell (Fig. 1.1).
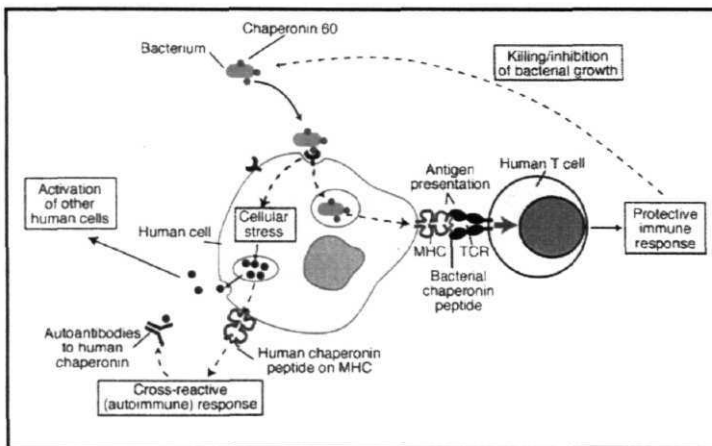
Figure 1.1 Interaction of Hsp60 proteins of bacteria with host cells During bacterial infection, host cells are activated by bacterial Hsp60 proteins that are secreted or expressed on the outer surface of bacteria. In response, the host cells produce cytokines and other intercellular signals. Presentation of pan of the Hsp on the MHC complex activates T cells via the T-cell receptor (TCR). Additionally, B cells are activated by Hsps to produce antibodies. Adapted with permission from Pockley, 2001.

The Hsp-peptide complexes are taken up by the antigen presenting cells through the cell surface receptors. The peptides thus get presented by the Major-Histocompatibility Complex (MHC) class I and class II molecules, which in turn stimulate the CD8⁺ and CD4⁺ T cells (Srivastava, 2002). These observations have led to the interesting role of Hsps as antigen presenters and carriers.

1.2.1.3 *Role as Chaperones of Antigen Presenters*

Hsps are involved in the folding and assembly of molecules that play important roles in the immune system. These include the immunoglobulins, T-cell receptors, and the MHC complex (Melnick and Argon, 1995; Williams and Watts, 1995). Members of the 70kDa heat shock protein (Hsp70) family are critically involved in the processing and presentation of antigens (Jacquier *et al.,* 1994). Bip, a member of the Hsp70 family of stress proteins and another endoplasmic chaperone, calnexin, promote the assembly of both MHC class I and class II molecules in the Endoplasmic Reticulum (ER). Calnexin facilitates MHC class I-Transport Associated Protein (TAP) interactions and thus controls peptide binding to MHC class I molecules (Jackson *et al.*, 1994; Ortmann *et at.,* 1994; Melnick and Argon, 1995).

1.2.1.4 *Role as Carrier Molecules*

Peptide transport from the proteasome to the ER and subsequent peptide loading of MHC class I molecules in the ER depends on a variety of Hsps including the Hsp70 and 90 kDa heat shock protein (Hsp90) families (Srivastava *et al.,* 1994). Recent studies have revealed that gp96, a member of the Hsp90 family in the ER, acts as a peptide acceptor. This protein receives peptides of cytosolic origin after their transport through the ER membrane by TAP molecules (Lammert *et al.,* 1997). Subsequently, gp96-peptide complexes bind to MHC and the peptides are then translocated from gp96 to MHC class I molecules in an ATP-dependent manner.

The capacity of Hsps to serve as carrier molecules has also been studied extensively in murine tumor models. Transfection with mycobacterial 60 kDa heat shock protein (Hsp60) reduces the tumorogenicity of a murine macrophage tumor cell line suggesting the role of Hsp60 in delivery of immunodominant tumor antigens to the cell surface (Lukacs *et al.,* 1993).

1.2.1.5   *Role as Adjuvants*

Hsps have unequivocally been demonstrated to be associated with a large variety of peptides including tumor, viral, cytosolic, nuclear and secreted antigens *in vivo.* When tested for their ability to complex with peptides *in vitro,* Hsps were shown to effectively present peptides of synthetic origin to the immune system. Blachere *et al.* (1997) showed that Hsps and a few synthetic peptides, when administered alone, were non-immunogenic. However, a strong immunogenic response in a MHC class I restricted manner was elicited upon administration of the Hsp70-peptide and gp96-peptide complexes. The studies thus demonstrated that Hsps act as powerful adjuvants for generation of CD8$^+$ responses (Blachere *et al.,* 1997).

Use of either foreign or self-Hsps as carrier molecules for antigenic determinants thus provides a basis for applying Hsps in conjugate vaccines.

1.2.2 Heat Shock Proteins as Antigens

Members of the Hsps family have been described as dominant antigens in several infectious diseases. They have been typically classified according to their molecular weights, the 60kDa Hsp being denoted as Hp60, the 70kDa Hsp as Hsp70 etc.

1.2.2.1   *Hsp60*

Immune responses to Hsp60 are frequently found in a wide range of microbial infections. This Hsp family has been found to elicit humoral as well as cell mediated

immune responses. For example, direct involvement of Hsp60-specific T cells has been demonstrated in a murine model of yersiniosis (Noll and Autenrieth, 1996). In this case, CD4$^+$ αβT cells specific for Hsps mediate significant protection. Similarly, levels of antibodies against Hsp60 were shown to increase significantly after vaccination with a trivalent vaccine against tetanus, diphtheria, and pertussis in infants (Del Giudice *et a/.,* 1993). Therefore, Hsp60 family is classified as an important immunological determinant in several pathogenic organisms.

Hsp60 from various bacteria induce the release of certain proinflammatory cytokines and adhesion molecules (Fig. 1.2).
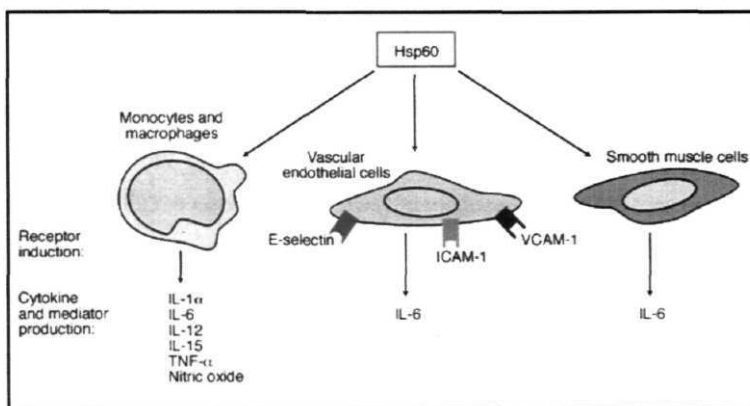
Figure 1.2 Intercellular signaling by Hsp60 Hsp60 induces the secretion of pro-inflammatory cytokines and also induces the expression of adhesion molecules on myeloid and vascular cell types, including smooth muscle cells. Adapted with permission from Pockley, 2001.

Chlamydia pneumoniae induces the expression of adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1) in human endothelial cells (Kaukoranta *et al.,* 1996a). *C. pneumoniae* also stimulates the production of proinflammatory cytokines such as Tumor Necrosis Factor-α (TNF-α) and interleukin-6 (IL-6) upon infection of human monocyte-

derived macrophages (Kaukoranta *et al,* 1996b). Moreover, human Hsp60 has been shown to elicit a potent proinflammatory response. The proinflammatory response results in the induction of TNF-α and Nitric oxide (NO) formation in cells of the innate immune system in a Toll like receptor 4 (TLR4) dependent manner, thereby suggesting that Hsp60 also serves as a cell signalling molecule (Ohashi *et al,* 2000).

1.2.2.2 *Hsp70*

Another member of the Hsp family that has been shown to play a major role in immunity is the 70kDa heat shock protein, Hsp70. Increased antibody levels to Hsp70 have been identified in sera of patients suffering from malaria, leishmaniasis, schistosomiasis, filiariasis, and candidiasis (Shinnick, 1991). Responses to pathogen-derived Hsp70 are more restricted and are sometimes exclusively species specific. Moreover, recombinant Hsp70 has recently been shown to upregulate the expression of pro-inflammatory cytokines TNF-α, interleukin-1β (IL-1β) and IL-6 in human monocytes in a CD 14 dependent pathway, transduced by Toll like receptor 2 (TLR2) and TLR4 (Asea *et al,* 2002). Thus, Hsp70 also constitutes an important immunological determinant.

1.2.2.3 *Hsp90*

Similar to Hsp60 and Hsp70 families, Hsp90 has also been shown to be involved in elicitation of the immune response. Hsp90-spedfic antibodies contribute to protection against *Candida albicans* infection (Matthews and Burnie, 1992). Hsp90 has also been shown to play a crucial role in LPS-mediated macrophage activation (Byrd *et al.*, 1999).

1.2.3 Vaccination with Pathogen Heat Shock Proteins

As Hsps represent dominant antigens in numerous microbial infections, a potential use of pathogen-derived Hsps for vaccination has been suggested. Different

vaccination strategies using Hsps have been successful in inducing significant protection in various infectious disease models. Immunization of mice with recombinant HsplO and Hsp60 from *Helicobacter pylori* protected the animals against subsequent infection and development of gastroduodenal disease (Ferrero *et al.,* 1995). Similarly, protection against pulmonary histoplasmosis was induced upon vaccination of mice with recombinant Hsp60 from *Histoplasma capsulatum* (Gomez *et al.,* 1995). Protection has also been achieved by administration of naked DNA encoding the different Hsps. Mice receiving plasmid DNA encoding mycobacterial Hsp60 showed partial protection against subsequent challenge with *Mycobacterium tuberculosis* (Bonato *et al.,* 1998). Similar protection was achieved with plasmid DNA encoding mycobacterial Hsp70 (Lowrie *et al.,* 1997).

1.2.4 *Mycobacterium tuberculosis* and the Immune System

Tuberculosis, caused by *M. tuberculosis,* is the single largest infectious cause of human mortality. Tuberculosis is responsible for ∼3 million deaths and ~8 million new cases every year. Approximately one-third of the world's population harbors *M. tuberculosis* and is at a risk of developing the disease. In humans, the majority of *Ai. tuberculosis* infections are initially controlled and a latent infection, without clinical signs of disease, is established. During the latent phase, low numbers of tubercle bacilli persist in dormancy inside a granulomatous containment. It is the interaction between the pathogen and the host immune system that results in protection against mycobacterial disease via the cell-mediated immunity (Andersen, 1997).

1.2.4.1 *Mycobacterial Heat Shock Proteins as Antigens*

In mycobacterial infections, reactivity to Hsps has been shown to predominate. Most of the antigens identified upon mycobacterial infection or vaccination in mice or humans are involved in the T-cell response.

HsplO

Although the *Escherichia coli* and human 10 kDa Hsps (Hsp10) do not elicit strong immune response (Richardson *et al.*, 2001), the mycobacterial homologue has been found to be strongly immunogenic. The *Mycobacterium leprae* and *M. tuberculosis* HsplOs have previously been shown to be important T-cell antigens (Mehra *et al*, 1989; 1992). Approximately one third of the *M. leprae* reactive T cells cross react with HsplO. Specific immune responses to HsplO include production of antibodies, T-cell proliferation and delayed-type hypersensitivity (Young and Garbe, 1991). Moreover, immunodominant T-cell epitopes have recently been mapped to regions of the *M. leprae and M. tuberculosis* HsplO (Young *et al.*, 1992; Kim *et al.*, 1997; Chua-Intra *et al.*, 1998).

Hsp60

Hsp60 has been shown to be an immunodominant target of the humoral and T-cell response in mice and humans (Young, 1990). Hsp60-specific antibodies have been detected in patients with tuberculosis and leprosy, and also in mice after infection with *M. tuberculosis* (Young *et al.,* 1988; Shinnick, 1991). Moreover, $CD4^+$ $\alpha\beta T$ cells specific for the mycobacterial Hsp60 have been found in patients with leprosy or those vaccinated with *M. bovis* BCG (Mustafa *et al.,* 1993). Surprisingly, about 20% of all mycobacterium-reactive $CD4^+$ $\alpha\beta T$ cells in mice immunized with killed *M. tuberculosis* are specific for Hsp60 (Kaufmann *et al.,* 1987). These studies suggest a protective role for Hsp60-specific T cells in mycobacterial infection.

Hsp70

The mycobacterial Hsp70 is a dominant antigen during the human T-cell response to mycobacterial infection despite the conserved sequence with the human homologue. T-cell recognition of the *M. leprae* Hsp70 antigen occurs in the context of multiple HLA-DR phenotypes and has been shown to be species specific (Adams *et*

al., 1997). Mycobacterial Hsp70 has been proposed to be utilized in subunit vaccine design since it contains a variety of T-cell epitopes presented with multiple HLA-DR molecules (Oftung *et al,* 1994).

1.2.4.2 *Mycobacterial Heat Shock Proteins as Potential Vaccine Candidates*

Currently, the only available vaccine against tuberculosis is *Mycobacterium bovis* Bacille Calmette-Guerin, the BCG vaccine. The extent of protection offered by BCG against tuberculosis is currently intensely debated, thus soliciting search for a better vaccine. Hsps of mycobacteria have been shown to be potential candidates in the development of subunit vaccines. Recently, promiscuous epitopes from the mycobacterial Hsp60 have been identified. These epitopes were responsible for recognition by the $CD4^+$ T cells in association with the HLA-DR molecules (Mustafa *et al,* 1999). These have thus been implicated in the design and development of vaccine against mycobacterial diseases. Immunization with the mycobacterial Hsp65 antigen induces protection against *M. leprae* and *M. tuberculosis* in mouse models of infections (Gelber *et al.,* 1994; Silva *et al.,* 1994). In addition, DNA vaccination in mice with *M. tuberculosis* Hsp65 antigen has been shown to provide protection against challenge with *M. tuberculosis* (Tascon *et al,* 1996). Mycobacterial Hsp70 has also been shown to act as a vaccine vehicle capable of eliciting both humoral and cell-mediated immune response. Importantly, Hsp70 fusion proteins induced these immune responses without adjuvants thus suggesting that Hsp70 functions as an exceptionally powerful carrier, capable of eliciting both T cell and B cell responses (Suzue and Young, 1996).

Hsps are abundant intracellular molecules possessing a range of housekeeping and cytoprotective functions. However, under certain circumstances, these proteins are released from cells into the extracellular environment. As secreted proteins the Hsps possess a range of immunoregulatory activities. Bacterial Hsps induce expression of pro-inflammatory cytokines and intercellular cell adhesion molecules on host endothelial cells. In addition, these proteins promote antigen presentation to

the immune system by chaperoning peptides to the antigen presenting cells. Their role as antigen presenters is a result of a very different role that the Hsps play in the intracellular milieu, that of molecular chaperones. As chaperones, the Hsps promote the correct folding and assembly of other cellular proteins under normal and stressed conditions. This they perform by interacting with the unfolded or misfolded polypeptides and preventing their aggregation thus providing the nascent polypeptides a chance to fold into the correct conformation.

1.3 HEAT SHOCK PROTEINS AS MOLECULAR CHAPERONES

1.3.1 Protein Folding *In vivo*

Anfinsen and co-workers established that the primary structure of a protein contains all the information necessary to direct the native secondary and tertiary fold (Anfinsen, 1973). *In vitro* many proteins can fold spontaneously as long as they do not undergo aggregation or interact inappropriately with themselves or other proteins. Under physiologic conditions inside the cell, however, the folding process for many proteins, particularly those with multi-domain structures, is prone to production of a variety of misfolded species and aggregates. It is thus necessary to prevent the aggregation of newly synthesized proteins and circumvent any off-pathway folding intermediates. An important step towards the molecular understanding of how proteins fold inside cells came with the discovery of specialized protein components, molecular chaperones, that play essential roles in enabling polypeptides to reach biologically active forms in a variety of cellular compartments (Horwich *et al.*, 1993; Hartl, 1996; Bukau and Horwich, 1998).

1.3.2 Molecular Chaperones

Although a protein's tertiary structure is determined by the primary sequence of the polypeptide, it is achieved, in most cases with the aid of helper proteins. Molecular chaperones are protein machines that recognize non-native states of other proteins and, by controlled binding and release, assist these substrate proteins to fold properly (Hard, 1996). While molecular chaperones assist the non-covalent assembly of proteins *in vivo*, they themselves are not permanent components of these proteins (Ellis, 1997). The functions of these proteins as molecular chaperones depend upon their ability to recognize and bind to hydrophobic regions of proteins that might

become exposed during the synthesis and folding of proteins under normal growth conditions (Gottesman and Hendrickson, 2000).

1.3.3 Heat Shock Proteins as Molecular Chaperones

Many molecular chaperones are members of the heat shock family of proteins. Under normal conditions these proteins play an important role in folding, assembly and transport of proteins. Under conditions of stress the Hsps bind and stabilize cellular proteins at intermediate stages of folding and assembly, thereby preventing their misfolding and aggregation.

1.3.3.1 *Hsp60*

The Hsp60 chaperones are essential for the survival of cells in all conditions of growth. These proteins occur as large cylindrical protein complexes composed of two stacked rings of 7-9 identical or closely related subunits. These proteins promote the proper folding and assembly of cellular proteins by enclosing the substrate polypeptide in their large cylindrical cavity in an ATP dependent manner. Elegant conformational changes are known to occur during the binding and release of substrate polypeptides.

1.3.3.2 *Hsp70*

The Hsp70 chaperones are a family of highly conserved ATPases found in prokaryotes as well as in eukaryotes. Hsp70s play essential roles under normal cellular conditions, including *de novo* protein folding, membrane translocation, disassembly of oligomeric protein structures, facilitating proteolytic degradation of unstable proteins as well as controlling the biological activity of certain regulator)' proteins and transcription factors (Bukau and Horwich, 1998). Under conditions of stress the Hsp70 proteins prevent aggregation and assist refolding of misfolded proteins (Martin and Hard, 1997). Hsp70 is aided in its role as a molecular chaperone by the co-chaperones, Hsp40 and GrpE. The major function of the Hsp70-Hsp40-

GrpE system is to prevent the premature **misfolding** of growing polypeptide chains on ribosomes.

1.3.3.3   *Hsp90*

Hsp90 is an abundant cytosolic protein present in both prokaryotes and eukaryotes. Levels of Hsp90 amount to as much as ~ 1 % of soluble proteins in the cell even in unstressed conditions. This protein is postulated to have an ATP dependent role in the refolding of misfolded proteins that accumulate in response to various stress treatments. Hsp90 has a high selectivity for its targets and is associated with the folding of a **defined** set of signal transducing proteins such as steroid hormone receptors and protein kinases (Buchner, 1999; Caplan, 1999).

1.3.4 The Chaperonins

Chaperonins are a subgroup of molecular chaperones found in all domains of life. These proteins have been shown to mediate the ATP-dependent folding of many proteins *in vivo* and *in vitro* (Frydman, 2001). Chaperonins consist of ~60kDa subunits arranged in two stacked rings forming a large central cavity where unfolded polypeptides bind and undergo productive folding (Ranson *et al.,* 1998) (Fig. 1.3).

Based on their evolutionary origin, chaperonins have been classified as - Type I chaperonins found in the bacterial cytosol and in endosymbiont derived organelles of eukaryotes. These form **heptameric** rings and are represented by bacterial GroEL, **mitochondrial** Cpn60 and chloroplast Cpn60 (**Kusmierczyk** and Martin, 2001). Type II chaperonins, present in the eukaryotic cytoplasm and in archaea, include the archaeal **thermosome** and the eukaryotic chaperonin containing TCP-1 (CCT) (Trent *et al.,* 1991; Frydman *et a/.,* 1992). Thermosome rings are eight- or nine- **membered,** and consist of one to three subunit types (Trent *et al.,* 1991; Ditzel *et al.,* 1998). CCT rings consist of eight subunits, each the product of an essential gene (Kubota *et a/.,* 1999). The crystal structures of *Eschericbia coli* GroEL and the thermosome from

Thermoplasma   acidophilum   reveal that all chaperonins share a similar domain architecture (Braig *et al.,* 1994; Xu *et al.,* 1997; Ditzel *et al.,* 1998). Each subunit of the chaperonin is divided into three domains designated: apical, intermediate and equatorial. The equatorial domain forms the interface between individual rings and is the site of ATP binding and hydrolysis. Substrate protein binds at the central face of the apical domain while a short intermediate domain connects it to the equatorial domain (Braig *et al.,* 1994; Xu *et al.,* 1997).

a   b
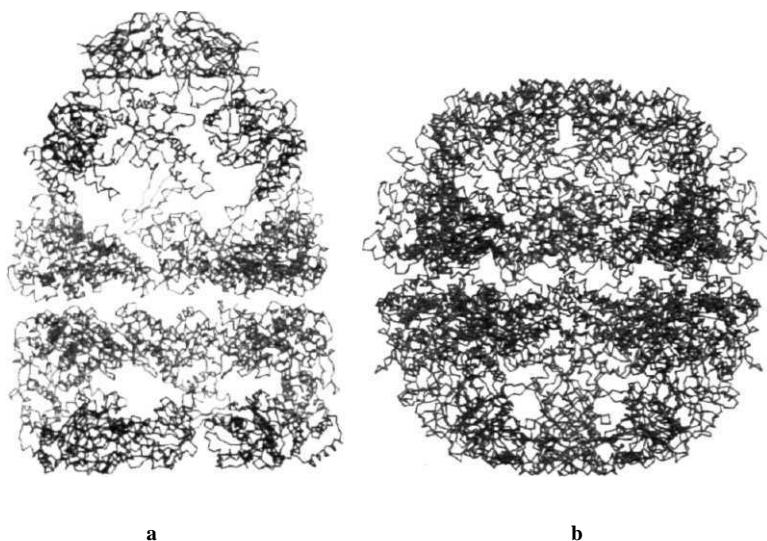
Figure 1.3 Structures of the Type I and Type II chaperonins (a) Structure of the GroEL$_{14}$-GroES7-ADPT complex. GroES is colored green. (b) Closed form of the thermosome comprising 8 subunits per ring. The subunits of both the chaperonins have essentially the same fold. The difference lies in the lid extension in the group II apical domain. The group II lid extension replaces GroES in closing the chamber. Figure generated using coordinates of 1aon and 1a6d from PDB (Xu *et al.,* 1997; Ditzel *et al.,* 1998).

1.3.4.1   *GroEL as a Molecular Chaperone*

The bacterial chaperonin, GroEL, is one of the best characterized molecular chaperones. X-ray studies combined with electron microscopy studies have provided a wealth of information into the functional cycle of this chaperonin (Fenton and Horwich, 1997; Xu *et al.,* 1997; Fukami *et al.,* 2001). GroEL is active as a double heptameric ring with each ring containing a large central cavity in which substrate protein binds and undergoes proper folding (Hendrix, 1979; Langer *et al.,* 1992; Braig *et al.,* 1993). GroEL, like all the type I chaperonins, is assisted in its function by a 10kDa co-chaperonin, GroES. The co-chaperonin exists as a hep tamer and adopts a dome-like structure that can bind to either GroEL ring to form a cap on the central cavity (Chandrasekhar *et al,* 1986; Saibil *et al.,* 1991; Langer *et al,* 1992; Hunt *et al,* 1996).

GroEL performs two actions in assisting protein folding. One is the prevention of aggregation of polypeptides by binding to collapsed, partially structured folding intermediates in the central channel. The other action is facilitation of folding, occurring inside the central cavity after enclosure by the co-chaperonin, GroES, in the presence of ATP (Fenton and Horwich, 1997). Within the central channel, polypeptides initiate folding, and a fraction reaches either the native form or a conformation committed to achieving the native state. The fraction of polypeptides that fails to reach a committed state then rebinds to GroEL for another attempt at folding.

About 10-15% of all the newly synthesized polypeptide chains interact with GroEL under normal conditions and transit the chaperonin very rapidly. This number, however, rises to more than 30% upon exposure to stress (Ewalt *et al,* 1997).

1.3.4.2 *Architecture of GroEL-GroES*

The architecture of GroEL, the GroEL-GroES complex and its polypeptide and nucleotide complexes were first described at low resolution (30A or lower) by negative stain and cryo-Electron Microscopy (EM) (Fenton and Horwich, 1997). The crystal structures of GroEL and GroES have since been determined at atomic resolution as separate assemblies and in complex with one another (Braig *et al.,* 1994; Boisvert *et at.,* 1996; Hunt *et al.,* 1996; Mande *et ai,* 1996; Wang and Boisvert, 2003; Xu *et at.,* 1997; Taneja and Mande, 2002; Roberts *et al.,* 2003).

GroEL

Crystal structures of the unliganded GroEL and the GroEL-GroES complex show a cylindrical complex with subunits of GroEL assembled into two heptameric rings stacked back-to-back to form the native 14-mer (Fig. 1.4). Each subunit of GroEL is folded into three distinct domains (Braig *et al.,* 1994; Xu *et ai,* 1997).

The equatorial domain is the largest of the three domains containing the N and C-terminal residues projecting into the central channel. This domain is highly α-helical and well ordered. The equatorial domain provides most of the intra-ring and all of the inter-ring contacts across the equatorial plane. This domain also provides the ATP binding site as established by mutational analysis and the crystal structure.

The apical domain forms the opening of the central channel. This domain is involved in interactions with GroES and substrate protein. The binding of non-native proteins to GroEL is largely nonspecific and hydrophobic in nature (Viitanen *et ai,* 1992; Lin *et ai,* 1995). The binding occurs at a ring of hydrophobic residues lining the walls of the central cavity (Langer *et ai,* 1992; Fenton *et ai,* 1994). A more detailed view of this region emerged from the crystal structure (1.7Å) of a monomeric mini-chaperone that comprised only the apical domain (Buckle *et ai,* 1997). The apical

domain possesses an intrinsic flexibility that is functionally necessary to accommodate binding of a large variety of different polypeptides in the central channel.

a   b
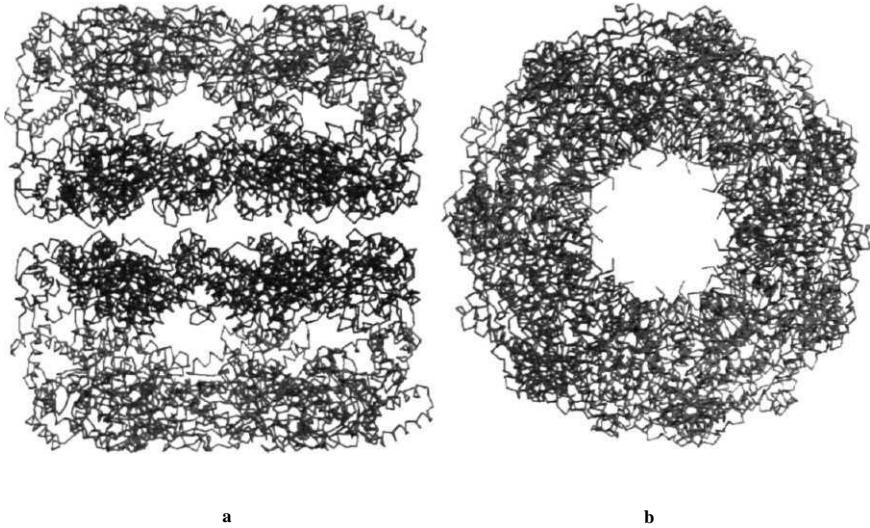
Figure 1.4 Structure of GroEL (a) The apical, intermediate and equatorial domains are colored green, purple and blue, respectively. Non-native polypeptide binds to the apical domain of GroEL shown in green. (b) The top view of GroEL showing the central channel. It is in the central cavity that the polypeptide binds. Figure generated using coordinates of 1kp8 from PDB (Wang and Boisvert, 2003).

The intermediate domain is the smallest and forms the link between the apical and equatorial domains. The covalent connections of this domain to the equatorial and apical domain are through short antiparallel segments that serve as hinges in allosteric adjustments. Inspection of GroEL in different conformational states both by EM (Roseman *et al.,* 1996) and by comparison of X-ray structures of unliganded GroEL and a GroEL-GroES complex suggest that hinge-like motions centre on this domain.   The intermediate domain thus forms a major scaffold for allosteric

communication between the equatorial and apical domains, either within a single subunit or through non-covalent contacts between different subunits.

GroES

GroEL's co-chaperonin partner, GroES, is a dome-shaped structure, 70-80Å in diameter. The protein exists as a single seven-membered ring of 10kDa subunits (Hunt *et al.,* 1996; Mande *et al.,* 1996; Taneja and Mande, 2002; Roberts *et aL,* 2003). Each subunit has a core β-barrel structure with two large protrusions extending from it. These loops are highly disordered and flexible in the unbound GroES (Landry *et aL,* 1993). One of the loops, the dome loop, contributes to the top of the dome. The second, known as the mobile loop, interacts with a GroEL subunit, in a 1:1 stoichiometry with residues in the apical domain. This loop becomes highly structured upon interaction with GroEL and forms the GroEL-GroES interface contributed by mainly aliphatic side chains from GroEL and GroES.

The GroESLComplex

Interaction between GroEL and GroES is necessary for the folding of a variety of polypeptides in an ATP-dependent manner. Interaction of GroES with one of the GroEL rings leads to the formation of the asymmetric complex, with the GroES bound ring referred to as the *cis* ring and the unbound ring known as the *trans* ring (Fig. 1.5).

The GroES ring caps the apical surface of the GroEL *cis* ring closing off the end of the central channel (Mayhew *et a/.,* 1996). Upon this interaction the interior cavity of GroES becomes continuous with that of GroEL. The binding of GroES leads to large *en bloc* movements of the apical and intermediate domains of GroEL and hence an elevation of the apical domain. The reorganized GroEL domains dramatically reshape the central cavity formed by the *cis* ring and GroES. Moreover, the polypeptide binding properties of the cavity lining are completely altered. The

cavity, now double the original size, is capable of accommodating a globular protein or a molten globule intermediate of >70kDa.

Figure 1.5 Structure of the asymmetric GroEL-GroES complex GroES is shown in green. The GroES bound ring is known as the *as* ring while the other is referred to as the *trans* ring.

The GroES mobile loops bind to and compete for the same surface of GroEL as the polypeptide (Chen *et al.,* 1994; Fenton *et al.,* 1994; Roseman *et al.,* 1996). Hydrophobic residues that bind the non-native polypeptide in the cavity of the *trans* ring are thus used to stabilize interfaces that support the GroES-GroEL complex. As a result, polar residues replace the surface of the expanded *as* cavity. This switch in the chemical character of the cavity triggers dissociation of the non-native polypeptide from the wall of the cavity. The released polypeptide is now free to initiate folding as an isolated molecule in a hydrophilic environment conducive to bury the substrate polypeptide's hydrophobic residues. While major conformation

changes are observed in the apical domains of the GroEL *as* ring, the equatorial domains in this ring undergo little change except for small but potentially important local shifts in the nucleotide-binding pocket. The domain movements in the GroEL *as* ring are driven by the binding and hydrolysis of ATP and are essential for the completion of the chaperonin reaction cycle.

1.3.4.3 *The Protein Folding Cycle*

GroEL functions as a two-stroke engine whereby the binding and hydrolysis of ATP by one chaperonin ring controls the release of the co-chaperonin GroES from the opposite ring (Fig. 1.6).
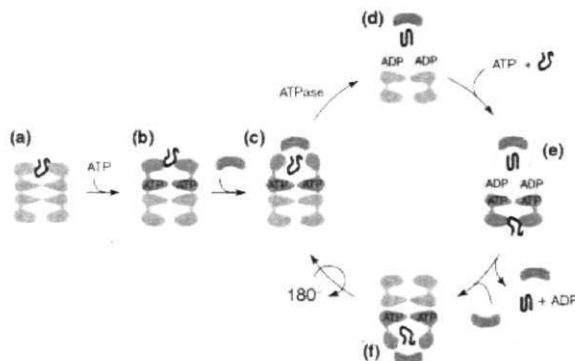
Figure 1.6 The ATP dependent protein folding cycle of GroEL The schematic shows the GroE complexes as sections through the stacked-ring complexes. (a) GroEL has high affinity for non-native polypeptide substrate (black curved line). (b) ATP binds with positive co-operativity to one ring producing an altered conformation with reduced substrate affinity. (c) The ATP-bound ring rapidly binds GroES resulting in a massive conformation change in the GroES-bound ring. (d) The substrate folds inside the chamber and ATP is hydrolysed. (e) ATP binding to the opposite ring primes the release of GroES and the trapped substrate. (f) A new substrate gets encapsulated in the opposite ring. Adapted with permission from Saibil and Ranson, 2002.

ATP binds cooperatively to the subunits of one ring in GroEL triggering a conformational change that reduces substrate affinity in the ATP-bound ring (Bochkareva *et al.,* 1992; Jackson *et al,* 1993). Immediate binding of GroES to the ATP-bound ring seals the cavity enclosing the substrate polypeptide (Roseman *et al.,* 1996). While inside the cavity the protein attempts to refold in a time period of 10-20seconds following which the seven ATP molecules in the *cis* ring are hydrolysed (Todd *et al.,* 1994; Kusmierczyk and Martin, 2001).

ATP hydrolysis in the *cis* ring primes the ring to release its ligands. The actual release is triggered as a result of negative cooperativity between the two GroEL rings. Binding of ATP to the *trans* ring results in a reduced GroES affinity in the *cis* ring thereby allowing the substrate protein to be released from the *cis* ring (Yifrach and Horovitz, 1995, 1996).

Communication between the two rings is responsible for the release of GroES and the bound substrates. This was shown by a single ring mutant of GroEL, which upon formation of a complex with GroES, performs only one round of ATP hydrolysis. This ring is now poised to release the bound polypeptide. However, due to the lack of the distal ring, it does not receive the signal to release GroES and hence the bound polypeptide (Weissman *et al.,* 1995, 1996; Llorca *et al.,* 1997; Rye *et al,* 1997).

1.3.5 *groEL*: Arrangement on the Genome

A single copy of the *grvEL* gene has been reported in most of the eubacteria studied. The GroES and GroEL proteins are encoded by genes arranged on the bicistronic *groESL* operon. This chromosomal arrangement allows for co-ordinated expression of the *groES* and *groEL* genes, which is in accordance with their biological function as molecular chaperones. Mazodier *et al.* however first reported an exception to this rule (1991). Their study revealed the presence of two *groEL*-like genes in *Streptomyces albus* of which only one, named *groEL1,* was adjacent to a *groES*-like gene.

Such multiple copies of *groEL*-like genes were confirmed in 12 Streptomyces species. Later studies revealed that the presence of multiple copies of *groEL* was not restricted to Streptomyces alone. Two *groEL* genes were found to exist in the genome of *M. leprae* (Rinke de Wit *et al.,* 1992) as also in *M. tuberculosis* (Kong *et al.,* 1993). This existence of a duplicate set of *groEL* genes on the genome is common to the mycobacterial class.

1.4 CHAPERONINS OF *Mycobacterium* *tuberculosis*

The genes encoding the two chaperonins in *M. tuberculosis* exist on separate parts of the genome. The genes have been referred to as *cpn60.1* and *cpn60.2*. While *cpn60.1* is arranged on an operon with the *groES* (*cpn10*), the second gene, *cpn60.2,* is present as a separate gene. The two *cpn60* genes bear ~66% sequence identity, with the gene products displaying 61% sequence identity and 76% sequence similarity. Upon comparison with other members of the Cpn60 family, *M. tuberculosis* Cpn60.2, product of *cpn60.2,* consistently shows 5-10% greater sequence identity and similarity than Cpn60.1, product of, *cpn60.1.* Cpn60.1 and Cpn60.2 display 53% and 59% sequence identity and 70% and 76% similarity to the *E. coli* GroEL, respectively.

Much attention has focused on the *M. tuberculosis* molecular chaperonins in the pathology of tuberculosis because of their immunogenicity and the ability to directly activate human monocytes and vascular endothelial cells.

1.4.1 Co-Chaperonin, Cpn10

cpn10, present upstream of *cpn60.1,* codes for the 10kDa co-chaperonin partner protein, CpnlO. CpnlO elicits a strong humoral and cell-mediated immune response upon infection. This protein has been shown to be an important T-cell antigen (Young and Garbe, 1991). CpnlO has been postulated to be a major factor responsible for bone resorption in Pott's disease (Meghji *et al.,* 1997). Biophysical characterization and crystal structure of the *M. tuberculosis* CpnlO have revealed regions that may play crucial roles in bone resorption (Taneja and Mande, 2001a, b).

1.4.2 Chaperonin, Cpn60

The content of the Cpn60 proteins of *M. tuberculosis* increases from 1-10% or more under conditions of stress such as are likely to occur during infection (Young and Garbe, 1991). Both Cpn60s of *M. tuberculosis* have a role to play as important

virulence factors in tuberculosis. In fact **Cpn60.2**, also known as the 65kDa antigen was first identified for its role in **autoimmunity**. The two proteins however, differ in their cytokine inducing potency and efficacy.

Cpn60.1 was recently shown to result in the production of a large number of potent pro-inflammatory cytokines including **IL-1β, TNF-α**, IL-6 and IL-12. This protein is able to stimulate human peripheral blood mononuclear cells to synthesize and secrete cytokines in a CD14 dependent manner (Lewthwaite *et al.,* 2001).

Cpn60.2 is also a very potent **cytokine** inducer and stimulates the synthesis and secretion of a range of pro-inflammatory cytokines and the anti-inflammatory cytokine **IL-10** (Lewthwaite *et a/.,* 2001). Moreover, Cpn60.2 plays a major role in autoimmunity and arthritis and as an antigen of **γδ-T** cells (van Eden *et al.,* 1998). This protein has been shown to activate human peripheral blood mononuclear cells in a **CD14** independent mechanism.

While previous studies have established the role of the two chaperonins of *M. tuberculosis* as potent immunogens, reasons for their strong **immunogenicity** in comparison to the homologous *E. coli* or human chaperonins remain unknown. Moreover, these proteins have not been characterized at the biochemical and structural level and hence the mechanisms of cytokine induction by these proteins remain to be elucidated. The advantage that the duplication of the chaperonin genes offers to the pathogen also remains unknown. With both the chaperonins acting as potent virulence factors, the structural and biochemical studies of these proteins become very crucial to the understanding of their function. Furthermore, given the high sequence **homology** that the *M. tuberculosis* Cpn60s share with other members of this protein family, the study becomes very important to understand the dual role of these proteins, as immunogens and chaperonins, and the relationship and advantage the two functions have to offer. Biochemical and structural studies of *M. tuberculosis* Chaperonin 60s were hence undertaken to understand the properties of these proteins. The thesis reports the results of this work.

1.5   REFERENCES

1. Adams, E., Basten, A., Rodda, S., and Britton, W.J. (1997). Human T-cell clones to the 70-kilodalton heat shock protein of *Mycobacterium leprae* define mycobacterium-specific epitopes rather than shared epitopes. *Infect Immun.* 65, 1061-1070.

2. Andersen, P. (1997). Host responses and antigens involved in protective immunity to *Mycobacterium tuberculosis. Scand J. Immunol AS,* 115-131.

3. Anfinsen, C. (1973). Principles that govern protein folding. *Science* 181, 223-230.

4. Asea, A., Rehli, M., Kabingu, E., Boch, J.A., Bare, O., Auron, P.E., Stevenson, M.A., Calderwood, S.K. (2002). Novel signal transduction pathway utilized by extracellular HSP70: role of toll-like receptor TLR2 and TLR4. *JBiol Chem.* 277, 15028-15034.

5. Blachere, N.E., Li, Z., Chandawarkar, R.Y., Suto, R., Jaikaria, N.S., Basu, S., Udono, H., and Srivastava, P.K. (1997). Heat shock protein-peptide complexes, reconstituted *in vitro,* elicit peptide-specific cytotoxic T lymphocyte response and tumor immunity. *J. Exp. Med.* 186,1315-1322.

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

7. Boisvert, D.C., Wang, J., Otwinowski, Z., Horwich, A.L., and Sigler, P.B. (1996). The 2.4A crystal structure of the bacterial chaperonin GroEL complex with ATP-Y-S. *Nature Struct. Biol.* 3, 170-177.

8. Bonato, V.L.D., Lima, V.M.F., Tascon, R.E., Lowrie, D.B., and Silva, C.L. (1998). Identification and characterization of protective T cells in hsp65 DNA-vaccinated and *Mycobacterium tuberculosis-infected* mice. *Infect. Immun.* 66, 169-175.

9. Braig, K., Simon, M., Furuya, F., Hainfeld, J.F., and Horwich, A.L. (1993). A polypeptide bound by the chaperonin GroEL is localized within a central cavity. *Proc. Natl. Acad. Sci. USA* 90, 3978-3982.

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

11. Buchncr, J. (1999). Hsp90 and Co. - a holding for folding. *Trends Biochem. Sci.* 24, 136-141.

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

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

14. Byrd, C.A., Bornmann, W., Erdjument-Bromage, H., Tempst, P., Pavletich, N., Rosen, N., Nathan, C.F., and Ding, A. (1999). Heat shock protein 90 mediates macrophage activation by Taxol and bacterial lipopolysaccharide. *Proc. Natl. Acad. Sci. USA* 96, 5645-5650.

15. Caplan, A.J. (1999). Hsp90's secrets unfold: new insights from structural and functional studies. *Trends Cell Biol.* 9, 262-268.

16. Chandrasekhar, G.N., Tilly, K., Woolford, C, Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the *groES* morphogenetic protein of *Escherichia coli. J Biol Chem.* **261,** 12414-12419.

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

18. Chua-Intra, B., Peerapakorn, S., Davey, N., Jurcevic, S., Busson, M., Vordermeier, H.M., Pirayavaraporn, C, and Ivanyi, J. (1998). T-cell recognition of mycobacterial GroES peptides in Thai leprosy patients and contacts. *Infect. Immun.* **66,** 4903-4909.

19. Del Giudice, G., Gervaix, A., Costantino, P., Wyler, C.A., Tougne, C, De Graeff-Meeder, E. R., van Embden, J., Van der Zee, R., Nencioni, L., Rappuoli,

R., Suter, S., and Lambert, P.H. (1993). Priming to heat shock proteins in infants vaccinated against pertussis. *J. Immunol.* **150,** 2025-2032.

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

21. Ellis, R.J. (1997). Do molecular chaperones have to be proteins? *Biochem. Biopbys. Res. Commun.* **238,** 687-692.

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

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

24. Fenton, W.A., and Horwich, A.L. (1997). GroEL-mediated protein folding. *Protein Sci.* 6, 743-760.

25. Ferrero, R.L., Thilberge, J.M., Kansau, I., Wuscher, N., Huerre, M., and Labigne, A. (1995). The GroES homologue of *Helicobacter pylori* confers protective immunity against mucosal infection in mice. *Proc. Nat/. Acad. Sci. USA*92, 6499-6503.

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

27. Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J.S., Tempst, P., and Hartl, F.U. (1992). Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11,** 4767-4778.

28. Fukami, T.A., Yohda, M., Taguchi, H., Yoshida, M., and Miki, K. (2001). Crystal structure of chaperonin-60 from *Paracoccus denitrificans. J. Mol. Biol.* **312,** 501-509.

29. Gelber, R.H., Mehra, V., Bloom, R.R., Murray, L.P., Siu, P., Tsang, M., and Brennan, P.J. (1994). Vaccination with pure *Mycobacterium leprae* proteins inhibits *M. leprae* multiplication in mouse foot pads. *Infect. Immun.* **62,** 4250-4255.

30. Gomez, F.J., Allendoerfer, R., and Deepe, G.S. Jr. (1995). Vaccination with recombinant heat shock protein 60 from *Histoplasma capsulatum* protects mice against pulmonary histoplasmosis. *Infect. Immun.* 63, 2587-2595.

31. Gottesman, M.E., and Hendrickson, W.A. (2000). Protein folding and unfolding by *Escherichia coli* chaperones and chaperonins. *Current Opi. Microbiol.* **3,** 197-202.

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

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

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

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

36. Jackson, G.S., Staniforth, R.A., Halsall, D.J., Atkinson, T., Holbrook, J.J., Clarke, A.R., and Burston, S.G. (1993). Binding and hydrolysis of nucleotides in the chaperonin catalytic *cycle:* implications for the mechanism of assisted protein folding. *Biochemistry* **32,** 2554-2263.

37. Jackson, M., Cohen-Doyle, M., Peterson, P., and Williams, D. (1994). Regulation of MHC class I transport by the molecular chaperone, calnexin (p88/IP90). *Science* **263,** 348-387.

38. Jacquier-Sarlin, M.R., Fuller, K., Dinh-Xuan, A.T., Richard, M.J., and Polla, B.S. (1994). Protective effects of Hsp70 in inflammation. *Experientia* 50, 1031-1038.

39. Kaufmann, S.H.E., Vath, U., Thole, J.E.R., van Embden, J.D.A., and Emmrich, F. (1987). Enumeration of T cells reactive with *Mycobacterium tuberculosis* organisms

and specific for the recombinant **mycobacterial** 65 kilodalton protein. *Eur. J. Immunol.* **178,** 351-357.

40. **Kaufmann,** S.H.E. (1991). Heat shock proteins and pathogenesis of bacterial infections. *Springer Semin. Immunopathol.* 13, 25-36.

41. Kaufmann, S.H.E., and Schoel, B. (1994). Heat shock proteins as antigens in immunity against infection and self, *In* R. I. **Morimoto,** A. Tissieres, and C. Georgopoulos (ed.), The biology of heat shock proteins and molecular chaperones. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 495-531.

42. Kaukoranta-Tolvanen, S.S., Ronni, T., Leinonen, M., Saikku, P., and Laitinen, **K.** (1996a). Expression of adhesion molecules on endothelial cells stimulated by *Chlamydia pneumoniae. Microb. Pathog.* 21, 407-411.

43. Kaukoranta-Tolvanen, S.S., Teppo, A.M., Laitinen, **K.,** Saikku, P., Linnavuori, **K.,** and Leinonen, M. (1996b). Growth of *Chlamydia pneumoniae* in cultured human peripheral blood mononuclear cells and induction of a cytokine response. *Microb. Pathog.* 21, 215-221.

44. Kim, J., Sette, A., Rodda, S., Southwood, S., Sieling, P.A., Mehra, V., **Ohmen,** J.D., **Oliveros,** J., Appella, E., Higashimoto, Y., Rea, T.H., Bloom, B.R., and Modlin, R.L. (1997). Determinants of T cell reactivity' to the *Mycobacterium leprae* GroES **homologue.** *J. Immunol.* **159,** 335-343.

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

46. Kubota, H., Yokota, S., Yanagi, H., and Yura, T. (1999). Structures and co-regulated expression of the genes encoding mouse cytosolic chaperonin CCT subunits. *Eur. J. Biochem.* **262,** 492-500.

47. Kusmierczyk, A.R., and Martin, J. (2001). Chaperonins — keeping a lid on folding proteins *FEBS Utters* **505,** 343-347.

48. **Lammert,** E., Arnold, D., Nijenhuis, M., **Momburg,** F., Hammerling, G.J., Brunner, J., Stefanovic, S., **Rammensee,** H.G., and Schild, H. (1997). The endoplasmic reticulum-resident stress protein gp96 binds peptides translocated by TAP. *Eur.J. Immunol.* **27,** 923-927.

49. **Landry,** S.J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C, and Gierasch, L.M. (1993). Characterization of a functionally important mobile domain of GroES. *Nature* **364,** 255-258.

50. **Langer,** T., Pfeifer, G., Martin, J., **Baumeister,** W., and **Hartl,** F.U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein within its central cavity. *EMBO J.* **11,** 4757-4765.

51. Lewthwaite, J.C., Coates, A.R.M., **Tormay,** P., Singh, M., Mascagni, P., **Poole,** S., Roberts, M., Sharp, L., and Henderson, B. (2001). *Mycobacterium tuberculosis* Chaperonin60.1 is a more potent cytokine stimulator man Chaperonin60.2 (Hsp65) and contains a CD14-binding domain. *Infect. immun.* 69, 7349-7355.

52. Lin, S., Schwarz, F.P., and Eisenstein, E. (1995). The hydrophobic nature of GroEL-substrate binding./. *Biol. Chem.* **270,** 1011-1014.

53. **Llorca, O.,** Perez, J., Carrascosa, J.L., **Galan,** A., and Muga, A. (1997). Effects of inter-ring communication in GroEL structural and functional asymmetry. /. *Bio/. Chem.* **272,** 32925-32932.

54. Lowrie, D.B., **Silva,** C.L., Colston, M.J., Ragno, S. and Tascon, R.E. (1997). Protection against tuberculosis by a plasmid DNA vaccine. *Vaccine* 15, 834-838.

55. Lukacs, K.V., Lowrie, D.B., Stokes, **R.W.,** and Colston, M.J. (1993). Tumor cells transfected with a bacterial heat-shock gene lose tumorigenicity and induce protection against tumors. *J. Exp. Med.* **178,** 343-348.

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

57. Martin, J., and Hard, F.U. (1997). Chaperone-assisted protein folding. *Curr. Opin. Struct. Biol.* 7, 41-52.

58. Matthews, R., and Burnie, J. (1992). **The** role of Hsp90 in fungal infection. *Immunol. Today* **133,** 345-348.

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

60. Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C.J. (1991). Characterization of the *groEL*-like genes in Streptomyces albus. /. *Bacteriol.* **173,** 7382-7386.

61. Meghji, S., White, P.A., Nair, S.P., Reddi, K., Heron, K., Henderson, B., Zaliani, A., Fossati, G., Mascagni, P., Hunt, J.F., Roberts, M.M., and Coates, A.R.M. (1997). *Mycobacterium tuberculosis* chaperonin 10 stimulates bone resorption: a potential contributory factor in Pott's disease./. *Exp. Med.* **186,** 1241-1246.

62. Mehra, V., Bloom, B.R., Torigian, V.K., Mandich, D., Reichel, M., Young, S.M., Salgame, P., Convit, J., Hunter, S.W., McNeil, M., *et al.* (1989). Characterization of *Mycobacterium leprae* cell wall associated proteins with the use of T lymphocyte clones./. *Immunol.* **142,** 2873-2878.

63. Mehra, V., Bloom, B.R., and Bajardi, A.C. (1992). A major t cell antigen of *Mycobacterium leprae* is a 10kDa heat shock cognate protein./. *Exp. Med.* **175,** 275-284.

64. Melnick, J., and Argon, Y. (1995). Molecular chaperones and the biosynthesis of antigen receptors. *Immunol. Today* **16,** 243-250.

65. Morimoto, R.I., and Milarski, K.L. (1990). Expression and function of vertebrate Hsp70 genes, In R.I. Morimoto, A. Tissieres, C. Georgopoulos (ed.), Stress proteins in biology and medicine. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY., 323-359.

66. Mustafa, A.S., Lundin, K.E., and Oftung, F. (1993). Human T cells recognize mycobacterial heat shock proteins in the context of multiple HLA-DR molecules: studies with healthy subjects vaccinated with *Mycobacterium bovis* BCG and *Mycobacterium leprae. Infect. Immun.* **61,** 5294-5301.

67. Mustafa, A.S., Lundin, K.A., Meloen, R.H., Shinnick, T.M., and Oftung, F. (1999). Identification of promiscuous epitopes from the mycobacterial 65-kilodalton heat shock protein recognized by human CD4+ T cells of the *Mycobacterium leprae* memory repertoire. *Infect. Immun.* **67,** 5683-5689.

68. Noll, A., and I. B. Autenrieth. (1996). Immunity against *Yersinia enterocolitica* by vaccination with Yersinia Hsp60 immunostimulating complexes or Yersinia Hsp60 plus interleukin-12. *Infect. Immun.* 64, 2955-2961.

69. Oftung, F., Geluk, A., Lundin, K.E., Meloen, R.H., Thole, J.E., Mustafa, A.S., and Ottenhoff, T.H. (1994). Mapping of multiple HLA class II-restricted T-cell epitopes of the mycobacterial 70-kilodalton heat shock protein *Infect. Immun.* 62, 5411-5418.

70. Ohashi, K., Burkart, V., Flohe, S., and Kolb, H. (2000). Cutting edge: heat shock protein 60 is a putative endogenous ligand of the toll-like receptor-4 complex *J Immunol.* **164,** 558-561.

71. Ortmann, B., Androlewicz, M., and Cresswell, P. (1994). MHC class I/β_2 microglobulin complexes associate with the TAP *transporter* before peptide binding. *Nature* **368,** 864-867.

72. Pockley, A.G. (2001). Heat shock proteins in health and disease: therapeutic targets or therapeutic agents? *Exp. Rev. Mol. Med.* 3, 1-21.

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

74. Richardson, A., Schwager, F., Landry, S.J., and Georgopoulos, C. (2001). The importance of a mobile loop in regulating chaperonin/ co-chaperonin interaction: humans versus *Escherichia coli.* *J Biol Chem.* **276,** 4981-4987.

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

76. Roberts, M.M., Coker, A.R., Fossati, G., Mascagni, P., **Coates,** A.R.M., and Wood, S.P. (2003). *Mycobacterium tuberculosis* chaperonin 10 heptamers self-associate through their biologically active loops. *JBacteriol.* **185,** 4172-4185.

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

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

79. Saibil, **H.,** Dong, Z., Wood, S., and auf der Mauer, A. (1991). Binding of chaperonins. *Nature* **353,** 25-26.

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

81. Shinnick, T.M. (1991). Heat shock proteins as antigens of bacterial and parasitic pathogens. *Curr. Top. Microbiol. Immunol.* **167,** 145-160.

82. **Silva,** C.L., **Silva,** M.F., Pietro, R.C., and Lowrie, D.B. (1994). Protection against tuberculosis by passive *transfer* with T-cell clones recognizing mycobacterial heat shock protein 65. *Immunology* **83,** 341-346.

83. Srivastava, P.K., Udono, H., Blachere, N.E., and **Li,** Z. (1994). Heat shock proteins *transfer* peptides during antigen processing and CTL priming. *Immunogenetics* 39, 93-98.

84. Srivastava, P. (2002). Interaction of heat shock proteins with peptides and antigen presenting cells: chaperoning of the innate and adaptive immune responses. *Annu Rev Immunol.* 20, 395-425.

85. Suzue, **K.,** and Young, R.A.. (1996). Heat shock proteins as immunological carriers and vaccines. *EXS* **77,** 451-465.

86. Taneja, B., and Mande, **S.C.** (2001a). Three-dimensional structure of *Mycobacterium tuberculosis* chaperonin-10 reveals a partially stable conformation of its mobile loop. *Curr. Sri.* **81,** 87-91.

87. Taneja, **B.**, and Mande, **S.C.** (2001b). Metal ions modulate the plastic nature of *Mycobacterium tuberculosis* chaperonin-10. *Protein Eng.* 14, 391-395.

88. Taneja, B., and Mande, S.C. (2002). Structure of *Mycobacterium tuberculosis* chaperonin-10 at 3.5A resolution. *Acta. Crystall.* **D58,** 260.

89. Tascon, R.E., Colston, M.J., Ragno, S., Stavropoulos, E., Gregory, D., and Lowrie, D.B. (1996). Vaccination against tuberculosis by DNA injection. *Nat Med. 2,* 888-892.

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

91. Trent, J.D., Nimmesgern, E., Wall, J.S., Hartl, F.U., and Horwich, A.L. (1991). A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354,** 490-493.

92. van Eden, W., van der Zee, R., Paul, A.G., Prakken, B.J., Wendling, U., Anderton, S.M., and Wauben, M.H. (1998). Do heat shock proteins control the balance of T-cell regulation in inflammatory diseases? *Immunol Today.* 19, 303-307.

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

94. Wang, J., and Boisvert, D.C. (2003). Structural Basis for GroEL-Assisted Protein Folding from the Crystal Structure of (GroEL-KMgATP)₁₄ at 2.0A resolution/. *Mol. Biol.* **327,** 843-855.

95. Weissman, J.S., Hohl, C.M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H.R., Fenton, W.A., and Horwich, A.L. (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell* **3,** 577-587.

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

97. Williams, **D.B.,** and Watts, T.H. (1995). Molecular chaperones in antigen processing. *Curr. Opin. Immunol.* 7, 77-84.

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

99. Young, D.B., Lathringa, R.B., Hendrix, R.W., **Sweetser,** D., and Young, R.A. (1988). Stress proteins are immune targets in leprosy and tuberculosis. *Proc. Natl. Acad. Sci. USA* 85, 4267-4270.

100. Young, R.A., and **Elliot,** T.J. (1989). Stress proteins, infection, and immune surveillance. G?//59, 5-8.

101. Young, R.A. (1990). Stress proteins and immunology. *Annu. Rev. Immunol.* 8, 401-420.

102. Young, D.B., and **Garbe,** T.R. (1991). Heat shock proteins and antigens of *Mycobacterium tuberculosis. Infect. Immun.* 59, 3086-3093.

103. Young, D.B., **Kaufmann,** S.H., Hermans, P.W., and Thole, J.E. (1992). Mycobacterial protein antigens: a compilation. *Mol. Microbiol.* **6,** 133-145.

104. Yifrach, O., and Horovitz, A. (1995). Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry* 34, 5303-5308.

105. Yifrach, O., and Horovitz, A. (1996). Allosteric control by ATP of non-folded protein binding to GroEL. *J. Mol. Biol.* 255, 356-361.

106. **Zügel,** U., and Kaufmann, S.H.E. (1999). Role of heat shock proteins in protection from and pathogenesis of infectious diseases. *Clin. Microbiol. Reviews* 12, 19-39.

Chapter 2

Cloning, Overexpression and Purification of

Mycobacterium tuberculosis Chaperonin60s

2.1   INTRODUCTION

Complete understanding of biological functions that proteins play *in vivo* involves extensive *in vitro* characterization of these macromolecules. Study of the biochemical and biophysical properties in combination with information on the structural organization, as judged by X-ray crystallography, gives us an opportunity to explore the mechanism of action of proteins.

Availability of proteins in their pure form and in sufficiently large quantities becomes a prime requirement for accomplishment of the different *in vitro* studies. Recombinant DNA technology has made it possible to obtain large amounts of any desired protein. Cloning of the desired gene in a suitable expression vector and its overexpression in a bacterial or eukaryotic host allows the production of the protein of interest. Moreover, the current protein expression methods have been highly effective in production of recombinant proteins, thus making possible their availability in large quantities for the various studies intended.

Expression vectors often allow introduction of different affinity tags to recombinant proteins thus simplifying and speeding up the process of protein purification. Among the different affinity tags available are the maltose binding protein, glutathione-S-transferase or a small 6-Histidine residue tag. Recombinant proteins expressed from such expression vectors are purified as fusion proteins over affinity chromatography columns, properties of which are based on the tag utilized. Cleavage of the affinity tag from the purified fusion proteins yields the pure proteins of interest in large amounts.

Advances in recombinant DNA technology further offer the study of specific amino acid residues of a protein. Site directed mutagenesis of DNA at the desired location results in a gene product altered at a single amino acid residue. The altered protein thus obtained can then be subjected to different comparative studies along with the native protein.

M. tuberculosis Chaperonin60s were overexpressed as Histidine-tagged ((His)6-tag) fusions from the *E. coli* expression vectors. The purified proteins hence obtained were used for the various biochemical and structural studies undertaken.

2.2 MATERIALS

Chemicals and reagents used for the cloning, purification, enzymatic characterization and crystallization of *M. tuberculosis* Cpn60s were obtained from several commercial sources.

2.2.1 Reagents used in DNA Cloning

Oligonucleotides were procured from Biobasic, Inc, Microsynth, Integrated DNA Technologies, Inc. and MWG Oligos. The Quick Change site directed mutagenesis kit was procured from Stratagene. Different DNA purification kits were procured from Qiagen (Chatsworth CA).

Deep vent polymerase, T4 DNA ligase, dNTPs, restriction endonucleases and protein and DNA gel markers were purchased from New England Biolabs (Beverly, MA).

2.2.2 Reagents used for Protein Purification and Characterization

Chemicals including ATP, ammonium molybdate, polyvinyl alcohol, citrate synthase, Tris, NaCl, EDTA, gel filtration chromatography markers, reagents for crystallization and others various chemicals were procured from either Sigma Inc. and Amersham Pharmacia. Ni-NTA-agarose was purchased from Qiagen. Q-sepharose and pre-packed FPLC columns were from Amersham Pharmacia.

Spectroscopic studies were done using the UV-Vis Perkin Elmer spectrophotometer. *E. coli* gene pulser used for electroporation was from Biorad. DNA sequencing was performed by ABI Prism automated DNA sequencer.

2.2.3 Recipes for Reagents used

All media and buffers used in protein expression and purification were prepared in deionized double distilled water (DDW) with low conductivity.

Compositions of the media and antibiotics used are as in Tables 2.1 and 2.2. Unless otherwise mentioned, all solutions were prepared by standard procedures as described in Sambrook *et al.* (1989).

Table 2.1 Composition of Culture media

Medium	Composition
Luria-Bertani (LB)	10g bactotryptone + 5g yeast extract + 10g NaCl per lt in DDW. pH was adjusted to 7.2 with NaOH. The medium was sterilized by autoclaving.
Terrific Broth (TB)	A. 12g bactotryptone + 24g yeast extract + 4ml glycerol in 900ml DDW B. 0.17M KH_2PO_4 + 0.72M K_2HPO_4 A and B were autoclaved separately and mixed later
5X M9 Minimal	30g Na_2HPO_4 + 15g KH_2PO_4 + 5gNH_4Cl + 1.5g NaCl per lt in DDW. The medium was sterilized by autoclaving.
1X M9 Minimal	200ml 5X Minimal medium + 1ml of 1M $MgSO_4$ + 10ml of 40% Glucose + 100µl 0.5% thiamine + 1ml of 4.2g/lt Fe(II)SO_4. The final volume was adjusted to 1lt with DDW. All reagents were sterilized separately before mixing

Table 2.2 Composition of antibiotics used

Antibiotic	Stock solution	Working concentration
Ampicillin	100mg/ml in DDW	100µg/ml
Kanamycin	30mg/ml in DDW	30µg/ml
Chloramphenicol	34mg/ml in ethanol	12.5|ig/ml
Tetracycline	10mg/ml in ethanol	10µg/ml

Antibiotics prepared in double distilled water were sterilized by passing through a 0.22µm filter.

Protein, expression and purification were assessed on 10% SDS-Polyacrylamide gels with elecrophoresis (PAGE) at a constant current of 30mA for 1hour (Laemmli, 1970) (Table 2.3).

Table 2.3 Composition of solutions used for SDS-PAGE

Reagents	Composition
30% acrylamide	29.2% acrylamide + 0.8% bis-acrylamide
4X separating gel	1.5M Tris-Cl, pH 8.8 + 0.4% SDS
4X stacking gel buffer	0.5M Tris-Cl, pH 6.8 + 0.4% SDS
1X Laemmli sample buffer	10% glycerol + 1% 2-mercaptoethanol + 2% SDS + 0.1% bromophenol blue in 1X separating gel buffer
1X Running buffer	3g Tris-Cl + 14.4 g glycine + 1g SDS per litre
Destaining solution	Methanol : Acetic acid : Water :: 5:1:4
Staining solution	0.1g/l of Coomassie brilliant blue R in destaining solution

Purity of the different plasmid vectors used and constructed was assessed by agarose gel electrophoresis on 0.8% agarose gels (Table 2.4).

Table 2.4 Composition of solutions used for agarose gel electrophoresis

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

2.2.4 Vectors and Bacterial Strains

The cloning vector, pBluescript SK (+) was purchased from Stratagene. Expression plasmids, pET3a, pET23d, pET28a and *E. coli* BL21 (DE3) were purchased from Novagen. *E. coli* DH5(X strain was from Promega.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Cloning of *M. tuberculosis cpn60s*

2.3.1.1 *Cloning of cpn60.1*

The gene coding for the *M. tuberculosis cpn60.1* (Rv3417c) was PCR amplified and cloned into the *E. coli* expression vector through an intermediate step of sub-cloning in pBluescript (SK+).

cpn60.1 was previously cloned in our laboratory in the expression vector, pKK223-2 and designated pKKGLl. In order to introduce a (His)6-tag in the gene product *cpn60.1* was amplified using the plasmid pKKGLl as the template and the primers MTGRL1H6N.FOR and MTGRL1H6N.REV (Table 2.5). These primers introduced the NdeI and BamHI restriction sites into the amplified product at the 5' and 3' end, respectively. The optimized conditions for PCR amplification are listed in Tables 2.6 and 2.7.

Table 2.5 Primers used for cloning of *cpn60s* Restriction endonuclease (RE) sites are highlighted as underlined sequences. Insertion of the six histidine codons in primers is shown in italics.

Primer Name	RE site	Primer Sequence
MTGRL2.FORA	NcoI	5' TCATTACCATGGCCAAGACAATTG 3'
MTGRL2.REVD	HindIII	5' AACCGCAAGCTTTCAGAAATCCATG 3'
MTGRL2H6C.FOR	NcoI	5' TCATTACCATGGCCAAGACAATTGC 3'
MTGRL2H6C.REV	HindllI	5' A<u>AAAGCTT</u>CTA*GTGATGATGATGATGATG*AAAATCCA TACCACCCATGTCGCC 3'
MTGRL1H6N.FOR	NdeI	5' <u>TCATATG</u>*CACCACCACCACCACCAC*ATGAGCAAGCT GATCGAATACGACG 3'
MTGRL1H6N.REV	BamHI	5' TGGATCCTCAGTGCGCGTGCCCGTG 3'

Table 2.6 Optimized conditions for PCR amplification

Components	pET3aGL1.2	pETGL2	**pET28.3GL2**
Template	20ng	3ng	3.5ng
Forward Primer	200nM	10μM	200nM
Reverse Primer	200nM	10μM	200nM
dNTPs	200μM	1mM	200μM
MgSO4	4mM	3.5mM	3.5mM
Deep Vent DNA Polymerase	1.5units	1.5units	1.5units

Table 2.7 Optimized PCR cycling parameters

Cycle	pET3aGL1.2	pETGL2	pET28.3GL2
Denaturation	95°C, 5min	95°C, 5min	95°C, 5min
Denaturation / Extension / Annealing (30 cycles)	95°C, 5min	95°C, 5min	95°C, 5min
	55°C, 2min	50°C, 2min	50°C, 2min
	72°C, 3min	72°C, 3min	72°C, 3min
Extension	72°C, 10min	72°C, 10min	72°C, l0min

The amplified product was ligated at the SmaI site of pBluescript (SK+) and the transformants in *E. coli* DH5α selected on the basis of blue-white selection. From the positive clone, designated pBSGLl, *cpn60.1* was subcloned in the Amp[r] expression vector, pET3a using the restriction enzymes, NdeI and BamHI, and was designated pET3aGL1.2 (Fig. 2.1).
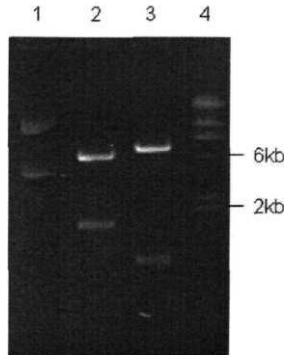
Figure 2.1 Agarose gel scan showing cloning of *cpn60.1* into the *E. coli* expression vector pET3a Lane 1 shows the undigested clone in pET3a corresponding to the supercoiled form of the plasmid. Lane 2: Digestion with NdeI and BamHI yielding the 1.6kb fall out of *cpn60.1*. Lane 3: Digestion with HindIII releases a 0.6kb fall out confirming the clone. Lane 4 shows the λ-HindIII DNA digest as marker.

2.3.1.2 *Cloning of cpn60.2*

The 1.6kb gene, *cpn60.2* (Rv0440), was amplified from the *M. tuberculosis* H37Rv cosmid, A10, kindly provided by Stewart Cole using the primers MTGRL2.FORA and MTGRL2.REVD (Table 2.5) (Cole *et al.*, 1998). The amplified gene incorporating the restriction sites, NcoI and Hindlll at the 5' and 3' ends was cloned into pBluescript (SK+) as for *cpn60.1* and designated pBSGL3. The gene was then subcloned into the Amp^r expression vector, pET23d and designated pETGL2.

pBSGL3 served as a template for re-amplification of *cpn60.2* in order to introduce a (His)$_6$-tag into the protein. MTGRL2H6C.FOR and MTGRL2H6C.REV served as primers for incorporation of the (His)6-tag (Table 2.5). Cloning of the amplified product into pBluescript (SK+) yielded the plasmid, pBSGL2HC. *cpn60.2* was subcloned into the Kan^r expression vector, pET28a, using the restriction

enzymes NcoI and Hindlll. The clone hence obtained was designated pET28.3GL2 (Fig. 2.2).

a   b
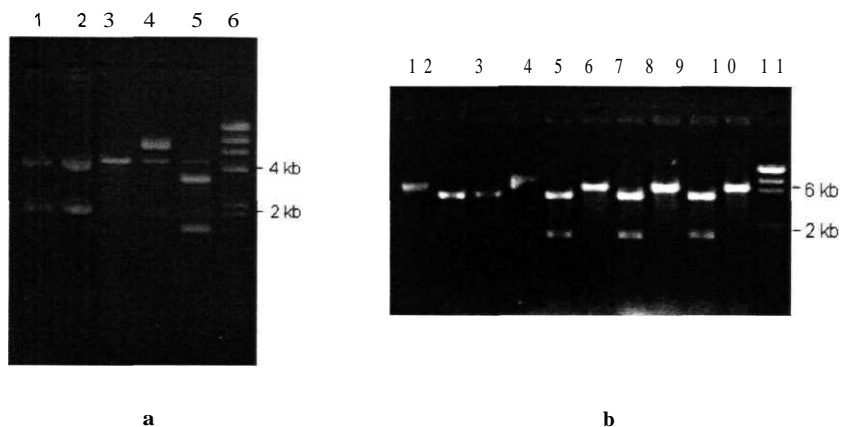
Figure 2.2 Agarose gel scans showing cloning of *cpn60.2* into *E. coli* expression vectors (a) Cloning into pET23d Lanes 1 and 2 show the uncut and Ncol/Hindlll digested control digestions of pET23d, respectively corresponding to the supercoiled form of the parent vector. Lane 3 is the SacI linearized pETGL2 clone showing the size of 5.2kb. Lane 4 is the control Ncol/Hindlll digestion of pETGL2 and lane 5 shows the 1.6kb fall out after the Ncol/Hindlll digestion. Lane 6 shows the λ-HindIII DNA digest as marker. **(b) Cloning into pET28a** Lanes 1 and 4 show the undigested vector and the positive clone, respectively. Lanes 5, 7 and 9 show the 1.6kb fall out of *cpn60.2* from the three positive clones upon Ncol/Hindlll digestion while the parent vector, pET28a, does not give this fall out (Lane 2). Lanes 3, 6, 8 and 10 correspond to the linearized vector and the three positive clones. Lane 11 has the λ-HindIII DNA digest as marker.

The different clones obtained were confirmed by automated DNA sequencing.

Dr. K. Ito, Kyoto University kindly provided the Tet^r plasmid pKY206 containing the *E. coli groEL* and *groES* genes.

2.3.2 Site Directed Mutagenesis

Point mutants of the *M. tuberculosis cpn60.2* and *E. colt groEL* were generated using the Quick-Change site directed mutagenesis kit as per the manufacturer's protocol.

Mutagenesis for *M. tuberculosis cpn60.2* was carried out in the plasmid pET28.3GL2 while that of *E. coli groEL* was carried out in pKY206. Two mutants of *E. coli groEL,* namely *E. colt* GroEL A2S and a double mutant, *E. coli* GroEL A2S/E76S were generated. The point mutant, *E. coli* GroEL A2S, was generated and used as a template for generation of the double mutant, *E. coli* GroEL A2S/E76S. The different primers used to achieve the mutants are described in Table 2.8. Mutants were confirmed by automated DNA sequencing.

Table 2.8 Primers used for mutagenesis of *cpn60.2* and *groEL* Codons altered for mutagenesis are highlighted as underlined text.

Name	Mutation	Sequence
GL2KS.FOR	Lys42→Ser	5^1 CGTCCTGGAAAAGTCGTGGGGTGC 3'
GL2KS.REV	Lys42→Ser	5' GCACCCCAC<u>GAC</u>TTTTCCAGGACG 3'
GL2KA.FOR	Lys42→Ala	5^1 CGTCCTGGAAAA<u>GGC</u>GTGGGGTGC 3'
GL2KA.REV	Lys42→Ala	5' GCACCCCAC<u>GCC</u>TTTTCCAGGACG 3'
ECGLA2S.FOR	Ala2→Ser	5' TAAAGATAATGGCA<u>TCT</u>AAAGACGTAAAATTC 3'
ECGLA2S.REV	Ala2→Ser	5' GAATTTTACGTCTTTAGATGCCATTATCTTTA 3'
ECGLE76S.FOR	Glu76→Ser	5' GCAGATGGTGAAATCAGTTGCCTCTAAA 3'
ECGLE76S.REV	Glu76→Ser	5' TTTAGAGGCAAC<u>TGA</u>TTTCACCATCTGC 3'

2.3.3 Protein Overexpression and Purification

2.3.3.1 *Overexpression and Purification of Cpn60.1*

The expression vector carrying *cpn60.1*, pET3aGL1.2 was transformed into *E. coli* BL21 (DE3) strain. The culture expressing *cpn60.1* in Terrific Broth (TB) was induced with 1mM IPTG at an O.D.$_{600}$ of 0.5. The induced cultures were grown at 37°C for 5-6hrs. Overexpressed Cpn60.1 fractionated into inclusion bodies upon cell lysis (Fig. 2.3a).

a   b
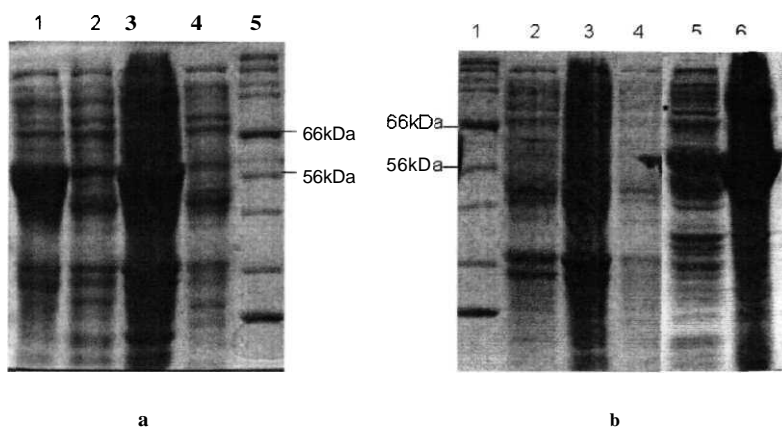
Figure 2.3 SDS-PAGE scans showing expression and solubilization of Cpn60.1 (a) Overexpression of Cpn60.1 at 37°C Lane 1 shows the overexpression of Cpn60.1 upon induction with 1mM IPTG. Lane 2 shows the uninduced culture as a control. Lanes 3 and 4 show the pellet and supernatant fractions obtained after sonication of the overexpressing culture, respectively. Presence of Cpn60.1 in the pellet suggests its existence in the insoluble form as inclusion bodies. Marker is shown in lane 5. **(b) Overexpression of Cpn60.1 attempted at 18°C and 25°C** Lanes 2, 3 and 4 show the induced culture, and the pellet and supernatant obtained after sonication when the cells were grown at 18°C. Lanes 5 and 6 show the supernatant and pellet when the culture expressing *cpn60.1* was induced at 25°C. Lane 1 shows the protein marker.

Attempts were made to promote solubilization by different growth conditions of cells harbouring the pET3aGL1.2 plasmid. Growth of cells expressing *cpn60.1* and

induction was carried out at 18°C and 25°C. While no expression was obtained at 18°C only partial solubilization of Cpn60.1 was attained at 25°C (Fig. 2.3b).

E. coli GroEL has earlier been shown to promote solubilization of several proteins (Cole, 1996; Caspers *et al.,* 1994; Thomas and Baneyx, 1996). Thus in a further attempt to promote protein solubilization, *cpn60.1* was co-expressed with *E. coli groEL.* This was achieved by co-transformation of *E. coli* BL21 (DE3) with plasmids pET3aGL1.2 and pKY206. While, *E. coli* GroEL was obtained in the soluble fractions, Cpn60.1 still was confined to the inclusion bodies (Fig. 2.4).
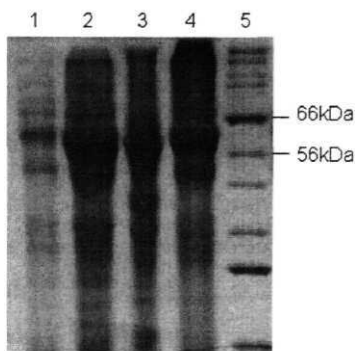
Figure 2.4 SDS-PAGE scan showing co-expression of Cpn60.1 and GroEL Lanes 1 and 2 show the uninduced and induced cultures of *E. coli* BL21 (DE3) co-expressing Cpn60.1 and GroEL. Lanes 3 and 4 show the pellet and supernatant fractions upon cell lysis. Lane 3 clearly indicates the existence of Cpn60.1 in the pellet.

The *E. coli* strain, C41 (DE3), a mutant of *E. coli* BL21 (DE3), has been shown to promote solubilization of proteins (Miroux and Walker, 1996). pET3aGL1.2 was thus transformed into this strain and tested for its effect on solubility. Fig. 2.5a shows the presence of the overexpressed protein in the pellet. Thus, Cpn60.1 could not be obtained in the soluble fraction in a strain recommended for solubilization of proteins.
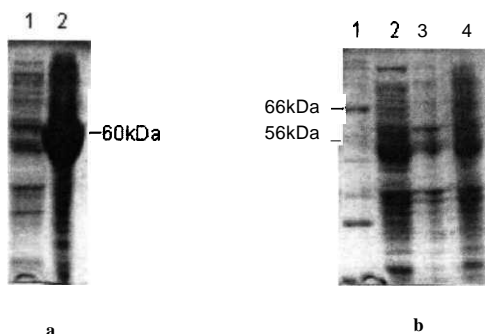
Figure 2.5 SDS-PAGE scans showing solubility trials performed for Cpn60.1 (a) Solubility in
***E. coli* C41 (DE3)** Lanes 1 and 2 show the supernatant and pellet fractions obtained upon lysis of
cells overexpressing the protein. Presence of Cpn60.1 in the pellet clearly shows that the protein
could not be obtained in the soluble fraction. **(b) Solubility in *E. coli* GJ1158** Lanes 2 and 4 show
the induced cultures obtained upon growth at 37°C and 30°C, respectively. Lane 3 shows the
uninduced culture at 37°C. Protein marker is shown in lane 1.

A salt inducible strain, *E. coli* GJ1158, kindly provided by Dr. J. Gowrishankar,
developed in his laboratory for increasing the solubility of proteins was tested for
solubilization of Cpn60.1. Cells expressing the gene, *cpn60.1,* were grown in LB
medium without sodium chloride (LBON) supplemented with ampicillin. Cells were
induced at an O.D.$_{600}$ of 0.8 with 0.3M NaCl at 30°C and 37°C. However, no
induction was obtained in either case (Fig. 2.5b).

Solubilization of Cpn60.1 was finally achieved by including sarcosyl in the lysis
buffer (50mM Tris-Cl, pH8.0, 150mM NaCl, 15mM imidazole) at a concentration of
10mM. Presence of sarcosyl resulted in a complete recovery of the overexpressed
Cpn60.1 in the soluble fraction (Fig. 2.6). The protein was purified using the Ni-NTA
affinity chromatography. *E. coli* BL21 (DE3) over-expressing *cpn60.1* were lysed by
sonication. The resuspension buffer, 50mM Tris-Cl, pH 8.0; 150mM NaCl; 15mM
imidazole, was supplemented with 10mM sarcosyl, and 1mM PMSF (phenyl methyl
sulfonyl fluoride). Upon centrifugation of the cell lysate the supernatant obtained was
loaded onto the Ni-NTA column equilibrated with the resuspension buffer. The
column was washed with 8-10 bed volumes of the same buffer. Protein was eluted

with 50mM Tris-Cl, pH 8.0; 150mM NaCl; 100mM imidazole. The purified protein was concentrated using the Amicon concentration assembly and dialysed against the appropriate buffers. This procedure was undertaken for all further purifications of Cpn60.1 used for the various assays and crystallization.
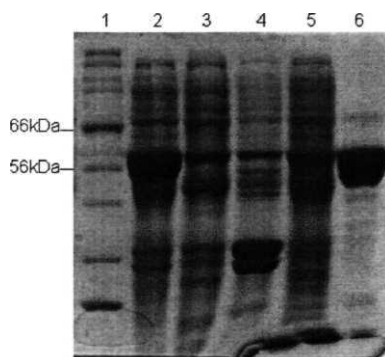
Figure 2.6 SDS-PAGE scan showing solubilization and purification of Cpn60.1 Solubilization of Cpn60.1 was achieved by sarcosyl used at concentrations much below its critical micellar concentration. Lanes 2 and 3 show the induced and uninduced cultures of *E. coli* BL21 (DE3) cells expressing the plasmid carrying the *cpn60.1* gene. Lanes 4 and 5 show the pellet and supernatant fractions of the overexpressing culture obtained. Sarcosyl clearly seems to promote the solubilization of Cpn60.1 as seen by its presence in the supernatant. Lane 6 shows the purified Cpn60.1. The protein marker is shown in lane 1.

Cpn60.1, without (His)6-tag was expressed from the plasmid, pKKGL1. Cell lysate was subjected to ammonium sulphate precipitation with 30% ammonium sulphate. The pellet containing Cpn60.1 was dialysed against 50mM Tris-Cl, pH8.0 supplemented with 1mM EDTA. The dialysed protein was loaded on the anion exchanger, Q-sepharose and washed with 50mM Tris-Cl, pH8.0 supplemented with 150mM NaCl and 1mM EDTA. Elution of Cpn60.1 was achieved by increasing the salt concentration to 300mM.

2.3.3.2 *Overexpression and Purification* of *Cpn60.2*

The expression vector carrying *cpn60.2*, pET28.3GL2, was transformed into *E. coli* BL21 (DE3). The culture expressing *cpn60.2* in Terrific Broth was induced with 0.2mM IPTG at an O.D.$_{600}$ of 0.5. Induction was allowed for 5-6hrs at 37°C. The protein was purified using Ni-NTA affinity chromatography as described for Cpn60.1 (Fig. 2.7).

a   **b**
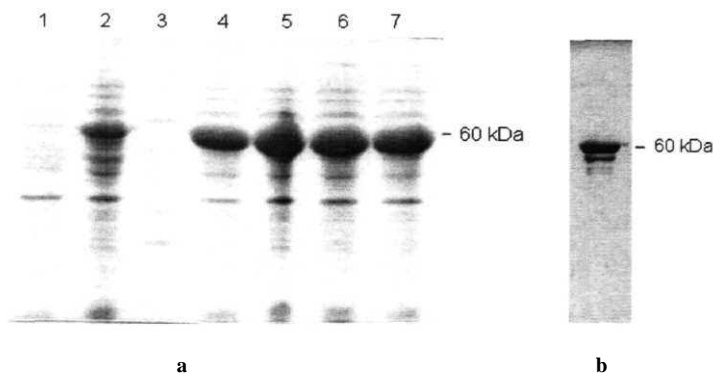
Figure 2.7 SDS-PAGE scans showing the overexpression profile and purification of Cpn60.2 **(a)** Lanes 1 and 2 show the uninduced and induced cultures of *E. coli* BL21 (DE3) harboring the plasmid pET28.3GL2. Lane 3 is the broad range marker from NEB. Lanes 4-7 show the cultures induced with 0.2mM, 0.4mM, 0.6mM and 0.8mM IPTG, respectively. (b) Purified Cpn60.2 after Ni-NTA chromatography. The lower band shows the degraded product of Cpn60.2 obtained upon protein purification.

As is shown in Fig. 2.7b Cpn60.2 invariably eluted as a doublet. N-terminal sequencing of the two bands showed that the higher molecular weight band corresponded to the intact protein, while the lower band was Cpn60.2 truncated at the N-terminal residue K41 (discussed in Chapter 5). In an attempt to avoid protein degradation K41A and K41S mutants of Cpn60.2 were generated. These mutants were overexpressed and purified as the wild-type recombinant Cpn60.2 using the Ni-NTA affinity chromatography. However, even the mutations K41S and K41A did not seem to prevent degradation of the protein. Hence, the protein used for all the

various biochemical assays and crystallizations was the wild-type Cpn60.2 and was a mix of the intact and truncated proteins.

Cpn60.2, without (His)6-tag was expressed from the plasmid, pETGL2. Purification was performed by salting out non-specific proteins with 15% ammonium sulphate followed by precipitation of Cpn60.2 with 30% ammonium sulphate. The pellet containing Cpn60.2 was dialyzed against 50mM Tris-Cl, pH 8.0 supplemented with 1mM EDTA. The dialyzed protein was loaded onto the anion-exchanger, Q-sepharose, and eluted with 50mM Tris-Cl, pH 8.0, 150mM NaCl, 1mM EDTA.

2.3.3.3 *Purification of E. colt GroEL and GroES*

Overexpression of *E. coli* GroEL and GroES was achieved by transformation of pKY206 in *E. coli* BL21 (DE3) and induction at 37°C with 0.3mM IPTG. Purification was essentially done as described earlier (Clarke *et al.,* 1998). The same protocol was followed for purification of the mutant, *E. coli* GroEL A2S/E76S.

2.3.4 Quaternary Structure Determination of Cpn60s

Oligomeric state of the Cpn60s was tested by size exclusion chromatography. Size exclusion chromatography was performed at room temperature using the FPLC system (Pharmacia Amersham) equipped with Superdex-200 HR 10/30. The column was equilibrated with at least 3-bed volumes of 50mM Tris-Cl (pH 8.0) and 150mM NaCl prior to each run. A typical flow rate of 0.5ml/min was maintained. Absorbance at 280nm was measured to monitor elution of the protein from the column.

2.3.5   Generation of Multiple Sequence Alignment

Homologous protein sequences of *M. tuberculosis* Cpn60s were obtained by PSI-BLAST (Altschul *et al.,* 1997). The multiple sequence alignment was generated using the program ClustalW (Thompson *et al,* 1994). The alignment showed

conservation of most of the residues among the 43 sequences with a few remarkable differences in *M. tuberculosis* Cpn60.1 (Fig. 2.8). In an attempt to understand the reasons for the non-canonical behavior of Cpn60.1 (discussed in Chapter 4) mutants of *E. coli* GroEL were generated at residue positions 2 and 76.

Figure 2.8 Multiple Sequence Alignment of the various Cpn60 sequences Residue positions, at the protein interface, with a very high conservation in all Cpn60 sequences, but different in *M. tuberculosis* Cpn60.1 have been highlighted. These positions correspond to the residue numbers — 2, 11, 76, 102, 117, 388, 445 and 452 in *E. coli* GroEL.

```
CH60_BACHD   ERLKKE--EAGFGFN-AATGENVNMVEAGIVDPTKVTRSALQHAASVSAMFLTTEAVIADKPEENE-GGG-GMPDMGGMGGMGGMM--------------
CH60_BACSU   ERLKNE--EIGVGFN-AATGENVNMIEAGIVDPTKVTRSALQNAASVAAMFLTTEAVVADKPEENG-GGA-GMPDMGGMGGMGGMM--------------
CH60_LISIN   ERLKHE--AVGVGFN-AANGENVNMIDAGIVDPTKVTRSALQNASSVAALLLTTEAVVADQPDEN--GPA-AGPDMG-MGGMGGMM--------------
CH60_LISMO   ERLKHE--AVGVGFN-AANGENVNMIDAGIVDPTKVTRSALQNASSVAALLLTTEAVVADKPDEN--GPA-AVPDMG-MGGMGGMM--------------
CH60_BACP3   ERLKHE--KPGIGFN-AATGENVDMIEAGIVDPTKVTRSALQNAASVAAMVLTTEACVADKPEENK-GNN-HWPDMG-----GMM--------------
CH60_BACST   ERLKTE--KPGIGFN-AATGENVDMIEAGIVDPTKVTRSALQNAASVAAMFLTTEAVVADKPEENK-GGNPGMPDMG-----GMM--------------
CH60_STRBP   ERLKHA--EAGVGFN-AATHENVNMLEAGIVDPTKVTRSALQNAASVAAMFLTTEAVVASIPEPEN-NEG---POMG---GMPGMM-------------
CH60_LACLA   DKLRSE--EVGTGFN-AATGQNVNMIEEGIVDPAKVTRSALQNAASVAGLILTTEAVVANKPEPA---APAMPP-MDPSMGMGGMM------------
CH60_STRPY   DKLKNS--PAGTGFN-AATGENVDMIKTGIIDPVKVTRSALQNAASVASLILTTEAVVANKPEPAT-PAPAMPAGMDPGM-MGGF-------------
CH60_CLOAB   EKIINS--EKGIGFD-ALHEKYVDMLSVGIVDPTKVTRSALQNAASVASTFLTTECAVADIPEKDKPEMPGGAPGMG----MGGMY-----------
CH60_CLOPE   EKVKNS---DAGVGFD-ALMKAGIVDPTKVTRSALQNAASVASTFLTTEAAVADIPEK---EMPQGA-GMG----MGGMY-----------
CH60_CLOTM   EKVKAS---EPGIGFD-AYNEKYVNMIEAGIVDPAKVTRSALQNAASVASMVLTTESVVADIPEK---EISGGPGGLG----MGGMY--------
CH60_THEBR   EKIKAAK-DPNFGYD-AYKEEFIDMFKAGIVDPTKVTRSALQNAASIASMILTTEAIVVDIPEK---NTGMPNPGLG------MIMM-----------
CH60_CHLMU   QQVMSRS-AN-EGYD-ALRDAYIDMIEAGILDPAKVTRSALESAASVAGLLLTTEALIAKIPEEKPAAAPA-----NPGAGMDT-----------
CH60_CHLTR   QQVMSRS-AN-EGYD-ALRDAYIDMIEAGILDPAKVTRSALESAASVAGLLLTTEALIAKIPEEKPAAAPA-----NPGAGMDT-----------
CH60_CHLPN   QQVLSRS-AN-EGYD-ALRDAYIDMIDAGILDPTKVTRSALQNAASVAGLLLTTEALIADIPEEKSSSAPA-----NPGAGMDT-----------
CH60_CHLPS   QQVLSRS-SS-EGYD-ALRDAYIDMIEAGILDPTKVTRCALESAASVAGLLLTTEALIADIPEEKSSSAPA-----NPGAGMDT-----------
CH60_PARDE   GKVRESS-DKAPGFN-AQTEHYGDMFKPGVIDPAKVVRTALEDAASVASLLITTEAMIAEKPEPKA-PAGG----NPDMGGMGGMM---------
CH61_RHOSH   GKVRESN-EKSPGFN-AQTEHYGDMFKPGVIDPAKVVRTALEDAASVASLLITTEAMIADKPEPKS-PAGGPG--MGGMGGMGGMM--------
CH60_RHOCA   GKVRESA-DPAPGFN-AQTEHYGDMFGPGVIDPAKVVRTALEDAASIAGLLITTECMIAMKPEPKGAPAGG----MGGMGGMDMM--------
CH61_BRAJA   GKLLENS-STHNGFN-AASGEYGDLAKAGVIDPAKVVRTALQDAASVAALLITTEALIAEKP-KYSKPAPA----APP---MDF----------
CH65_RHIME   GKLRECT-EPSYGNN-AQTNEYGDLTAMGVIDPAKVVRSGLQDAASVAGLLVTTEAVIAEKG-KK-RAAPA----LPAGGGMDF---------
CH60_ECOLI   MTVVK-GG-DGNYGYNMAATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGMGGMGGMGGMM---M---
CH60_LEPIN   HQIKTEK-KG-LGYD-ASSFKNVHMIESGIIDPAKVTRSALQNAASIAGLLTTEGAIIDIKEENTSGGGG---YPMDPGMGMM-----------
CH60_THEMA   EHAKAKK-GN-EGFN-ALDMVHEDMIQAGVVDPTKVTRSALQNAASIGSMILTTEVTIIDKPDKDAPNPHAGMG-GGGMGGMGGMM--------
CH60_THETH   QQILAETKNPRYGFN-AATGEFVDMVEAGIVDPAKVTRSALQNAASIGALILTTEAVVANGEKEKESTPASLG----AGDMDF----------
CH61_SYNVU   ERVKEKP--FDVGYD-AAKDEYVNMFDAGIVDPAKVTRSALQNAASIAGMVLTTECIIVDKPEPKENNPAGSGLGMGGDFDY----------
CH61_SYNY3   ERVKEKE--FNVGYN-AASLEYVDMLAAGIVDPAKVTRSALQNAASIAGMVLTTECIIVDKPB-KEKAPAG---APGGDFDT----------
CH60_CYAPA   EQVKDKP--FNIGYN-AANAQFVDMFEAGIVDPAKVTRSGLQNAASIAGMVLTTECIIVDKAESQKKAASVN----RNSFDY---------
CH65_RHIME   DEIKHLD--PSIGYD-ASTSKFVNMYESGIIDPAKVTRSALQNASSIAGMILTTECLVVDEMNRHMEVEK--------------
CH60_PORPU   EKVKTND--PCIGYD-ASNGKIVNMYDAGIIDPAKVARSALQNATSIAAMVLTTECIVVDKLEA--------------
CH62_SYNVU   ENVRAGD--FNYGFN-AATGAYEDLVSAGIIDPAKVVRSALQNAGSIAGMVLTTEALVVEKPEPK-PAAPANGGMGGMG-----GMM------
CH62_SYNY3   EKVKEAT--GNQGYN-VITGKIEDLIAAGIIDPAKVVRSALQNASIAGMVLTTEALVVEKPEPAAPAMPDMGGMGGMGGMGGMGMM------
CH62_MYCLE   EKVRNLS--VGHGLN-AATGEYEDLLKAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKTAAPASDPTGGMGGMDF----------
CH62_MYCTU   EKVRNLP--AGHGLN-AQTGVYEDLLAAGIIDPAKVVRSALQNAASIAGLFLTTEAVVADKPEKEKASVPG-GGDMGGMDF---------
CH60_MYCPA   EKVRNSP--AGTGLN-AATGEYEDLLKAGIADPVKVTRSALQNAASTAGLFLTTEAVVADKPEKAAPAGDPTGGMGGMDF---------
CH62_STRAL   EKVRNLS--VGHGLN-AATGQYVDMIABGILDPAKVVRSALQNAASIAALFLTTEAVIADKPEKAAAAPG-GMPGGDMDF---------
CH61_STRLI   SKVADLD--KGQGFN-AATGEYGDLVKAGVIDPVKVTRSALENPASIASLLLTTEVLVVEKKEEHEPAAGG-HSHLGHSH----------
CH61_STRCO   SKVADLD--KGQGFN-AATGEYGDLVKAGVIDPVKVTRSALENASSIASLLLTTEVLVVEKKEEHEPAAGG-HSH-GHSH---------
CH61_STRAL   TKVAELD--KGQGFN-AATGEYGDLVKAGVIDPVKVTRSALENASIASLLLTTEVLVVEKKAEEHEPAAG--HGH-GHSH---------
CH61_MYCLE   DKVSGLP--AGHGLN-ASTLGYGDLVADGVVDPVKVTRSAVLNAASVARMMLTTETAVVDDKPAKTEEHHD----HHGHAH---------
CH61_MYCTU   NKVSELP--AGHGLN-VNTLSYGDLAADGVIDPVKVTRSAVLNASSVARMMLTTETVVVDKPAKAEDHDH----HHGHAH---------
                  * :.         ::   *:.**,**.*,:  ,*:.  ,:**** :,.
```

2.4   CONCLUSIONS

The *M. tuberculosis* genes, *cpn60.1* and *cpn60.2,* were cloned into suitable expression vectors and the resultant proteins purified over the Ni-NTA affinity column. Size exclusion chromatography of Cpn60.1 and Cpn60.2 established the oligomeric state of the proteins.

Sequence comparison of the *M. tuberculosis* Cpn60s with the other known homologues identified a few natural amino acid occurrences in Cpn60.1 that were otherwise conserved among all the Cpn60s. Mutagenesis of these residues in *E. coli* was carried out in order to assess their importance to the Cpn60.1 structure.

Exhaustive attempts to solubilize Cpn60.1 were made by expression of the protein in the presence of different Cpn60s and other known expression strains. Solubility was finally achieved by inclusion of sarcosyl during cell lysis. Both, Cpn60.1 and Cpn60.2 were purified to ~98% homogeneity.

2.5 REFERENCES

1. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc. Acids Res.* 25, 3389-3402.

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

3. Cole, P.A. (1996). Chaperone-assisted protein expression. *Structure* 4, 239-242.

4. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C, Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, *et al.* (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393,** 537-544.

5. Caspers, P., Stieger, M., and Burn, P. (1994). Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli. Cell. Mol. Biol.* 40, 635-644.

6. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

7. Miroux, B., and Walker, J.E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260,** 289-298.

8. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *In Molecular Cloning: A laboratory manual.* (Cold Spring Harbor Laboratory Press, New York)

9. Thomas, J.G., and Baneyx, F. (1996). Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing Heat-shock proteins. *J. Biol. Chem.* **271,** 11141-11147.

10. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc. Acids Res.* **22,** 4673-4680.

Chapter 3

Crystallization, Data Collection and

Structure Solution of Chaperonin60.2

3.1   INTRODUCTION

Proteins play a large variety of important roles in the maintenance of life. Structures and dynamics of these **macromolecules** provide information about their molecular function. Crystallography is a tool that can be used to determine the three-dimensional structure of proteins - an information that provides valuable insights into the specific biological roles that these molecules play.

The first step towards X-ray structure determination is to obtain suitable single crystals of the protein. Crystal growth can be achieved by obtaining a right combination of different factors including protein concentration, ionic strength, temperature, organic solvents, pH, cofactors etc. (Blundell and Johnson, 1976). Crystal growth is followed by the measurement of intensity diffraction data.

The next major challenge in structural determination using X-ray crystallography is known as the "phase problem". The phase problem arises from the fact that the diffraction data contains information only about the amplitudes but not the phases of structure factors. Over the years many methods have been developed to deduce the phases for reflections, including direct methods, **isomorphous** replacement, multi-wavelength anomalous dispersion (MAD) and molecular replacement. The isomorphous replacement and MAD methods require collection of additional experimental data that includes diffraction data on heavy metal soaked derivative crystals, or collection of diffraction at different wavelengths on crystals containing anomalous scatterers.   The molecular replacement method utilizes homologous structures to provide a set of initial, approximate phases.

Initial structure information can be obtained by calculation of electron density by Fourier transforming the structure factor amplitudes along with their phases. For more reliable information and finer details, however, the structure must be refined. Refinement is the process of adjusting the model to find a closer agreement between

the calculated and observed structure factors. The various refinement techniques available are thus used to approach the final structure.

Structural studies of *M. tuberculosis* chaperonins combined with biochemical analysis would enable us to understand the multiple functional roles of these proteins. Structural analysis of *M. tuberculosis* Cpn60s was thus undertaken as a part of this study. A combination of techniques was attempted for solving of the crystal structure of Cpn60.2, which are described in this chapter.

3.2 CRYSTALLIZATION OF *Mycobacterium* tuberculosis CHAPERONIN60s

Preliminary crystallization trials with the *M. tuberculosis* Cpn60s were carried out with different arrays of random conditions. 96 random conditions of Magic-96 (Hol and Sarfaty, personal communication) were screened at 4°C and 25°C. In addition, the 96 random conditions of Straight-96 (Hol and Sarfaty, personal communication) were also explored at the two temperatures.

Prior to crystallization, the proteins were suspended in a low ionic strength buffer, 10mM Tris-Cl, pH 8.0. Crystallizations were set up using the hanging drop vapour diffusion method wherein 2μl of the protein was mixed with an equal volume of the reservoir solution. The plates were left undisturbed for 7 days for crystal growth and then observed under the microscope.

3.2.1 Crystallization of Cpn60.1

Crystallization of Cpn60.1 was attempted at protein concentrations of 6mg/ml and 15mg/ml. Initial trials resulted in extensive precipitation in a large number of conditions.

As part of the TB Structural Genomics Consortium, crystallization of Cpn60.1 was sought at their facility. Crystallization of Cpn60.1 at a concentration of 10mg/ml was attempted. While the protein precipitated in most of the conditions, crystals were obtained in the following condition (Fig. 3.1).

Tris-Cl, pH 8.5	100mM
Lauryl di-methylamine N-oxide (LDAO)	0.5%
PEG3350	14.5%
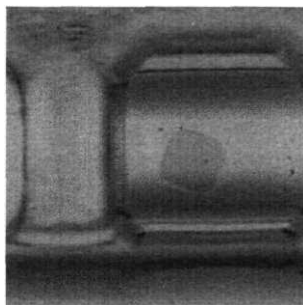
Figure 3.1 Crystal of Cpn60.1 obtained at the TB Structural Genomics Consortium Facility
Crystals appeared after a long incubation period of more than two months.

In an attempt to reproduce these crystals several crystallizations were set up with minor variations of this condition at 25°C. The conditions tested, however, resulted in precipitation of the protein. Further crystallization trials of Cpn60.1 are underway.

3.2.2   Crystallization of Cpn60.2

Crystallization of Cpn60.2 was attempted at a concentration of 10-25mg/ml. Preliminary trials did not yield any promising result with the protein getting precipitated in most of the conditions of Magic-96 and Straight-96.   However, promising crystals were obtained in the following conditions with 14mg/ml of Cpn60.2

HEPES, pH 7.5   100mM

PEG3350   25%

n-propanol   10%

Similar crystallization conditions were then screened varying pH, precipitant and co-solvent concentrations and the plates left undisturbed for 3-4 days. Crystals were obtained in 2 days that further grew to their optimal size in a period of 7 days

(Fig. 3.2). Similar crystallization conditions have also been reported by Adir *et al.* (2002).
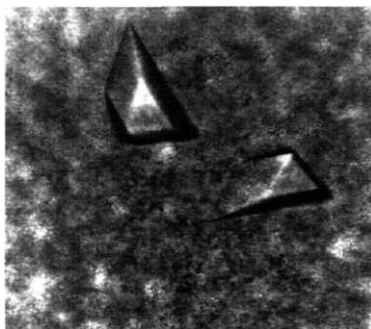
Figure 3.2 Crystals of *M. tuberculosis* Cpn60.2 Crystals were obtained in 25% PEG3350 and 10% n-propanol at pH 7.5. These crystals were used for data collection at the BESSY synchrotron.

The crystals of Cpn60.2 diffracted to a resolution of 4A when exposed to the in-house X-ray source. The resolution remarkably improved to 3.2A when exposed to the synchrotron X-ray beam at BESSY, Berlin.

3.3   X-RAY DIFFRACTION AND DATA COLLECTION OF Cpn60.2 CRYSTALS

3.3.1 Data Collection at Room Temperature

Crystals of Cpn60.2 grown as described in the previous section were mounted in glass capillaries and were exposed to X-rays. The intensity data were collected on a MAR345 image plate scanner with the monochromatic CuKα radiation from a Rigaku RU300 rotating anode generator operated at 40kV, 100mA. The crystals diffracted to a modest resolution of 4A.

3.3.2 Data Collection at Cryo Temperature

Collection of X-ray data at cryo temperature of 100K offers a number of benefits including minimization of radiation damage, an effective increase in resolution limit and reduction of background scatter and absorption (Rodgers, 1997). A large number of reagents have been suggested as cryoprotectants including ethylene glycol, glycerol, glucose, MPD, PEG400 etc.

Crystals of Cpn60.2 were soaked in the artificial mother liquor supplemented with PEG400 or glycerol as cryoprotectants. The minimum concentration of the cryoprotectant was determined by flash cooling the cryoprotective harvest solution. Visual assessment of the glassiness of the frozen solution was used to define the optimum concentration of the cryoprotective agent. Further, the stability of crystals in the cryoprotectant was assessed by their examination under the microscope.

In order to minimize possible non-isomorphism, crystals were allowed to equilibrate in the artificial mother liquor supplemented with 35% PEG400 for a period of 1-2 hours. The crystals diffracted to a resolution of 3.8A but with a high mosaic spread. Flash annealing of crystals has been shown to improve diffraction quality and reduce mosaicity of crystals (Yeh and Hol, 1998). Alternate freeze-thaw of

the Cpn60.2 crystals, however, resulted in a highly mosaic and poor diffraction. In an attempt to reduce crystal damage, gradual increase in the cryoprotectant concentration was performed. This was done by a serial transfer of crystals in increasing concentrations of PEG400 ranging from 5-35% in steps of 5%. This step-wise increase in concentration of the cryoprotectant improved the quality of diffraction with a much reduced mosaic spread. However, these crystals too diffracted only upto a maximum resolution of 3.8A.

Further, cryoprotection of Cpn60.2 crystals was attempted by a brief soak in artificial mother liquor supplemented with glycerol as the cryoprotectant. A short soak of a few seconds in 10% glycerol proved to be an ideal condition resulting in diffraction upto 3.2A resolution on a synchrotron source with a reasonable mosaicity (Fig. 3.3). X-ray diffraction data were thus collected using crystals soaked in 10% glycerol for a few seconds prior to data collection. Data were collected at the BESSY synchrotron using MAR CCD detector.
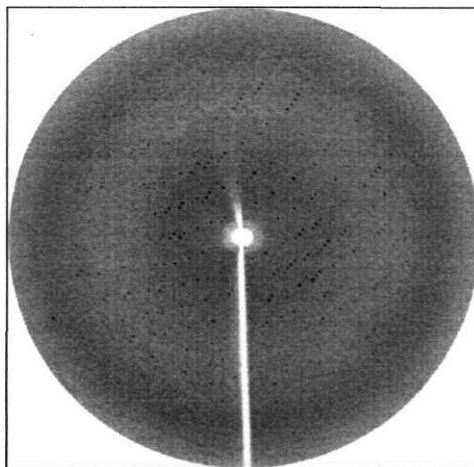
Figure 3.3 Diffraction pattern of one of the Cpn60.2 crystals covering 1° oscillation range The crystals diffract modestly upto a resolution of 3.2A.

3.3.3 Data Collection at Synchrotron

X-ray diffraction data for two Cpn60.2 crystals, Crystal 1 and Crystal 2, were collected at 100K using the fast scanning MAR CCD-detector of 165 mm diameter. The crystal to detector distance was maintained at 250mm for both the data sets. 1° oscillation frames were collected covering a range of 180° and 150° for the two crystals. The exposure time for each frame was 15 seconds. The data were processed, scaled and merged with DENZO and SCALEPACK (Otwinowski and Minor, 1997). When processed independently, useful data could be obtained upto a resolution of 3.5A for Crystal 1, and 3.1 A for Crystal 2. Although Crystal 2 diffracted to a higher resolution than Crystal 1, the data completeness was only about 90% through all resolution ranges (Table 3.1). Therefore, both the data sets were integrated upto 3.0A resolution and then scaled together. The merged and scaled data were found to be useful upto a resolution of 3.2A. Data collection and reduction statistics of the two crystals and the merged data set are listed in Table 3.1.

Table 3.1 Data collection statistics of native crystals of Cpn60.2 Numbers in parentheses represent statistics in the last resolution bin.

	Crystal 1	*Crystal 2*	*Merged Da ta*
Space group	P2$_1$	P2$_1$	P2$_1$
a (A)	58.7	58.8	58.7
b(A)	113.8	114.0	113.8
c(A)	79.5	80.0	79.5
β (˚)	94.6	94.6	94.6
Maximum resolution (A)	3.5	3.1	3.2
Matthews number	**2.4**	2.4	2.4
No. of unique reflections	14,325	17,404	17,489
Redundancy	3.3 (2.2)	2.1 (1.9)	5.0 (3.0)
Completeness (%)	97.5 (84.2)	90.9 (88.6)	99.5 (98.0)
Average I/σ(I)	15.7 (2.0)	29.3 (1.8)	20.2 (2.6)
R$_{merge}$ (%)	6.6 (36.1)	4.9 (29.6)	10.0 (41.5)
Range of the last resolution bin (A)	3.48-3.4	3.17-3.1	3.31-3.2
$R_{merge} = \Sigma \mid I_{hkl} - <I_{hkl}> \mid / \Sigma \mid I_{hkl} \mid$			

Inspection of (0 k 0) reflections showed very weak intensity values for k=odd reflections, thus suggesting that the correct space group was $P2_1$ (Table 3.2). The Matthews coefficient as calculated from unit cell parameters, suggested presence of two molecules in the crystal asymmetric unit.

Table 3.2 Intensities of (0 k 0), k=odd reflections, in the merged data set

h	k	l	Intensity	Sigma	I/Sigma
0	3	0	0.4	1.9	0.2
0	5	0	-2.2	4.7	-0.5
0	7	0	-1.8	2.4	-0.8
0	9	0	-6.9	3.4	-2
0	11	0	0.9	3.4	0.3
0	13	0	5.9	4.6	1.3
0	15	0	-5.7	3.6	-1.6
0	17	0	6	4.4	1.3
0	19	0	-6.4	7.8	-0.8
0	21	0	-10.4	7.9	-1.3
0	23	0	-10.1	6.8	-1.5
0	25	0	1.4	7	0.2
0	27	0	-22.8	12.8	-1.8
0	29	0	9.3	41.5	0.2
0	31	0	5.1	9.5	0.5
0	33	0	-2.8	9.6	-0.3
0	35	0	1.9	11.7	0.2

The possibility that the data could be twinned was ruled out by inspection of cumulative intensity distribution, as shown in Fig 3.4. The cumulative intensity distribution for both acentric and centric reflections matched well with the theoretical pattern. That the data were not twinned was further confirmed by 2nd and 4th moments tests.
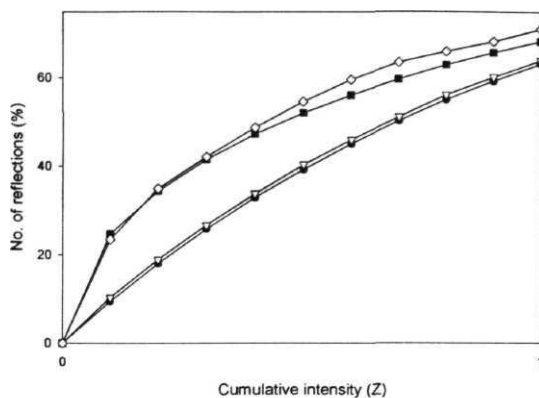
Figure 3.4 N-Z distribution of the centric and acentric reflections observed in the Cpn60.2 data The cumulative intensity is plotted against the number of reflections obtained for theoretical acentric (—•—), theoretical centric (—•-) and the observed centric (—0—) and acentric (—V—)

Wilson plot analysis showed a typical distribution of intensities as observed for proteins with a characteristic trough at 5.5A and a peak at 4.5A (Fig. 3.5). The overall B-factor was estimated to be $88A^2$, although admittedly this estimation is inaccurate due to the low resolution of the data.
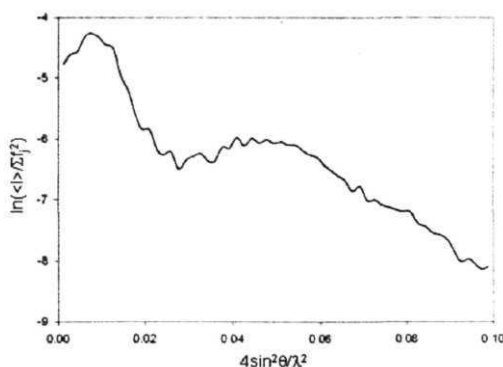
Figure 3.5 Wilson plot for Cpn60.2 with data to 3.2A resolution The fall off of intensity values is as expected for **protein crystals.**

The data quality therefore appeared to be reasonable with no evidence for twinning. All further work, including structure solution and refinement, was carried out with the data merged from the two crystals.

3.4   STRUCTURE DETERMINATION

One of the central difficulties in structure determination using X-ray crystallography is the "phase problem". Different techniques are available for solution of the phase problem. These include:

1. Isomorphous Replacement, requiring the attachment of heavy atoms to the protein molecules in the crystal.

2. Multiple wavelength Anomalous Dispersion (MAD), requiring the presence of anomalous scattering atoms within the protein structure. Incorporation of seleno-methionine in protein provides the necessary anomalous scatterer for determination of the correct phase angles.

3. Molecular Replacement, requiring the availability of the structure of a homologous protein.

Simultaneous attempts to solve the X-ray structure of Cpn60.2 by the above mentioned methods were made. Since Cpn60.2 bears high homology to GroEL, molecular replacement was the method of choice for structure determination. Attempts to solve the structure by molecular replacement were made with the different crystal structures of GroEL available in the Protein Data Bank. Initial attempts to get the correct solution, however, did not yield satisfactory results, as discussed later. As an alternative, therefore, structure solution was attempted by the methods of isomorphous replacement and MAD.

3.4.1 Isomorphous Replacement

3.4.1.1 *Heavy Atom Derivative Screening*

Determination of the initial phases of the crystal was sought by the method of single or multiple isomorphous replacement. Numerous heavy atom derivatisations of

the Cpn60.2 crystals were attempted. Heavy atom derivatisation was performed at room temperature by transferring the crystal from the mother liquor drop to the artificial mother liquor supplemented with the heavy atom. Concentrations varying from 3mM to 10mM were tried for most of the heavy atoms with the exception of Potassium Iodide that was tested at 500mM (Table 3.3). The soak time varied from 2 minutes to 2 days depending on the heavy atom and the concentration used. In cases where high concentration of heavy atoms resulted in poor diffraction, lower concentrations or shorter soak periods were attempted.

Typically for each heavy atom soaked crystal, intensity data were collected for 10 frames of 2° oscillation each. The data were processed with DENZO/ SCALEPACK, and merged with the native data. χ^2 analysis was used to assess if there were significant intensity differences between the native and heavy atom soaked crystals. While χ^2 values close to 1 suggested no heavy atom substitution, values greater than 10 suggested non-isomorphism.

Results of diffraction obtained from the different heavy atom soaked crystals are summarized in Table 3.3. Crystals that diffracted to a resolution less than 5A were not pursued any further.

Table 3.3 Conditions of various heavy atom soaks The remarks column lists if the heavy atom data were pursued for further analysis, and the reasons for abandoning (poor resolution, i.e. less than 5 A; non-isomorphism, i.e. high χ^2 values; or poor substitution of heavy atoms, i.e. $\chi^2 \leq 2$).

Heavy metal	*Concentration (mM)*	*Soak time*	*Maximum resolution (A)*	*Remarks*
Cesium chloride	1	5hrs	*1*	**Poor resolution**
Europium chloride	**5**	30min	7	**Poor resolution**
Europium nitrate	5	5hrs	4.5	$\chi^2 < 1$
Gadolinium acetate	5	8hrs	4	Non-isomorphous
Gold tetra-cyano platinate	5	90min	3.6	$\chi^2 < 2$
Lanthanum chloride	5	24hrs	No diffraction	Poor **resolution**
Lead nitrate	1	4hrs	Poor diffraction	**High** mosaicity, poor resolution
Mercuric acetate	1	5hrs	5	**Non-isomorphous**
Mercuric chloride	5	2 days	7	Poor **resolution**
Mercuric iodide	5	28hrs	4	Non-isomorphous
Mersalyl acid	2.5	30min	**5**	$\chi^2 < 2$
Palladium chloride	5	40min	7	Poor resolution
PCMBA	1	4hrs	3.5	Non-isomorphous
PCMBS	5	3hrs	3.6	$\chi^2 < 2$
Phenyl mercuric acetate	1	5hrs	3.5	Non-isomorphous
Potassium hexa-chloro platinate	1	5hrs	6.5	Poor resolution
Potassium iodide	500	1min	5	$\chi^2 < 1$
Potassium tellurite	5	4hrs	4.5	High mosaicity
Potassium tetra-chloro aurate	**5**	1hr	4	$\chi^2 < 1$
Potassium tetra-chloro platinate	5	28hrs	6	High mosaicity, poor resolution
Samarium chloride	10	24hrs	No diffraction	Poor **resolution**
Samarium nitrate	1	4hrs	3.5	Non-isomorphous
Samarium oxide	**5**	7hrs	4	High mosaicity
Silver nitrate	1	30min	6.5	High mosaicity
Sodium tungstate	**10**	24hrs	4.5	Not a derivative
Stannous chloride	**5**	3.5hrs	No diffraction	Poor resolution
Thallium chloride	**5**	4hrs	3.5	$\chi^2 = 1$
Thallium nitrate	**5**	8hrs	7	Poor resolution
Uranyl acetate	1	30min	No diffraction	Poor resolution
Uranyl nitrate	1	30min	No diffraction	Poor resolution
Uranyl sulphate	5	5hrs	No diffraction	**Poor resolution**
PCMBA: **Phenyl Chloro Mercury Benzoic Acid;** PCMBS: **Phenyl Chloro Mercury Benzene Sulfonic acid**				

Cpn60.2 crystals soaked in mercuric acetate diffracted to a resolution of 5A. Merging of the derivative data set with that of the native resulted in R_{iso} of 20.1% and an overall χ^2 value of 5.5 suggesting a potential derivative. Difference Patterson with $|FPH — Fp|^2$ as coefficients, where FPH and Fp are the derivative and the native structure factor amplitudes respectively, was computed. The Harker section in the difference Patterson is shown in Fig. 3.6.
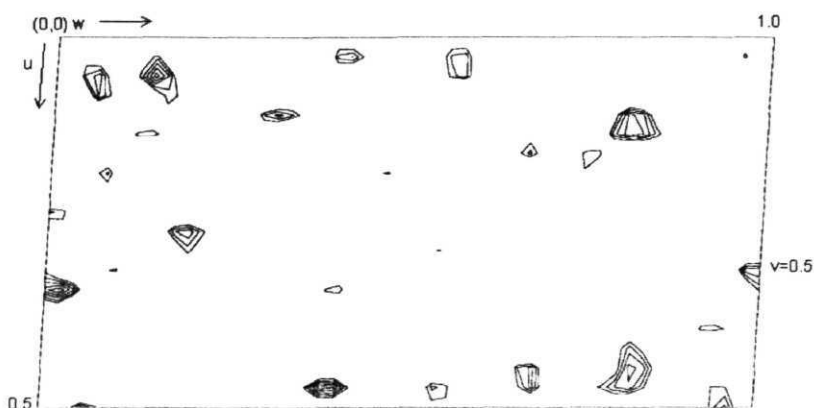
Figure 3.6 | FPH — F_P |² Patterson Harker section Contours are drawn at arbitrary intervals.

A consistent set of heavy atom positions could not be obtained by analysing the Patterson. Moreover, the heavy atom positions identified could not be refined well in MLPHARE (CCP4, 1994). It was therefore inferred that these crystals were either non-isomorphous resulting in a high R_{iso}, or the heavy atom substitution in the crystals was not sufficiently high to be able to detect the sites in the difference Patterson.

As indicated in Table 3.3, for a large number of compounds or conditions attempted, useful data could not be obtained for further analysis. Therefore, even as the attempts to find more derivatives continued, structure solution was sought simultaneously through MAD and molecular replacement techniques.

3.4.2 Multiwavelength Anomalous Dispersion

The multiwavelength anomalous dispersion method for structure determination offers some advantages over the isomorphous replacement method. In this method all the required data can be measured from a single crystal and therefore problems arising due to lack of isomorphism do not arise. Thus, phase determination by the method of multiwavelength anomalous dispersion, using Selenium incorporated crystals was undertaken.

3.4.2.1 *Generation of Selenomethionyl Cpn60.2*

Expression of selenomethionyl (Se-Met) Cpn60.2 was achieved by growth of *E. coli* BL21 (DE3) harboring pET28.3GL2, in M9 minimal medium supplemented with selenomethionine (as listed in Table 2.1 of Chapter 2). Single colony of the transformants was grown overnight in LB medium supplemented with $30\mu g/ml$ of kanamycin. 1 ml of the overnight culture was gently centrifuged for 2 minutes at room temperature and the cell pellet resuspended in 1ml of M9 minimal medium. The resuspended culture was used to inoculate 1 litre of M9 medium supplemented with all amino acids except methionine at a concentration of 50mg/l each. Methionine was instead replaced with selenomethionine at the same concentration. Growth was continued at 37°C and the culture induced with 0.2mM IPTG. Growth of the induced culture was allowed for 16hrs.

Purification process of the Se-Met protein was similar to that for the native Cpn60.2 as described in Chapter 2.

3.4.2.2 *Crystallization of Selenomethionyl Cpn60.2*

Crystals of Se-Met Cpn60.2 were grown at conditions similar to those of the native protein. The crystals of Se-Met derivatized Cpn60.2 were more fragile than the native crystals and readily developed cracks when transferred to artificial mother liquor.

3.4.2.3 *Absorption Edge of Selenomethionyl   Cpn60.2*

A fluorescence scan of the Se-Met Cpn60.2 crystal was measured in order to determine the absorption edge of the anomalous scatterer. Comparison of the fluorescence scan with the theoretical anomalous scattering coefficient of Selenium showed that maximum in f'' occurred at 12,660eV as expected. The fluorescence scan thus confirmed the incorporation of selenomethionine into the protein (Fig. 3.7).

a   b
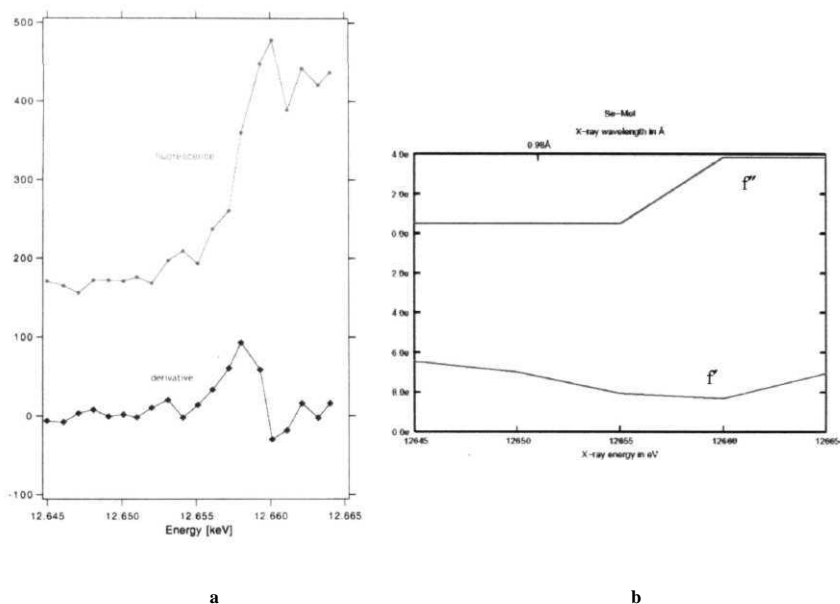
Figure 3.7 Fluorescence scan of selenium (a) Experimental curve of Se-Met Cpn60.2 The red curve shows fluorescence of selenium as a function of incident energy, while the blue curve represents the first derivative of the fluorescence curve. Maximum in the derivative curve represents the inflection point, where the f component of the absorption curve would be minimum. The maximum in the red curve represents the maximum absorption coefficient, where f'' would be maximum. **(b) Theoretical curve** Theoretical scattering coefficients, f and f'', of selenium showing the maximum f'' at 12,660eV.

Se-Met **derivatized** crystals of Cpn60.2 were exposed to the **XRD1** beam at ELETTRA, Trieste. The Se-Met crystals, however, did not yield a good diffraction, resolution being limited to 6A. No useful data could thus be collected with these crystals. Structure determination through MAD was thus not pursued further.

3.4.3 Molecular Replacement

While structure solution was being attempted by **isomorphous** replacement, exhaustive attempts were also made to determine the structure by molecular replacement. Matthews coefficient had suggested presence of two molecules in an asymmetric **unit.** Calculations of self-rotation function to establish the non-crystallographic symmetry were performed in different resolution ranges and using different radii of integration. The self-rotation functions showed distinct peaks for $\kappa=180°$ along the a- or c- directions (Fig. 3.8). These peaks typically were 5(7 levels above the mean value.
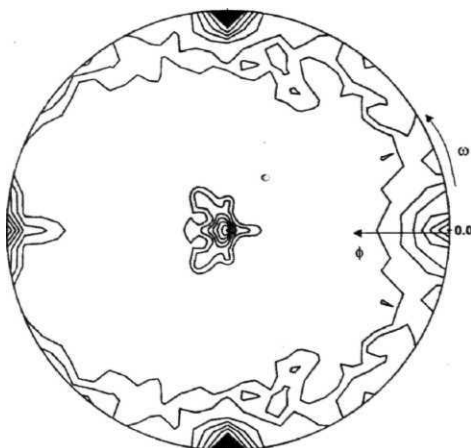
Figure 3.8 A representative self-rotation function plot at $\kappa=180°$ The rotation function was calculated with a radius of integration of 27 A in the resolution range of 8 to 3.2A. The map is contoured at equal intervals of 0.5a starting from 2σ. The strong peak at $\omega=90°$ and $\phi=90°$ corresponds to the crystallographic two-fold axis along the b-direction.

The two peaks correspond to the two pairs of **non-crystallographically** related molecules in the unit cell. Thus, the **non-crystallographic** symmetry could be established through self-rotation function.

M. tuberculosis Cpn60.2 shares 59% sequence identity with the *E. coli* and *Paracoccus denitrificans* Cpn60s. Structure solution of Cpn60.2 was attempted by molecular replacement using the AMoRe package in the CCP4 suite (Navaza, 1994). **Poly-Ala** coordinates of the monomers of available crystal structures of *E. coli* Cpn60 (GroEL) and that of the *P. denitrificans* Cpn60 were used as the search models. Temperature factors of the available coordinates were left unaltered.

The GroEL molecule has approximate dimensions of 72 X 60 X 54 A. Molecular replacement calculations were attempted with radii of integration ranging from 25 to 35 A. The models were placed in primitive triclinic unit cell of dimensions 100 X 100 X 100 A, in order to prevent overlaps between translation related molecules. Different resolution ranges were tried in rotation function calculations. Typically the best solutions were obtained in the resolution range 20 to 4 A. As suggested by Navaza (1994), correlation coefficient between F_{obs} and F_{calc} (in **P1** unit cell) is a good indicator of the rotation function. The solutions were hence sorted on the basis of correlation coefficient between F_{obs} and F_{calc}. The top 100 solutions from each rotation function were subjected to translation searches.

Translation function calculations were performed using either the Crowther and **Blow** method, or correlation coefficient searches as available in AMoRe. Translation searches were attempted in different resolution ranges varying from a lower limit of 30A and upper limit of 4A. In many of the translation searches, the structure factor amplitudes were sharpened with B-factor upto $90A^2$. The space group of Cpn60.2 being $P2_1$, placement of the first molecule was sought only in the x-z plane. Due to the possibility of choosing multiple origins in this space group, searches were limited to 0-0.5 in the x-, and 0-0.5 in the z-directions. Position of the second molecule was searched in the entire asymmetric unit, since the origin was

determined by the prior placement of the first molecule. The translation searches for the second molecule were performed for all the remaining rotation function solutions. Validity of translation search was assessed by discrimination between the highest and the next highest peaks, and also by inspection of packing of molecules in the unit cell.

3.4.3.1 *Search Models for Molecular Replacement*

Molecular replacement was attempted using the different models available in the PDB (Table 3.4). The models vary in the relative orientation across the various domains in GroEL subunits.

Table 3.4 Crystal structures used as search models for molecular replacement

Model	*PDB ID*	*Resolution* (A)	*Details of the model*	*Reference*
E. coli GroEL	1aon	**3.0**	Asymmetric complex with GroES and ADP. GroEL coordinates available in two forms: *as* (GroES bound state) and *trans* (uncomplexed form)	(Xu *et al.*, 1997)
E. coli GroEL	1grl	**2.8**	Uncomplexed GroEL	Braig *et al.*, 1994
E. coli GroEL	1oel	**2.8**	Refined coordinates of uncomplexed GroEL	Braig *et al.*, 1995
E. coli GroEL	1kp8	**2.0**	**Complex of GroEL with ATP-K$^+$-Mg^{2+}**	Wang and Boisvert, 2003
P. denitrificans Cpn60	1iok	**3.2**		Fukami *et al.*, 2001

Cross rotation searches were performed in a primitive triclinic unit cell as described above. However, for all the models used in these searches, the signal/noise ratio for the different solutions was indistinguishable. Identification of the correct solution was therefore difficult. The best cross rotation solutions obtained with the different models are listed in Table 3.5.

Table 3.5 Cross rotation solutions from AMoRe for the different search models a, β, γ represent the orientation as defined by the three Eulerian angles. The solutions for each model are sorted according to the correlation coefficient (CC_F) as explained in the text. RF_F is the R-factor of the solution. The corresponding signal to noise ratio is listed in the last **column,** while the highest signal/noise ratio of the rotation function is indicated in the second column. It is apparent that the highest signal to noise ratio peak did not correspond to the best correlation peak.

Model	Highest signal	α(°)	PC)	γ(°)	CC_F	RF_F	Signal/noise
trans-GtoEL. (laon)	4.2	122	84	187	44.5	55.9	3.2
		236	90	50	44.5	56.3	2.8
		62	90	18	44.4	56.0	2.3
		307	90	229	44.2	56.2	2.8
		116	88	30	44.1	55.9	3.1
cis-GroEL (laon)	4.4	119	84	205	44.0	55.6	2.4
		79	67	213	44.0	55.7	3.5
		296	75	209	44.0	55.3	2.9
		111	88	218	43.9	56.1	3.1
		104	68	36	43.7	55.6	2.5
1grl	4.9	80	79	25	44.9	55.5	3.8
		43	64	68	44.6	55.8	3.2
		263	69	24	44.6	55.7	4.0
		142	83	215	44.5	55.9	3.5
		132	71	189	44.1	56.2	3.9
1oel	4.4	67	90	19	44.7	55.3	3.5
		111	90	196	43.8	55.7	3.5
		251	78	20	43.5	55.4	3.7
		148	67	176	43.3	55.5	3.6
		234	87	49	43.1	55.5	3.8
1iok	4.8	121	90	219	44.8	56.5	3.0
		65	85	33	42.4	57.2	3.5
		110	90	210	41.7	57.3	3.0
		252	75	30	41.6	57.3	3.1
		214	90	52	41.5	57.6	2.5
1kp8	5.3	259	84	23	48.4	56.8	4.1
		281	90	203	48.1	56.9	2.7
		68	88	28	47.4	56.9	3.3
		112	90	210	47.0	57.1	3.2
		242	83	46	46.4	57.4	5.3

One of the reasons for not obtaining a distinct rotation solution could be large deviations in the relative movements across the domains in the structure in comparison to the search model. Thus, an individual domain-wise search was attempted using the apical and equatorial domains of the *E. coli* and *P. denitrificans* Cpn60s. The highest resolution structure of GroEL was taken as a representative from *E. coli*. Cross rotation searches were performed as described for complete subunits (Table 3.6).

Table 3.6 Cross rotation solutions from AMoRe for different domains as search models α, β, γ represent the orientation as defined by the three Eulerian angles. The solutions for each model are sorted according to the correlation coefficient (CC_F).

Model	Highest Signal	α(°)	β(°)	γ(°)	CC_F	RF_F	Signal/noise
1iok							
Equatorial domain	3.8	92	55	37	44.8	55.6	3.7
		68	90	28	44.3	55.2	3.7
		112	90	209	44.3	55.2	3.7
		121	78	221	44.3	55.2	3.1
		251	81	32	44.1	55.0	2.7
Apical domain	3.6	309	90	239	47.3	60.3	1.9
		233	80	*1.2*	46.3	60.7	2.5
		51	76	54	45.7	61.3	3.6
		202	90	56	45.2	61.5	2.6
		153	73	186	45.0	61.4	2.8
1kp8							
Equatorial domain	4.4	253	21	330	44.5	55.1	2.6
		59	30	355	44.4	55.1	2.5
		60	61	310	44.2	56.0	3.8
		137	36	111	44.0	55.5	3.8
		306	45	106	44.0	55.3	2.6
Apical domain	3.5	3	**90**	124	40.9	57.3	3.0
		335	**64**	251	39.4	57.7	2.9
		176	**79**	305	39.3	57.6	3.5
		313	77	310	39.1	58.1	2.3
		188	90	122	38.7	58.2	2.9

Cross rotation search using the equatorial domain of 1iok resulted in the highest solution 3.8a above the mean (Table 3.6). The solution was indistinguishable from the other rotation solutions that varied from 2.7-3.7σ. Moreover, translation search with the highest peak resulted in the correlation coefficient of 0.442, which was not remarkably different than the other solutions. The solution was thus not pursued any further.

Cross rotation search with apical domain of liok as the search model resulted in the highest peak, 3.6a above the mean. When sorted according to correlation coefficient, however, the best solution was only 1.9(7 above the mean (Table 3.6). Moreover, a good solution could not be obtained upon translation search for either of the solutions. Similarly rotation and translation searches using individual domains of 1kp8 did not yield any satisfactory solution (Table 3.6).

Since none of the models attempted for molecular replacement seemed to yield any promising translation solution, accuracy of the cross rotation solution was re-assessed. This was done by bringing the various search models to a common frame of reference and a comparison of the various rotation solutions thereby obtained. Comparison of the cross rotation solutions obtained for the various models yielded similar orientation parameters suggesting that the rotation solution was essentially correct.

3.4.3.2   *E. coli (GroEL-ATP-K$^+$-Mg^{2+})$_{14}$ as the search model*

The crystal structure of *E. coli* GroEL in complex with 7 ATP molecules is available at 2A resolution (Wang and Boisvert, 2003). Molecular replacement was attempted with one GroEL subunit from the (GroEL-ATP-K$^+$-Mg^{2+})$_{14}$ complex as the search model. The highest peak in the cross rotation function was 5.3a above the mean. The Eulerian angles of this solution were very similar to the solution with the highest correlation coefficient (Table 3.5). The highest solution in rotation function is shown in Fig. 3.9.
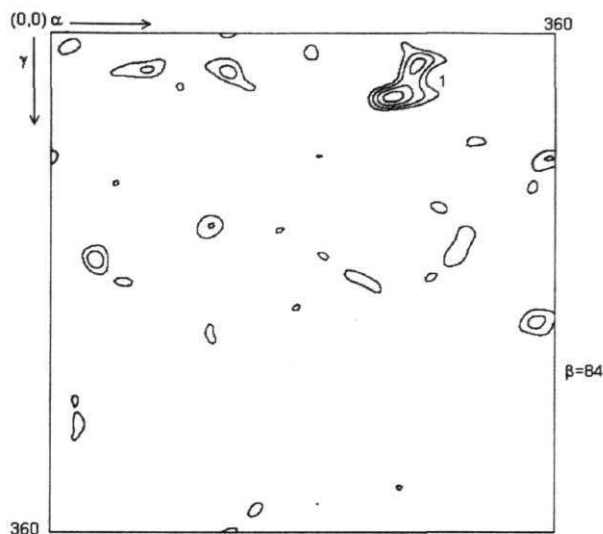
Figure 3.9 β=84° section of the cross rotation search performed in AMoRe with 1kp8 as the search model The highest peak listed in Table 3.5 is marked as 1.

Translation search was performed for all the obtained cross rotation function solutions. For every rotation function solution, the best translation solution defined by its correlation coefficient is listed in Table 3.7.

Table 3.7 Solutions from AMoRe obtained upon translation search

α(°)	PC)	Y(°)	Tx	Ty	Tz	CC_F	RF_F
259	84	23	0.297	0	0.238	64.7	55.8
281	90	203	0.047	0	0.441	63.7	56.1
68	88	28	0.484	0	0.083	62.4	56.4
112	90	210	0.016	0	0.417	62.1	56.0
54	53	88	0.453	0	0.107	61.9	57.1
242	83	46	0.031	0	0.083	61.8	57.1
109	59	218	0.250	0	0.262	61.4	56.8
186	90	21	0.469	0	0.441	61.3	56.6
30	65	234	0.234	0	0.214	61.3	56.8
22	70	35	0.469	0	0.238	61.3	57.4

As is apparent, a significantly distinct translation solution could not be identified on the basis of correlation coefficient, or the R-factor. As described earlier, different models used in molecular replacement calculations had yielded similar orientation parameters in cross rotation functions suggesting that the solution with $\alpha=259°$, $\beta=84°$ and $\gamma=23°$, in cross rotation function might represent the correct orientation of the model. Translation function was therefore once again calculated only for this orientation. Table 3.8 and Fig. 3.10 show the different solutions obtained upon translation search for this cross rotation solution.

Table 3.8 Different translation solutions for the highest cross rotation solution

$\alpha(°)$	*ßin*	$\gamma(°)$	*Tx*	*Ty*	*Tz*	*CC_F*	*RF_F*
259	84	23	0.297	0	0.238	64.7	55.8
			0.484	0	0.048	**64.3**	56.3
			0.156	0	0.274	63.6	56.3
			0.359	0	0.155	63.5	56.1
			0.484	0	0.298	63.2	56.7
			0.453	0	0.131	62.9	56.5
			0.281	0	0.321	62.7	56.5
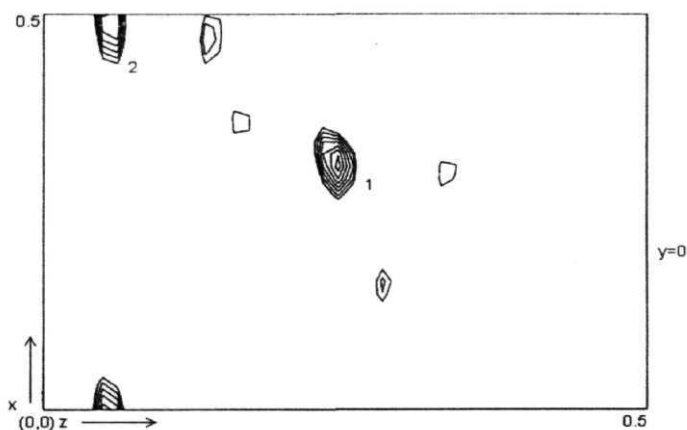
Figure 3.10 y=0 section obtained after the translation search of the highest cross rotation solution The **two** peaks labeled 1 and 2 represent the **two** most **probable translation solutions.**

As is apparent, the difference in correlation coefficients and R-factors is once again not significantly high for the obtained peaks.

Search for the second molecule was initiated by fixing the first molecule in the **asymmetric** unit according to the different solutions in translation function described above (Table 3.8). Fixing the molecule with the topmost translation vector listed in Table 3.8 did not yield a distinct translation solution for the second molecule. A reasonable solution was however obtained when the second translation vector (Table 3.8) was used. A remarkable improvement in the correlation coefficient to a value of 0.68 and a simultaneous drop in the R-factor to 0.525, was observed (Table 3.9).

Table 3.9 Translation search solutions obtained for the second molecule in the asymmetric unit The two sets of searches were performed using the two different translation solutions for the first molecule. As is apparent, while there is no significant difference seen in the correlation coefficients for the first solution, a distinct improvement in the correlation coefficient is obtained for the second one. Rows labeled Molecule 1 represent the parameters used for orientation and position of the first molecule.

	$a(°)$	$P(°)$	$Y(°)$	Tx	Ty	Tz	CC_F	RF_F
Molecule 1	259	84	23	0.297	0.000	0.238		
	68	88	28	0.986	0.105	0.088	65.3	54.3
	105	62	18	0.825	0.173	0.367	64.7	54.0
	230	33	102	0.567	0.861	0.006	64.5	54.6
	191	73	**163**	0.809	0.748	0.812	64.5	55.1
	315	53	**171**	0.944	0.420	0.038	64.4	54.9
Molecule 1	259	84	23	0.484	0.000	0.048		
	68	88	28	0.982	0.721	0.585	68.1	52.5
	112	90	210	0.011	0.223	0.417	67.5	52.7
	179	19	309	0.927	0.528	0.042	65.8	53.9
	281	90	203	0.525	0.495	0.951	65.5	54.3
	315	53	**171**	0.406	0.153	0.105	65.3	53.9

Fig. 3.11 shows the translation solution for the second molecule obtained upon fixing the first molecule.
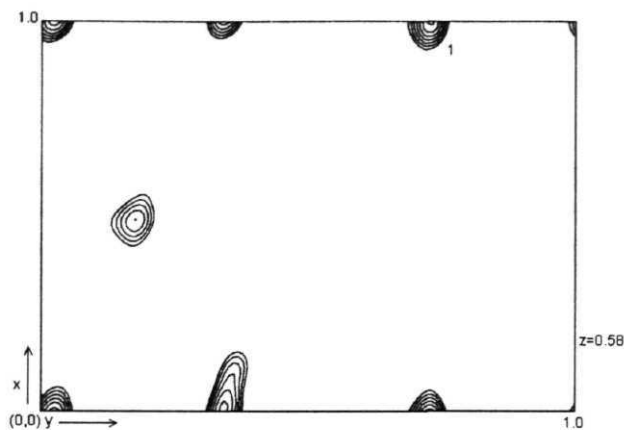
Figure 3.11 z=0.58 section obtained after the translation search of the highest cross rotation solution The peak labeled 1 represents the translation solution for the second molecule.

Packing of the two molecules was checked on graphics for the different translation solutions. The orientation parameters of (259°, 84°, 23°) and (68°, 88°, 28°) of the two molecules and their positions of (0.484, 0.0, 0.048) and (0.982, 0.721, 0.585) yielded a reasonable packing without any overlaps between symmetry related molecules. The resulting solution was therefore pursued for refinement of Cpn60.2.

3.5 STRUCTURE REFINEMENT OF **CHAPERONIN60.2**

The molecular replacement model obtained using a subunit of GroEL from **1kp8 (GroEL-K$^+$-Mg^{2+}-ATP)$_{14}$** was refined using REFMAC5 (Murshudov *et a/.,* 1997) and CNS (Brünger *et a/.,* 1998). 5% of the randomly chosen reflections were set aside for calculation of R_{free}. The same set of test reflections was maintained in either of the refinement programs. **Non-crystallographic** symmetry constraints between the monomers were not enforced in order to ensure that potential structural differences between monomers were not obscured.

Refinement was initiated considering the two molecules in the asymmetric unit as distinct rigid bodies. Rigid body refinement upon definition of the apical, intermediate and equatorial domain as rigid entities dramatically reduced the R and R_{free} from 0.526 and 0.509 to 0.477 and 0.480, respectively. A key to this reduction was refinement with the low resolution data, the data being restricted to 3.8A. Inclusion of high resolution data upto 3.2A did not lower the R and R_{free}.

Rigid body refinement was followed by several cycles of restrained positional refinement. All reflections in the resolution range 50-3.2A were used for the refinement. A tight molecular geometry was enforced by maintaining a high weight on the geometric terms throughout the refinement. Individual isotropic B-factors were refined for all the atoms. Refinement cycles were interspersed with model building in O (Jones *et a/.,* 1991) using σ_A weighted 2Fo-Fc and Fo-Fc electron density maps.

The Fo-Fc map calculated in the initial stages of refinement indicated a major conformation change in the first helix present in the equatorial domain (Chapter 5). This could be detected upon omission of the coordinates of this helix from the molecular replacement model. Fitting of the helix as indicated by the electron density dramatically reduced the R and R_{free} from a value of 0.395 and 0.452 to 0.323 and

0.413, respectively (Table 3.10). Further sessions of model building and restrained refinement gradually reduced the R and R_{free} to 0.267 and 0.385, respectively.

At this stage, conjugate gradient minimization and restrained, individual B-factor refinement in CNS was initiated. During the initial stages of refinement in CNS, the lower resolution data was truncated at 6A and no bulk solvent corrections were applied. Moreover, reflections with an observed magnitude less than *2X5* were rejected during refinement. In the final stages, however, the lower resolution data upto 50A were included and simulated annealing refinement initiated with 100 initial steps of minimization. The molecule was heated to 3000K and slowly cooled to 300K at a rate of 25K per cycle. This was followed by 1000 steps of constant temperature annealing and 100 steps of minimization performed at 300K. The refinement was completed with REFMAC5 using target maximum likelihood function, and with bulk solvent correction. The final R and R_{free} are 24.1% and 28.5%, respectively.

Table 3.10 Improvement of R and R_{free} at the different steps of refinement

Stages	Different refinement methods undertaken	R(%)	R_{free}(%)
1	Refinement of molecules as a single rigid body	54.5	52.8
2	Rigid body refinement of each molecule divided into 5 distinct domains	47.7	48.0
3	Restrained refinement with 65 residues at N-terminus deleted	40.8	47.6
4	Restrained refinement with 90 residues at N-terminus deleted	39.7	46.2
5	Model adjustment and restrained refinement	35.8	44.0
6	Restrained refinement with 11 residues at C-terminus deleted	35.4	43.3
7	Addition of helix 60-79 in chains A and B in electron density followed by rigid body refinement with the two helices as rigid bodies	39.5	45.2
8	Model adjustment and restrained refinement	32.8	42.6
9	Model building with help of omit maps at region 383-403 in chain B	32.3	41.3
10	Model adjustment and restrained refinement	32.7	41.0
11	Loop 80-87 deleted in A and B	32.2	41.7
12	Side chains mutated to match the *M. tuberculosis* sequence in chains A and B	29.6	42.4
13	Model building followed by restrained refinement	27.5	39.5
14	Fit helix 60-79 better in electron density	26.9	38.4
15	Fit region 175-181 in electron density in chain A	26.7	38.5
16	Several cycles of model building and refinement	24.1	28.5

3.6 STRUCTURE VALIDATION OF CHAPERONIN60.2

The refined model of Cpn60.2 was subjected to a variety of checks in order to assess its quality. Among the checks included were PROCHECK (Laskowski *et al.,* 1993) and WHATCHECK. Deviations in bond lengths and bond angles as obtained from the check results were manually removed in O. Side-chain flips and assessment of the H-bond acceptors and donors were manually performed. Residues lying in the disallowed regions of the Ramachandran plot were corrected in O. Short contacts within a molecule of across the symmetry related molecules were manually removed in O. The final model showed an overall good geometry as assessed by the Ramachandran plot (Fig. 3.12). The overall statistics of the final structure from PROCHECK is given in Table 3.11.

Table 3.11 Summary of PROCHECK results of the final Cpn60.2 structure

Property	Statistics
Ramachandran plot	
Core	87.2%
Generously allowed	0.3%
Allowed	12.5%
Disallowed	0.0%
Residue properties	
Bad contacts	0
Maximum standard deviation in bond lengths and bond angles from mean value	3.1
Main chain bond lengths and angles	99.9% within limits
Planar groups	100.0% within limits
G-factors	
Dihedrals	-0.09
Covalent	0.58
Overall	0.17
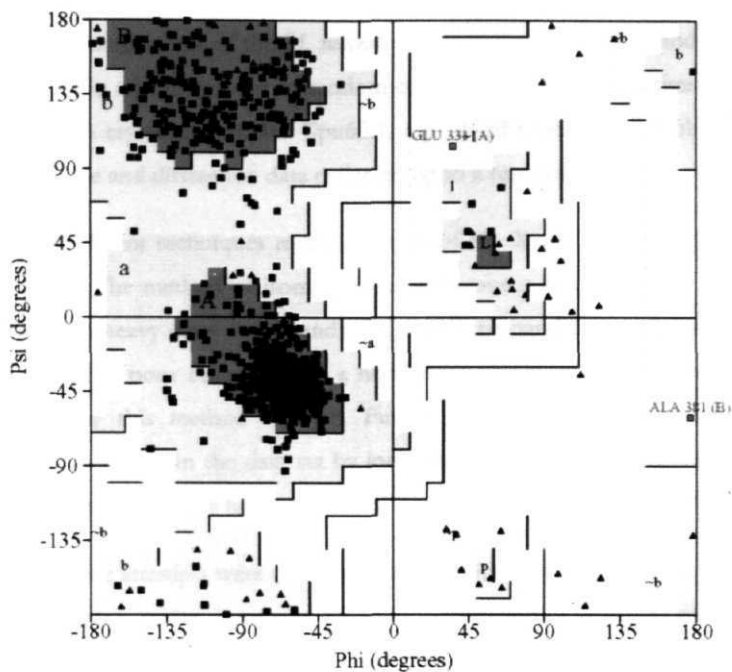
Figure 3.12 Ramachandran plot of the final structure of Cpn60.2 The residues E331 in chain A and A381 in chain B lie in the high temperature factor regions of the structure.

The final structure of Cpn60.2 is deposited in the Protein Data Bank with the accession code 1SJP.

3.7 CONCLUSIONS

Crystallizations of the *M. tuberculosis* chaperonins, Cpn60.1 and Cpn60.2, were attempted in a large number of random conditions. While only a limited success was obtained in crystal growth for Cpn60.1, crystals of Cpn60.2 were obtained at room temperature and diffraction data collected up to a resolution of 3.2A.

Different techniques to the solution of the X-ray structure of Cpn60.2 were employed. The method of isomorphous replacement was attempted with a number of different heavy atom compounds. Almost all the conditions and compounds tested yielded either poor diffraction or a non-isomorphous crystal thus making structure solution by this method difficult. Further, attempts were also made to obtain anomalous signals in the data set by incorporation of selenomethionine in Cpn60.2. Poor diffraction quality however limited its use in structure determination.

While attempts were on for structure solution by isomorphous replacement or MAD phasing, major stress was laid on determination of structure by the method of molecular replacement. Considering a high sequence homology of Cpn60.2 with the *E. coli* GroEL structure solution was attempted by molecular replacement using GroEL as the search model. Solution of molecular replacement seemed reasonable and was pursued for refinement.

In the crystal *M. tuberculosis* Cpn60.2 exists as a dimer with two molecules in the asymmetric unit. The two molecules are related by an approximate 2-fold symmetry with the rotation angle of 173°. Assessment of molecular packing in the crystal showed absence of any short contacts between the different symmetry related molecules. The structure was refined with REFMAC5 and CNS.

The structure refined to 3.2A resolution has the final R and R_{free} of 24.1% and 28.5%, respectively. The structure coordinates have been deposited to the PDB.

3.8   REFERENCES

1. **Adir,** N., Dobrovetsky, E., **Shafat,** I., Cohen, C, and **Kashi,** Y. (2002). Isolation, purification and preliminary X-ray characterization of Cpn60-2 (65 kDa heat-shock protein) from *Mycobacterium   tuberculosis. Acta Crystallogr.* **D58,** 1474-1475.

2. Blundell, T.L., and Johnson, L.N. (1976). Crystallization of proteins. In Protein Crystallography (Academic Press Inc., New York) pp. 59-82.

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

4. Braig, K., Adams, P.D., and Brunger, A.T. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8A resolution. *Nat. Struct. Biol.* 2, 1083-1094.

5. **Brünger,** A.T., Adams, P.D., **Clore,** G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., **Simonson,** T., and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for **macromolecular** structure determination. *Acta Crystallogr.* **D54,** 905-921.

6. Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *ActaCrystallogr.*   **D50,** 760-763.

7. Drenth, J. (1994). Principles of protein X-ray crystallography. Springer advanced texts in chemistry.

8. Fukami,T.A., Yohda, M., Taguchi, H., Yoshida, M., and **Miki,** K. (2001). Crystal structure of chaperonin-60 from *Paracoccus denitrificans. J. Mol. Biol.* **312,** 501-509.

9. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved method for building protein models in electron density maps and the location of errors in these models. *Ada Crystallogr.*   **A47,** 110-119.

10. **Laskowski,** R.A., MacArthur, **M.W.,** Moss, D.S., and Thornton, J.M. (1993). PROCHECK — a program to check the **stereochemical** quality of protein structures./. *Appl. Cryst.* **26,** 283-291.

11. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of **macromolecular** structures by the maximum likelihood method. *Acta Crystallogr.* **D53,** 240-255.

12. Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. *Acta Crystallogr.* **A50,** 157-163.

13. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol.* **276:** Macromolecular crystallography Part A. C.W. Carter Jr. and R. M. Sweet, eds. (New York: Academic Press), pp. 307-326.

14. Rodgers, D.W. (1997). Practical cryocrystallography. In *Methods Enzymol.* **276** (Carter, Jr., C.W. and Sweet, R.M., eds.) Academic Press Inc., New York) pp. 183-203.

15. Wang, J., and Boisvert, D.C. (2003). Structural Basis for GroEL-Assisted Protein Folding from the Crystal Structure of (GroEL-K-Mg-ATP)$_{14}$ at 2.0A resolution/. *Mol. Biol.* **327,** 843-855.

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

17. Yeh, J.I., and Hol, W.G. (1998). A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallogr.* **D54,** 479-480.

Chapter 4

Mycobacterium tuberculosis GroEL Homologues Unusually Exist as Dimers and Retain the Ability to Suppress Aggregation of Substrate Proteins

4.1   INTRODUCTION

Molecular chaperones are a diverse set of proteins that mediate the correct folding, assembly, transport and degradation of other proteins *in vivo* (Saibil and Ranson, 2002).   Chaperonins form a sub-group of molecular chaperones that are found in all domains of life. Group I chaperonins comprise a family of highly conserved proteins, approximately 60kDa in molecular weight that form a cylindrical assembly of two heptameric rings. Unfolded polypeptides bind and undergo productive folding within a large central cavity of this gigantic cylindrical assembly (Langer *et al,* 1992). The *Escherichia coli* chaperonin, GroEL, has provided a paradigm in understanding protein folding mechanisms mediated by the chaperonins (Xu *et a/.,* 1997). GroEL promotes *de novo* folding of ~10-15% of all proteins in bacterial cytosol in co-ordination with the heptameric co-chaperonin, GroES (Ewalt *et al.,* 1997). ATP dependent conformational changes in GroEL have been shown to be necessary for proper chaperonin function *in vivo* (Saibil and Ranson, 2002; Rye *et al.,* 1997).

Chaperonins in all bacteria are encoded by the essential *groEL* and *groES* genes, also called *cpn60* and *cpn10* arranged on the bicistronic *groESL* operon (Bachman, 1990). The operon arrangement of the two genes is highly conserved among all known bacterial species. Interestingly, *Mycobacterium tuberculosis* contains two copies of the *cpn60* genes (Kong *et al,* 1993). One of these genes, *cpn60.1,* is organized on the operon with *cpn10,* while the second copy, *cpn60.2,* is arranged separately on the genome. A similar arrangement of *cpn10* and *cpn60* genes has been described in *Streptomyces spp.* (Mazodier *et al,* 1991) and *Mycobacterium leprae* (Rinke de Wit *et al.,* 1992).

The *E. coli* chaperonin, GroEL, is known to be essential at all temperatures of growth (Fayet *et al.,* 1989).   Ubiquitous presence of chaperonins, and their high conservation suggests a similar indispensable role that these proteins play in all life forms. A high sequence similarity that *M. tuberculosis* Chaperonin60s (Cpn60s) exhibit

with *E. coli* GroEL, suggests that these might also perform similar role in this organism. The existence of a duplicate set of *cpn60* genes in *M. tuberculosis,* however, has been perplexing. Plant chloroplasts also possess two copies of Cpn60s that are known to form a hetero-oligomeric assembly (Dickson *et al.,* 2000). The two *M. tuberculosis* GroEL homologues may similarly be interdependent for the assembly of the protein into a functional hetero-tetradecamer. Duplication therefore might suggest either an interdependence of the gene products or alternatively, the necessity for a redundant function.

While reasons for gene duplication remain unclear, both the Cpn60s of *M. tuberculosis* have been shown to be highly antigenic in nature, eliciting strong B-cell and T-cell immune responses. Cpn60s have also been proposed to play roles as virulence determinants and as inducers of host inflammatory responses (Lewthwaite *et al.,* 1998). Moreover, both Cpn60.1 and Cpn60.2 of *M. tuberculosis* have been shown to be potent cytokine inducers (Lewthwaite *et al.,* 2001). The Cpn60s might therefore represent important components of *M. tuberculosis,* playing roles as immunomodulators, and perhaps also required for proper protein folding and transport.

Biochemical properties of *M. tuberculosis* Cpn60s, especially their role as molecular chaperones, have not been characterized yet Moreover, presence of two copies of these genes in the *M. tuberculosis* genome is intriguing. To address these questions and to elucidate their role in protein folding, the *M. tuberculosis cpn60* genes were expressed in *E. coli.*

4.2   EXPERIMENTAL PROCEDURES

4.2.1 Purification of *M. tuberculosis* Cpn60s

cpn60.1 and *cpn60.2* were cloned in suitable expression vectors and the gene products purified by affinity chromatography over Ni-NTA agarose column as described in Chapter 2. Purification of Cpn60s was also carried out in the presence of 10% glycerol, 1mM ATP-γ-S or 10μM *M. tuberculosis* Cpn10. These supplements were present in the buffers throughout the purification procedure. *E. coli* GroEL and GroES were overexpressed in *E. coli* BL21 (DE3) harbouring the plasmid pKY206 and purified using minor modifications of the published procedure (Clark *et al.*, 1998). Purification of (His)6-tag-less chaperonins of *M. tuberculosis* was performed as described in Chapter 2. Purification of *M. tuberculosis* CpnlO was performed as previously described (Taneja and Mande, 2001).

4.2.2 Reconstitution of Cpn60 Oligomers

Reconstitution was performed essentially as described earlier with slight modifications (Ybarra and Horowitz, 1995). Briefly, 10μM Cpn60 (protomers) was incubated with 4M urea on ice for 90 minutes. Reconstitution was initiated by a rapid 10-fold dilution of the monomeric Cpn60 in the buffer containing 50mM Tris-Cl, pH 8.0, 10mM $MgCl_2$, 5mM ATP, 1M ammonium sulphate or 10% glycerol and *E. coli or M. tuberculosis* CpnlO in a 1:1 molar ratio. The reconstitution mix was incubated at 25°C for 2 hrs and then analyzed on a 6% Native-PAGE.

4.2.3 Circular Dichroism Measurements

Circular Dichroism (CD) spectra of the *M. tuberculosis* Cpn60s were recorded using a Jasco J-715 spectropolarimeter at room temperature. The proteins in lOmM Tris-Cl buffer, pH 8.0 were used at a concentration of 1mg/ml. Far UV-CD spectrum was recorded using a 0.01cm path length cuvette.

4.2.4 Protein Analysis by Native-PAGE and Immunoblotting

Native proteins of *M. tuberculosis* were resolved on a 6% Native-PAGE. The proteins were then transferred onto HybondC (Amersham) nitrocellulose sheet by the method of Towbin *et al.* (1979). Monoclonal antibodies, IT-56 (anti-Cpn60.1) and mAb67-2 (anti-Cpn60.2), kindly provided by John Belisle (Colorado State University) and A. H. Kolk respectively, were used for detection of proteins. HRP-labeled goat anti-rabbit IgG (anti-anti-Cpn60.1) and goat anti-mouse IgG (anti-anti-Cpn60.2) were used at dilutions of 1:2000 and 1:10,000, respectively. The proteins were visualized using standard protocols.

4.2.5 Analysis of Urea-Promoted Dissociation of the *E. coli* GroEL Mutant

Point mutant of *E. coli groEL* was generated using the Quick-Change site directed mutagenesis kit as described in Chapter 2. GroEL mutant was purified using similar protocol as that for the native GroEL.

Urea induced dissociation of *E. coli* GroEL and its mutant was performed as previously reported (Horowitz *et al.*, 1993). Briefly, 10ul samples containing 10.6µg of the native or mutant GroEL (13.3pmol of oligomer) in 100mM Tris-Cl, pH 8.0 supplemented with 10mM MgCl$_2$, 100mM KC1 and 4.5mM dithiothreitol were pre-incubated for 5min at 25°C. This protein was then mixed with an equal volume of 0 to 9M urea. The samples were incubated for 40min at room temperature and then analyzed on a 6% Native-PAGE.

4.2.6 ATP Hydrolysis by Cpn60s

The Cpn60 ATPase activity was quantitated by a colorimetric assay performed in microtiter plates as described earlier (Henkel *et a/.,* 1988). Briefly, the reaction buffer containing lOOmM Tris-Cl (pH 8.0), 10mM KC1, 10mM MgCl$_2$, 2.5uM GroEL

and varying concentrations of ATP was incubated at 37°C for 15 minutes. The enzymatic reaction was terminated by addition of 200µl of the acidic solution of malachite green, ammonium molybdate and polyvinylalcohol. The activity was measured as the amount of inorganic phosphate (Pi) liberated at 655nm in an ELISA plate reader. The values obtained were corrected by subtracting the blank readings for non-enzymatic release of Pi due to hydrolysis of ATP and Pi contamination in the absence of enzyme as well as substrate. A standard curve with monobasic potassium phosphate was run concurrently with each experiment and thus nanomoles of Pi released were calculated.

4.2.7 bis-ANS Fluorescence Assay

Binding of bis-ANS to *M. tuberculosis* Cpn60s was monitored by exciting the probe at 395nm and recording the emission spectra in the range of 400-600nm. The protein and bis-ANS were used at a concentration of 20µM. Fluorescence intensity measurements were carried out at room temperature on a Varian Eclipse spectrofluorimeter. The fluorescence of buffer (100mM Tris-Cl, pH 8.0) and bis-ANS alone were measured as controls.

4.2.8 Aggregation of Citrate Synthase

Citrate synthase (0.015mg/ml) was incubated at 43°C in 50mM HEPES-KOH buffer, pH 7.5, in the presence or absence of *M. tuberculosis* Cpn60.1 and Cpn60.2. Aggregation was monitored for 20min on a Hitachi F-4000 spectrofluorimeter with emission and excitation wavelengths set at 465nm and corresponding band passes set at 3.0nm. Temperature of the sample was maintained using a Julabo circulating water-bath. Internal temperature of the cuvette was monitored using a Physitemp type T microcouple.

4.2.9 Chemical Denaturation and Refolding of Citrate Synthase

Denaturation and refolding of citrate synthase was carried out as previously described (Zhi *et a/.,* 1992). Briefly, 15µM citrate synthase from pig heart was denatured by 6M guanidine-HCl in lOOmM Tris-Cl buffer, pH 8.0 containing 20mM DTT. The enzyme was incubated in the denaturant for 2hrs at room temperature. Renaturation was carried by a 100-fold dilution into lOOmM Tris-Cl buffer, pH 8.0, 10mM KCl and 10mM MgCl$_2$ containing the Cpn60, Cpn10 and ATP to final concentrations of 1µM, 2uM and 2mM, respectively. Aliquots were withdrawn at different time points and tested for the recovery of activity at room temperature.

Citrate synthase activity was measured as described by Srere (1966). The reaction was monitored as a decrease in absorbance at 233nm due to cleavage of Acetyl-CoA and utilization of oxaloacetate at 25°C. The reaction mix contained lOOmM Tris, pH 8.0, 0.15mM Acetyl CoA and 0.1 mM oxaloacetate. The reaction was initiated by addition citrate synthase to a final concentration of 3nM.

4.2.10 Chemical Denaturation and Refolding of Rhodanese

Denaturation of rhodanese to a final concentration of 9µM was carried out in lOOmM Tris-Cl, pH 8.0 containing 6M guanidine-HCl and 1mM dithiothreitol (DTT) for 2hrs at room temperature. Refolding was initiated at 37˚ C by a rapid dilution in 50mM Tris-Cl, pH 8.0 supplemented with 20mM MgCl$_2$, lOmM KCl, 50mM Na$_2$S$_2$O$_3$ and 5mM DTT in the presence or absence of the different Cpn60s. The final concentration of rhodanese in the refolding mix was 108nM. The protomer concentrations of Cpn60s and CpnlO were maintained at 2.5uM and 2.4µM, respectively. 2mM ATP was added to the refolding solution, containing the chaperonins, just before initiation of protein refolding. An aliquot from the refolding mix was withdrawn at different time points and assayed for the recovery of activity as described earlier (Sorbo, 1953).

4.3   RESULTS

4.3.1 Purification of *M. tuberculosis* Cpn60s

As described in Chapter 2, the full length chaperonin *cpn60.1* and *cpn60.2* clones were obtained and overexpressed in *E. coli.* The two chaperonins were expressed in large quantities and exhibited a protomer molecular mass of ~60kDa. As judged by SDS-PAGE (Fig. 2.6 and 2.7, Chapter 2), both the proteins were more than 98% pure after the metal affinity purification.

4.3.2 Quaternary Structure Determination

A tetradecameric structure for native Cpn60 has been established for a number of chaperonins from different species (Hubner *et al.*, 1996; Fukami *et a/.,* 2001). In a few cases, however, Cpn60 homologues have also been reported to exist as heptamers (Viitanen *et a/.,* 1992). The native molecular mass of the purified *M. tuberculosis* proteins, Cpn60.1 and Cpn60.2, was established by size exclusion chromatography. Intriguingly, the purified Cpn60.1 and Cpn60.2 eluted as sharp homogenous peaks at volumes corresponding to molecular mass of approximately 110kDa when compared to standard molecular weight markers (Fig 4.1).
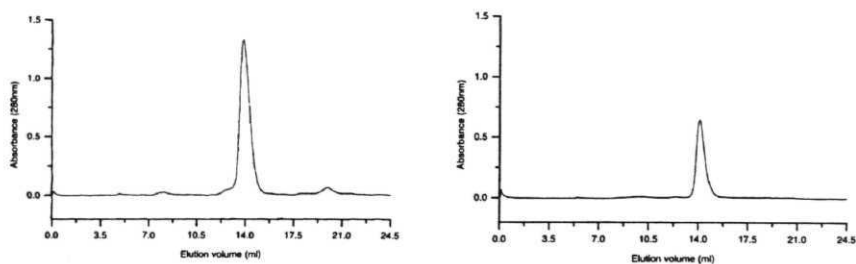
Figure 4.1 Gel filtration chromatograms of *M. tuberculosis* Cpn60s Elution profile of the purified (a) Cpn60.1 and (b) Cpn60.2 on a Superdex-200 HR 10/30 column. Comparison with standard molecular-mass markers showed that Cpn60.1 and Cpn60.2 elute as homogenous dimers.

Neither Cpn60.1 nor Cpn60.2 eluted at the position where *E. coli* GroEL and canonical Cpn60 14-mers would appear. Elution profile on the size exclusion chromatography of the two Cpn60s lacking the (His)6-tag was consistent with that of the (His)6-tagged proteins, suggesting that presence of (His)6-tag did not interfere in oligomerization.

Osmolyte stabilizers, such as glycerol, are known to promote stability of oligomeric state of proteins. Similarly, presence of ATP and the co-chaperonin, Cpn10 may also be crucial to the oligomeric assembly of the *M. tuberculosis* Cpn60s. Purification of Cpn60.1 and Cpn60.2 was therefore performed in the presence of glycerol, or in presence of ATP-γ-S and *M. tuberculosis* co-chaperonin, CpnlO. Their oligomeric state was then tested on Native PAGE. Proteins purified under these conditions, however, did not exhibit oligomerization (Fig. 4.2).

	Cpn60.1			Cpn60.2		
M. tuberculosis Cpn10			+		+	
Glycerol	+			+		
ATP-γ-S		+				+
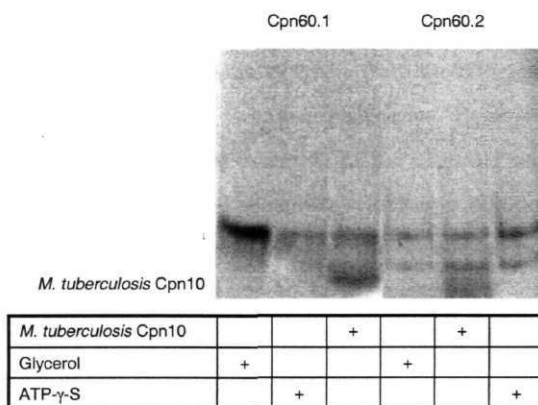
Figure 4.2 Native gel electrophoresis of purified *M. tuberculosis* Cpn60s demonstrating their dimeric state Cpn60.1 and Cpn60.2 were purified in presence of glycerol, ATP-γ-S and *M. tuberculosis* CpnlO. The purified Cpn60s migrate as low molecular weight proteins showing their inability to assemble into oligomers in the presence of these supplements.

Thus, purification of *M. tuberculosis* Cpn60s, either in presence of co-chaperonin or other osmolyte stabilizers did not seem to promote the oligomeric assembly of these proteins. The size exclusion chromatography and Native-PAGE data hence suggest that the *M. tuberculosis* GroEL homologues might intrinsically exist as dimers.

4.3.3 Reconstitution of *M. tuberculosis* Cpn60s

4.3.3.1 *Reconstitution in Presence of Protein Stabilizers*

In order to functionally reconstitute the chaperonins in their canonical oligomeric forms the purified Cpn60.1 and Cpn60.2 were subjected to *in vitro* reconstitution experiments under a variety of conditions. It has previously been reported that urea-dissociated *E. coli* GroEL 14-mers are capable of reassembling into the tetradecameric state upon removal of the chaotrope (Iissin *et al.*, 1990). Reconstitution of the *E. coli* GroEL tetradecamers from the disassembled monomers requires the presence of Mg^{2+} and adenine nucleotides. Presence of ammonium sulphate has also been reported to be essential for the oligomer reassembly (Ybarra and Horowitz, 1995). When subjected to similar refolding conditions, neither of the two *M. tuberculosis* GroEL homologues spontaneously assembled into tetradecamers (Fig 4.3). Reconstitution was also attempted in the presence of glycerol. However, none of the various combinations of nucleotides, GroES, ammonium sulphate, glycerol, K^+ or Mg^{2+}, promoted oligomeric assembly of the *M. tuberculosis* chaperonins (Fig 4.3).
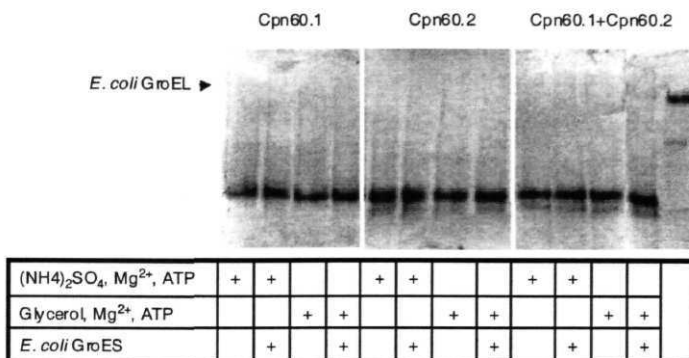
Figure 4.3 Reconstitution of *M. tuberculosis* Cpn60s demonstrating their dimeric state *In-vitro* reconstitution of Cpn60 oligomer was attempted in the presence of ammonium sulphate and glycerol as described in "Experimental Procedures". Presence of either ammonium sulphate or glycerol in the reconstitution mix does not alter the oligomeric state of Cpn60.1 (lanes 1-4) or Cpn60.2 (lanes 5-8) irrespective of the presence or absence of *E. coli* GroES. Simultaneous presence of the two Cpn60s (lanes 9-12) also does not promote the oligomeric assembly of the *M. tuberculosis* Cpn60s. Lane 13 indicates the oligomeric state of the native *E. coli* Cpn60.

Plant chloroplast chaperonins are known to oligomerize as heterotetradecamers of two different polypeptide chains (Hubner *et al.,* 1996). In order to test the possibility whether the two *M. tuberculosis* Cpn60s were similarly interdependent on each other for reassembly into oligomeric state, reconstitution of the two proteins was attempted in presence of each other. Incorporation of both Cpn60.1 and Cpn60.2 in the reaction mix, however, did not alter the oligomeric state of the reconstituted proteins (Fig 4.3). Moreover, the oligomeric state remained unaltered irrespective of the presence or absence of the nucleotides or GroES. These results thus further corroborate that the M. *tuberculosis* GroEL homologues are dimeric proteins, unlike their *E. coli* counterpart.

4.3.3.2 *Reconstitution in Presence of co-Chaperonin*

GroES has previously been shown to facilitate the assembly of GroEL (Seale *et al.*, 1996). Reassembly of the *M. tuberculosis* Cpn60s was thus performed in the presence of the cognate Cpn10. Presence of the *M. tuberculosis* co-chaperonin, Cpn10 in the reconstdtution mix however, had no effect on the oligomeric assembly of the purified proteins (Fig. 4.4). Failure to reconstitute the *M. tuberculosis* chaperonins in higher oligomeric states, thus suggested that their existence as dimers might be an intrinsic property of the two Cpn60s.

	E. Coli GroEL		Cpn60.1			Cpn60.2		
$(NH4)_2SO_4$, Mg^{2+}, ATP	+	+	+	+	+	+	+	+
E. coli GroES				+		+		
M. tuberculosis Cpn10		+			+			+
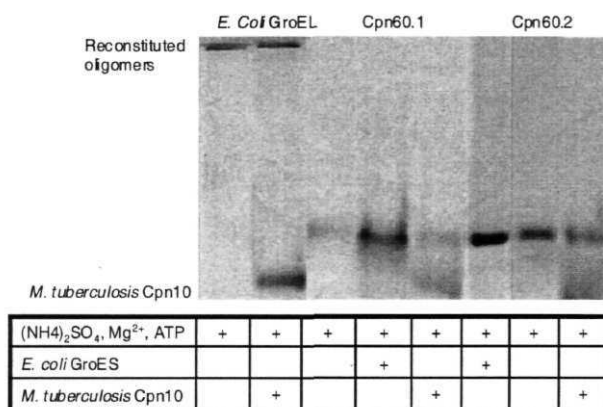
Figure 4.4 Native gel electrophoresis of *M. tuberculosis* Cpn60s reconstituted in the presence of CpnlO *In vitro* reconstitution of Cpn60 oligomers was attempted in the presence of the cognate co-chaperonin, *M. tuberculosis* CpnlO. *E. coli* GroEL monomers subjected to reconstitution in the presence of ammonium sulphate (lane 1) and *M. tuberculosis* CpnlO (lane 2) readily reassemble as tetradecamers. On the other hand, *M. tuberculosis* Cpn60.1 and Cpn60.2, subjected to reconstitution in the presence of ammonium sulphate (lanes 3 and 7), *E. coli* GroES (lanes 4 and 6) and *M. tuberculosis* CpnlO (lanes 5 and 8), clearly are unable to reconstitute as tetradecamers.

4.3.4 Circular Dichroism Measurements of *M. tuberculosis* Cpn60s

The above data indicate that the two *M. tuberculosis* GroEL **homologues** possess an unusual quaternary structure. The possibility that their dimeric structure might have arisen due to the loss of secondary and tertiary structure was tested by measuring the CD spectrum of the proteins. The far UV-CD spectrum of *M. tuberculosis* Cpn60.1 and Cpn60.2 (Fig. 4.5) is characteristic of highly helical proteins with signature bands for helical structure at 208 and 222nm (Johnson, 1990).
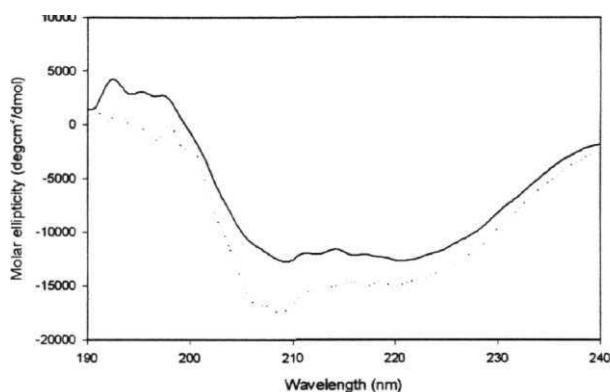
Figure 4.5 Far UV-CD spectrum of *M. tuberculosis* Cpn60.1 and Cpn60.2 Mean residue ellipticity versus wavelength was measured at 25°C in 10mM Tris-Cl, pH 8.0. Signature bands for helical structure are seen at 208 and 222nm suggesting a well-formed secondary structure for the Cpn60 dimers.

The CD data suggest that the *M. tuberculosis* chaperonins are in a significantly folded conformation. The unusual properties of the *M. tuberculosis* chaperonins thus do not appear to be due to absence of secondary structure of the proteins.

4.3.5 **Oligomeric** State **of** *M. tuberculosis* Cpn60s *in vivo*

As mentioned above, Cpn60s from other prokaryotes exist as tetradecamers in their cellular environment and the same oligomeric state is also observed *in vitro. In vitro* purification and reconstitution results show that the *M. tuberculosis* Cpn60s exist as dimers, a state surprisingly different from that observed for other Cpn60s. To discount the possibility that the unusual behavior of chaperonins as dimers was a result of their heterologous expression, the oligomeric state of these proteins in their homologous cellular environment was assessed. This was tested by resolving the cellular fraction of *M. tuberculosis* proteins on Native-PAGE followed by Western blot analysis (Fig. 4.6). Upon immunoblotting, both, Cpn60.1 and Cpn60.2 appeared as lower oligomers, and not tetradecamers, reconfirming the *in vitro* data (Fig. 4.6). In conclusion therefore, absence of the oligomeric state of the *M. tuberculosis* Cpn60s, *in vivo* and *in vitro,* indicates the natural tendency of these proteins to exist as dimers.
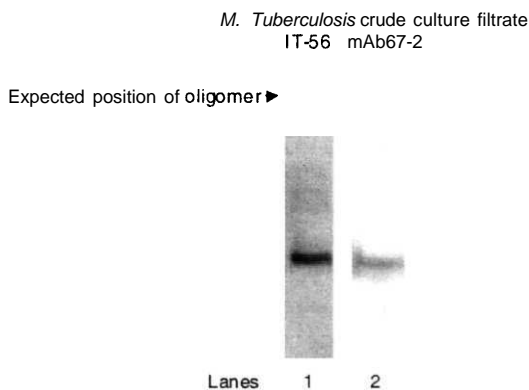
Figure 4.6 Western blot analysis of *M. tuberculosis* **culture proteins demonstrating dimeric state of the mycobacterial Cpn60s** Culture extracts of *M. tuberculosis* were subjected to 6% Native-PAGE and probed with monoclonal antibodies, IT-56 and mAb67-2, against Cpn60.1 (lane 1) and Cpn60.2 (lane 2) respectively. Western blot analysis indicates that Cpn60.1 and Cpn60.2 migrate at a molecular weight much lower than that of the expected oligomer thus suggesting that Cpn60s of *M. tuberculosis* exist as dimers *in vivo*. Expected position of the oligomeric species is indicated.

4.3.6 Importance of A2 and E76 in Oligomerization

Multiple sequence alignment of 43 Cpn60 protein sequences was generated by ClustalW as described in Chapter 2. The alignment suggested that residues at positions 2 and 76, which occur at inter-monomer interface in the tetradecameric GroEL structure, are strictly conserved as Alanine and Glutamate (Fig. 4.7 and 2.8, Chapter 2). These positions are, however, occupied by Serine in *M. tuberculosis* Cpn60.1. To assess the importance of these residues in maintenance of the oligomeric structure an A2S/E76S mutant of *E. coli* GroEL was generated.

Figure 4.7 Crucial side chain — main chain interaction at *E. coli* GroEL's oligomeric interface involving E76 Interaction between the side-chain carboxyl of E76 and main-chain amide of E386 across two subunits in *E. coli* GroEL has been suggested to be important for the oligomerisation of the protein. A natural occurrence of Serine at position 76 in *M. tuberculosis* Cpn60.1 may disrupt this interaction and is hypothesized to contribute to the loss of its oligomeric structure. The two polypeptide chains have been colored in different shades of gray. Other crucial interface residues important for maintenance of the oligomeric state are indicated in Fig. 2.8, Chapter 2. This figure was generated using MOLSCRIPT (Kraulis, 1991).

The monomer - tetradecamer equilibrium of the native as well as that of A2S/E76S mutant was analyzed as a function of urea concentration on Native-PAGE. In accordance with an earlier report, the native protein dissociates into monomers at a urea concentration greater than 3M (Fig. 4.8). The mutant, however, disassociated into monomers at a much lower concentration of urea than the wild type GroEL (Fig. 4.8) (Arai *et a/.,* 2003). The results thus indicate that the thermodynamic stability of the GroEL oligomeric assembly is significantly reduced upon mutation of A2S and E76S. A natural occurrence of Serine at residue positions 2 and 76 in Cpn60.1 might thus be responsible for the lower oligomeric state of this protein.
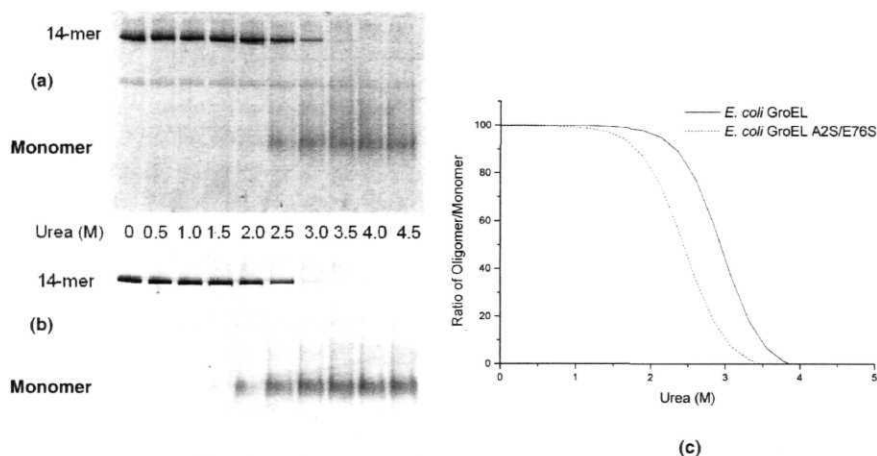
Figure 4.8 Effect of urea concentration on dissociation of *E. Coli* GroEL Urea induced dissociation of the (a) native and (b) A2S/E76S GroEL mutant was analysed on a 6% Native-PAGE. (c) Ratio of the oligomeric to monomeric forms of GroEL plotted against urea concentration indicates that the mutant dissociates into monomers at 2.5M urea concentration, in comparison to 3M for the native.

4.3.7 ATPase Activity of *M. tuberculosis* Cpn60s

E. *coli* GroEL possesses a weak K^+ stimulated ATPase activity, an essential component of the chaperonin mediated protein folding reaction (Viitanen *et al.,* 1990). Due to their unusual quaternary structure, measuring the ATPase activity of the recombinant *M. tuberculosis* Cpn60 proteins was an important test for their functional characterization. Fig. 4.9 compares the ATPase activity of *M. tuberculosis* Cpn60.1 and Cpn60.2 with that of the *E. coli* GroEL. *E. coli* GroEL catalyzed the hydrolysis of ATP with a k_{cat} of $2.6min^{-1}$ at 37°C. Cpn60.1 and Cpn60.2 exhibited a k_{cat} of 0.16 and $0.28min^{-1}$ respectively. The data thus suggest absence of ATPase activity in the GroEL homologues of *M. tuberculosis.*
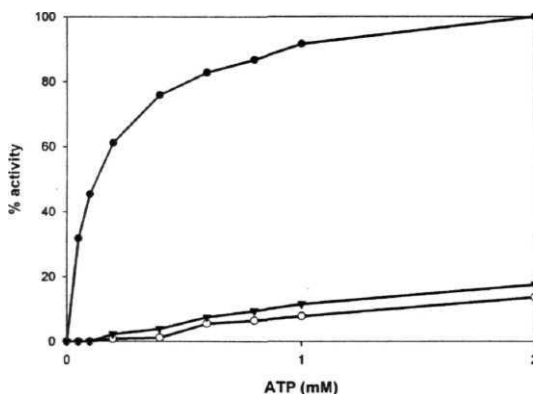
Figure 4.9 ATPase activity of *M. tuberculosis* **Cpn60s and** *E. coli* GroEL ATP hydrolysis by *E. coli* GroEL, *M. tuberculosis* Cpn60.1 and Cpn60.2 was measured as a function of ATP concentration at 37°C in the presence of 10mM $MgCl_2$ and 100mM KC1 for 15 minutes. Cpn60.1 (-o-) and Cpn60.2 (-T-) clearly show very little ATPase activity compared to GroEL (-•-).

4.3.8 bis-ANS Fluorescence Assay for Cpn60s

Loss of oligomerization would have led to the exposure of large hydrophobic regions on the surface of *M. tuberculosis* Cpn60s that were otherwise buried. Also the apical domain in a GroEL oligomer is known to possess exposed hydrophobic area

where substrate polypeptides are known to bind in the central cavity (Xu *et al.*, 1997). Presence of such hydrophobic patches on the surface of *M. tuberculosis* Cpn60s was probed by binding to 1,1-bis (4-anilino) naphthalene-5, 5'-disulfonic acid (bis-ANS).
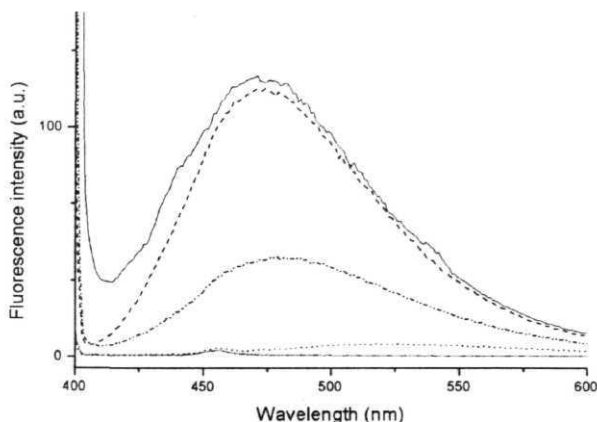
Figure 4.10 Fluorescence intensity enhancement of bis-ANS upon binding to *M. tuberculosis* Cpn60s The fluorescence intensity of bis-ANS is observed to increase significantly in presence of both Cpn60.1 (dashed line) and Cpn60.2 (dashed-dotted line), suggesting presence of hydrophobic surfaces on the proteins. Solid line represents the enhancement seen for *E. coli* GroEL in presence of bis-ANS. Fluorescence intensity for the buffer (dotted line) and bis-ANS (short dotted line) were measured as controls.

As shown in Fig. 4.10, interaction between bis-ANS and Cpn60s is significantly enhanced in presence of Cpn60s suggesting binding of bis-ANS to the hydrophobic surfaces of these proteins. The fluorescence enhancement in Cpn60.1 is comparable to that observed for *E. coli* GroEL. In comparison, the increase observed for Cpn60.2 is much lower, suggesting smaller percentage of hydrophobic surface exposed in Cpn60.2. The results clearly indicate the presence of hydrophobic surfaces on the Cpn60s, a larger surface in Cpn60.1 than in Cpn60.2.

4.3.9 Effect of *M. tuberculosis* Cpn60s on Aggregation of Citrate Synthase

E. colt GroEL is known to facilitate refolding of denatured citrate synthase by suppressing its aggregation (Buchner *et at,* 1991). Exposed hydrophobic patches on *M. tuberculosis* Cpn60s suggested that these might exert similar effect on substrate proteins. Fig. **4.11** compares the ability of Cpn60.1 and Cpn60.2 to prevent the aggregation of citrate synthase at 43°C measured by light scattering.
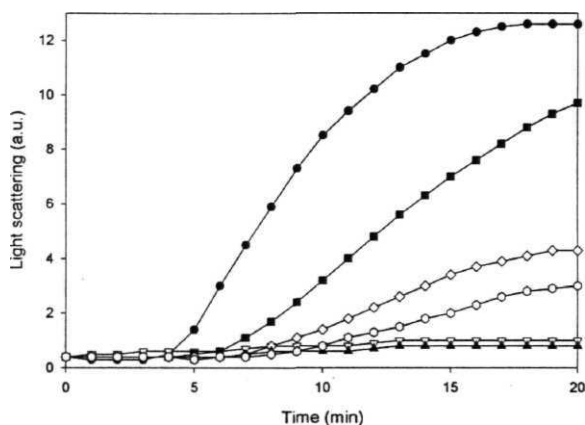
Figure 4.11 Aggregation of citrate synthase as a function of time Aggregation of citrate synthase at **43°C** in the absence (-•-) and presence of equimolar ratios of *M. tuberculosis* Cpn60.1 (-V-); and Cpn60.2 (—0—) measured as a function of light scattered at 465nm. Increase in molar excess ratio of Cpn60.2 to 2-fold (—o—) or 14-fold (—A—) resulted in suppression of aggregation of citrate synthase to a better extent. Both the Cpn60s are clearly able to prevent light scattering by citrate synthase when present in molar excess. Citrate synthase in the presence of BSA (-•-) was taken as a control.

Cpn60.1 was more potent in suppressing aggregation when used in equimolar ratio (Fig. **4.11**). In molar excess, however, both the chaperonins were observed to completely prevent aggregation of citrate synthase.

4.3.10 Effect of *M. tuberculosis* Cpn60s on Refolding of Rhodanese

In order to assess the ability of the *M. tuberculosis* Cpn60s to facilitate refolding of substrate proteins we tested *in vitro* refolding of rhodanese. Cpn60.2 was seen to be able to promote the refolding of guanidine-HCl denatured rhodanese (Fig. 4.12). The percentage recovery was however less than that obtained by spontaneously refolded rhodanese. Intriguingly, this ability of Cpn60.2 was independent of the presence or absence of the co-chaperonin or ATP. Cpn60.1 did not seem to promote the refolding of rhodanese. This might have been due to the tight binding of Cpn60.1 to the substrate preventing its release. It is known that a GroEL bound polypeptide is initiated to refold by hydrolysis of ATP. In absence of ATP, however, the polypeptide remains bound to GroEL tightly. Cpn60.1 might similarly bind denatured rhodanese, thus preventing its successful refolding.
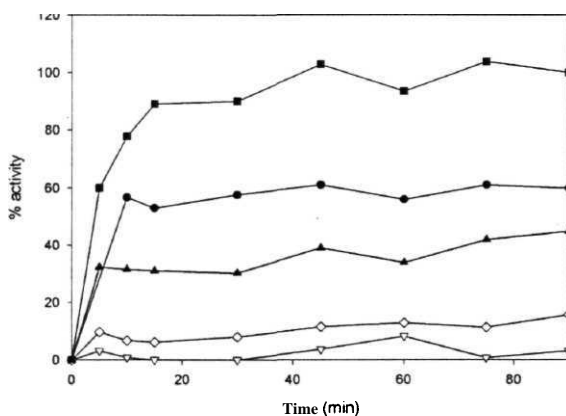
Figure 4.12 Effect of *M. tuberculosis* Cpn60s on refolding of chemically denatured rhodanese
Refolding of denatured rhodanese was monitored in the absence (—•—) or presence of *E. coli* GroEL, GroES and ATP (-•-); GroEL, in absence of GroES or ATP (-V-); *M. tuberculosis* Cpn60.1 (—◊—) and *M. tuberculosis* Cpn60.2 (—A-). Activity of *M. tuberculosis* Cpn60s was measured in the presence of cognate *M. tuberculosis* Cpn10 and ATP as mentioned in the text. Activity obtained when rhodanese was refolded in presence of *E. coli* GroEL for 90 minutes was considered to be 100%.

4.3.11 Effect of *M. tuberculosis* Cpn60s on Refolding of Citrate Synthase

The ability of *M. tuberculosis* chaperonins to refold a wider range of substrate proteins was tested by *in vitro* refolding of citrate synthase. Cpn60.1 was capable of promoting refolding of citrate synthase, to an extent comparable to that obtained by Cpn60.2 (Fig. 4.13). Moreover, the presence or absence of co-chaperonin or ATP did not affect the extent of recovery of citrate synthase activity.
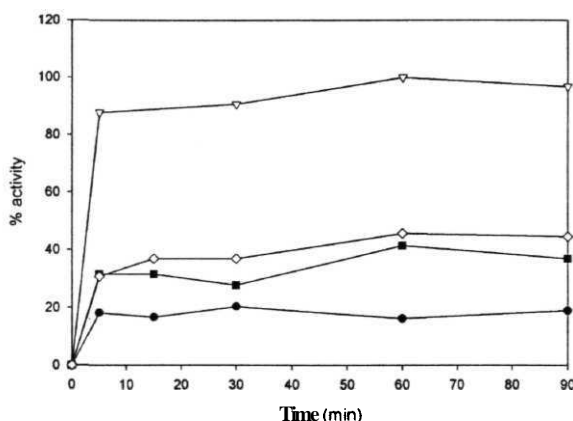
Figure 4.13 Effect of *M. tuberculosis* Cpn60s on refolding of chemically denatured citrate synthase Refolding of denatured citrate synthase was monitored in the presence of *E. coli* GroEL, GroES and ATP (-V-); *M. tuberculosis* Cpn60.1 (–◇–) and Cpn60.2 (-•-). Spontaneous refolding of unfolded citrate synthase was monitored similarly (-•-) as described in Experimental Procedures. Refolding of denatured citrate synthase by *M. tuberculosis* Cpn60s was performed in presence of *M. tuberculosis* Cpn10 and ATP. Activity obtained when denatured citrate synthase was refolded in presence of *E. coli* GroEL for 60 minutes was considered as 100%.

The results therefore suggest that the *M. tuberculosis* Cpn60s are efficient in preventing aggregation of denatured proteins, but do not possess ATP-dependent chaperoning activity.

4.4 CONCLUSIONS

GroEL is the major heat shock protein present in all forms of life guiding several essential steps during synthesis, folding, transport and degradation of proteins. In *E. coli*, *groEL* is known to be an essential gene for growth at all temperatures (Fayet *et a/.,* 1989). Upon thermal stress, nearly 15% of the normal protein mass of the cell consists of the GroEL and GroES proteins (Herendeen *et a/.,* 1979). Over 150 homologues of Cpn60 sequences are currently available with a pair-wise similarity extending from 40 to 100% at the amino acid level (Brocchieri and Karlin, 2000). High conservation of Cpn60s across species suggests that they play an important role in physiology of all species.

Heat shock proteins, including Cpn60s, are not only induced under thermal stress, but also show elevated expression levels under a variety of other unnatural conditions. Not surprisingly therefore, heat shock proteins are induced within pathogenic organisms upon invasion of host cells, presumably contributing to their survival within the hosts (Monahan *et al.,* 2001). Study of regulation of heat-shock proteins in *M. tuberculosis* has shown enhanced expression of both the Cpn60s upon thermal shock to the bacteria (Stewart *et at,* 2002), as well as upon phagocytosis by macrophages (Monahan *et at,* 2001). It is therefore reasonable to believe that Cpn60s of *M. tuberculosis* contribute to its defensive response against external stress conditions.

The study shows interesting biochemical characteristics of the two Cpn60s of *M. tuberculosis.* A major surprising outcome of the study has been that the two Cpn60s are distinct from the canonical GroEL homologues. The most distinctive feature of *M. tuberculosis* Cpn60s is their oligomeric nature, where unlike the known Cpn60s, they do not form a 14-mer oligomeric assembly. The proteins rather exist as dimers irrespective of the presence or absence of nucleotides. Several attempts to reconstitute the *M. tuberculosis* Cpn60.1 or Cpn60.2 under a variety of conditions did

not yield higher oligomers. The lower oligomeric state of *M. tuberculosis* Cpn60s also appears to be their natural state of existence as shown by our Western blot analysis on Native PAGE. Current understanding of the chaperonin function assumes a strict oligomeric assembly of GroEL, a **tetradecameric** state, which is the functional unit for protein refolding (White *et al.,* 1995; Weber *et al.,* 1998). Thus, the functional relevance of the **dimeric** state of **mycobacterial** chaperonins is still not clear.

Interestingly, many of the highly conserved residues, listed with "conservation index" better than 0.4 of Cpn60s (Brocchieri and Karlin, 2000) are observed to be different in *M. tuberculosis* chaperonins. For example, a crucial **Glutamate** at position 76, involved in an important side chain - main chain interaction across two monomers, is conserved in all Cpn60 sequences (Fig. 4.7, see Fig. 2.8 for the multiple sequence alignment). The conserved Glutamate is, however, replaced by a Serine in *M. tuberculosis* Cpn60.1. Several other alterations have been observed to occur at the interface of Cpn60.1 (Fig. 2.8, Chapter 2). Similarly, presence of an Alanine at position 2 has earlier been reported to be crucial for maintenance of the oligomeric state of GroEL. Mutation of Alanine to Serine at this position was shown to weaken the intersubunit interactions in GroEL destabilizing its oligomeric structure (Horowitz *et al.,* 1993). Interestingly, this position is occupied by a Serine in *M. tuberculosis* Cpn60.1. It is thus hypothesized that crucial changes in the interface residues might have resulted in the loss of oligomerization of the *M. tuberculosis* Cpn60.1. In an attempt to verify this hypothesis, site directed mutagenesis of these residues was performed in the *E. coli* GroEL. Urea induced unfolding transition show that the mutation A2S/E76S indeed destabilizes the oligomeric structure of GroEL. Similarly, other identified positions in Cpn60.1, which largely occur at oligomeric interface in GroEL, might further contribute to the loss of 14-meric state of Cpn60.1.

Another interesting observation in the study is the loss of the ATPase activity of the *M. tuberculosis* Cpn60s. The protein folding cycle of GroEL has been reported to be largely dependent on the ATPase activity of the protein. Evidence indicating

that the *M. tuberculosis* chaperonins have lost their ATPase activity is presented. Interestingly however, despite the loss of canonical oligomeric state and ATPase activity, the *M. tuberculosis* Cpn60s retain their ability to suppress aggregation of substrate **proteins. However, the extent to which the** two Cpn60s suppress aggregation is different, possibly due to variation in the exposed hydrophobic surfaces of the two proteins. The results indicate that Cpn60.1 is more potent in preventing aggregation than Cpn60.2. For substrate proteins such rhodanese, Cpn60.2 is able to bind these proteins reversibly giving them a chance to fold to their native conformation. Cpn60.1, on the other hand, remains tightly bound to the substrate protein. The substrate protein is thus prevented from aggregating but does not get refolded. Substrates such as citrate synthase on the other hand by reversible binding to Cpn60s are refolded more efficiently.

The ability of *M. tuberculosis* Cpn60s to promote refolding in the absence of ATP is similar to that observed for other known chaperones. For example, functionally active monomeric minichaperones containing part of the polypeptide binding domain of GroEL are known to be effective *in vitro* and *in vivo* (Zahn *et al.,* 1996; Chatellier *et al.,* 1998). Also, monomeric Cpn60 from *Thermus thermophilus* has been shown to possess the ability to suppress aggregation and promote protein folding with no requirement for ATP or GroES (Taguchi *et al.,* 1994). Thus, *M. tuberculosis* Cpn60s too, in a similar fashion might have evolved to function in lower oligomeric state with no ATP requirement.

A recent study surprisingly shows the presence of the mycobacterial Cpn60s in the plasma membrane fraction although the major functional role of these proteins lies in the cytosol (Sinha *et al.,* 2002). Moreover, various other reports suggest alternative roles for the GroEL homologues. For example, a GroEL homologue in Buchnera, an intracellular symbiotic bacterium of aphids, is not only a molecular chaperone but also a phosphocarrier protein, suggesting that the protein plays a role in a signal transducing system (Matsumoto *et al.,* 1999). The *M. leprae* Hsp65

(Cpn60.2) has been shown to display proteolytic activity that is catalytically related to the HslVU protease (Portaro *et al.,* 2002). By similarity, the *M. tuberculosis* Cpn60.2 too may have gained such a role.

Judicial utilization of energy sources in a cell is important to the organism for its survival. The protein folding cycle by the Cpn60s in all life forms is an ATP dependent pathway utilizing seven ATP molecules in each reaction cycle (Todd *et al.,* 1994). Ability of an organism to promote protein folding without energy utilization would prove to be economical. It is hypothesized that high demands of energy resulting from its extremely slow rate of metabolism have led *M. tuberculosis* to devise an altered route to protein folding, avoiding the usual energy dependent pathway. Cpn60s, through such a route, thus continue to fold substrate proteins within the cytosol sparing ATP for other crucial processes of the cell. Thus loss of an oligomeric state of Cpn60s of *M. tuberculosis,* yet retaining their chaperoning role is suggestive of a design in evolution to save the energy sources of the bacterium. It is pertinent to note in this context that other ATP-dependent enzymes, such as RecA, also have reduced ATPase activity in *M. tuberculosis* (Datta *et al.,* 2000). Further analysis of the role of GroEL homologues would lead to a better understanding of the importance of the presence of two Cpn60s in *M. tuberculosis.*

4.5 REFERENCES

1. Arai, M., Inobe, T., Maki, K., Ikura, T., Kihara, H., Amemiya, Y., and Kuwajima, K. (2003). Denaturation and reassembly of chaperonin GroEL studied by solution X-ray scattering. *Protein Sci.* 12, 672-680.

2. Bachman, B.J. (1990). Linkage map of *Escherichia coli* K-12, Edition 8. *Microbiol. Rev.* 54, 130-197.

3. Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R, Rudolph, R., Schmid, F.X., and Kiefhaber, T. (1991). GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* 30, 1586-1591.

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

5. Chatellier, J., Hill, F., Lund, P., and Fersht, A.R. (1998). *In vivo* activities of GroEL minichaperones. *Proc. Nat/. Acad. Sci. USA.* 95, 9861-9866.

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

7. Datta, S., Prabu, M.M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2000). Crystal structures of Mycobacterium tuberculosis RecA and its complex with ADP-AlF4: implications for decreased ATPase activity and molecular aggregation. *Nucl. Acids Res.* 28, 4964-4973.

8. Dickson, R., Weiss, C., Howard, R.J., Alldrick, S.P., Ellis, R.J., Lorimer, G., Azem, A., and Viitanen, P.V. (2000). Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. /. *Bio/. Chem.* 275, 11829-11835.

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

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

11. Fukami,T.A., Yohda, M., Taguchi, H., Yoshida, M., and Miki, K. (2001). Crystal structure of chaperonin-60 from *Paracoccus denitrificans*. *J. Mol. Biol.* **312,** 501-509.

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

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

14. Horovitz, A., Bochkareva, E. S., Kovalenko, O., and Girshovich, A.S. (1993). Mutation Ala2—>Ser destabilizes intersubunit interactions in the molecular chaperone GroEL. *J.Mol. Biol.* **231,** 58-64.

15. Hubner, P., Dame, G., Sandmeier, U., Vandekerckhove, J., Beyer, P., and Tadros, M.H. (1996). Molecular analysis of the *Rhodobacter capsulatus* chaperonin (*groESL*) operon: purification and characterization of Cpn60. *Arch. Microbiol.* 166, 193-203.

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

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

18. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallography* 24, 946-950.

19. Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and 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. *EMBOJ.* 11,4757-4765.

20. Lewthwaite, Jo C., Skinner, A., and Henderson, B. (1998). Are molecular chaperones microbial virulence factors? *Trends Microbiol.* 6, 426-428.

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

22. Lissin, N.M., **Venyaminov,** S.Yu., and Girshovich, A.S. (1990). (Mg-ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature* **348,** 339-342.

23. Matsumoto, **K.,** Morioka, M., and Ishikawa, H. (1999). **Phosphocarrier** proteins in an intracellular symbiotic bacterium of **aphids.** *J. Biochem. (Tokyo)* **126,** 578-583.

24. Mazodier, P., **Guglielme,** G., Davies, J., and Thompson, C.J. (1991). Characterization of the *groEL*-like genes in *Streptomyces a/bus. J. Bacteriol.* **173,** 7382-7386.

25. **Monahan,** L, Betts, J., **Banerjee,** D., and Butcher, P. (2001). Differential expression of mycobacterial proteins following phagocytosis by **macrophages.** *Microbiology* **147,** 459-471.

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

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

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

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

30. **Seale, J.W.**, Gorovits, B.M., Ybarra, J., and Horowitz, P.M. (1996). Reversible oligomerization and **denaturation** of the chaperonin GroES. *Biochemistry* 35, 4079-4083.

31. Sinha, **S.**, Arora, S., Kosalai, K., **Namane,** A., **Pym,** A.S., and Cole, S. T. (2002). **Proteome** analysis of the plasma membrane of *Mjcobacterium tuberculosis.* *Comp. Funct. Genom.* 3, 470-483.

32. Sorbo, B.H. (1953). Crystalline rhodanese. I. Purification and **physiochemical** examination. *Acta.Chemica.* *Scan.* 7, 1129-1136.

33. Srere, P.A. (1966). Citrate-condensing **enzyme-oxalacetate** binary complex. Studies on its physical and chemical properties. /. *Bio/. Chem.* **241,** 2157-2165

34. Stewart, G.R., **Wernisch,** L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B., and Butcher, P.D. (2002). Dissection of the heat-shock response in *Mjcobacterium tuberculosis* using mutants and microarrays. *Microbiology* **148,** 3129-3138.

35. Taguchi, H., **Makino,** Y., and Yoshida, M. (1994). Monomeric chaperonin-60 and its 50-kDa fragment possess the ability to interact with non-native proteins, to suppress aggregation, and to promote protein folding. *J. Biol.* *Chem.* **269,** 8529-8534.

36. Taneja, B., and Mande, S. C. (2001). Metal ions modulate the plastic nature of *Mycobacterium* tuberculosis **chaperonin-10.** *Prat. Eng.* **14,** 391-395.

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

38. Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from **polyacrylamide** gels to nitrocellulose sheets: procedure and some applications. *Proc.* *Nat/. Acad. Sci.* *USA.* 76, 4350-4354.

39. Viitanen, P.V., Lubben, T.H., Reed, J., Goloubinoff, P., O'Keefe, D.P., and Lorimer, G.H. (1990). Chaperonin-facilitated refolding of ribulosebisphosphate

carboxylase and ATP hydrolysis by Chaperonin 60 (GroEL) are K⁺ dependent. *Biochemistry* 29, 5665-5671.

40. Viitanen, P.V., Lorimer, G.H., Seetharam, R., Gupta, R.S., Oppenheim, J., Thomas, J.O., and Cowan, N.J. (1992). Mammalian mitochondrial chaperonin 60 functions as a single toroidal ring./. *Bio/. Chem.* **267,** 695-698.

41. White, Z.W., Fisher, K.E., and Eisenstein, E. (1995). A monomeric variant of GroEL binds nucleotides but is inactive as a molecular chaperone. *J. Bio/. Chem.* **270,** 20404-20409.

42. Weber, F., Keppel, F., Georgopoulos, C, Hayer-Hartl, M.K., and Hartl, F.U. (1998). The oligomeric structure of GroEL/GroES is required for biologically significant chaperonin function in protein folding. *Nat. Struct. Bio/.* 5, 977-985.

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

44. Ybarra, J., and Horowitz, P.M. (1995). Refolding and reassembly of active chaperonin GroEL after denaturation.*J. Bio/. Chem.* **270,** 22113-22115.

45. Zahn, R., Buckle, A.M., Perrett, S., Johnson, C.M., Corrales, F.J., Golbik, R., and Fersht, A.R. (1996). Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proc. Nat/. Acad. Sci. USA.* 93, 15024-15029.

46. Zhi, W., Landry, S.J., Gierasch, L.M. & Srere, P.A. (1992). Renaturation of citrate synthase: influence of denaturant and folding assistants. *Protein Sci.* 7, 522-529

Chapter 5

Crystal Structure of Chaperonin60.2

of Mycobacterium tuberculosis

5.1   INTRODUCTION

Chaperonin60 (Cpn60), also commonly referred to as heat shock protein 60 (Hsp60), is one of the major molecular chaperones present ubiquitously in all forms of life.   These molecular chaperones are known to assist the folding, assembly and transport of several cellular proteins (Hartl, 1996). Cpn60s have been shown to be over-expressed under a variety of unnatural conditions such as thermal stress, hypoxia, nutrient deprivation, phagocytosis etc. Invasion of a host being an apparent form of a stress, induction of Cpn60s has also been observed within pathogenic organisms, presumably contributing to their survival within the host. The overexpressed pathogen derived Cpn60s act as major antigens resulting in strong immune responses from the host (Zügel and Kaufmann, 1999).

In *Escherichia coli,* the chaperonin, GroEL, has been shown to be essential for growth at all temperatures (Fayet *et a/.,* 1989). The *E. coli* chaperonin has provided a paradigm in understanding protein folding mechanisms mediated by the chaperonins (Xu *et al.,* 1997). GroEL promotes *de novo* folding of $\sim$ 10-15% of all proteins in bacterial cytosol in co-ordination with the heptameric co-chaperonin, GroES (Ewalt *et al.,* **1997).**

X-ray studies combined with electron microscopy studies have provided valuable insights into the functional cycle of this chaperonin. (Braig *et al.,* 1994, Xu *et aL,* 1997 ; Chen *et al.,* 1994). Crystal structures of the unliganded GroEL and the GroEL-GroES complex show a cylindrical complex with subunits of GroEL assembled into two heptameric rings stacked back-to-back to form the native 14-mer. The two rings enclose a large central channel that facilitates proper protein folding in an ATP-dependent manner (Braig *et aL,* 1994; Boisvert *et aL,* 1996; Xu *et aL,* 1997). Crystal structure of Cpn60 from *Paracoccus denitrificans* also shows a similar arrangement of Cpn60 subunits as a tetradecamer (Fukami *et aL,* 2001).

GroEL is assisted in its function by a 10kDa co-chaperonin, GroES. The co-chaperonin exists as a heptamer and adopts a dome-like structure that can bind to either GroEL ring to enclose the central cavity (Hunt *et al.*, 1996; Langer *et al.*, 1992). GroES acts as a lid to seal off the folding chamber and helps displace bound substrate protein into the cavity where the protein can undergo productive folding. Binding of GroES to GroEL is dependent on adenine nucleotide. ATP dependent conformational changes in GroEL have been shown to be necessary for proper chaperonin function *in vivo* (Rye *et a/.,* 1997).

In most prokaryotic organisms, a single copy of *cpn60* gene usually occurs on the chromosome and is found on the *groESL*operon along with *cpn10*. Some of the actinomycete species, however, are known to express multiple *cpn60* genes. For example, *Mycobacterium tuberculosis* contains two copies of the *cpn60* genes (Kong *et al.*, 1993). One of these, *cpn60.1*, is organized on an operon with *cpn10*, while the second, *cpn60.2,* is arranged separately on the chromosome. Study of regulation of heat-shock proteins in *M. tuberculosis* has shown overexpression of the two Cpn60s upon thermal shock to the bacterium (Stewart *et al.*, 2002), as well as upon phagocytosis by macrophages (Monahan *et al.*, 2001). It is therefore reasonable to believe that Cpn60s of *M. tuberculosis* contribute to its defensive response against external stress conditions. Moreover, both Cpn60.1 and Cpn60.2 of *M. tuberculosis* have been shown to be highly antigenic and potent cytokine inducers (Lewthwaite *et al.*, 2001; Young and Garbe, 1991). The Cpn60s therefore might represent important components of *M. tuberculosis,* playing roles as immunomodulators, and perhaps also required for proper protein folding and transport.

Although the biochemical and functional properties of gene product of *cpn60* occurring on the *grvESL* operon have been studied extensively, the necessity of a duplicate *cpn60* gene in certain bacteria is not known. Functional role of the second copy of Cpn60 as a chaperonin is poorly understood. A recent study suggests a proteolytic role for the second Cpn60 in *M. leprae* (Portaro *et al.*, 2002). The second

copy of Cpn60 might thus possibly play alternate roles in the physiology of bacteria possessing multiple *cpn60* genes. Structural studies on the second Cpn60 would therefore provide valuable insights into its functional properties. In this study the crystal structure of the *M. tuberculosis* Cpn60.2 and its relevance to the functional properties is presented.

5.2   EXPERIMENTAL PROCEDURES

5.2.1 Cloning, Expression and Purification of Cpn60.2

The full-length gene coding for the *M. tuberculosis* GroEL homologue, Cpn60.2, was PCR amplified as described in Chapter 2. The gene was over-expressed in *E. coli* BL21 (DE3) and purified by Ni-NTA affinity chromatography. Protein for crystallization was dialyzed extensively against 10mM Tris-Cl, pH 8.0 and concentrated to a final concentration of 14mg/ml.

5.2.2 Protein Analysis by Immunoblotting

Native proteins of *M. tuberculosis* were resolved on a 10% SDS PAGE. The proteins were then transferred onto HybondC (Amersham) nitrocellulose sheet by the method of Towbin *et al.* (1979). The anti-Cpn60.2 monoclonal antibody, mAb67-2, was used for the detection of proteins. HRP-labelled goat anti-mouse IgG was used at a dilution of 1:10,000. The proteins were visualized using standard protocols.

5.2.3 Crystallization and Data Collection

Crystals were grown at room temperature by hanging drop vapour diffusion method against a reservoir solution containing 100mM HEPES, pH 7.5, 25% PEG 3350 and 10% n-propanol. These conditions were similar to those reported by Adir *et al.* (2002). Crystals were frozen after cryoprotection with the addition of 10% glycerol to the mother liquor. X-ray diffraction data for Cpn60.2 were recorded, for two crystals, under standard cryogenic conditions at Beamline BL1 at the Protein Structure Factory, BESSY. The data were processed, scaled and merged with DENZO and SCALEPACK (Otwinowski and Minor, 1997).

5.2.4 Structure Determination and Refinement

The structure of Cpn60.2 was solved by molecular replacement method using the program AMoRe (Navaza, 1994) with crystal structure of *E.* rt>//GroEL (PDB ID: 1 kp8) as the search model. The search model was a monomer of the *E. coli* GroEL. The rotation and translation parameters of the second molecule were obtained by fixing those of the first molecule. Refinement was undertaken using REFMAC5 (Murshudov *et a/.,* 1997) and CNS (Brünger *et al.,* 1998). Initial refinement consisted of rigid body, positional and B factor refinement. Models were rebuilt in O (Jones *et a/.,* 1991) using σ_A weighted 2Fo-Fc and Fo-Fc electron density maps. The progress of refinement was monitored by crystallographic R-factor, and parameterization at each stage was monitored throughout using R_{free} from a set of randomly selected 5% reflections. The quality of the structure was monitored using PROCHECK (Laskowski *et a/.,* 1993). Refined coordinates and structure factors have been submitted to the Protein Data Bank with the accession code 1SJP.

5.3 RESULTS

5.3.1 Purification and Determination of Quaternary Structure of Cpn60.2

Cpn60.2 was expressed in large quantities and exhibited a protomer molecular mass of ~60kDa (Fig. 2.7b, Chapter 2). Intriguingly during size exclusion chromatography, the purified Cpn60.2 eluted as a homogenous dimer, unlike the tetradecameric *E. coli* GroEL (Fig. **4.1b**, Chapter 4). This finding was surprising since Cpn60.2 shares a high sequence homology with other chaperonins that exist as higher oligomers. Attempts to reconstitute native-like tetradecamers *in vitro,* or by following different purification procedures described in Chapter 4, did not yield the tetradecameric protein suggesting that Cpn60.2 might exist as a dimer physiologically.

SDS-PAGE analysis of the purified Cpn60.2 had earlier shown that the protein migrated as a doublet with the high molecular weight band corresponding to ~60kDa (Fig. 2.7b, Chapter 2). N-terminal sequencing revealed the higher molecular weight band sequence to be AKTIAYD, while that of the low molecular weight band to be WGAPTITNTG. The former corresponds to the intact protein, while the latter corresponds to Cpn60.2 truncated at K41. In an attempt to prevent degradation of Cpn60.2, several combinations of protease inhibitors, such as leupeptin, aprotinin, soyabean trypsin inhibitor, PMSF or EDTA, were included during purification. However, none of the inhibitors, used separately or in a cocktail, prevented the proteolytic cleavage

The observation that cleavage of Cpn60.2 could not be prevented even in the presence of standard protease inhibitors was suggestive of its intrinsic instability. In an attempt to assess the stability of Cpn60.2 in its native environment, *M. tuberculosis* extracts were probed with anti-Cpn60.2 monoclonal antibody. Appearance of

degraded products in Western blot analysis of *M. tuberculosis* extracts confirmed its tendency to degrade (Fig. 5.1).

– 60kDa
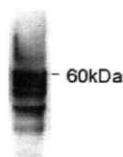
Figure 5.1 Western blot analysis of native *M. tuberculosis* proteins Soluble cell wall fraction of *M. tuberculosis* was resolved on SDS PAGE and probed with anti-Cpn60.2 antibody. Western blot analysis indicates the degradation of Cpn60.2 in the native extracts.

These results are in accordance with a previous study showing degradation of mycobacterial Cpn60.2 (De Bruyn *et al.*, 2000). A high propensity of Cpn60.2 to degrade suggests that the protein is intrinsically unstable.

The altered quaternary state of Cpn60.2, different than the expected 14-meric association, probably arose because of the protein's degradation. In the tetradecameric GroEL, inter-subunit contacts are formed by a (3-sheet contributed by the N- and C- termini of neighboring subunits. Earlier reports have confirmed that mutations at the N-terminus of GroEL destabilize the oligomer (Horowitz *et al.,* 1993a; 1993b). Cleavage of Cpn60.2 at K41 possibly led to the loss of these crucial interactions and hence its failure to form a tetradecamer.

Although Cpn60.2 appeared to have lost its canonical oligomeric state, it was able to act as a molecular chaperone. As described in Chapter 4, when present in molar excess ratio Cpn60.2 was able to completely prevent the aggregation of citrate synthase at 43°C.

Prevention of aggregation of citrate synthase suggested exposure of hydrophobic surface in the Cpn60.2 structure. Presence of such hydrophobic

patches on the surface of Cpn60.2 was probed by binding to bis-ANS. Interaction between bis-ANS and Cpn60.2 was accompanied by a significant fluorescence enhancement clearly suggesting the presence of surface exposed hydrophobic patches in Cpn60.2 (Fig. 4.10, Chapter 4). These hydrophobic patches in Cpn60.2 thus might serve as binding sites for non-polar molecules.

5.3.2 Crystallization, Structure Determination and Refinement

Cpn60.2 was crystallized in space group P2$_1$. MALDI-TOF analysis of the crystals indicated a molecular mass of ~53kDa, smaller than the expected size and corresponded to the fragment truncated at K41. Thus, although Cpn60.2 used for crystallizations was a mixed population of the intact and the truncated protein, it was the latter that selectively crystallized. Matthews number suggested the presence of a **dimer** in the crystallographic asymmetric unit Useful diffraction data could be collected on crystals soaked in glycerol for a short period and frozen in liquid N2. Long soaks in glycerol, or other cryoprotectants, worsened the diffraction quality significantly. **Data** from two crystals could be merged well with an overall completeness of more than 99% (Table 3.1, Chapter 3 and Table 5.1).

Table 5.1 Data collection statistics of the merged Cpn60.2 crystals

Data collection	
Space group	P2,
a (A)	58.7
b(A)	113.8
c(A)	79.5
β(°)	94.6
Matthews number	2.4
No. of unique reflections	17,489
Redundancy	5(3)
Completeness (%)	99.5 (97)
Average I/σ(I)	20.2 (2.6)
R$_{merge}$	10.0(41.5)

Molecular replacement calculations using *E. coli* GroEL monomers as the search models were attempted. *E. coli* GroEL is known to exist in two distinct conformational states - one corresponding to *cis,* the GroES bound state, and *trans,* the uncomplexed form. Several crystal structures of GroEL are available in the Protein Data Bank. However, none of these yielded a convincing molecular replacement solution. Finally, the best solution in molecular replacement was obtained using the *E. coli* (GroEL-K$^+$-Mg^{2+}-ATP)$_{14}$ complex structure (1kp8) as the model (Wang and Boisvert, 2003). The rotation function showed a weak signal with the highest solution being only 4(7 above the background, and the best translation peak with a correlation coefficient and R-factor of 0.64 and 0.56, respectively. This solution was only marginally better than the other possible solutions. The solution improved when the second molecule was also placed appropriately in the unit cell, with the correlation coefficient and R-factor improving to 0.68 and 0.53, respectively. The two molecules were seen to be related by approximately 173° rotation along the a-axis.

The first round of rigid body refinement showed a major conformational change from the starting coordinates. The apical and intermediate domains, as defined for the *E. coli* GroEL structure, showed a dramatic conformational change with respect to the equatorial domain. The movement was observed in both the molecules present in the asymmetric unit After several iterative rounds of refinement and manual rebuilding in O (Jones *et a/.,* 1991), the R and R$_{free}$ converged to 24.1% and 28.5%, respectively. The final structure showed an excellent geometry with 87.2% of the residues within the most favored regions of the Ramachandran plot (Table 5.2).

Table 5.2 Refinement statistics of Cpn60.2

Refinement statistics	
No. of protein atoms	6595
Resolution range (A)	50-3.2
No. of reflections used in refinement	**16,335**
No. of reflections used for calculation of R$_{free}$	870
R$_{work}$ (%)	24.1
R$_{free}$ (%)	28.5
Root mean square deviation	
Bond distances (A)	0.008
Bond angles (°)	1.003
Torsion angles (°)	3.017
Planarity (A)	0.004
Average B factor (A^2)	82
Main chain (A^2)	80
Side chain (A^2)	84
Ramachandran plot most favored (%)	87.2
PDB ID code	1SJP

R$_{work}$ = Σ | F$_{obs}$ - F$_{calc}$ | / Σ | F$_{obs}$ |; R$_{free}$ = R factor calculated from a subset of reflections excluded from refinement

5.3.3 Overall Structure

The final refined structure consists of residues 60-514 in chain A and 62-514 in chain B. No interpretable density could be obtained for residues 42-59 in chain A, 42-61 in chain B and residues 80-87 and 515-519 in both the chains. These regions were hence not modeled. The final model therefore comprises 447 **amino** acid residues in chain A and 445 residues in chain B.

The overall molecular architecture of the protein, as shown in Fig. 5.2, was similar to that of *E. coli* GroEL with the protein comprising of three distinct domains, equatorial, intermediate and apical, with the domain nomenclature as in *E. coli* GroEL (Braig et al, 1994.).
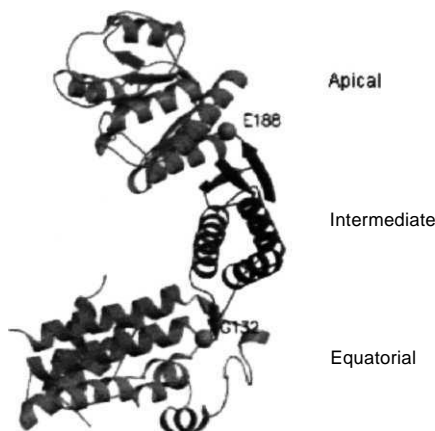
Figure 5.2 Overall molecular architecture of Cpn60.2 Cpn60.2 shows a fold similar to *E. coli* GroEL with the protein divided into three domains, equatorial, intermediate and apical. The hinge residues resulting in the movement across the domains with respect to GroEL are shown as spheres. Figures 5.2, 5.3 and 5.7 were generated using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

Size exclusion chromatography had earlier suggested the molecule to be a homogenous dimer (Fig. 4.1b, Chapter 4). This observation was also confirmed using Native PAGE (discussed in Chapter 4). Buried accessible area between two molecules in asymmetric unit was estimated to be 1384A^2. This value, although not very high for the formation of a strong dimer, appears to be comparable to the typical values observed in protein oligomers. The two molecules in the asymmetric unit interact with each other principally through their apical domains. Dimerization is mediated by several main chain as well as side chain charge - charge interactions across the two subunits. Several van der Waals, hydrophobic and hydrogen bonding interactions are also observed. Thus, the two Cpn60.2 molecules might represent a physiological dimer.

Intriguingly, whereas the two molecules in the asymmetric unit share a large solvent inaccessible interface, unlike anticipated they are not related by closed point group symmetry. The two monomers are related by a 173° rotation. The quaternary structure therefore can be defined as "open". Although not favored, the "open" quaternary structure has earlier been observed in a few protein structures (Banerjee *et a/.,* 1994; Skordalakes and Berger, 2003). As a consequence of lack of closed point group symmetry two types of non-crystallographically related pairs of dimers exist. Interactions between the two different dimer pairs are hence not identical. One of the two pairs buries an area of $1384A^2$ area, while the other pair buries $1039A^2$. The pair burying a larger accessible area was considered as the physiological dimer (Fig. 5.3).

Figure 5.3 Association of Cpn60.2 subunits in the crystal lattice Chain A is coloured in green. Area buried between the subunits green and pink is $1384Å^2$ while that between green and blue is $1039A^2$. The dimer constituted of the green and pink coloured chains was considered as the physiological dimer.

5.3.4 Comparison Between Two Cpn60.2 Molecules

An overall superposition between the two molecules in the dimer showed a somewhat high rms deviation of 1.71 A for 439 equivalent C^α atoms (Fig. 5.4). The high rms value resulted from differences in relative orientations of the three domains with respect to one another, with the largest deviations occurring in the apical domain.   Of the regions that deviate the most, residues 349-359 are involved in crystal packing.   Crystal packing requirements thus seem to enforce local conformational changes in the two molecules within the asymmetric unit   The individual domain-wise superposition resulted in lower rms deviations.   The rms deviation for the equatorial domain was 0.74A for 170 atoms, 0.62A for 92 atoms in the intermediate domain and 0.73A for 183 atoms in the apical domain.   The individual domains of Cpn60.2 thus seem to retain a high degree of similarity to each other.
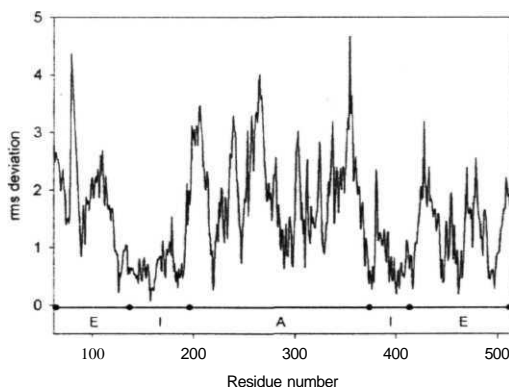
Figure 5.4 Plot of rms deviation (in A) between chains A and B of Cpn60.2 The apical, intermediate and equatorial domains are indicated as A, I and E, respectively.

The overall B-factor of the two chains is $82A^2$. The high temperature factor corresponds approximately to the B-value obtained from Wilson scaling. Analysis of domain-wise temperature factors showed interesting differences between the two molecules (Fig. 5.5, Table 5.3).
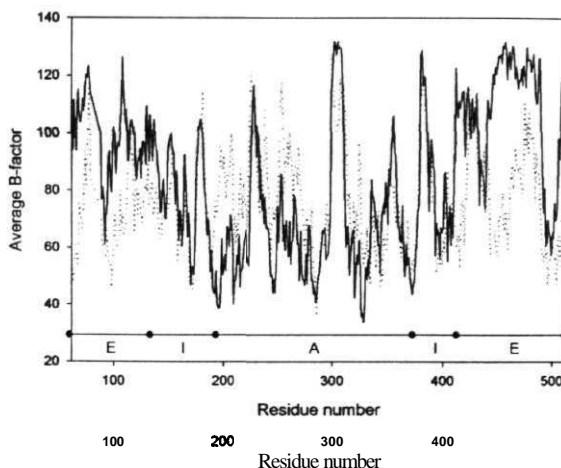
Figure 5.5 Plot of average residue temperature factor (in A^2) of chains A and B of Cpn60.2 The dotted line represents Chain A while the solid line represents Chain B. The apical, intermediate and equatorial domains are represented by A, I and E, respectively.

Temperature factors of all the three domains of chain A are similar to the average temperature factor (Table 5.3). On the other hand in chain B, the apical domain shows a lower average B-value of $69A^2$ and the equatorial domain shows a higher value $104A^2$. The reason for this difference in temperature factors is, however, not obvious. The higher disorder in equatorial domain as compared to the apical domain is perplexing, since in all the previous structures of GroEL and its homologues, the apical domain was found to be more dynamic than the equatorial domain.

Table 5.3 Domain-wise average temperature factors for the two Cpn60.2 chains

Domain	Chain A	Chain B
Apical	79	69
Intermediate	81	85
Equatorial	78	104

5.3.5 Comparison of Cpn60.2 with *E. coli* GroEL

When compared with the GroEL structure in the (GroEL-K$^+$-Mg^{2+}-ATP)$_{14}$ complex, Cpn60.2 showed rms deviation of 1.81 A for 339 equivalent C$^\alpha$ atoms. Comparison of individual domains of the two proteins showed better superposition with rms deviations of 0.86A for the equatorial domain (146 atoms), 1.28A for the intermediate domain (88 atoms) and 0.78A for the apical domain (181 atoms). The largest deviation was observed in the helix 60-79. Orientation of this helix was dramatically changed when compared to the *E. coli* GroEL. This change was evident in the early stages of refinement with clear indication from the electron density maps (Fig. 5.6). The possible reasons for this conformational difference are discussed later.
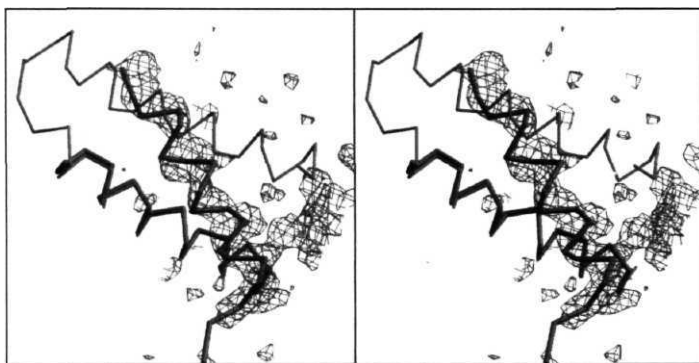
Figure 5.6 Conformational change in the 60-79 helix of Cpn60.2 | | Fo |- | Fc| | map contoured at 1(7 level clearly show a different orientation for this helix. The C$^\alpha$ trace for GroEL is shown in green, while that for Cpn60.2 is in magenta. This Figure was generated using Bobscript (Esnouf, 1999) and Raster3D (Merritt and Bacon, 1997).

Cpn60.2 has a more open tertiary conformation than the *E. coli* GroEL (Fig 5.7). This open conformation resulted from a large *en bloc* motion of the apical

domain away from the equatorial domain and a concomitant inward movement of the intermediate domain.
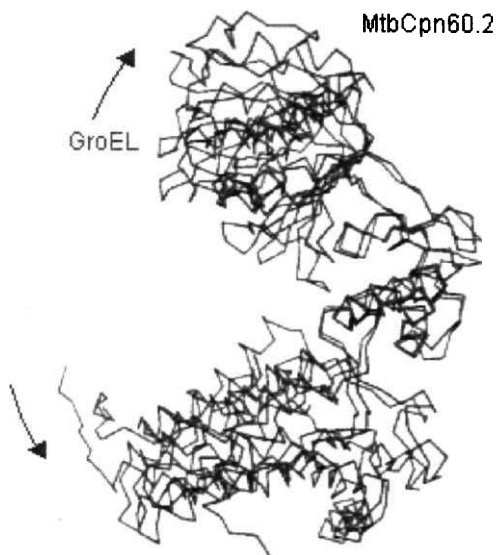
Figure 5.7 Superposition of Cpn60.2 with *E. coli* GroEL The two molecules were superposed with respect to the intermediate domain. Comparison reveals the outward movements of the apical and equatorial domains in Cpn60.2 with respect to GroEL. The outward movement is more prominent for the apical domain.

These movements were considered as rigid body motions across two distinct hinge points, one at the interface of the apical and intermediate domains, and the other at the interface of the intermediate and equatorial domains. The hinge points were identified by inspection of difference dihedral angle plots (Fig. 5.8).
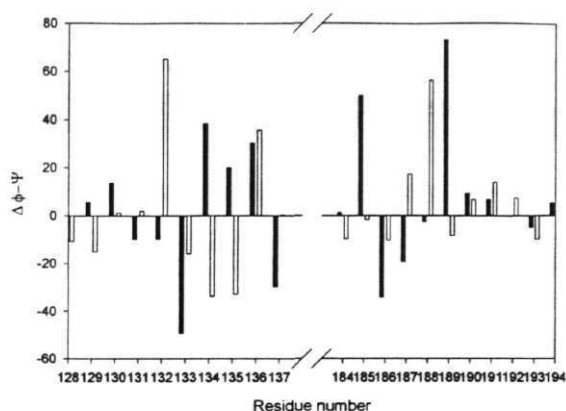
Figure 5.8 Difference φ-ψ plot of Cpn60.2 upon comparison with GroEL The maximal **deviations** occur at residues 132 and 133 between the equatorial and intermediate domains, and 188 and 189 between intermediate and apical domains. Solid bar indicates the Δφ while the open bar shows Δψ.

At the interface of the intermediate and equatorial domains, the plot showed maximal deviation at the ψ and φ of G132 and A133, respectively (Fig. 5.8). Changes in the dihedral angles at residues G132 and A133 may be considered as a rigid body rotation across the virtual C^α- C^α bond between these residues. These alterations in the dihedral angles resulted in a rotation of 14° of the intermediate domain towards the equatorial domain. A concomitant 19° swing of the apical domain relative to the intermediate domain resulted in an outward movement of the apical domain at residues E188 and G189. These differences once again may be considered as a rotation across the virtual C - C bond between E188 and G189. Corresponding residues in the *E. coli* GroEL-GroES complex bring about similar movements across the three domains. Differences in the relative arrangement of different domains in Cpn60.2, when compared with GroEL, might have arisen due to the different quaternary associations of subunits.

The interfaces between the domains are remarkably populated with charged residues. Sequence comparison of the known Cpn60s indicated high conservation of these charged residues suggesting the importance of the charged interfaces (Brocchieri and Karlin, 2000). In the Cpn60.2 structure, these residues exhibit extensive interaction networks. R318 of the apical domain is involved in interdomain salt bridges with E185 and E175 of the intermediate domain. The orientations of the guanidino group of R318 and carboxylates of glutamates are determined by their interactions with E331 and K377, respectively. The other region, where extensive charge-charge interactions are observed involves residues E394 and K390 of the intermediate domain and R207 and E209 of the apical domain. Though K390 does not interact directly with E209, these residues juxtapose the charge groups of E394 and R207 in an appropriate orientation for strong interactions. The equivalent residues in *E. coli,* however, do not show similar interactions. Loss of these interactions in *E. coli* seems to have arisen due to the closed tertiary conformation of the protein. Interestingly, the charge network in Cpn60.2 continues across the other monomer with the involvement of R281 and D358 from the other subunit. The extensive charge-charge interaction networks therefore might stabilize the open tertiary conformation of Cpn60.2.

5.3.6 Consequences of Loss of N-terminal Residues on the Cpn60.2 Structure

Loss of the first 41 residues in the Cpn60.2 might have led to serious consequences on its structure due to exposure of large hydrophobic surface of the protein. However, an elegant conformational change results in an alteration in the orientation of the N-terminal helix (residues 60-79) (Fig. 5.6). This structural alteration, in both the molecules in the asymmetric unit, thus shields hydrophobic residues of the equatorial domain that would otherwise have been exposed. In addition, this conformational change appears to promote the packing of the helices spanning residues 87-107 and 60-79.

The equatorial domain also spans the **nucleotide-binding** pocket in *E. coli* GroEL and therefore the loss of first 41 residues in Cpn60.2 might affect binding of nucleotides (Boisvert *et a/.,* 1996). Several charged and small hydrophobic residues from region 480-500 (residue numbers as in *E. coli)* contribute to the ATP-binding site in GroEL. In addition, residues at the N-terminus of the protein also contribute to this site. For example, G31 is crucial for interaction with the α-Phosphate of ATP while P33 is involved in hydrophobic interactions with the adenosine moiety. Moreover, residues D87 and T91 in GroEL are involved in hydrogen bonding with **Mg²⁺** and **β-phosphate** of ATP. Absence of **N-terminal** 41 residues leads to the loss of the nucleotide-binding pocket in Cpn60.2. Furthermore, electron density of the loop (residues 83-90) involved in **β-phosphate** binding was not interpretable. The absence of ATP-binding site resulting from deletion at the N-terminus in the protein, and disorder in the 83-90 loop therefore might lead to the loss of ATPase activity of Cpn60.2.

Characterization of ATPase activity of Cpn60.2 had earlier shown that, in comparison with *E. coli* GroEL, Cpn60.2 did not exhibit any ATPase activity (Described in Chapter 4). Thus, the loss of N-terminal residues of Cpn60.2 seems to abolish its ATPase activity.

5.3.7 Surface Properties **of** Cpn60.2

Prevention of citrate synthase aggregation and binding to bis-ANS suggested presence of hydrophobic areas on Cpn60.2 surface. The crystal structure reveals that Cpn60.2 has at least two hydrophobic patches exposed on its surface. One of these is in the equatorial domain, and possibly arises due to the loss of first 41 residues (Fig. 5.9).
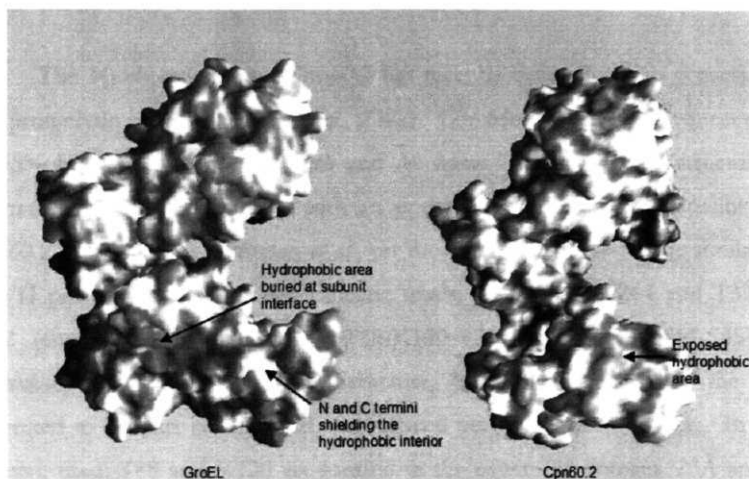
Figure 5.9 Comparison of surface properties of Cpn60.2 monomer with those of the GroEL monomer Hydrophobic regions exposed on the surface of GroEL and Cpn60.2 are indicated in green. While in GroEL, hydrophobic surfaces of a monomer bury upon oligomeric assembly of the protein, in Cpn60.2 these regions remain exposed even upon dimerization. The regions might thus serve to bind non-native polypeptides preventing their aggregation. This Figure was generated using GRASP (Nicholls, 1991).

As mentioned previously, conformational change involving the N-terminal helix (residues 60-79) partially shields this patch. However, some amount of hydrophobic surface still remains exposed. Another hydrophobic patch occurs in the apical domain. Its equivalent region in GroEL has been shown to bind substrate proteins. Moreover, light scattering experiments showed that Cpn60.2 prevents aggregation of citrate synthase when used in molar excess. Thus, the two hydrophobic patches in each of the Cpn60.2 monomers might serve a role in binding unfolded polypeptides thereby preventing their misfolding and aggregation.

5.3.8 Structural Comparison with **HslVU** Protease

The *Mycobacterium   leprae* Cpn60.2 has recently been shown to possess **HslVU-like** proteolytic activity (Portaro *et al.,* 2002). Considering the high sequence identity of 95% between the *M. tuberculosis* and *M. leprae* Cpn60.2, crystal structure of *M. tuberculosis* Cpn60.2 presents us with an opportunity to probe the possible role of Cpn60.2 in proteolysis. Portaro *et al.,* on the basis of 20% sequence similarity with HslVU protease, suggested two catalytic triads, **T136-K168-Y264** and T375-K409-S502 (corresponding to residues **T88-K120-Y216** and T327-K361-S454 in *M. tuberculosis).* The Cpn60.2 crystal structure, however, reveals that the residues suggested to be part of the triad were placed very far apart in space. In the first putative triad, T88 and K120 are located in the equatorial domain, 23A apart from one another, while Y216 is present on the apical domain and is 44A and 54A away from residues T88 and K120, respectively. The structure does not seem to suggest any possible conformational movement to bring these three residues together for the activity. In the second suggested triad, while residues T327 and K361 are present on the apical domain, 18A apart from one another, residue S454, in the equatorial domain, is 42A and 50A away from T327 and K361, respectively. Thus, although the suggested triads might not be responsible for the proteolytic role of Cpn60.2, some other residues might contribute to this activity.

Structural comparison of Cpn60.2 with that of the HslVU protease (Bochtler *et al.,* 2000) revealed that S177-T179-K377 in Cpn60.2 were spatially arranged as in the HslVU protease (T1-K33-S124). Structural superposition of the triads in the two crystal structures suggests that the proteolytic activity may be a consequence of this triad. Further interpretations of the role of Cpn60.2 and the possible residues involved in its activity' can be drawn upon mutagenesis of the potential residues and their analysis experimentally in conjunction with the crystal structure data.

5.4 CONCLUSIONS

E. coli GroEL has been one of the best characterized **ATP-dependent** molecular chaperones, with the chaperoning ability intricately related to its oligomeric state. *M. tuberculosis* possesses two copies of GroEL **homologues** known as Cpn60.1 and Cpn60.2. Sequence identity of more than 50% of both the proteins with GroEL suggests that these might also function as molecular chaperones in *M. tuberculosis.* Interestingly, the **mycobacterial** chaperonins have previously been shown to be secreted in the extracellular environment although their role as molecular chaperones is limited to the cytosol (De Bruyn *et a/.,* 2000). Their existence in extracellular environment suggests a possible alternate functional role. Both the proteins have long been known to be strong antigens responsible for inducing humoral and cellular immune responses (**Zügel** and Kaufmann, 1999). The two chaperonins are also known to be involved in cell signaling at very low concentrations (Ranford *et a/.,* 2000). It is thus possible that both the *M. tuberculosis* chaperonins have moonlighting functions.

Cpn60.2, also known as Hsp65, or the 65kDa antigen, can elicit a strong **delayed-type** hypersensitivity reaction in experimental animals infected with *M. tuberculosis* (**Silva** *et a/.,* 1995). Being one of the major immunoreactive proteins of *M. tuberculosis,* its **immunodominant** epitopes have recently been identified. It has therefore also been explored for its possible inclusion in a **recombinant** vaccine (Tascon *et a/.,* 1996). Moreover, Cpn60.2 has been shown to enhance cross presentation of exogenous proteins by dendritic cells to the **CD8[+] T-cells** (Chen *et al.,* 2004). Crystal structure of Cpn60.2 thus presents an opportunity to correlate the interesting structure-function properties of this protein.

Western blot analysis of *M. tuberculosis* extracts and difficulties in obtaining intact homogeneous recombinant protein suggested an intrinsic instability of Cpn60.2. SDS PAGE invariably showed two major bands in the purified

recombinant protein. **N-terminal** sequencing identified one of them as the native protein, while the other was a **proteolytic** product cleaved at Lysine 41. The truncated protein, which exists as a **dimer** for reasons discussed above, might represent a partially stable form. Crystal structure of the truncated Cpn60.2 of *M. tuberculosis* is presented in this chapter.

The structure shows interesting conformational changes that might be needed to accommodate the loss of first 41 residues. Despite these changes, however, hydrophobic regions of the equatorial domain remain exposed to the solvent. Moreover, hydrophobic patches were also exposed in the apical domain of Cpn60.2. Since it is unusual for proteins to expose hydrophobic regions, it is speculated that these exposed regions might have functional significance. In *E. coli* GroEL, the hydrophobic residues of the apical domain are known to bind unfolded polypeptides in the central cavity. The corresponding exposed residues in Cpn60.2 might similarly bind unfolded polypeptides. Interestingly, Cpn60.2 does not possess an ATPase activity. Prevention of aggregation of citrate synthase by binding through the exposed hydrophobic residues, but in absence of any discernible ATPase activity, therefore suggests that Cpn60.2 acts as an ATP-independent chaperone.

In addition to their role as molecular chaperones, heat shock proteins are known to be strong stimulators of the immune system. Cpn60.2 of *M. tuberculosis* has been shown to be a potent antigen capable of inducing strong immune responses in the host. The response achieved against this protein has been shown to be largely T-cell mediated. Regions of Cpn60.2 from *M. tuberculosis* and *M. leprae* have been identified that serve as potent T-cell epitopes. Intriguingly, when mapped onto the three-dimensional structure, the T-cell epitopes of Cpn60.2 lie largely on the equatorial domain. The apical domain on the other hand is essentially devoid of these **immunogenic** determinants. Structural significance of such a distribution on the Cpn60.2 surface, however, cannot be inferred easily.

One of the functions attributed to Hsps is their ability to bind cellular peptides and present them to the T-cell receptors. Presentation of antigenic peptides requires their uptake and intracellular processing by the antigen presenting cells. Hsps in complex with the bound peptides are taken up by the antigen presenting cells and processed similarly. Prevention of citrate synthase aggregation had suggested that Cpn60.2 is capable of binding non-polar peptides. This was further corroborated by surface properties of its structure. Cpn60.2 might thus act as a vehicle of antigenic peptides presenting these to the T-cell receptors.

Hsps in addition to their role as carriers might also present bound peptides to the T-cell receptor directly in a manner similar to CD1 mediated presentation of glycolipids. The crystal structure of Cpn60.2 reveals exposed hydrophobic patches on its surface thereby suggesting that the protein may act as an antigen presenter by binding to peptides on its hydrophobic surfaces.

Interestingly, Cpn60s of *M. tuberculosis* have been shown to be potent cytokine stimulators. This property of Cpn60s is also observed at very low concentrations of the proteins. Presence of these proteins in extracellular milieu, and the very small concentrations at which these exist in the extracellular medium, suggests that the proteins can exist in lower oligomeric form. As the area buried between different monomers in the dimeric Cpn60.2 is very small, it is conceivable that Cpn60.2 exists as a monomer at very low concentrations. The monomeric or dimeric states might be important in the cytokine stimulation ability. Since it would be difficult to crystallize proteins at such low concentrations, the dimeric form that was obtained fortuitously might serve as a good model to understand cell signaling properties of these proteins.

5.5 REFERENCES

1. Adir, N., Dobrovetsky, E., Shafat, I, Cohen, C. & Kashi, Y. (2002). Isolation, purification and preliminary X-ray characterization of Cpn60-2 (65 kDa heat-shock protein) from *Mycobacterium tuberculosis. Acta Crystallogr.* D58, 1474- 1475.

2. Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A., and Vijayan, M. (1994). Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc Natl Acad Sri USA.* 91, 227-231.

3. Bochtler, M., Hartmann, C, Song, H.K., Bourenkov, G.P., Bartunik, H.D., and Huber, R. (2000). The structures of HsIU and the ATP-dependent protease HsIU-HsIV. *Nature* 403, 800-805.

4. Boisvert, D.C., Wang, J., Otwinowski, Z., Horwich, A.L. and Sigler, P.B. (1996). The 2.4A crystal structure of the bacterial chaperonin GroEL complex with ATP-γ-S. *Nature Struct. Biol.* 3, 170-177.

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

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

7. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr.* 54, 905-921.

8. Chen, K.J. Lu, J., Wang, L., and Gan, Y.H. (2004). Mycobacterial heat shock protein 65 enhances antigen cross-presentation in dendritic cells independent of Toll-like receptor 4 signaling. /. *Leukoc Biol.* 75, 260-266.

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

10. De Bruyn, J., Soetaert, K., Buyssens, P., Calonne, I., De Coene, J.L., Gallet, X., Brasseur, R., Wattiez, R., Falmagne, P., Montrozier, H., Laneelle, M.A. and Daffe, M. (2000). Evidence for specific and non-covalent binding of lipids to natural and recombinant *Mycobacterium bovis* BCG Hsp60 proteins, and to the *Escherichia coli* homologue GroEL. *Microbiology* **146, 1513-1524.**

11. Esnouf, R.M. (1999). Further additions to Molscript version 1.4 including reading and contouring of electron density maps. *Acta Crystallogr.* **D55,** 938-940.

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

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

14. Fukami,T.A., Yohda, M., Taguchi, H., Yoshida, M., and Miki, K. (2001). Crystal structure of chaperonin-60 from *Paracoccus denitrificans.* *J. Mol. Biol. 312,* 501-509.

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

16. Horowitz, A., Bochkareva, E.S., and Girshovich, A.S. (1993a). The N terminus of the molecular chaperonin GroEL is a crucial structural element for its assembly. *J.Biol. Chem.* **268,** 9957-9959.

17. Horovitz, A., Bochkareva, E.S., Kovalenko, O., and Girshovich, A.S. (1993b). Mutation Ala2>Ser destabilizes intersubunit interactions in the molecular chaperone GroEL. J. *Mol. Biol.* **231,** 58-64.

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

19. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved method for building protein models in electron density maps and the location of errors in these models. *ActaCrystallogr.* **A47,** 110-119.

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

21. Kraulis, P.J. (1991). "MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures"./. *Appl. Crystallography* 24, 946-950.

22. Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F.U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein within its central cavity. *EMBOJ.* **11,** 4757-4765.

23. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures./. *Appl. Cryst.* 26, 283-291.

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

25. Merritt, E.A., and Bacon, D.J. (1997). Raster3D: Photorealistic Molecular Graphics. *Methods Enzymol.211,* 505-524.

26. Monahan, I., Betts, J., Banerjee, D., and Butcher, P. (2001). Differential expression of mycobacterial proteins following phagocytosis by macrophages. *Microbiology* **147,** 459-471.

27. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr.* **D53,** 240-255.

28. Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. *ActaCryst.* A50, 157-163.

29. Nicholls, A., Sharp, K., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struc. Func. Genetics.* **11,** 281-296.

30. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol,* **276:** Macromolecular crystallography Part A. C.W. Carter Jr. and R. M. Sweet, eds. (New York: Academic Press), pp. 307-326.

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

32. Ranford, J.C., Coates, A.R.M., and Henderson, B. (2000). Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones. *Exp. Rev. Mol. Med.* **September 15,** 1-17.

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

34. Silva, C.L., Pietro, R.L., Januario, A., Bonato, V.L., Lima, V.M., da Silva, M.F., and Lowrie, D.B. (1995) Protection against tuberculosis by bone marrow cells expressing mycobacterial hsp65. *Immunology.* 86, 519-524.

35. Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho Transcription Terminator: Mechanism of mRNA Recognition and Helicase Loading. *Cell* **114,** 135-146.

36. Stewart, G.R., Wernisch, L., Stabler, R., Mangan, J.A., Hinds, J., Laing, K.G., Young, D.B., and Butcher, P.D. (2002). Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology* **148,** 3129-3138.

37. Tascon, R.E., Colston, M.J., Ragno, S., Stavropoulos, E., Gregory, D., and Lowrie, D.B. (1996). Vaccination against tuberculosis by DNA injection. *Nat Med. 2,* 888-892.

38. Towbin H., Staehelin T., and Gordon J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl AcaaSci USA.* 76, 4350-4354.

39. Wang, J., and Boisvert, D.C. (2003). Structural Basis for GroEL-Assisted Protein Folding from the Crystal Structure of $(GroEL-K^+-Mg^{2+}-ATP)_{14}$ at 2.0A resolution/. *Mol. Biol.* **327,** 843-855.

40. Xu, Z., Horwich, A.L., and Sigler, P.B. (1997). The crystal structure of the asymmetric $GroEL_{14}-GroES_7-ADP_7$ chaperonin complex. *Nature* 388, 741-750.

41. Young, D.B., and Garbe, T.R. (1991). Heat shock proteins and antigens of *Mycobacterium tuberculosis. Infect. Immun.* 59, 3086-3093.

42. Ziigel, U., and Kaufmann, S.H.E. (1999). Role of heat shock proteins in protection from and pathogenesis of infectious diseases. *Clin. Microbiol. Reviews* 12, 19-39.

Appendix   A

In   vivo   complementation   studies   of

Mycobacterium   tuberculosis   Chaperonin60s

A1.1 INTRODUCTION

Chaperonin60s are a class of molecular chaperones that mediate the correct folding of newly synthesized polypeptides in an ATP-dependent manner (Hartl, 1996). The GroE complex of *Escherichia coli*, encoding the chaperonins GroEL and GroES, has been shown to be essential for bacterial growth at all temperatures (Fayet *et al.*, 1989). The chaperonin GroEL is involved in the *de novo* folding of 10-30% of all the proteins in the bacterial cytosol. It is in the central cavity of GroEL that the unfolded polypeptides undergo productive folding. This central cavity has been shown to be essential for its role as a molecular chaperone (Weber *et al.*, 1998).

GroEL was first identified by the isolation of temperature sensitive *E. coli* mutants that were defective for the assembly of the head proteins of bacteriophage A. and T4 and the tail proteins of bacteriophage T5 and were also affected for cellular growth at elevated temperatures (Sternberg, 1973a; 1973b). One such mutant *groELA4* represents the E191G substitution in GroEL (Zielstra-Ryalls *et al.*, 1993). The mutant exhibits normal growth at 30°C however survival at 43°C requires the external contribution of a functional *groEL*-homologue. The strain thus becomes suitable for testing the ability of various genes to complement the Ts phenotype of the mutant.

Biochemical analysis of the *M. tuberculosis* Cpn60s showed that the proteins do not exist as tetradecamers, the canonical form of Cpn60s, but rather as dimers (Described in Chapter 4). Moreover, the activity of the *M. tuberculosis* Cpn60s when compared with *E. coli* GroEL showed that these proteins do not exhibit ATPase activity unlike GroEL. The proteins, however, were capable of complete prevention of substrate aggregation suggesting their role as ATP-independent molecular chaperones. Since the *M. tuberculosis* Cpn60s behaved in a manner much different from the usual chaperonins their role as chaperones was studied by *in vivo* complementation using the *groELA4* mutant strain.

A1.2 EXPERIMENTAL PROCEDURES

A1.2.1 Cloning of *E. coli* and *M. tuberculosis cpn60s*

The genes coding for the *E. coli groEL* and *M. tuberculosis cpn60.1* and *cpn60.2*, were PCR amplified and cloned in the low-copy-number plasmid vector, pCL1920, through an intermediate sub-cloning step in pBluescript (SK+). The previously described expression plasmid, pKKGL1, cosmid A10 and vector, pKY206 served as the templates for PCR amplification of *cpn60.1*, *cpn60.2* and *groEL*, respectively. The amplified products were ligated at the SmaI site of pBluescript (SK+) and the transformants in *E. coli* DH5α selected on the basis of blue-white selection. From the positive clones obtained the three genes were subcloned in the spectinomycin resistant (Sp^r) plasmid, pCL1920 (Fig. A1.1). Table A1.1 describes the different primers used for PCR amplification and the designation of the different clones.

Table A1.1 Primers used for cloning of *cpn60s* in pCL1920 Restriction endonuclease (RE) sites are highlighted as underlined sequences.

Primer name	RE site	Primer sequence	Designation
ECGLPCL.FOR	SalI	5' TACTATAGTCGACTATGGCAGCTAAAGACG 3'	pCLECGL
ECGLA2S.REV	SacI	5' ATATACGAGCTCTCATGCCGCCCATGCC 3[1]	
GL1PCL.FOR	PstI	5'ATAGATCTGCAGGATGAGTAAGCTGATCGAAT ACG 3'	pCLGL1
GL1PCL.REV	BamHI	5' TACTATAGGATCCTCAGTGCGCGTGCCC 3'	
GL2PCL.FOR	HindIII	5' TAGATGAAGCTTAATGGCCAAGACAATTGCG3'	pCLGL2
GL2PCL.REV	SmaI	5' AGCTAGCCCGGGTCAGAAATCCATGCCACC 3'	
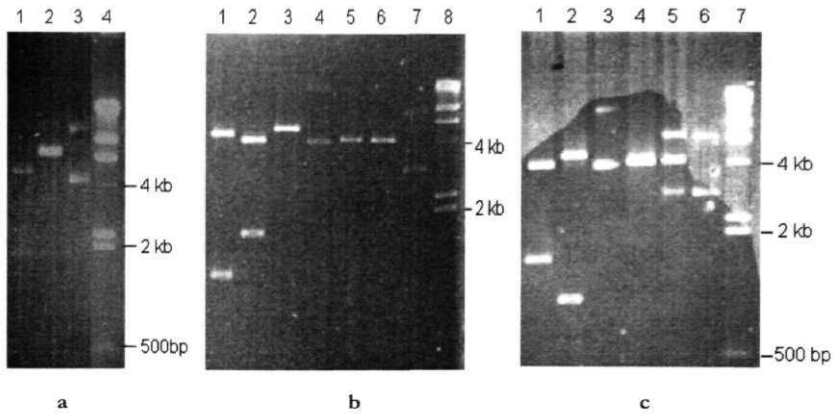
Figure A1.l Agarose gel scans showing cloning of *cpn60*s into pCL1920 (a) Cloning of *E. coli groEL* Lane 1: Sall/SacI digested clone showing the 1.6kb fall out of *groEL*. Lane 2: Positive clone linearized with BamHI. Lane 3: Undigested clone as a control. Lane4: λ-HindIII DNA digest as the marker. **(b) Cloning of *cpn60.1*** Lane 1: HindIII digestion of positive clone yielding the 1.1kb fall out. Lane *2:* BamHI/PstI digest giving the 1.6kb fall out confirming the positive clone. Lane 3: Positive clone linearized with SalI. Lanes 4 and 7 correspond to the undigested positive clone and pCL1920 as controls. Lanes 5 and 6 correspond to the linearized pCL1920 as control. Lane 8: X-HindIII DNA digest as marker. **(c) Cloning of *cpn60.2*** Lanes 1 and 4: SmaI/HindIII digested positive clone and pCL1920. The 1.6kb fall out confirms the clone. Lanes 2 and 5: SacI digested positive clone and the vector as a control. The positive clone releases the 1.1kb fall out. Lanes 3 and 6: Undigested positive clone and pCL1920 as controls. Lane 7: λ-HindIII DNA digest as marker.

A1.2.2   *In vivo* Complementation in *E. coli*

Complementation experiments were performed at 43°C as described by Chatellier *et al.* (1998). Plasmids pCLECGL, pCLGLl and pCLGL2 containing the *E. coli* and *M. tuberculosis cpn60* genes, along with the plasmid vector pCL1920 as negative control, were transformed into the temperature sensitive (Ts) *E. coli* strain, SV2 (kindly provided by Jean Chatellier and Alan Fersht). SV2 has the *groEL44* allele, representing an E191G substitution in GroEL (Zeilstra-Ryalls *et al.*, 1993). Sp^r transformants were selected at 30°C and the efficiencies of plating (EOP) for each determined at 43°C, relative to that at 30°C. 3ml cultures of appropriately

transformed cells were grown overnight at 30°C in LB medium supplemented with spectinomycin. Cells were serially diluted 10-fold in LB medium and from each dilution 100μl aliquot was spread on LB plates containing spectinomycin. Similarly, 5μl from each 10-fold dilution was spotted onto two LB plates, one of which was incubated overnight at 30°C and the other at 43°C. The number of viable cells/ml of culture was deduced from the number of colonies obtained at the different dilutions.

A1.3 RESULTS

In order to investigate the functional properties of *M. tuberculosis* Cpn60.1 and Cpn60.2, we sought complementation of a Ts *grvEL E. coli* mutant, SV2. The results, presented in Table A 1.2, indicate that the SV2 derivative carrying the plasmid, pCLECGL (with *E. coli groEL*$^+$)exhibited an EOP of 0.05 at 43°C (relative to that at 30°C). The SV2 derivatives carrying pCLGLl and pCLGL2 (with *M. tuberculosis* genes for *cpn60.1* and *cpn60.2,* respectively) behaved like that carrying the vector pCL1920, with an EOP of $<10^{-5}$.

Table A1.2 Test for complementation of the Ts *E. coli groEL44* mutant Complementation was performed with plasmids expressing *E. coli groEL* or *M. tuberculosis cpn60.1* or *cpn60.2*

Plasmid (Description)	Colony Forming Units		EOP (B/A)
	30°C (A)	43°C (B)	
pCL1920 (Vector)	1×10^8	$< 10^3$	$< 10^{-5}$
pCLECGL (*E. coli groEL*)	1.3×10^8	6×10^6	0.05
pCLGLl (*M.* tuberculosis *cpn60.1*)	4×10^8	$< 10^3$	$< 10^{-5}$
pCLGL2 (*AT.* tuberculosis *cpn60.2*)	5×10^8	$< 10^3$	$< 10^{-5}$

Fig. A 1.2 clearly shows the ability of the *E. coli grvEL* to complement the mutant *grvEL* gene in the Ts mutant. The presence of neither *cpn60.1* nor *cpn60.2* could promote the growth of the Ts *groEL44 E. coli* mutant at 43°C.
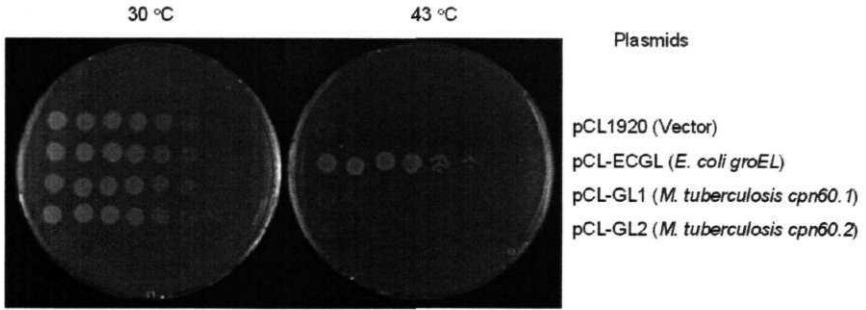
Figure A1.2 Test for complementation of the Ts *groEL44 E. coli* mutant Complementation was carried out with plasmids expressing *E. coli groEL* or *M. tuberculosis cpn60.1* or *cpn60.2.* Sp^r transformants of strain SV2 carrying plasmids as indicated at the right were grown overnight at 30°C in LB medium supplemented with spectinomycin, serially diluted and five microlitres from each 10-fold dilution spotted onto two LB-spectinomycin plates (from left to right in each row of the figure). One plate was incubated overnight at 30°C and the other at 43°C.

A1.4 CONCLUSIONS

The inability to achieve *in vivo* complementation of a Ts *groEL E. coli* mutant with the *M. tuberculosis cpn60* genes suggests that the mycobacterial counterparts of GroEL might have lost their function as molecular chaperones. Intriguingly, however, biochemical characterization of these proteins had suggested otherwise (Chapter 4). Thus, mere suppression of aggregation of substrate protein *in vitro* does not seem to impart the ability of GroEL-like function to the *M. tuberculosis* Cpn60s *in vivo*.

I propose that the anomaly between the *in vitro* and *in vivo* results can be explained by two possible means: (a) Cpn60s of *M. tuberculosis* require their cognate co-chaperonin for the role as molecular chaperones *in vivo* or (b) the two proteins, Cpn60.1 and Cpn60.2, are required simultaneously for effective *in vivo* complementation. In an attempt to resolve these questions, expression of the *M. tuberculosis cpn60.1* as a part of the *groESL* operon and co-expression of the *cpn60.1* and *cpn60.2* genes needs to be studied.

A1.5 REFERENCES

1. Chatellier, J., Hill, F., Lund, P.A., and Fersht, A. (1998). *In vivo* activities of GroEL minichaperones. *Proc. Natl. Acad. Sci. USA.* 95, 9861-9866.

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

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

4. Sternberg, N. (1973a). Properties of a mutant of *Escberichia coli* defective in bacteriophage *X* head formation (*groE*). I. Initial characterization. /. *Mol. Bio/.* 76, 1-23.

5. Sternberg, N. (1973b). Properties of a mutant of *Eschericbia coli* defective in bacteriophage *X* head formation *(groE).* II. The propagation of phage *X. J. Mol. Biol.* 76, 25-44.

6. Weber, F., Keppel, F., Georgopoulos, C, Hayer-Hartl, M.K., and Hartl, F.U. (1998). The oligomeric structure of GroEL/GroES is required for biologically significant chaperonin function in protein folding. *Nat. Struct. Biol.* 5, 977-985.

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

Appendix B

Identification of Conserved Residue Patterns in Small fi-barrel Proteins

This work has been published as:

Qamra, R., Taneja, B., and Mande, S. C. (2002). Identification of conserved residue patterns in small β–barrel proteins. *Protein Engg.* 15, 967-977

Identification of conserved residue patterns in small β-barrel proteins

Rohini Qamra, Bhupesh Taneja[1] and Shekhar C.Mande[2]

Centre for DNA Fingerprinting and Diagnostics, ECIL Road. Nacharam, Hyderabad 500 076, India

[1]Present address: Northwestern University. Chicago, IL, USA

[2]To whom correspondence should be addressed.
E-mail: shekhar@cdfd.org.in

Our abilities to predict three-dimensional conformation of a polypeptide, given its amino acid sequence, remain limited despite advances in structure analysis. Analysis of structures and sequences of protein families with similar secondary structural elements, but varying topologies, might help in addressing this problem. We have studied the small β-barrel class of proteins characterized by four strands in = 4) and a shear number of 8 (S = 8) to understand the principles of barrel formation. Multiple alignments of the various protein sequences were generated for the analysis. Positional entropy, as a measure of residue conservation, indicated conservation of non-polar residues at the core positions. The presence of a type II β-turn among the various barrel proteins considered was another strikingly invariant feature. A conserved glycyl-aspartyl dipeptide at the β-turn appeared to be important in guiding the protein sequence into the barrel fold. Molecular dynamics simulations of the type II β-turn peptide suggested that aspartate is a key residue in the folding of the protein sequence into the barrel. Our study suggests that the conserved type II P-turn and the non-polar residues in the barrel core are crucial for the folding of the protein's primary sequence into the β-barrel conformation.
Keywords: β-barrel/molecular dynamics/protein folding/SH3/type II β-turn

Introduction

The landmark work of Anfinsen indicated for the first time that the primary structure of a protein dictates its tertiary structure (Anfinsen, 1973). In a sequential protein folding model the primary structure of a protein initially yields α-helices, β-sheets and turns, the predominant secondary structural elements in proteins. By arranging these simple elements in precise patterns, complex protein structures assemble to achieve the diversity of protein functions. A major goal in understanding how the amino acid sequence of a protein specifies its structure is to understand how these elements of secondary structure are organized onto a tertiary scaffold. This requires learning how properties of individual amino acids are exploited in guiding an amino acid sequence into a particular fold. Much progress has been made in the last decade towards understanding the relationship between a protein's sequence and structure, yet the protein folding problem remains a captivating puzzle.

Researchers have learnt several rules governing the formation of helices and turns. However, the principles behind β-sheet formation are much less understood (Serrano, 2000). It is therefore especially intriguing to speculate how β-sheet proteins, having complex topologies and involving numerous contacts between residues distant in sequence, acquire their native structure. Several features of (5-sheet proteins have been suggested to be important for efficient folding and stability. The overall hydrophobic and polar pattern of amino acids may be a dominant driving force for defining a protein's topology (Eisenberg et ai, 1984; Bowie et al., 1990; Kamtekar et ai, 1993). Recognition between amino acid side chains on neighboring β-strands may guide a correct strand register and hence stabilize the resulting β-sheets (Merkel et al., 1999; Mandel-Gutfreund et ai, 2001). Another possibility is the formation of turns at critical locations in the protein structure. Turns may be particularly important for anti-parallel sheet formation and hence defining the protein topology. Supporting this hypothesis, recent studies indicate that the residues in the distal loop in the SH3 domain are important for nucleation of protein folding (Martinez and Serrano, 1999; Riddle et al., 1999). Different combinations of these possible interactions are the most likely determinant of β-sheet topology and hence protein stability.

In the course of evolution, three-dimensional structures of proteins are conserved to a greater degree than their sequences, which determine their structure. Residue substitutions, which tend to destabilize a particular site, would probably be compensated by other substitutions that confer greater stability on the structure. For example, if volume conservation were important to structure and function, a substitution involving a reduction of volume in the protein core might result in a destabilizing pocket in the core. In this case, it might become necessary to substitute another residue at a position distant in the sequence but near in space. This second substitution should then have a larger side chain in order to conserve the overall volume of the core and therefore the overall folded structure. Thus, if structural compensation is a general phenomenon, neighbouring sites in the three-dimensional structure will tend to evolve in a correlated fashion owing to the compensation process. In the past decade there has been a great deal of progress in the development of methods for predicting interactions in protein structures by analysis of correlated changes in sequence evolution (Altschuh et al., 1988; Shindyalov et al, 1994; Pollock and Taylor, 1997).

In this study, we have undertaken a comprehensive analysis of the sequence and structural variation seen in the small β-barrel proteins. A β-barrel is essentially identified by two geometric characteristics: the number of β-strands in the barrel (n) and the number of β-bridge staggers across the β-sheet (the shear number, S) (Murzin et al., 1994). Within the all-(5 protein class in the Structural Classification of Proteins (SCOP) database there exist five folds which can be grouped together as small β-barrels (Murzin et al., 1995). These barrels are characterized by the presence of four p-strands (n = 4) and a shear number of 8 (S = 8) (Murzin et ai, 1994). Although the five folds have similar secondary structure composition,

each has a distinct topology. The goal of this study was to identify conserved features across these β-barrel folds, which may also be important in the initial steps of the folding pathway and in guiding the protein's primary sequence into a P-barrel with the specific topology. In the work described here, we constructed and analyzed multiple sequence alignments for protein sequences in each of these five barrel folds. We also aligned structures of the different proteins, within and across the folds. In order to determine certain structural features common to the barrel folds at both the sequence and structural level, we studied the conservation and covariation in the SH3-like barrel, GroES-like and the PDZ domain-like folds. Molecular dynamics (MD) simulations on a GroES peptide, derived from a conserved β-turn, were also carried out in order to address its role as a possible nucleation site in the folding pathway. By combining sequence and structural analysis it was possible to interpret the pattern of conservation seen in the three protein folds.

Materials and methods

The SCOP database classifies all-β proteins into 93 folds according to their topology and evolutionary relationships (Murzin *et al,* 1995). Each fold is divided into superfamilies which are further classified into different families, that is, a group consisting of proteins with residue identities of 30% and greater or those having similar structure and function. The five folds include the SH3-like barrel, GroES-like, PDZ domain-like, N-terminal domains of the minor coat protein g3p and the Sm motif of the small nuclear ribonucleoproteins, SNRNP. A total of 42 different protein structures are grouped together in the SH3-like barrel fold. In the GroES-like and PDZ domain-like folds a total of nine and eight different structures have been solved, respectively. Only one structure each has been solved for both the N-terminal domains of the minor coat protein, g3p and the Sm motif of small nuclear ribonucleoproteins, SNRNP fold. A representative structure from each family within a fold was considered for the sequence and structure comparisons carried out in this study. Table I lists the initial target sequences considered for initiating the analysis.

Structure comparison

A total of 20 structures were considered for structural comparison of proteins (Table I). Of these, 13 belonged to the SH3-like barrel fold, three to the GroES-like barrel and two to the PDZ domain-like barrel fold. One structure was selected for each of the N-terminal domains of the minor coat protein g3p and the Sm motif of small nuclear ribonucleoproteins, SNRNP fold. Coordinates for each of these proteins were retrieved from the PDB (Bernstein *et al,* 1977). Superimpositions were done among the structures within the same and also across the various β-barrel folds. For inter-fold superimpositions, one representative from each fold was taken. The representative structure corresponded to any one protein in a fold for which complete sequence analysis was done as per the criteria of more than 30 sequences in the multiple sequence alignment (addressed later). Hence, the structures chosen for the inter-fold comparison were the α-spectrin SH3 domain protein, 1shg; the *Escherichia coli* GroES, 1aon and the rat neuronal nitric oxide synthase, 1qav from the SH3-like barrel, GroES-like and the PDZ domain-like fold, respectively. The structures were superposed by visualization, followed by least-squares fitting using the lsq commands of O (Jones *et al.,* 1991).

Residue conservation

To measure the level of conservation at each position in the alignment, the frequency of occurrence of an amino acid at each position was determined. This was achieved by the calculation of the positional entropy at each position in the alignments obtained. A positional entropy of *n* is equivalent to the diversity of n residues occurring at the position with a frequency of I/n. A position that is completely conserved will thus have a positional entropy of 1. For position *i,* with residues r = (A, C, D, ..., V, W, Y) occurring at frequencies $Pi(r)$, the entropy $H(i)$ is defined as

$$H(i) = -p_i(r)\ln[p_i(r)]$$

This entropy is known as the Shannon informational entropy (Shenkin *et al,* 1991).

The positional entropy is expressed as

$$N(i) = e^{H(i)}$$

Volume correlation

The correlation coefficient at each residue position in the alignment was calculated as a measure of covariation in the volumes of the side chains. The side-chain volumes were taken from Harpaz *et al.* (Harpaz *et al.* 1994). A pairwise correlation coefficient, $r(x,y)$ determined the correlation between two residue positions and was expressed as

$$r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[(n\sum x^2) - (\sum x)^2\right]\left[((n\sum y^2) - (\sum y)^2)\right]}}$$

where r = correlation coefficient, n = number of sequences, x = volume at residue position i and y = volume at residue position j.

Sequence alignment

A total of 20 initial target sequences corresponding to the representative protein in each family were considered for the analysis (Table I). The chosen target sequence was used for a BLAST search *(E* < 0.001) of the non-redundant database compilation (Altschul *et al.,* 1997). Homologous sequences were retrieved and the stretch of residues, aligned to the initial target domain, extracted from each protein sequence. Two or more domains within a protein sequence were considered as separate sequences. Thus, a sequence having two domains was split into two, each corresponding to a different domain, within the protein. These sequences were aligned using the ClustalW program (Thomson *et al.,* 1994). Once the initial alignment was constructed, sequences with ClustalW score of >90 were removed in order to remove any bias in the sequence analysis due to high degree of similarity. To avoid artifactual results arising out of inaccurate sequence alignments, sequences with a score of <25 were also removed. The remaining sequences were realigned such that, in the final alignment, no two sequences had a score of <25 or >90. Families where, after the editing, the number of sequences in the alignment was <30 were not considered for further analysis.

Only five of a total of 20 families in the n = 4, 5 = 8 β-barrel protein folds fulfilled the criterion of >30 sequences in the multiple alignment. These included the SH3 domain and the C-terminal domain of ribosomal protein L2 in the SH3-like barrel fold, GroES and alcohol dehydrogenase-like, N-terminal

Table I. β-Barrel proteins considered for sequence and structure comparisons

β-Barrel fold	Classification within the fold	Proteins used for sequence alignment construction	PDB ID
SH3-like barrel	C-Terminal domain of biotin and diphtheria toxin repressors		
	Biotin repressor (BirA)	Biotin repressor/biotin holoenzyme synthetase, C-terminal domain	1bia
	Diphtheria toxin repressor (DtxR)	Diphtheria toxin repressor (DtxR)	2dtr
	SH3 domain	a-Spectrin SH3 domain	1shg
	Myosin S1 fragment, N-terminal domain		
	Myosin S1 fragment, N-terminal domain	Myosin S1 fragment	2mys
	Translation proteins - SH3 like domain		
	Ribosomal proteins L24p and L21e	Ribosomal proteins L24 (L24p)	1ffk
	N-Terminal domain of eukaryotic initiation translation factor 5a	N-Terminal domain of eukaryotic initiation translation factor Sa	2eif
	C-Terminal domain of ribosomal protein L2	C-Terminal domain of ribosomal protein L2	1rl2
	Electron transport accessory proteins		
	Photosystem I accessory protein E(PsaE)	Photosystem I accessory protein E (PsaE)	1psf
	Nitrile hydratase B-chain	Nitrile hydratase B-chain	2ahj
	Ferredoxin thioredoxin reductase (FTR), alpha (variable) chain	Ferredoxin thioredoxin reductase (FTR). alpha (variable) chain	1dj7
	R67 dihydrofolate reductase	R67 dihydrofolate reductase	1vie
	CcdB		
	CcdB	CcdB	2vub
	DNA binding domain of retroviral integrase		
	DNA binding domain of retroviral integrase	DNA binding domain of retroviral integrase	1ex4
GroES-like	GroES-like		
	GroES	Chaperonin-10 (GroES)	1aon
	Alcohol dehydrogenase-like, N-terminal domain	Alcohol dehydrogenase	3bto
	SacY-like RNA-binding domain	SacY	1auu
	BglG-like antiterminator proteins		
PDZ domain-like	PDZ domain-like		
	PDZ-domain	Neuronal nitric oxide synthase, NNOS	1qav
	Interleukin 16	Interleukin 16	1il16
N-Terminal domains of the minor coat protein g3p	N-Terminal domains of the minor coat protein, g3p		
	N-Terminal domains of the minor coat protein, g3p	N-Terminal domains of the minor coat protein, g3p	2g3p
Sm motif of small nuclear ribonucleoproleins, SNRNP	Sm motif of small nuclear ribonucleoproteins, SNRNP		
	Sm motif of small nuclear ribonucleoprotcins. SNRNP	D1 core SNRNP protein	1b34

domain in the GroES-like fold and PDZ-domain in the PDZ domain-like fold. Sequence alignment data corresponding to these families were considered for statistical analysis.

Molecular dynamics simulations

MD simulations for a small peptide of the *E.coli* GroES were performed using the Discover module in the Insightll molecular modelling package (MSI/Biosys, San Diego, CA, 1997). The simulations were performed with a cubic periodic boundary condition (box dimensions 25x25X25) and consisted of the peptide solvated with water molecules. The effective water density in the solvation box was 0.96 g/cm³. All atoms were considered explicitly and their interactions were computed using the CVFF force field. The time step in the MD simulations was 1 fs. All simulations began with 100 iterations of the energy minimizations of the peptide to relax the local forces. Subsequently, MD simulations were performed at 300 K for 500 ps. A seven-residue peptide with the original conformation as in the protein with an intact type II turn was the starting structure. Simulations were performed for the wild-type sequence of the peptide and also on two other peptides. In one of these, the aspartate was mutated to asparagine and in the second the aspartate was mutated to alanine. The native-like side chain-main chain hydrogen bond was retained in the aspartate to asparagine mutant.

Results

The study involved comparison of sequence and structure data for the different four-stranded β-barrel folds. According to the
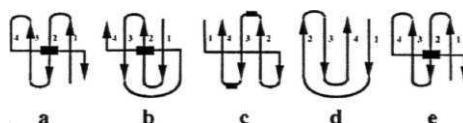
Fig. 1. Topologies of the β-barrel folds characterized by $n = 4$ and $S = 8$ according to the SCOP classification. The folds included in this category are (a) SH3-like barrel, (b) GroES-like. (c) PDZ domain-like, (d) N-terminal domains of the minor coat protein. g3p and (e) Sm motif of small nuclear ribonucleoproteins. SNRNP. The Sm motif of small nuclear ribonucleoproteins, SNRNP fold has the same topology as the SH3-like barrel. β-Strands are indicated by arrows. The boxes indicate the helices in the different folds.

number of strands forming a compact globular structure, these constitute the smallest barrels known. The difference among these different folds essentially lies in the manner in which the four β-strands are connected, thereby generating a unique topology (Figure 1). In each topological class multiple sequence alignments were generated for the sequence analysis and structures of proteins within a β-barrel fold superimposed on one another.

Comparisons within the SH3-like barrel fold

Superpositions were done among 13 structures in the SH3-like barrel fold (Table I). These structures superposed well on one another with a maximum r.m.s. deviation of 2.24 A for 29 atoms between the DNA binding domain of HIV-I integrase, 1ex4 and the diphtheria toxin repressor, 2dtr (Table IIa). The

Table II. Pair-wise r.m.s. deviations (A) when superimposing structurally equivalent Cα positions within a β-barrel fold. The numbers in parentheses represent the equivalent residues considered during the superpositions

(a) SH3-like barrel fold

	1vie	2dtr	1dj7	1psf	2vub	2mys	1shg	2ahj	1bia	1rl2	2eif	1ex4	1ffk
1vie	0.0	1.85(29)	1.49(46)	1.34 (40)	1.71 (47)	1.41 (37)	1.90(33)	1.98 (53)	1.90 (36)	1.74 (42)	1.70 (44)	1.71 (32)	1.53(42)
2dtr		0.0	1.99(36)	1.69(32)	1.87(34)	1.96(28)	1.81(35)	1.28(28)	1.73(37)	1.87(42)	1.66(21)	2.24(29)	1.47(28)
1dj7			0.0	1.45(53)	1.49(39)	1.83(46)	1.56(44)	0.91(46)	1.84(41)	1.72(42)	1.60(36)	1.56(37)	1.26 (45)
1psf				0.0	1.51 (38)	1.87(36)	1.88(41)	1.31(48)	1.82 (37)	1.69(38)	1.41(33)	1.65(38)	2.03(54)
2vub					0.0	1.29(35)	1.23(38)	1.72(47)	1.74(37)	1.94(40)	1.97 (34)	1.39(33)	1.69(37)
2mys						0.0	1.94(40)	1.75(43)	1.87(32)	1.41(34)	1.85(39)	1.72 (36)
1shg							0.0	1.61 (41)	1.89 (37)	2.21 (38)	1.94(30)	1.35(42)	1.81 (38)
2ahj								0.0	1.62(37)	1.46(41)	1.48(41)	1.54(38)	1.84(50)
1bia									0.0	1.89(39)	1.49(22)	1.99(35)	1.97(39)
1rl2										0.0	1.93(44)	1.96(42)	1.88(42)
2eif											0.0	1.50(33)	1.80(34)
1ex4												0.0	1.86 (37)
1ffk													0.0

(b) GroES-like fold

	1aon	3bto	1auu
1aon	0.0	1.35(51)	1.69(37)
3bto		0.0	1.61 (34)
1auu			0.0

minimum r.m.s. deviation of 1.23 A for 38 atoms was seen between the α-spectrin SH3 domain protein, 1shg and the CcdB protein, 2vub.

Of interest is the region at the type II β-turn of SH3-like barrel fold proteins. The turn, referred to as the diverging turn in the SH3 domain (Yi *et al.*, 1998), is present in the loop connecting strands 1 and 2 of the SH3-like barrel fold. Intriguingly, the occurrence of the type II turn in the SH3-like barrel fold proteins seems to be related to the length of the loop preceding the diverging turn. Of the 13 different SH3-like barrel fold structures studied (Table I), the type II turn was present in six. None of the remaining seven structures contained the type II turn. A common feature among these seven structures, lacking the type II turn, was the presence of a short loop between strands one and two of the barrel. The presence of a short intervening loop probably reduces the likelihood of the polypeptide chain from deviating from the folded barrel structure. Proteins containing the type II turn included the a-spectrin SH3 domain, 1shg; photosystem I accessory protein, 1psf; diphtheria toxin repressor, 2dtr, nitrile hydratase β-chain, 2ahj; the ribosomal protein L24, 1ffk and ferredoxin thioredoxin reductase, 1dj7. A stretch of >11 residues in the loop connecting strands 1 and 2 of the β-barrel necessitated the presence of the type II turn, as observed in five of these structures. The presence of the turn, in these structures, appears to guide the polypeptide into the β-barrel helping in the formation of the folded barrel structure.

Multiple sequence alignments for each of the 13 proteins considered for structural comparisons were generated as described in Materials and methods. A BLAST search with the amino acid sequence of the a-spectrin SH3 domain (SH3 domain family) gave 219 hits with *E* < 0.001. Splitting of multi-domain sequences augmented this number to 302. Exclusion of sequences with ClustalW scores of <25 and >90 drastically reduced the number of sequences in the final alignment to 30. A BLAST search for the sequence of the C-terminal domain of ribosomal protein L2 (translation proteins - SH3-like domain family) resulted in an initial number of 132

hits with *E* < 0.001. A total of 65 sequences homologous to the ribosomal protein w -e finally obtained by editing the sequences in a manner similar to that described above. For all the remaining sequences subjected to BLAST search, the number of sequences after editing was <30. These sequence alignments were hence not considered for further analysis for reasons described in the Materials and methods section.

The degree of conservation at each position in the multiple alignment generated was determined using the Shannon Informational entropy calculation (Shenkin *et al.*, 1991). Tables IIIa and b give the positional entropy values at the core residue positions for the SH3 domain and the C-terminal domain of ribosomal protein L2 families, respectively. Core residue positions for the representative proteins in each family were identified by calculating the percentage accessibility of side chains using the NACCESS program (Hubbard *et al.*, 1991). Residues with side chain accessibilities of <7% were considered part of the core. A total of nine core positions were identified in the a-spectrin SH3 domain. The positional entropies were <3 at all nine core positions in the SH3 domain protein (Table IIIa). In the case of the C-terminal domain of ribosomal protein L2, 10 of the 12 core positions showed high residue conservation as indicated by a positional entropy of <3 at these positions (Table IIIb). Tables IIIa and b also indicate the prevalence of amino acid residues at the core residue positions for the SH3-like barrel fold families, SH3 domain and the C-terminal domain of ribosomal protein L2. These core positions, as seen from the data, are predominantly occupied by valine, leucine or isoleucine in both the families. The other residues occupying positions in the barrel core are the non-polar residues including phenylalanine. methionine, alanine and glycine. The presence of these residues contributes to the high hydrophobicity at the core of the barrel in this fold. High conservation of non-polar residues at the core residue positions suggests the importance of a hydrophobic interior in maintaining the integrity of the fold.

A covariance analysis of residue volumes indicated a significantly high correlation between residues at the core positions

Table **III. Position-specific** statistics for occurrence and conservation of residues at the core positions in the five **β-barrel** families. The first column indicates residue position numbers corresponding to sequence of (a) the α-spectrin SH3 domain protein, **1shg**; (b) **C-terminal** domain of ribosomal protein L2, **1rl2**; (c) the *E.coli* GroES protein, **1aon;** (d) alcohol dehydrogenase-like, **N-terminal** domain protein, 3bto; and (e) the **PDZ-domain** protein, **1qav**. Occurrence of the most prevalent residue is indicated in the last column

	A	C	Y	W	S	M	D	N	E	Q	V	L	I	H	G	P	K	R	F	T	Positional entropy	Most prevalent residue
(a) SH3 domain: a total of 30 sequences																						
9	7		4								18								1		2.79	V
11	28													2							1.27	A
23											6	22	2								2.07	L
25						8													20		2.33	F
31						1					3	8	18								2.72	]
33											18	8	4								2.52	V
44		2									6				22						2.07	G
53											9		1						20		2.10	F
58		1									27	1	1								1.54	V
(b) C-Terminal domain of ribosomal protein L2: a total of 65 sequences																						
131							3				2	54	6								1.86	L
134		1					7				3	1	53								1.97	I
140											38	2	25								2.20	V
143											26	1	38								2.10	I
151		13			1										51						1.78	G
160	1	1	20	1	21		3	1			9					1		7			5.50	S
161	49	2				2						8		1	3						2.44	A
163											16	22	27								2.93	I
171	17		4		2						31		3	1						7	4.17	V
173											23	31	11								2.78	L
187	2	58			3			1													1.61	C
189	65																				1.00	A
(c) GroES: a total of 86 sequences																						
10											71	2	13								1.70	V
12											54	5	27								2.27	V
38		15													71						1.57	G
40											70		16								1.60	V
59						1					78	5	2								1.47	V
65											72	!	13								1.61	V
67			26			1					1	5	3						50		2.88	F
73		5			5			5	1	3	7	1			5					54	3.99	T
84			1		1	1					13	53	9			1			4		3.33	L
86						31					4	42	4					5			3.27	L
91											15	6	65								1.99	I
(d) Alcohol dehydrogenase-like, N-terminal domain: a total of 52 sequences																						
36											49		2							1	1.29	V
38				1							16	6	28				1				2.99	I
69	30					1							1	20			1				2.31	A
73											45		7								1.48	V
83						1					20	27							4		2.67	L
89	2										50										1.18	V
91	1				1	1						6				41	1			1	2.62	P
150	5				1	2							1							43	1.93	T
152	3	1			1	1					33	8	3							2	3.52	V
157	4	3						1			33	11									2.87	V
(e) PDZ domain: a total of 35 sequences																						
84											4	22	9								2.43	L
106											4		31								1.42	I
108											9		26								1.76	I
116	18										4	1	2		9		1				3.69	A
117	35																				1.00	A
123											2	33									1.24	L
128	11	3			1			10		2							1	7			5.05	A
129											3	9	23								2.30	I
132											33		2								1.24	V
137											12	21	2								2.30	L
140	17	2			1			3	1		10									1	4.00	A
145	31										4										1.42	A
152	20				4			2						1	1					7	3.57	A
156											30	2	3								1.65	V
158												29	5					1			1.70	L
160	7										21	4	3								2.96	V

in the SH3-like barrel (Figure 2). A high negative correlation of −0.97 was seen between residue positions 23 and 44 (residue numbers corresponding to the α-spectrin SH3 domain). An increase in volume of the core due to a larger side chain at residue position 23 (mostly leucine or valine) is compensated by a reduction in the side chain volume at the correlated position 44 (mostly valine or glycine) (Table IIIa). Similarly, a negative correlation is seen between the volumes of residues 25 and 44 (a correlation value of −0.64). Since the side chain volumes of positions 23 and 25 are negatively correlated with respect to position 44, it was anticipated that the correlation coefficient between them would be positive. Indeed, the correlation coefficient of side chain volumes at positions 23 and 25 shows a high positive value of 0.68. Another compensatory pair of residues is present at positions 44 and 53. Residue 44 exists in the third strand of the α-spectrin SH3 domain and residue 53 in the fourth strand. A negative correlation of −0.84 is seen between these core residue positions. Positive correlations of 0.75 between residues 23 and 53 and 0.42 between residues 25 and 53 result in compensation of the overall core volume. This observation strongly supports the belief that maintenance of the total volume of the core would be important to keep the barrel structure intact.

Comparisons within the GroES-like fold

Three representative proteins in the GroES-like fold were considered for the comparative study (Table I). Superposition of proteins within the GroES-like fold was done as in the case of the SH3-like barrel fold. The representative proteins included for analysis superposed well with one another with a minimum r.m.s. deviation of 1.35 A for 51 atoms between the horse alcohol dehydrogenase, 3bto and the *E.coli* GroES, 1aon (Table IIb). Different proteins within the alcohol dehydrogenase-like, N-terminal domain family also superimposed very well on one another (data not shown). Comparisons revealed an overall conservation of the β-barrel core in the representative proteins.

A BLAST search yielded 170 hits for the *E.coli* GroES protein. Splitting of the multi-domain sequences increased this number to 174. Further editing as described earlier for the SH3-like barrel fold, however, reduced the number to 86. An initial number of 412 hits in a BLAST search for the alcohol dehydrogenase reduced to 52 sequences in the final alignment after appropriate editing. In case of the SacY protein, the number of sequences in the final alignment was <30. This protein and the corresponding family were thus omitted from further sequence and structural comparisons.

Core residue positions were identified in the two GroES-like fold proteins, the *E.coli* GroES and the horse alcohol dehydrogenase. Positional entropy values for the corresponding families at these core positions are shown in Tables IIIc and d. Of the 11 core residue positions in the *E.coli* GroES, eight were highly conserved with positional entropies of <3. Two of the three high-entropy positions, 84 and 86, were largely occupied by non-polar residues, the most predominant being leucine. Another variable position in the core, 73, was mostly occupied by threonine. Ten core positions were identified in the alcohol dehydrogenase. High residue conservation is seen at nine of the 10 core positions, indicated by a positional entropy value of <3 (Table IIId). These positions were largely occupied by small hydrophobic amino acid residues. The predominant residue at position 152, the only high-entropy position in the alcohol dehydrogenase, was valine. Tables IIIc and d indicate the prevalence of non-polar amino acid residues, including valine, leucine and isoleucine, at the core residue positions for the GroES-like fold proteins. The presence of polar, uncharged residues at the core positions, however, is not uncommon. As reported earlier, valines at the core positions in these proteins are seen to be mutable into isoleucines but not to leucines (Table III) (Taneja and Mande, 1999). Unlike the SH3-like barrel fold proteins, proteins in the GroES-like fold did not show a high correlation between residue volumes in the barrel core.

Comparisons within the PDZ domain-like fold

The two representative structures of the PDZ domain-like fold interleukin 16, 1il16 and the neuronal nitric oxide synthase, 1qav superposed well on one another with an r.m.s. deviation of 1.90 A for 77 atoms. Proteins within the PDZ domain family when compared among themselves superposed well on one another with an overall conservation of the β-barrel (data not shown).

A BLAST search with the amino acid sequence of the neuronal nitric oxide synthase (representative of the PDZ-domain family) gave 230 hits. This initial number first rose to 386 owing to splitting of multi-domain sequences, but a final number of 35 sequences was obtained after editing. The number of sequences in the final alignment obtained from the protein interleukin 16 was <30. This protein and the corresponding family were thus not considered for further analysis.

A total of 16 core positions were identified in the neuronal nitric oxide synthase. Of these, 12 positions show high residue
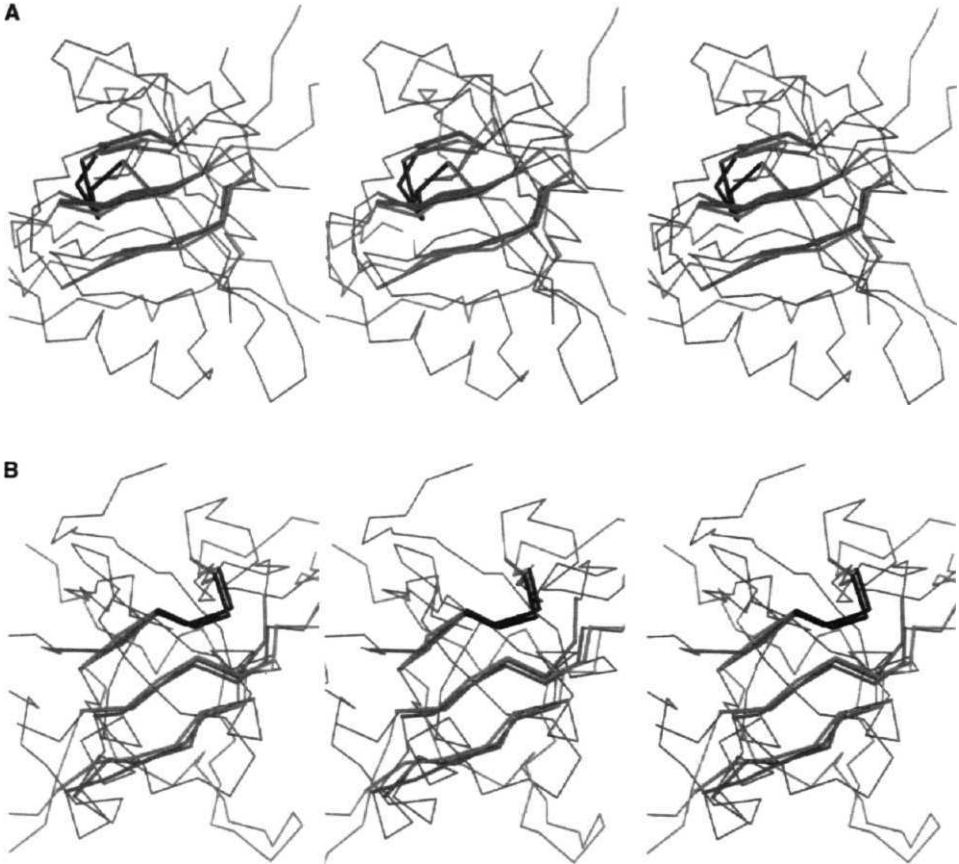
Fig. 3. (A) Superposed structures of the representative proteins of the SH3-like barrel. GroES-like and PDZ domain-like folds. The α-spectrin SH3 domain protein (1shg) is shown in pink, the *E.coli* GroES (1aon) in blue and the neuronal nitric oxide synthase (1qav) in green. The superposed 3₁₀ helix is highlighted in dark blue. **(B)** Superposed structures of the representative proteins of the SH3-like barrel. GroES-like and PDZ domain-like folds. The α-spectrin SH3 domain protein (1shg) is shown in pink, the *E.coli* GroES (1aon) in blue and the neuronal nitric oxide synthase (1qav) in green. The superposed type II β-turn is highlighted in dark blue.

conservation with a positional entropy <3 at each of these positions (Table IIIe). These core positions are predominantly occupied by valine, leucine or isoleucine. Alanine seems to be the residue of choice for the remaining four positions. As in the GroES-like fold, the valines appear to be mutable to leucines rather than to isoleucines (Table IIIe). Core residue positions did not show a significant correlation among residues in proteins considered in this fold.

The final number of sequences in the multiple sequence alignments generated for the representative proteins in the N-terminal domains of the minor coat protein, g3p and Sm motif of small ribonucleoproteins, SNRNP families was <30. Since there were no sequence data for the two families owing to lack of fulfillment of the set criteria for sequence analysis, the

two families and hence the corresponding folds were excluded from the study.

Comparisons across the β-barrel folds

One of the objectives of the study was to identify similarities and dissimilarities across the β-barrel folds. One representative structure from the three β-barrel folds, viz. a-spectrin SH3 domain (SH3-like barrel fold), *E.coli* GroES (GroES-like fold) and the neuronal nitric oxide synthase (PDZ domain-like fold) were therefore considered for the comparisons. The topologies of the three representative structures are different from one another as shown in Figure 1. Comparison of topologies of the SH3-like barrel and GroES-like fold shows the presence of a 3₁₀ helix interrupting the fourth strand in both the β-

973

Table IV. Pair-wise r.m.s. deviations (A) when superimposing structurally equivalent Cot positions across the β-barrel folds. The numbers in parentheses represent the equivalent residues considered during the superpositions

	1shg (SH3 domain)	1aon (*E.coli* GroES)	1qav (PDZ domain)
(a) Superimposition at the 3_{10} helix			
1shg (SH3 domain)	0.0	1.79(38)	1.48 (23)
1aon (*E.coli* GroES)		0.0	1.24 (23)
1qav (PDZ domain)			0.0
(b) Superimposition at the type II β-turn			
1shg (SH3 domain)	0.0	1.62(34)	1.48(23)
1aon (*E.coli* GroES)		0.0	1.50(31)
1qav (PDZ domain)			0.0

barrel folds. The first three strands of the barrel form a similar anti-parallel p-sheet in the two protein folds, yet the two have distinct topologies. The difference lies in the way in which the fourth β-strand hydrogen bonds with the other strands forming the barrel. In the case of the SH3-like barrel fold, the fourth strand runs anti-parallel to the third β-strand followed by the 3_{10} helix. This short helix juxtaposes the fourth strand to the first resulting in the formation of the barrel. The 3_{10} helix in GroES-like fold, however, juxtaposes the fourth strand to the third, for the formation of the complete barrel. Figure 1 also indicates the topology of the PDZ domain-like fold. This fold consists of two helices, one in the region connecting the p-strands two and three and the other between the third and the fourth strands.

The difference in topology of the three representative proteins thus makes it difficult to superpose the corresponding structures. The presence of a common β-barrel structural core, however, may allow comparison of the secondary structural elements forming the barrel in these proteins. Hence, ignoring the topology of the three folds, P-strands of the representative proteins were superimposed on one another. Structural comparisons yielded two alternative ways in which P-strands of the three proteins could be superimposed on one another with minimal r.m.s. deviation values. In one of the superimpositions, the 3_{10} helix of the α-spectrin domain superposes very well on that in the *E.coli* GroES (Figure 3a). The superposition is such that residues of strands 1, 2 and 3 of the α-spectrin domain align with those in strands 3, 2 and 1 of the *E.coli* GroES respectively. The r.m.s. deviation data are as shown in Table IVa. In the case of the PDZ domain-like fold, the structural alignment superposes strands 2, 3 and 4 of the neuronal nitric oxide synthase onto strands 1, 2 and 3 of the α-spectrin SH3 domain, respectively. In an alternative superposition, P-strands of the representative structures align such that strands 4, 1 and 2 of the a-spectrin SH3 domain align with strands 1, 2 and 3 of the *E.coli* GroES, respectively. The r.m.s. deviation data for this superimposition are given in Table IVb. Interestingly, this alternative superposition superimposes a type II (3-tum present in the three structures (Figure 3b).

The β-turn, referred to as the diverging turn in the SH3 domain, occurs at the intervening loop connecting strands 1 and 2. The turn has previously been reported to play a role in protein folding (Riddle *et al.*, 1999; Larson and Davidson, 2000). A similar P-tum is present just before the third strand in the PDZ domain-like fold. The φ–ψ values at this turn correctly place this turn in the type II category (ϕ_{i+1} = -63, ψ_{i+} = 140; ϕ_{i+2} = 98.6, ψ_{i+2} = −8.6).In the GroES-like

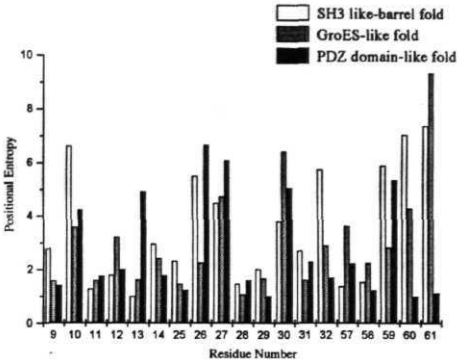**Fig. 4.** Positional entropies at the structurally aligned positions of the three β-barrel folds: SH3-like barrel fold, PDZ domain-like fold and GroES-like fold. The position numbers correspond to that of the α-spectrin SH3 domain protein. Position numbers 9, 11, 25, 31 and 58 are the core positions in all the three protein barrels. Residues 28 and 29 correspond to the $i+2$ and $i+3$ positions in the type II turn in the three-dimensional structure.

fold the turn (ϕ_{i+}, = -58.8, ψ_{i+}, = 134.6; ϕ_{i+2} = 96.3, ψ_{i+2} = -104) is present at the initiation of the third P-strand following the dome loop. The presence of the type II turn in the GroES-like and PDZ domain-like folds suggests that the turn in these protein folds may play a role similar to that observed in case of the SH3 domain. We considered this tum to be a crucial folding nucleus in the β-barrel folds treated in this study. Further analyses were therefore carried out in relation to the superimposition where the P-tum of all the three representative structures superposed on one another as indicated in Figure 3b.

Residue conservation

In order to assess the variability of ammo acid residues across the three P-barrel folds, positional entropies were compared at the structurally aligned positions of the three folds. Upon alignment of the representative structures in the SH3-like barrel, GroES-like and PDZ domain-like folds, three P-strands of each structure superimposed well on one another (Figure 3b). Sequences of the representative proteins were then aligned on the basis of the structural alignment. Residues spanning the first β-strand of the a-spectrin SH3 domain (9–14) aligned with positions 38–43 (strand 2) in the *E.coli* GroES and with residues 106–111 (strand 2) in the neuronal nitric oxide synthase. Residues 25-32, spanning the diverging type II turn (26–29) of the a-spectrin SH3 domain, aligned with positions 59–66 and 123–130 of the *E.coli* GroES and neuronal nitric oxide synthase, respectively. Residues 28 and 29 (numbers correspond to the a-spectrin SH3 domain) form the $i+2$ and $i+3$ positions of the type II turn. Residues spanning strand 4 of the a-spectrin SH3 domain (57-61) structurally aligned with those in the first β-strand of *E.coli* GroES (residues 11–15) as also in the neuronal nitric oxide synthase (residues 95-98). Figure 4 shows the positional entropies at the structurally aligned residues. Among these 19 structurally aligned positions, high conservation in each of the three proteins is seen at eight positions. The positional entropies at these positions are <3 in all the three protein sequence alignments. Remarkably, five of these eight highly conserved positions form the core of the
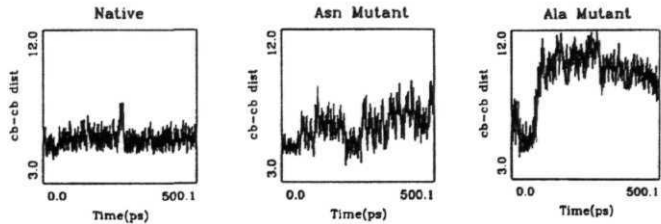
Fig. 5. Comparison of the Cβ–Cβ distance among the native and mutant peptides of *E.coli* GroES during molecular dynamics simulations. It can be seen clearly that the native peptide with aspartate at t + 3 position shows a stable Cβ–Cβ distance. The distance increases rapidly in the aspartate to alanine mutant peptide. See text for details.

Table V. Comparison of positional entropies of the β-barrel folds at the structurally aligned positions. Residue position numbers in the first column correspond to the α-spectrin SH3 domain protein, 1shg

Structurally aligned positions	Residue position in the structure	Positional entropy in SH3-like barrel fold	Positional entropy in GroES-like fold	Positional entropy in PDZ domain-like fold	Positional entropy <3 for all three folds
9	Core	2.8	1.6	1.4	Yes
10		6.6	3.6	4.2	
11	Core	1.3	1.6	1.8	Yes
12		1.8	3.2	2.0	
13		1.0	1.6	4.9	
14		2.9	2.4	1.8	Yes
25	Core	2.3	1.5	1.2	Yes
26		5.5	2.3	6.7	
27		4.5	4.7	6.1	
28	β-Turn	1.5	1.1	1.6	Yes
29	β-Turn	2.0	1.7	1.0	Yes
30		3.8	6.4	5.0	
31	Core	2.7	1.6	2.3	Yes
32		5.8	2.9	1.7	
57		1.4	3.6	2.2	
58	Core	1.5	2.3	1.2	Yes
59		6.0	2.9	5.4	
60		7.0	4.3	1.0	
61		7.4	9.4	1.1	

barrel in all the three protein structures (Table V). The core positions are largely occupied by small non-polar residues. High conservation at core positions in the protein barrel suggests the importance of core residues in the formation and maintenance of the barrel structure.

Of the remaining three highly conserved residue positions, two correspond to residues in the type II β-turn mentioned earlier. The $i + 2$ and $i + 3$ residue positions of the type II turn (corresponding to residues 28 and 29 in the α-spectrin SH3 domain) show a high residue conservation. The predominant residue at the $i + 2$ position is glycine and that at $i + 3$ is aspartate. A high residue conservation at the type II β-turn has earlier been reported for the GroES-like fold at the corresponding positions 62 and 63 of *E.coli* GroES. H-bonding between the side-chain carboxylate of aspartate and main chain amide of the first residue of the turn has been suggested to be important in juxtaposing the β-strands of the barrel, such that the barrel structure is maintained (Taneja and Mande, 1999). A similar side chain-main chain interaction in the SH3 domain has been suggested to stabilize the type II β-turn (Larson and Davidson, 2000). The high degree of conservation seen at the $i + 2$ and $i + 3$ residue positions of the type II turn in the three barrel folds suggests the importance of this turn in formation or maintenance of the β-barrel structure.

Molecular dynamics simulations

Since the side chain–main chain interaction in the type II β-turn appears to be important for the barrel structure formation and maintenance, disruption of this interaction should result in the disintegration of the type II turn. This would ultimately result in the loss of the barrel structure. To investigate the stability of the type II β-turn upon alteration of the aspartate, we performed extensive MD simulations for the GroES peptide and its mutants. The sequence of the peptide taken for MD simulations was VKVGDIV (corresponding to residues 59—65 in the *E.coli* GroES). The starting conformation for the MD simulations was as observed in the crystal structure of the protein. Additional MD simulations were done with the aspartate mutated to asparagine in one case and to alanine in the other. Any alteration in the conformation in the type II turn would immediately be evident from changes in values of the dihedral angles in the type II turn.

A comparison of the ϕ and ψ values was performed for the different residue positions in the β-turn during the 500 ps simulation. The ϕ and ψ values of the $i + 1$ residue remain more or less similar in all the three peptides (data not shown). However, major deviations occur in the ϕ-ψ values of the $i + 2$ residue, glycine, when the $i + 3$ residue is mutated from aspartate to asparagine or alanine. While in the native peptide,

the ϕ and ψ values at ι + 2 position fluctuate around the value of +100 and –40, respectively, $\phi_{i\,+\,2}$ changes to about +150 in the mutant peptides. The $\psi_{i\,+\,2}$ value also drops from -2 to about -100 for both the mutant forms. This, however, occurs after an initial sudden rise of $\psi_{i\,+\,2}$ from -2 to +70. A large deviation is seen at the fourth residue position in the P-turn in the mutant peptide. In the case of the aspartate to asparagine mutation, $\phi_{i\,+\,3}$ drops from about -60 to –141 while it remains stabilized in the native peptide. Comparison of the distance between $C\beta$ atoms of the ι and 1 + 3 residue further corroborates the disintegration of the type II turn upon mutation of the fourth residue in the turn (Figure 5). While this distance is maintained at around 5.6 A in the native peptide, it increases to about 8 A in the aspartate to asparagine mutant and to about 10 A in the aspartate to alanine mutant peptide. These results indicate the importance of the side chain-main chain H-bond interaction in the type II turn maintenance. Alteration of aspartate to asparagine is hence sufficient to cause the disruption of the type II turn conformation. A high conservation of the turn, as also the residues in the turn, thus might be of evolutionary importance in maintaining the structure of the P-barrel.

Discussion

Within various protein families such as serine proteases, cysteine proteases and globins, the three-dimensional structure is remarkably similar despite considerable variations in the amino acid sequences. To a certain extent, conserved residues or conservative changes account for the structural conservation. In addition, correlated pairs of residues have an important role in stabilizing the protein structure. Determination of these conserved features along with the compensatory substitution patterns helps in increasing our understanding of features that may determine the three-dimensional structure of a protein.

In this study, we have attempted to identify folding determinants in the small p-barrel proteins. Conservation patterns across these β-barrel folds reveal interesting similarities of residues at the core of the protein barrels. Irrespective of the topologies, these proteins show a high conservation of small non-polar amino acid residues at the core positions. The core residue positions are predominantly occupied by valine, leucine and isoleucine. Interestingly, valines at the core positions are seen to be mutable into isoleucine and not leucine, an observation reported earlier (Taneja and Mande, 1999). The higher frequency of substitution of isoleucine by valine has been attributed to a higher P-sheet propensity of isoleucine and valine than leucine (Wilmot and Thornton, 1988). Branching of side chains at the $C\beta$ positions in both valine and isoleucine, but not leucine has previously been suggested as a possible reason for such a mutation pattern (Taneja and Mande, 1999). The observed mutation pattern and a high conservation of non-polar side chains suggest that the overall hydrophobic pattern of amino acids may drive the protein sequence to collapse into the P-barrel conformation.

Correlation analysis of the SH3-like barrel fold suggests that maintenance of the total core volume occurs within the SH3 domain family of proteins. Amino acid substitutions resulting in an increase or a decrease in the volume of the core is compensated by replacement of another amino acid residue. This amino acid residue is present at a position that might be distant in sequence, but near in space to the mutated residue so as to conserve the total volume of the core and hence the overall folded structure. Interestingly, an earlier

analysis of 266 SH3 sequences did not find evidence for correlated substitutions (Larson and Davidson, 2000). We suggest that our criteria of choosing sequence identities between 25 and 90 generates a more accurate multiple alignment for a robust statistical analysis. Accuracy of the alignment is reflected in observation of the covarying mutations.

Of the common features, the presence of a type II P-turn is the most intriguing. This turn seems to be important in the formation of the β-barrel. Earlier studies have reported the importance of the β-turn in SH3 domain (Riddle et al., 1999; Larson and Davidson, 2000). Conservation of this turn, not only in proteins constituting one of the β-barrel folds but also across the various β-barrel folds considered in this study, suggests that this region might be an important nucleation site in the folding pathway of the β-barrel proteins (Riddle et al., 1999; Larson and Davidson, 2000). This nucleation appears to be guided by the residues present within the turn. High residue conservation has been seen at the ι + 2 and ι + 3 residue positions. While the presence of glycine at i + 2 guides the protein sequence into a turn, aspartate at ι + 3 is important for a unique side chain–main chain interaction. Simulation studies corroborate similar conclusions of independent work carried on an SH3 peptide (Krueger and Kollman, 2001). Furthermore, our analysis suggests that alteration of aspartate to asparagine or alanine destabilizes the P-turn conformation. The glycyl-aspartyl dipeptide hence appears to be a major factor in helping maintain the integrity of the barrel.

Our study shows interesting similarities among proteins in the different P-barrel folds. Despite large differences in sequence and function, the occurrence of a conserved glycyl-aspartyl dipeptide, intriguingly at a conserved type II turn, suggests the importance of the turn and the residues forming it in the formation of the P-barrel. In addition, a conserved hydrophobic core suggests its role in maintenance of the barrel structure. Further studies such as site-directed mutagenesis should confirm the importance of these conserved features in the formation and maintenance of protein structure.

References

Altschuh,D.. Vernet,T., Berti,P., Moras,D. and Nagai,K. (1988) *Protein Eng..* 2, 193-199.
Altschul,S.F., Madden,T.L., Schaffer,A.A.,Zhang,J., Zhang,Z., Miller,W and Lipman,D.J. (1997) *Nucleic Acids Res.*, 25, 3389-3402.
Anfinsen,C.B. (1973) *Science*, 181, 223-230.
Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T and Tasumi,M. (1977) *J.Moi BioL*, 112, 535-542.
Bowie,J.U., Reidhaar-OlsonJ.F. Lim,W.A. and Sauer,R.T. (1990) *Science*. 247. 1306–1310.
Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984) *J. Mot. Biol.*, 179, 125-142.
Harpaz,Y., Gerstein,M. and Chothia,C. (1994) *Structure*, 2. 641-649.
Hubbard,S.J., Campbell,S.F. and Thornton,J.M. (1991) / *Mol. Biol.*, 220, 507-530.
Jones,T.A., Zou,J.Y., Cowan,S.W and Kjeldgaard,M. (1991) *Acta Crystallogr.*, A47, 110-119.
Kamtekar,S., Schiffer,J.M., Xiong,H., Babik,J.M. and Hecht,M.H. (1993) *Science*, 262, 1680–1685.
Kraulis,P.J. (1991) *J.Appl. Crystallogr.*, 24, 946–950.

Krueger,B.P. and Kollman,P.A. (2001) *Proteins: Struct. Funct. Genet.*, 45, 4–15.

Larson,S.M. and Davidson,A.R. (2000) *Protein Sci.*, **9**, 2170-2180.

Mandel-Gutfreund,Y., Zaremba,S.M. and Gregoret,L.M (2001) *J. Mol. Biol.*, 305, 1145–1159.

Martinez,J C and Serrano.L. (1999) *Nature Struct. Biol., 6,* 1010-1016.

Merkel,J.S., Sturtevant,J.M. and ReganX. (1999) *Structure*, **7**, 1333-1343.

Murzin,A.G.,Lesk,A.M. and Chothia,C. (1994) J. *Mol. Biol.*, **236,** 1382-1400.

Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C (1995) *J. Mol. Biol.*, **247,** 536-540.

Pollock.D.D. and Taylor,W.R. (1997) *Protein Eng.*, **10**, 647–657.

Riddle,D.S., Grantcharova,V.P., Santiago,J. V., Alm,E., Ruczinski,I. and Baker,D (1999) *Nature Struct. Biol.*, **6**, 1016-1024.

Serrano,L. (2000) *Adv. Protein Chem.*, 53, 49–85

Shenkin,P.S., Erman,B. and Mastrendrea,L.D. (1991) *Proteins,* **11**, 297-313.

Shindyalov,I.N., Kolchanov,N.A. and Sander,C. (1994) *Protein Eng.*, 7, 349–358.

Taneja,B. and Mande,S.C. (1999) *Protein Eng.*, 12, 815-818.

Thomson,J.D., Higgins,D.G. and Gibson,TJ. (1994) *Nucleic Acids Res.,* 22, 4673-80.

Wilmot,C.M and Thornton,J.M (1988) *J. Mol Biol.*, 203, 221-232.

Yi,Q., Bystroff,C.,Rajagopal,P., Klevit,R.E. and Baker,D (1998) *J Mol. Biol..* 2*3, 293–300.