University of Southampton

Investigations into the Structure and Subunit Dispersal of the Pea Plant (*Pisum sativum*) Chloroplast Chaperonin Ch-cpn60

Terence Robinson

A thesis submitted for the degree of
DOCTOR OF PHILOSOPHY

January 2003

Division of Biochemistry and Molecular Biology

ABSTRACT

Preparations of the plastid chaperonin Cpn60 from pea leaves (*Pisum sativum*) contain equal amounts of two homologous subunit types, Cpn60α and Cpn60β. The organization of these subunits within the tetradecameric chaperonin complex and the functional significance of their distribution are unknown.

In this thesis I describe the preparation of both subunit types by overexpression in *E.coli* and confirm that while Cpn60β can assemble as an homooligomer, Cpn60α remains monomeric. The structural basis of this divergent behaviour was analysed by homology modeling with the X-ray structure of GroEL. This revealed a selection of sequence differences in Cpn60α that would be expected to destabilize intersubunit interactions.

A method was developed to crosslink adjacent plastid Cpn60 subunits with a bifunctional reagent, SPDP. By reacting with a single cpn60β Cys 519 residue and a lysine 22 residue, ~6Δ apart on an adjacent α or β subunit, the method has the potential to report a nearest-neighbour subunit identity in mixed complexes.

Rhombohedral crystals (R32, a=b=108Å, c=154Å γ=120°) were grown from the cpn60α preparation and X-ray analysis attempted to discover a structural basis for the behaviour of the protein. The asymmetric unit of the crystal was too small to accommodate an entire subunit, suggesting that proteolysis had occurred during crystallization. A weak molecular replacement solution employing the *E.coli* apical domain as the search model indicated that the crystal contained a portion of the intermediate domain also. In parallel to the molecular replacement experiments, a set of Seleno-methionine experiments were attempted with the intention of carrying out MAD structural analysis of the cpn60α subunit, via the bioincorporation of the SeMet analog. However, the lack of SeMet incorporated product potency, due to the toxicity of the SeMet to the cpn60α protein, was to become a major hinderance to completing this set of experiments.

Finally, steady-state K^+ dependent ATP hydrolysis rates were determined for isolated subunits and oligomeric complexes. Modulation of subunit activities by complex formation is likely in view of the known allosteric regulation of GroEL and such measurements may report on interactions between cpn60α and cpn60β subunits.

This thesis is dedicated to my wife Sandi

and

(In memory of my father)

"Knowledge is proud of what he knows

Wisdom is humble he knows no more"

W.Cowper (1785)

The '*Task*'

Table of Contents

FIGURES

TABLES

ABBREVIATIONS

ADP	Adenosine diphosphate
ATP	Adenosine 5'-triphosphate
ATPase	Enzyme hydrolysing ATP to ADP
Arg	Arginine (R)
Asp	Aspartic acid (D)
Ala	Alanine (A)
Asn	Asparagine (N)
Å	Ångstrom (10^{-10} metres)
Abs	Absorbance
BCA	Bicinchoninic Acid
bp	base pair
BSA	Bovine Serum Albumin
^{o}C	Degrees Celsius
CCT	Chaperonin-Containing TCP-1
Ch-cpn60	Chloroplast chaperonin 60
CCP4	Collaborative Computational Project, Number 4
cDNA	Complimentary DNA
cm	centimetre
CNBr	Cyanogen Bromide
Co-cpn	Co-chaperonin
Cryo-EM	Cryogenic-Electron Microscopy
Cys	Cysteine (C)
dH_2O	deionised water
DMSO	Dimethyl Sulphoxide
DNA	Dideoxyribonucleic Acid
DTNB	5,5'-dithiobis-(2-Nitrobenzoic Acid)
DTT	Dithiothreitol

E.coli	Escherichia coli
EDTA	Ethylenediamineteraacetic Acid
ESRF	European Synchrotron Research Facility
F	Structure Factor
F_{calc} (c)	Calculated structure factor amplitude
F_{obs}(o)	Observed structure factor amplitude
FFT	Fast Fourier Transform
FPLC	Fast Protein Liquid Chromatography
g	Gravity
GA	Gluteraldehyde
Gln	Glutamine (Q)
Glu	Glutamic acid (E)
Gly	Glycine (G)
HCl	Hydrogen Chloride
Hetero	Heterologous
His	Histidine (H)
His.Tag	Histidine.Tag
Homo	Homologous
HPLC	High Protein Liquid Chromatography
HSL	Homo Serine Lactone
Hsp	Heat shock protein
hsp	Heat shock protein (gene)
H8 + H9	Helix 8+ Helix 9
IEX	Ion Exchange
Ile	Isoleucine (I)
IPTG	Isopropyl β-D-thioglucopyranoside
K^o	Kelvin (Temperature)
Kb	Kilobase pair
KDa	Kilo Dalton
KNF	Koshland, Nemethy, Filmer
KSCN	Potassium Thiocyanide

LB	Luria Broth
Leu	Leucine (L)
LDH	Lactate Dehydrogenase
LSU-BP	Large Subunit-Binding Protein
Lys	Lysine (K)
M	Mole
mA	milliamp
MAD	Multi-wavelength Anomalous Dispersion
MDH	Malate DeHydrogenase
Mer	Oligomer
Met	Methionine (M)
mg	Milligram
ml	Millilitre
Mg^{2+}	Magnesium
MgCl	Magnesium Chloride
MOLREP	Molecular Replacement Program
mon	Monomer
mt-cpn	Mitochondrial-chaperonin
MR	Molecular Replacement
mM	milliMolar
MS-ES	Mass Spectroscopy-Electro Spray
mV	milli Volt
Mw	Molecular weight
MWC	Monod, Wyman, Changeux
NaCl	Sodium Chloride
NAD+	Nicotinamide Adenine Dinucleotide
NADH	NAD+ (in the reduced form)
NCS	Non Crystallographic Symmetry
ng	nanogram
NHS	N-Hydroxy Succinimide
nm	nanometre

NMM	New Minimal Media
NMR	Nuclear Magnetic Resonance
o/n	overnight
O.D.	Optical density
Pea rubb	Pea Rubisco β subunit
Pea ruba	Pea Rubisco α subunit
PDB	Protein Data Bank
PEP	Phosphoenolpyruvate
PCR	Polymerase Chain Reaction
PC	Patterson Correlation
PK	Pyruvate Kinase
Phe	Phenylalanine (F)
PMSF	Phenylmethanesulphonyl fluoride
Pro	Proline (P)
PVDF	Polyvinyl difluoride
RNA	Ribonucleic Acid
rpm	revolutions per minute
R/T	Room Temperatue
RUBISCO	Ribulose 1,5-bisphosphate Carboxylase-Oxygenase
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SeMet	Seleno-Methionine
SEC	Size exclusion Chromatography
Ser	Serine (S)
SG	Space Group
SGI	Silicon Graphics Interactive Workstation
SIB	Swiss Institute of Bioinformatics
SOC	Supplied growth medium
TCP-1	Tailless Complex Polypeptide-1
TFA	Trifluoroacetic Acid
TF55	Thermophilic Factor (55kDa)
Thr	Threonine (T)

TMD	Targeted Molecular Dynamics
TRiC	TCP-1 Ring Complex
Tris	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan (W)
Tyr	Tyrosine (Y)
µg	Microgram
µm	micrometer
UV	Ultra Violet
Val	Valine (V)
Ve	Elution Volume
v/v	volume/volume
w/v	weight/volume
%	Percentage
@	at
Σ	Sum Of
ε	Molar Extinction Coefficient
λ	Wavelength
[ ]	Concentration

Acknowledgements

I owe a special debt of gratitude to the following people. To my wife Sandi for her continued support, encouragement and personal sacrifices, which were selflessly given and greatly appreciated. To my daughters Maria and Jessica, again for their support and who, never once, questioned or doubted my sincerity. Also special thanks go to my sister Rebecca for her continued emotional and financial support.

I am especially grateful to three people without whose help, knowledge and encouragement, this thesis would not have been written. First, and foremost, to Professor Steve Wood for bestowing upon me his insight and knowledge and also the courage to doubt, reason and then question all that is before me. Secondly, I would like to thank Dr Alun Coker for his friendship and unswerving commitment to my education in crystallography. Finally, my thanks go to Dr Jon Cooper who imparted the knowledge of crystallography with patience and a quiet humility.

Further thanks go to Dr Maria Hidalgo Lara, for providing both the crystals to work on and the ch-cpn60 α and β protein starting products, so vital to my PhD. Thanks go to Dr Graham Dunn, for his help with the Selenomethionine experiments.

Thanks go to the following people for both their help and friendship; Dr Fiyaz Mohammed, Monty, Paul, Gordon, Dr Shu-Fen Coker, Dr Raj Gill, Dr Leighton Coates, Dr Sanjay Mall and Simon.

I wish to thank the BBSRC for funding me during this PhD

It just leaves me to say that, without the time, effort and personal sacrifices made by all of the above people on my behalf, my abilities to see further now than before I began this journey would have been greatly diminished. Thank you all.

CHAPTER ONE

1. Protein Folding

1.1. Introduction

The genetic code is linear and encodes a peptide sequence, which is also linear. However, before the linear protein chain produced on the ribosome can fulfil its functional role it must fold into its three-dimensional "native" configuration.

Protein folding poses two fundamental problems. Firstly, how can proteins reach a correctly folded state on a non-astronomical time-scale given the plethora of possible conformations available? This problem is often referred to as Levinthal's paradox (Karplus,1997., Levinthal, 1966). Secondly, what makes *in vivo* protein folding so efficient, considering the adverse conditions in the crowded environment of the cell where the concentrations of macromolecules exceed 30% (w/v) (Ellis & Hartl, 1996; Luby-Phelps, 1994).

The 'classic' experiments of Anfinsen (1973) on the refolding of unfolded ribonuclease demonstrated that the primary structure alone contained all of the information necessary for its folding into a stable and functional three-dimensional structure. Furthermore, the driving force for protein folding was the attainment of the lowest Gibbs free energy of the protein in its normal physiological milieu. This is known as the 'thermodynamic hypothesis' (Epstein *et al.*,1963)

Most *in vitro* studies of protein folding have used small, single-domain proteins that undergo cooperative and reversible folding. Such systems are "ideal" as the folding reaction is not complicated by off-pathway side reactions, such as aggregation. In addition, these studies have usually been performed under conditions of high dilution and low temperature, giving the folding protein chain sufficient space to explore alternative conformations in its quest for the "global" thermodynamically stable state.

However, the environment in the cell is very crowded; the effective protein concentration inside a typical cell has been estimated to be as high as 300mg/ml (Zimmerman and Trach, 1991). The effect of this "crowding" is to hinder the folding of the protein because of interactions with neighbouring molecules, resulting either in attainment of a structure with a Gibbs free energy higher than that of the native state (kinetic-trap) or aggregation.

Refolding experiments often lead to the formation of kinetically trapped intermediates that aggregate, even in dilute aqueous solutions and at low temperature (Jaenicke, 1998). As aggregation is at least partly driven by hydrophobic interactions, it is even more pronounced when folding is attempted under the physiological conditions prevalent in the cell. In particular, the very high concentration of macromolecules creates conditions of crowding that highly favour aggregation (Ellis, 1997).

1.2 The Discovery of Molecular Chaperones

It was not until the discovery of molecular chaperones that some insight was gained into how the cell overcomes the problem of unproductive folding of proteins. Chaperones are a ubiquitous class of proteins that sometimes interact with the nascent protein chain to help mediate correct folding within the crowded intracellular environment.

Laskey (Laskey *et al.*, 1978) first coined the term 'Molecular Chaperones' to describe the properties of nucleoplasmin, an abundant acidic nuclear protein in *Xenopus* oocyte extracts, which are required for the assembly of nucleosomes from histones and DNA. It was observed that the functional role of nucleoplasmin was to reduce the positive charge on the highly charged histone protomers to facilitate the correct interactions with DNA.
A second protein, rubisco (Ribulose 1,5-bisphosphate carboxylase-oxygenase) large subunit binding protein, was later postulated to fulfil a similar role in the pea as it was observed to facilitate the correct assembly of rubisco subunits in the chloroplast (Musgrove, 1987, Hemmingsen *et al.*, 1988).

Following the *in vitro* reconstitution of the chaperone GroEL/ES(GroES is a co-chaperone of GroEL)-dependent assembly of dimeric rubisco (Goloubinoff *et al.*, 1989), several studies using purified proteins have now established that the primary function of chaperones is indeed to assist in facilitating the correct folding of monomeric polypeptide chains (Laminet *et al.*, 1990., Martin *et al.*, 1991., Viitanen *et al.*, 1991., Zheng *et al.*, 1993). These observations shed new light on earlier experiments which had shown that enzymatically inactive proteins could be salvaged only in the presence of certain "accessory proteins".

The subsequent cloning and sequencing of the bacterial *HSP60* gene revealed that the encoded 60kDa heat shock protein (hsp60) was homologous to the bacterial stress chaperone protein, GroEL, and the rubisco subunit binding protein, Ch-cpn60 (Chloroplast-chaperonin 60), (Reading *et al.*, 1989). It was further demonstrated that mitochondria, chloroplasts and bacteria all possess closely related chaperones that are essential for the assembly of some oligomeric proteins *in vivo* (Barraclough and Ellis, 1980, Musgrove; 1987). Accordingly, all proteins assisting protein folding and preventing aggregation are now referred to as 'molecular chaperones'. A plethora of studies have now shown that molecular chaperones influence protein folding by interacting non-specifically and promiscuously with elements of non-native structure exposed in the unfolded protein.

1.3 The Concept of the Folding Pathway

The folding of a globular protein from its denatured state is very rapid, and often occurs in less than a second. Intuitively, the sampling of all conformations to attain the precise tertiary structure of lowest global energy on such a short time-scale, appears quite incredible. The paradox that such a phenomenon should indeed occur, was succinctly expressed by Cyrus Levinthal (1966) and is now known as the 'Levinthal Paradox'.

Given that about 10^{50} conformations can exist for a polypeptide chain such as ribonuclease, Levinthal calculated that even if the molecule could try a new conformation every 10^{-13} seconds, it would take 10^{30} years to sample a significant fraction of them; a period much longer than the

postulated age of the known universe. And yet ribonuclease was experimentally observed to fold in about one minute! The resolution to Levinthal's Paradodox came initially from kinetic studies, which demonstrated that folding took place via a series of intermediate states, leading to the concept of the protein folding "pathway".

The 'classic' notion of a protein folding pathway is one that includes a series of discrete intermediates (Kim and Baldwin,1990). To obtain a global overview of the folding process, the folding pathway may be pictured as a trajectory on an energy landscape, in which the x and y dimensions are spatial ordinates and the vertical z dimension is the free energy ordinate. 'Energy Landscape Theory' (Onuchic *et al* .,1995,Dill and Chan,1995), maintains that there is no unique folding pathway but rather that there exist a multiplicity of folding pathways which converge to the native state (Onuchic *et al.*,1995, Go,N.,1983 Bryngelson *et al.*,1987, Leopold *et al.*,1992). It is proposed that folding is initiated by a nucleation event which may start at a number of points on the energy landscape and that all of these partially folded structures will be 'funneled' by free energy minimisation toward the final state. The essence of this 'funnel landscape' concept is one of competition between the tendency toward the folded state and trapping in local minima, which can be described by the 'ruggedness' of the landscape. Good (evolutionary selected) folding sequences will by definition fold rapidly on minimally frustrated landscapes (e.g. a 'smooth' funnel; Onuchic *et al.*,1995), due to the existence of only a few kinetic traps between the fully unfolded and folded states. Minor variations in pH, temperature, or mutations in the protein primary structure will alter the native conformation by favouring other low-energy structures, which in a funnel-like landscape are very similar in energy to the native state.

The validity of the concept of the existence of multiple folding pathways on a "folding funnel" of varying ruggedness, has been demonstrated by several computational simulations using minimalist models (Onuchic *et al.*,1995, Dill,1997, Scheraga,1992), (Fig 1.3.1).

Further evidence suggests that 'off-pathway' states exist which correspond to local free energy minima on the funnel and which may temporarily trap the protein (Dill and Chan; 1997).

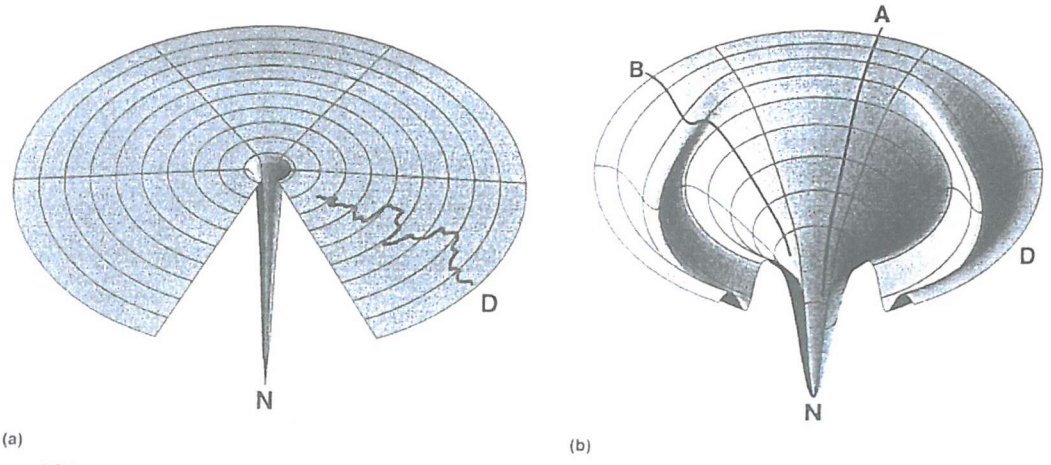


Fig 1.3.1 Energy surfaces to visualise protein conformations. (a) The Levinthal "golf-course" energy landscape. **N** is the native conformation and **D** the denatured state. The chain searches for **N** randomly, that is, on a level playing field of energies, and has very little chance of finding the tiny hole leading to the native state. **(b)** A more realistic energy landscape, which tends to "funnel" the protein toward the native state. Many paths are possible, some of which lead straight "downhill" like **A**, and give rapid folding. Others, like **B**, may lead to secondary energy minima, which may delay folding. (Dill and Chan, (1997).

1.4 *In vitro* and *In vivo* Protein Folding

To Anfinsen's great credit, studies of protein folding over the last fifty years have confirmed that the folding of proteins in optimal solvent conditions is indeed an autonomous and spontaneous process, governed exclusively by the amino acid sequence. However, most of these studies have been performed *in vitro* and this begs the obvious question of how far the results obtained can be extrapolated to the *in vivo* regime.

A series of experiments performed by Pelham, (1989) clearly disproved the then commonly held belief that *in vivo* protein folding yields 100% native protein. This observation challenged the dogma that spontaneous and correct protein folding required only the information present in the primary sequence and demonstrated that misfolding and misassembly do indeed occur.

In so doing, the investigators also demonstrated that cytosolic yield of native protein is not significantly different from the *in vitro* case if accessory proteins are used to overcome such obstacles as aggregation and large temporal folding differences. These observations suggest that the native folds achieved in *in vitro* studies are indeed significant, and consequently deemed valid for the study of protein folding *in vivo*.

1.5 The Function of Molecular Chaperones *in vivo*

Uncertainty stills exists about the *in vivo* role of molecular chaperones in protein folding processes. Undoubtedly, chaperones play important biological roles, for example, deletion of a number of *E.coli* chaperones such as GroEL/ES and Grp E results in loss of *in vivo* protein viability (Fayet *et al.*, 1986). Deletion of other chaperones, DnaK and DnaJ, leads to temperature-sensitivity (Ang *et al.*, 1991).

One of the initial pieces of evidence that chaperones mediate or facilitate folding reactions *in vivo* was obtained by studying the effects of GroEL/ES over-expression on the assembly of recombinant rubisco subunits in *E.coli* (Goloubinoff *et al.*, 1989).

It was further demonstrated that around one half of all *E.coli* proteins could bind to GroEL when in the denatured state (Viitanen *et al.*, 1992). It is a generally held belief that association with chaperones prevents aggregation of nascent chains during translation, thereby contributing to the correctly folded protein emerging from the ribosome.

1.5.1 Stress Response

As chaperones are synthesised at increased levels, when organisms are under heat shock stress or other stress, their role in the thermo-resistance of basic cellular processes is apparent.

The expression of one particular class of Hsps, the small Hsps (sHsps), is dramatically upregulated upon heat stress and these chaperones have been demonstrated to mediate thermo-resistance in plants and mammalian cells. The upregulation of host sHsps is also observed in the over-expression of heterologous proteins in recombinant systems (Goloubinoff *et al.*, 1989; Gragerov *et al.*, 1992; Escher and Szalay; 1993). The many roles played by chaperones under cellular stress conditions further demonstrate their ability to cooperate in flexible, functional networks (Buchberger *et al.*, 1996).

1.5.2 Chaperone Action *In vitro*

Although chaperones are known to transiently bind to misfolded proteins, which increases the proteins solubility to enable correct folding, chaperones may also facilitate the dissolution of misfolded protein aggregates. Indeed, Lindquist and colleagues have demonstrated the dissolution of protein aggregates in intact cells by the chaperone Hsp 104 (Parsell *et al.*, 1994).
The interior of a cell is a very crowded and dynamic environment and consequently the cytoplasm no longer behaves as an ideal folding fluid. Unassisted refolding to the native structure (i.e. in the absence of chaperones) under these conditions competes with protein aggregation (Zettlmeissl *et al.*, 1979).

Conversely, the inclusion of chaperones during renaturation experiments has been demonstrated to suppress aggregation (Holl-Neubebaur, *et al.*, 1991; Buchner *et al.*, 1991; Schroder *et al.*, 1993).

In vitro studies on Hsp 90 have shown that members of this chaperone family facilitate the renaturation of unfolded proteins by suppressing aggregation (Wiech *et al.*, 1992). Protein aggregates are thought to result from a heterogenic (sometimes homogenic, e.g. inclusion bodies in recombinant over-expression) accumulation of various unfolded conformations. Consequently, it is possible that certain protein aggregates may be better substrates for chaperone-mediated dissolution than others. A further example of chaperone assisted reactivation of protein aggregates has been presented Ziemienowicz *et al.*, 1993. This group has demonstrated that aggregates of DNA polymerase or RNA polymerase can be reactivated *in vitro* by both GroEL/ES and the chaperones of the DnaK system.

Folding intermediates trapped in one of the many states of energy just higher than that of the native state (molten globules) have also been demonstrated to bind to chaperones, the binding energy of the interaction being used to induce structural transformation to the native state (Corrales and Fersht, 1995).

Despite the promiscuous nature of their interactions with misfolded and aggregated proteins, chaperones do exhibit certain substrate specificity. Distinct features have begun to emerge for the individual chaperones, in terms of mechanistic aspects, substrate specificity, and the dependence on ATP.

Although the *in vitro* assays used are as near to representative of cytosolic conditions, they are, by any standards, regarded as artificial. However, there exist several attestations to the mechanistic principles determined *in vitro* also being just as valid when applied to the *in vivo* situation.

It is now generally accepted that proteins can attain their native conformations in an unaided and 'spontaneous manner' but that certain protein folding pathways and environments do require chaperone assistance. For example, the GroE chaperones are known to interact with the small protein, barnase, but this protein is perfectly capable of 'spontaneous' refolding to the native state, with yields of 100% in the absence of any chaperones (Gray and Fersht, 1993). Similarly,

β-actin and α-tubulin form stable complexes with the GroEL chaperone machinery, but it has been observed that the refolding of these proteins is not facilitated by the chaperone (Tian *et al.*, 1995). Finally, denatured succinyl-CoA synthetase appears unable to interact with the GroEL chaperone, in any form (Fong and Bridger, 1992).

1.6 The Utility of the Chaperone-Mediated Folding Phenomenon

With the discovery that *in vivo* protein folding can be facilitated by the simultaneous over-expression of molecular chaperones, this has now become a popular method to increase the yield of correctly folded recombinantly expressed proteins, in particular the GroE chaperones have been shown to facilitate the correct folding of a large number of structurally and functionally unrelated proteins (Mendoza *et al.*, 1991, Martin *et al.*, 1991, Zheng *et al.*, 1993, Grimm *et al.*, 1993, Goloubinoff *et al.*, 1989). Furthermore, a large number of studies have also examined the effects of GroEL/ES over-expression on the solubility of aggregation-prone recombinant proteins. Goloubinoff *et al* (1989) first demonstrated that co-over-expression of the GroE operon increased ten-fold the yield of active *A. nidulans* rubisco hetero-oligomers (Thomas and Baneyx, 1996; Fong and Bridger; 1992, Horwich *et al.*, 1993). However, proteins larger than ~60kDa cannot be accommodated within the central folding cavity of GroEL and are unable to undergo GroES-assisted folding (Weissman *et al.*, 1995, Mayhew *et al.*, 1996).

Although chaperones are currently thought to function by assisting aggregation-prone folding intermediates to reach a correctly folded native state via specific binding mechanisms, their effects may also be exerted indirectly. For example, over-production of a chaperone within a system may support not only the correct folding of host proteins which have co-evolved, but also that of heterologous recombinant proteins.

Recently the Clp ATPases, members of the *E.coli* Hsp 100 family, have also been shown to function in a similar way to DnaK-DnaJ chaperones (Wickner *et al.*, 1996, Wawryznow, 1995). More specifically, Clp ATPase co-over-expression appears to prevent the degradation of certain proteins, presumably by assisting the correct folding of unstable polypeptide molecules (Lee and Olins,1992, Battistoni *et al.*, 1993).

1.7 *In vitro* refolding of Inclusion Body Proteins

The cloning of several molecular chaperones in *E.coli* expression vectors has facilitated their *in vivo* use in recombinant systems to circumvent the problem of inclusion body formation. Since inclusion bodies are enriched with the target protein and are relatively easy to purify, *in vitro* protein refolding of over-expressed protein aggregates is often an attractive alternative to soluble protein production (Rudolph and Lilie, 1996, Georgiou and Bowden, 1990). The efficiency of refolding of inclusion body proteins varies according to the protein of interest, depending on the nature of aggregation side reactions that compete kinetically with correct folding; the yield can often be dramatically improved by optimising the refolding conditions (Mendoza *et al.*, 1991, Martin *et al.*, 1991, Peralta *et al.*, 1994, Hansen and Gafni, 1993).

It is important to note that the chaperone systems do not always effect a beneficial outcome on protein folding. There are monomeric variants of the *E.coli* GroEL chaperone that, whilst retaining the ability to bind nucleotides, are unable to suppress aggregation pathways or to promote the refolding of target proteins (Weich *et al.*, 1992).

Conversely, truncated versions of the GroEL protomer have been observed to function as ATP-independent "mini chaperones" (i.e. apical domains), although with a lower efficiency than that of the fully functional oligomeric complex (Makino *et al.*, 1993, Zahn *et al.*, 1996). Furthermore, truncated monomers of the GroEL homologue from *Thermus thermophilus* can increase the refolding yields of lactate dehydrogenase and rhodanese in the absence of adenine nucleotides (Taguchi *et al.*, 1994). Chaperone assisted refolding of target proteins without the necessity for co-factors has further increased the utility of chaperones. One such example is the abilities of the truncated protomers of the GroEL homologue of *T. thermophilus* to facilitate the ATP-independent refolding of rhodanese, when immobilised on a TSK matrix column (Taguchi *et al.*, 1994).

1.8 Chaperones and the Immune Response

Because of their ubiquity and abundance, the chaperones of microbial pathogens are powerful effectors of the host immune response (Kaufmann, 1994). Indeed, immunisation with pathogenic chaperones has been demonstrated to induce protective 'adaptive' immunity. For example, inoculation of laboratory mice with the chaperone cpn60 from *Yersina enterocolitica* protects them from infection by this gastro-enteritis causing pathogen (Noll *et al.*, 1994). However, pathogens entering the host undergo a stress response characterised by increased chaperone synthesis (Kaufmann, 1990;1991) and in turn the infected host cells undergo a corresponding stress response and also increase chaperone synthesis (Jäättelä ,1990, Kaufmann *et al.*, 1991, Steinhoff *et al.*, 1991). Consequently, although chaperones may serve to promote adaptive protective immunity against pathogens, they may also contribute to immune diseases via the unavoidable stress responses of both pathogen and host.

Nevertheless, chaperones have been used successfully as vaccine carriers for pathogen-specific antigens and this approach may eventually prove to be useful in the control of infectious disease (Ellis, 1996).

It was Fowler and Fowler (1976) who coined the term "multiplex" as signifying a process involving simultaneous transmission of several messages along a given channel of communication. That is to say, chaperonins act simultaneously on lymphocytes, macrophages, fibroblasts and other cells in response to stress, which in turn can result in simultaneous transmission of immune messages, such as cytokines to immune effector cells, which can then adapt to cope with the stress (Ellis, 1996).

In summation, the exploitation of the folding machinery 'naturally' provided in the form of the molecular chaperones, could be seen as an ideal step forward in the desired acquisition of facilitated folding processes. Researchers are frequently confronted with protein misfolding and aggregation upon over-expression of proteins in host cells, such as *E.coli*. In order to obtain active protein from the aggregated inclusion body material, *in vitro* protein refolding must be performed. This scenario presents an advantage of using molecular chaperones as endogenous

folding enhancers, which in turn may prove to be a most potent ally in the prevention of unwanted protein misfolding.

1.9 The Chaperone and Chaperonin systems

Genetic and biochemical analysis shows that several distinct and highly conserved chaperone systems exist; with many chaperones being designated as heat shock (stress) proteins.

Two major classes of ATP-dependent chaperones, Hsp70 and the cylindrical chaperonin complexes, have been implicated in *de novo* protein folding in the cytosol of eukaryotic and prokaryotic cells, as well as in organelles of presumed endosymbiotic origin, namely mitochondria and chloroplasts. For example, the ubiquitous Hsp70 and Hsp60 (chaperonin) chaperones recognise hydrophobic surfaces in the context of extended and collapsed (globular) conformations respectively, which are bound correspondingly either by 'local' enclosure of the chain or by 'global' enclosure of the polypeptide in a central cavity. Although both systems assist proteins to reach a functional conformation, in most cases through the expenditure of ATP, Hsp70 and the chaperonins are structurally and functionally distinct and represent radically different principles of chaperone action.

Chaperonins are a particular class of chaperones (Hemmingsen *et al.*, 1988); their multimeric toroidal structures allow them to sequester non-native proteins inside their central cavity and to direct the chemical energy derived from ATP hydrolysis into the promotion of the folding process.

The existence of molecular chaperones does not violate the principle of protein self-assembly. It can be said that although chaperones interact with a large number of different substrate polypeptides, they do not possess the information that would influence the final outcome of a folding reaction. Rather, by preventing (and perhaps reversing) unproductive intra- and intermolecular interactions, they enable the unfolded polypeptide to reach the definitive structure specified by its amino acid sequence (Hendrick and Hartl, 1993, Hartl *et al.*, 1994).

1.10 Chaperone Systems

1.10.1 Hsp70, Its Co-Chaperones And Homologs

Historically, Hsp70s were identified by induction under conditions of stress, during which they are now known to provide an essential action of preventing aggregation and assisting refolding of misfolded proteins. Hsp70s are a highly conserved family of chaperones, distributed ubiquitously in all prokaryotes and in the cellular compartments of eukaryotic organisms, some compartments containing multiple Hsp70 homologs with distinct cellular functions (Craig *et al.*, 1994).

Most Hsp70s are ~70kDa (Mw) and consist of two functionally coupled domains. Binding and release of the substrate rely on modulation of the intrinsic peptide affinity of Hsp70 by cycles of ATP binding and hydrolysis by the N-terminal domain (Bukau and Horwich; 1998; Hartl; 1996). The peptide-binding site of Hsp70 contains a cleft that binds the peptide in an extended conformation (Stuart *et al.*, 1994), making Hsp70 a chaperone corresponding to 'local' enclosure of its given substrates, unlike the 'global' enclosure stratagem observed in the oligomeric ring-like structures of the chaperonins.

Welch and co-workers (Beckmann *et al.*, 1990) have demonstrated that cytosolic forms of Hsp70 bind co-translationally to nascent polypeptide chains and to newly synthesised proteins in the normal (unstressed) cell in an ATP-dependent manner. The association of cytosolic Hsp70 with nascent polypeptide in translating ribosomes has subsequently been confirmed in a number of organisms (Nelson *et al.*, 1992). Hydrolysis of ATP is the rate-limiting step in the ATPase cycle of Hsp70 proteins in isolation (Gao *et al.*, 1993, McCarty *et al.*, 1995, Karzai and McMacken, 1996, Theyssen *et al.*, 1996) and almost certainly results in substantial conformational changes in Hsp70 that convert it to the high affinity, slow exchange state, which sequesters a substrate protein.

In *E.coli*, DnaK, DnaJ and GrpE proteins cooperate synergistically in a variety of biological functions, including interaction with the folding actions of the Hsp70 chaperones. The Hsp40, or DnaJ family, consists of over 100 members, all defined by the presence of a highly conserved 'J' domain of 78 residues (Laufen *et al.*, 1998). J-domain proteins play important regulatory roles as co-chaperones (co-cpns), recruiting Hsp70 partners and accelerating the ATP-hydrolysis step of the chaperone cycle.

In *E.coli*, cycling of the Hsp70 homolog, DnaK, between its different nucleotide-bound states is regulated by two cofactors, DnaJ and GrpE (Bukau and Horwich, 1998., Mayer and Bukau, 1998, McCarty *et al.*, 1995, Gomer *et al.*, 1996). The 41kDa DnaJ protein is, in its own right, a chaperone, which can bind to unfolded polypeptides and prevent their aggregation (Szabo *et al.*, 1996, Langer *et al.*, 1992). DnaJ binds to DnaK and stimulates its ATPase activity, generating the ADP-bound state of DnaK, which can then interact stably with the polypeptide substrate (Suh *et al.*, 1999). The 23kDa GrpE protein acts as a nucleotide exchange factor; it binds to the ATPase domain of DnaK and, by distorting the nucleotide-binding pocket, induces release of bound ADP (Harrison *et al.*, 1997).

The interaction of DnaJ with nascent ribosome-bound polypeptide chains (as short as 55 residues), with the use of firefly luciferase and chloramphenicol acetyltransferase in cross-linking experiments, have been investigated (Hendrick *et al.*, 1993). These investigations showed that both folding and subsequent mitochondrial translocation required DnaK, DnaJ, and GrpE and led to the proposal that DnaJ protects nascent polypeptide chains from aggregation and, in cooperation with Hsp70, controls their productive folding once a complete polypeptide, or a polypeptide domain, has been synthesised.

Elucidation of the DnaK reaction cycle has now provided a paradigm for all Hsp70s. In deed, homologs of bacterial DnaJ, collectively referred to as Hsp40 proteins, have since been identified in all cellular compartments that contain an Hsp70 (Kelley, 1998).

1.10.2 Hsp90

Hsp90 is, generally, a chaperone involved in the maturation of signal-transducing proteins, notably steroid hormone receptors and protein kinases (Pratt and Toft, 1997). Hsp90 is a cytosolic chaperone, highly conserved and essential to all organisms. Prokaryotic homologs have been characterised, and further, it is also known to be abundant in the endoplasmic reticulum of eukaryotes (grp94).

Recently solved crystal structures of the N-terminal domain of the yeast Hsp90 (Prodromou *et al.*, 1997) has shown it to be a dimeric protein (monomers ~80-90kDa) with an elongated structure. The dimerisation domain is at the C-terminal, thought to reflect an end-to-end arrangement, leaving the N-terminal opposed. A similar arrangement has been described for the cytosolic Hsp90 (Maruya *et al.*, 1999).

Hsp90 is frequently found in complexes with other chaperones. *In vitro*, Hsp90 exhibits chaperone activity with diverse proteins, suggesting a general function (Bose *et al.*, 1996, Burston and Clarke, 1995), and emphasised by its ubiquitous presence throughout all organisms observed to date.

Prokaryotic and eukaryotic Hsp90s are now known to protect non-native substrate proteins from inactivation and subsequent aggregation (Jakob *et al.*, 1995, Weich *et al.*, 1992, Freeman and Morimoto, 1996). Stable interaction with folding intermediates might be conferred by the partner proteins with which eukaryotic Hsp90 is known to cooperate. *In vivo*, eukaryotic Hsp90 appears to perform at least part of its function in complex with Hsp70 (Buchner *et al.*, 1998).

The *E.coli* protein that is homologous to eukaryotic Hsp90 was discovered by Bardweel and Craig (1987), also referred to as HtpG. At 71kDa (Mw), the prokaryotic version of the Hsp90 is significantly smaller than its eukaryotic relatives, due to two truncated, charged regions (Bardwell and Craig, 1987).

Little is known about the function of *E.coli* Hsp90 *in vivo* and sequence homology searches have revealed that most of the known partner proteins of Hsp90 in higher eukaryotes do not have homologs in *E.coli* (Jakob, 1996). *In vitro*, *E.coli* Hsp90 is known to function as a molecular chaperone, with an activity similar to Hsp90 from yeast or higher eukaryotes (Jakob *et al.*, 1995).

Loss of Hsp90 function in yeast, using a conditional temperature-sensitive mutant, did not generally affect protein folding or refolding, although modest defects were observed in the folding of some heterologously expressed proteins (Nathan *et al.*, 1997). Similarly, studies performed *in vitro* with the test substrates luciferase and β-galactosidase failed to find an essential role for Hsp90 in the folding reaction, although it was capable of stimulating the folding mechanisms of other molecular chaperones, such as Hsp70 and Hsp40 (DnaJ), (Schumacher *et al.*, 1996, Freeman and Morimoto, 1996).

The results have suggested that Hsp90 is not required for the *de novo* folding of most proteins, but is required for a specific subset of proteins that have greater difficulty reaching their native conformations (Nathan *et al.*, 1997). *In vitro*, in the absence of nucleotide, Hsp90 can maintain non-native substrate in a "folding-competent" state that refolds upon addition of Hsp70, DnaJ homolog, and nucleotide (Freeman and Morimoto, 1996).

A somewhat controversial viewpoint suggests that the Hsp90 heat shock protein goes beyond its accepted passive role in cell stress responses, to become an explicit molecular mechanism assisting in phenotypic evolutionary change (Rutherford and Lindquist, 1998). In this paper they report that the mutant form of the Hsp90 (Hsp83) heat shock protein (*Drosophila* fruit fly), effects the phenotypic variation affecting almost any adult structure, which is produced both in laboratory strains and in wild type populations.

In stating the conclusions drawn from their experimental data, Rutherford and Lindquist are providing first evidence for the existence of a molecular mechanism that assists the process of evolutionary change, in response to the environment. This would suggest that the Hsp90 protein is special, even unique, in its critical integrative position in the genetic architecture of development coupled to its role in stress response within the cell.

1.10.3 The Sec Protein Translocation System

Sec B is a molecular chaperone found in *E.coli* that is dedicated to the facilitation of the export of a number of proteins destined for the periplasmic space or the outer membrane. This role in export is demonstrated *in vivo* by the accumulation of related precursor species in a strain that is devoid of Sec B (Kumamoto and Beckwith, 1985) and *in vitro* by showing that Sec B is required for translocation of precursors into inverted vesicles of cytoplasmic membrane (Weiss *et al.*, 1988).

In order for proteins to pass through lipid bi-layers, the hydrophilic parts of the proteins must be shielded from the hydrophobic lipid tails. The Sec secretion complex performs this function through the formation of a pore that can be visualised using electron microscopy (Hanein *et al.*, 1996). The Sec proteins are ancient homologues in eukaryotes and prokaryotes and the mode of action, with signal sequences directing the proteins through the system, is also conserved. The Sec proteins perform a variety of functions chaperoning the correct folding of membrane proteins ensuring the correct orientation of possible multiple membrane spanning domains.

Interactions with other members of the chaperone super family are also seen, indicating that pathways of chaperones do exist to help fold proteins from their point of synthesis on the ribosome to the site of final folding (Miller,1993).

1.11 The chaperonins

The chaperonins are an example of a recently discovered evolutionary class of large protein complexes, which also includes the proteasome (Groll *et al.*, 1997). The term 'chaperonin' was coined by Hemmingsen and colleagues (1988) to describe one family of highly-sequence related proteins acting as molecular chaperones, found in chloroplasts, mitochondria, and eubacteria such as *E. coli*.

The chaperonins are now recognised to be essential components that provide an element of regulation to the folding pathways of proteins within the cell. This pathway is designed to stabilise labile protein folding intermediates and partition them towards successful isomerisation to the native state. Unlike Hsp70s, chaperonins appear to interact with non-linear hydrophobic polypeptides, exposed in compact intermediates (Frydman *et al.*, 1994; Rommelaere *et al.*, 1999; Hayer-Hartl *et al.*, 1994). They form enclosed, or partially enclosed, compartments that carry out passive reactions on sequestered polypeptide chains, in that they play no role in the final component or function of the protein involved.

The role of the ATPase cycle has recently become well defined in this mechanism. A combination of kinetics, crystallographic studies and cryo-electron microscopic pictures of different states has revealed the sequence of concerted hinge rotations that reorient polypeptide-binding sites and lead to massive conformational changes in the chaperonin complex.

The chaperonin family consists of two distinct subclasses: Group I, the GroE subclass found in chloroplasts and other plastids, mitochondria, and eubacteria; Group II, the chaperonin of the archaebacteria and eukaryotic cytosol subclass, termed TRic or CCT (for TCP-1 ring complex or chaperonin-containing TCP-1, respectively, where TCP-1 is tail-less complex polypeptide-1).

The chaperonins are large cylindrical protein complexes consisting of two stacked rings of seven to nine subunits each (reviewed Bukau and Horwich, 1998, Hartl, 1996). However, although the two subfamilies possess a similar cage-like structure, there are important differences in the regions that bind substrate proteins.

1.11.1 Group I Chaperonins

In the Group I (GroE) subfamily, the chaperonin complex comprises 14 subunits (with two different subunit types, α and β, existing in the chloroplast form) in two opposing seven-membered rings. Irrespective of source, each chaperonin possesses a similar domain organisation; with two large domains, equatorial and apical, linked by a hinging region called the intermediate domain.

GroEL (hsp60), from *E.coli*, is by far the most studied and best understood chaperonin, and as such has become the paradigm on which all Group I chaperonin analysis is based. The hsp60 chaperonins from chloroplasts and mitochondria were among the first complexes implicated in oligomeric assembly (Ellis *et al.*, 1994, Ostermann *et al.*, 1989), and the structures of the GroEL oligomer and its various complexes have now been determined by x-ray crystallography (Braig *et al.*, 1994, Boisvert *et al.*, 1996). Fig 1.11.1 portrays a single subunit of the GroEL chaperonin.

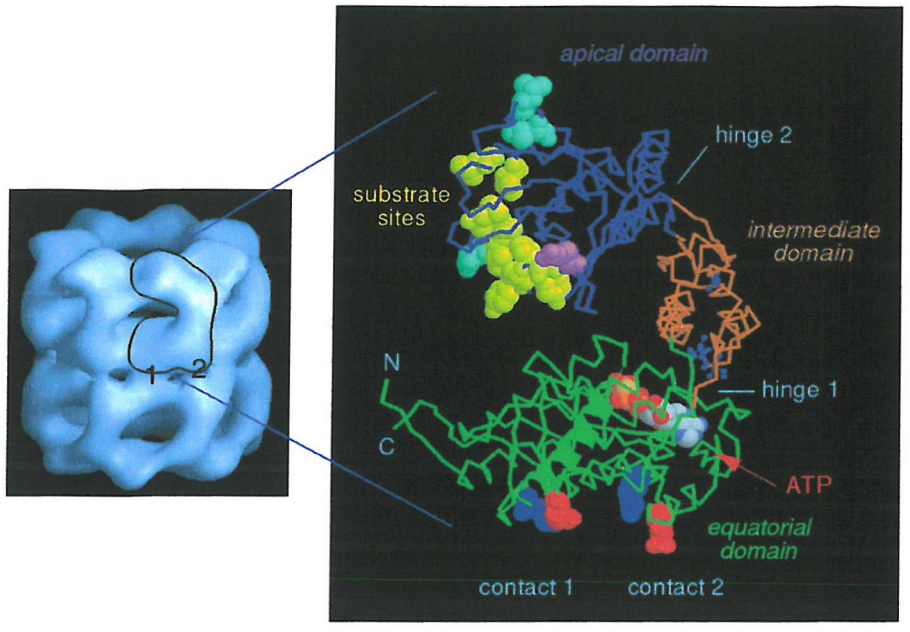


Fig 1.11.1 Structure of the GroEL oligomer and subunit. On the right is the crystal structure of GroEL to 2.5nm resolution, showing the double-ring structure and the location of an individual subunit within the oligomer. On the right is the back bone trace of the GroEL subunit (Roseman *et al.*, 1996).

GroEL works in unison with its co-chaperonin (co-cpn) GroES, which acts as a lid to the GroEL folding box, and uses an ATPase cycle to transiently create an enclosed and enlarged cavity in which protein folding takes place. GroEL (hsp60 or cpn60) is a tetradecamer of 58kDa subunits, and its co-chaperonin GroES (hsp10 or cpn10) is a heptameric ring of 10kDa subunits.

The ring-shaped structure of GroEL is essential for its folding activity, which allows it to promote the folding of substrates that the Hsp70 system is unable to fold. It has been demonstrated that a wide range of purified proteins in their non-native states can interact with GroEL; examples include rubisco (Goloubinoff *et al.*, 1989, Baneyx and Gatenby, 1992, van der Vies *et al.*, 1992), and pre-β-lactamase (Laminet *et al.*, 1990). From these, as well as other studies (Höll-Neugebauer *et al.*, 1991, Martin *et al.*, 1993), it has become apparent that chaperonins regulate protein folding by stabilising folding intermediates, thereby influencing the kinetic partitioning between aggregated (misfolded) and correctly folded proteins.

For GroEL, the folding reaction is driven by cycles of binding and release of the co-cpn GroES, which alternates with the binding and release of the nonnative protein substrate (Fisher and Yuan, 1994, Todd *et al.*, 1994). GroEL is a weak ATPase and when diluted into a requisite folding buffer, most denatured proteins form a stable binary complex with the chaperonin. Compelling evidence indicates that partially folded protein substrates bind within the central cavity of one ring of the GroEL toroid, usually at a stoichiometry of one molecule per chaperonin 14-mer (Hartl and Martin, 1995, Hartl, 1996). This action may arrest any legitimate refolding that would occur in the absence of GroEL, but also efficiently suppresses unproductive aggregated side reactions. Binary complexes between GroEL and partially folded proteins may remain stable for an inordinate time, but are easily disrupted, and release of substrate affected, with the addition of adenine nucleotides. That is to say, most substrates binding with high affinity to GroEL require hydrolysable ATP to affect release from the chaperonin complex.

However, other polypeptides can be discharged with lower efficiency by non-hydrolysable ATP analogs, ADP, or even shifts to low temperatures. (Brunschier *et al.*, 1993, Hansen and Gafni, 1993). Once released from the chaperonin, the protein may reach a native conformation, become trapped in an inactive or assembly-incompetent state, or become aggregated. However, it has become evident that, if necessary, the ejected, partially folded protein can rebind to the chaperonin and continue the cycle until folding is complete (Martin *et al.*, 1993).

The co-cpn GroES is known to bind one end of GroEL (Fig 1.11.2), in the presence of adenine nucleotides, in its functional form and has been proven to cap the polypeptide-bound (*cis*) end of the GroEL toroid, if the substrate size is smaller than 50-60 kDa (Weissman *et al.*, 1995, Mayhew *et al.*, 1996).

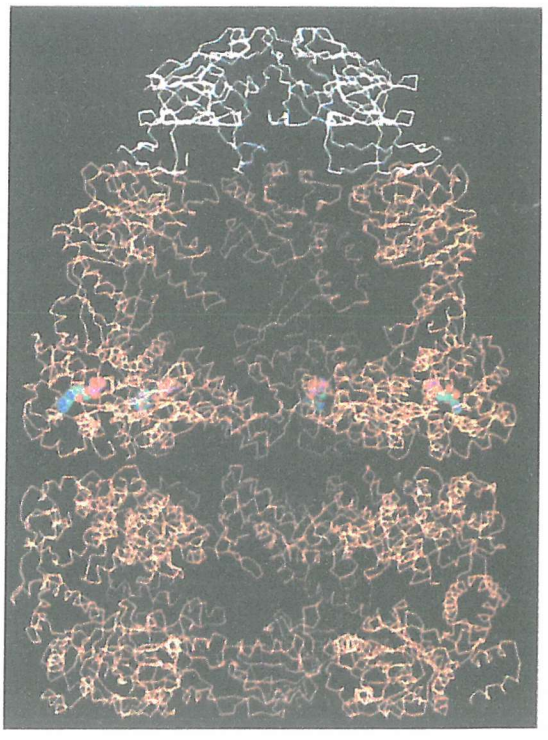


Fig 1.11.2 Cα skeleton drawing of GroEL-GroES-$(ADP)_7$ complex. This drawing shows the complex to be sliced vertically along the central axis. The interaction of GroES (white) with GroEL (orange) forms a continuous dome-shaped cavity of 2-fold increased volume relative to an unliganded structure. The nucleotide (coloured balls) is shown in its binding sites (Bukau and Horwich, 1998).

Although both symmetric (football) and asymmetric (bullet) complexes of GroEL, bound by GroES, have been observed (Todd *et al.*, 1994, Török *et al.*, 1996), only the latter are believed to be physiologically functional (Todd *et al.*, 1993).

In the asymmetric complexes the GroEL ring, with GroES attached, is referred to as the *cis*-ring (proximal), and the opposing (distal) ring as the *trans*-ring. The recently determined structure of the GroEL-GroES-$(ADP)_7$ complex revealed that the large rigid-block movements of the intermediate and apical domains in the *cis*-ring allowed bound GroES to stabilise a folding chamber with ADP confined to the *cis*-ring (Chen *et al.*, 1994, Roseman *et al.*, 1996, Xu *et al.*, 1997) (Fig 1.11.3). The conformational transition of a GroEL subunit can be divided into two stages; the first one is associated with the binding of nucleotide (t to r', t=tense r'=relaxed) and the second one involves the binding of the co-chaperonin GroES (r' to r'') (Boisvert *et al.*, 1996, Braig *et.al.*, 1994, Chen *et al.*, 1994, Roseman *et al.*, 1996, Xu *et al.*, 1997). The conformational changes in the apical domains doubled the volume of the central cavity and resulted in burial of the hydrophobic peptide-binding residues at the interface with GroES and between the GroEL subunits. These structural changes result in the enlarged central cavity having a polar surface that favours protein folding (Chen *et al.*, 1994, Xu *et al.*, 1997). The conformational changes induced in GroEL upon binding of ATP have been observed by the use of several techniques including Cryo-electron microscopy, x-ray crystallography and fluorescence labelling (Chen *et al.*, 1994, Llorca *et al.*, 1997, Torok *et al.*, 1996, Xu *et al.*, 1997). In addition, it has been shown that binding of nucleotide to one GroEL ring is strongly favoured by GroES binding to the other ring (Sparrer and Buchner, 1997), with the nucleotides found to bind to a site near to the top of the equatorial domain facing the cavity (Boisvert *et al.*, 1996).

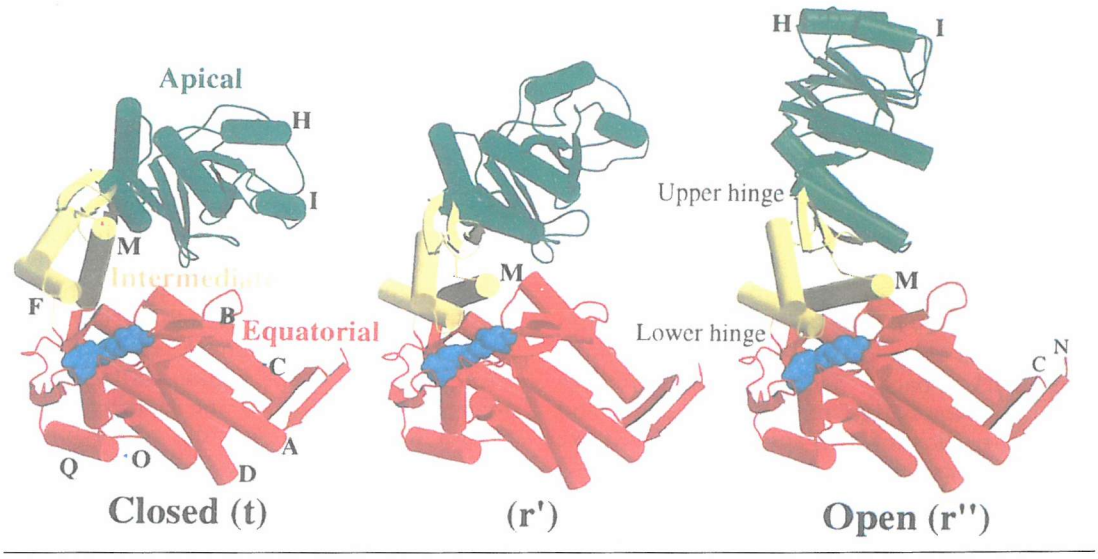


Fig 1.11.3 The apparent overall architecture and conformational change of GroEL.
The folding chamber ("Anfinsen cage") is created from the unliganded ("closed") structure by large, nearly rigid body movement of the intermediate and apical domains of the *cis*-ring subunits (Boisvert *et al.*, 1996). These movements are couple to the binding of ATP and GroES in a process with high positive cooperativity. Colour coding is: apical (green), intermediate (yellow), equatorial (red). The ATP nucleotide is shown as a blue space-filling model (Ma *et al.*,2000).

1.11.2 The Significance of Nested Cooperativity

In the case of GroEL, homotropic regulation by ATP is exhibited by nested cooperativity. This phenomenon manifests as 'positive' intra-ring cooperativity and 'negative' inter-ring cooperativity of ATP-binding, modulated by various hetero-tropic allosteric effectors. Cooperativity in ATP binding and hydrolysis by chaperonins reflects the switching of rings between protein binding and releasing states (Horovitz *et al.*, 2001). This form of ATP-driven binding and release is important for regulation of the reaction cycle, and most likely increases the mechanical force applied to bound proteins (Horovitz *et al.*, 2001).

A model, first put forward by Wyman in 1967, to describe certain allosteric phenomena exhibited by haemoglobin, proposed that nested cooperativity arises from interactions between allosteric units within the same molecule. Studies on the initial rates of ATP hydrolysis by GroEL, at differing ATP concentrations demonstrated that the molecule underwent two ATP-induced allosteric transitions (Yifrach and Horovitz, 1995). The data led the investigators to two levels of allostery; one occurring within each ring (intra-), and a second between the two heptameric rings (inter-). Fig 1.11.4 depicts the nested cooperativity pathway, in accordance with the standard MWC representation (Monod *et al.*, 1965) and the KNF-type allosteric transitions (Koshland *et al.*, 1966).

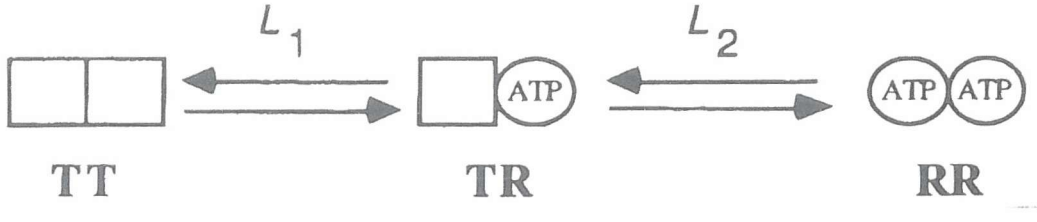


Fig 1.11.4 A Scheme for nested cooperativity in GroEL with respect to ATP.
In the absence of ligands, GroEL is predominantly in the **TT** state. Upon ATP binding the equilibrium is shifted toward the **TR** state. A further shift in the equlibrium toward the symmetrical **RR** state occurs at higher ATP concentrations. Models of cooperativity based on the MWC and KNF principles. (Horovitz *et al.*, 2001).

Insight into the structural basis and dynamics of the allosteric transitions of GroEL has been gained from the crystal structure of the GroEL-GroES-$(ADP)_7$ asymmetric complex (Xu *et al.*, 1997) and from comparisons with the structures of other stable states (Braig *et al.*, 1994, Boisvert *et al.*, 1996, Roseman *et al.*, 1996).

The trajectory of the conformational change of a GroEL subunit, upon ATP binding, was simulated using targeted molecular dynamics (TMD) (Ma *et al.*, 2000), (Fig 1.11.5). These simulations indicate that intra-ring positive cooperativity in GroEL is best described as a two-state process. Firstly, steric clashes that arise if one subunit changes conformation from the tense 'closed' state 't' to the relaxed 'open' state 'r', whilst its subunit neighbour to the right remains positionally unchanged. Further, each heptameric ring is recognised to be at equilibrium between these two states, with the ability to interconvert between the two states in a concerted manner. A further indication of positive cooperativity in GroEL was alluded to by Ma's simulated breaking of the Arg197-Glu386 interaction with both the left and the right neighbouring subunits (Fig 1.11.6).

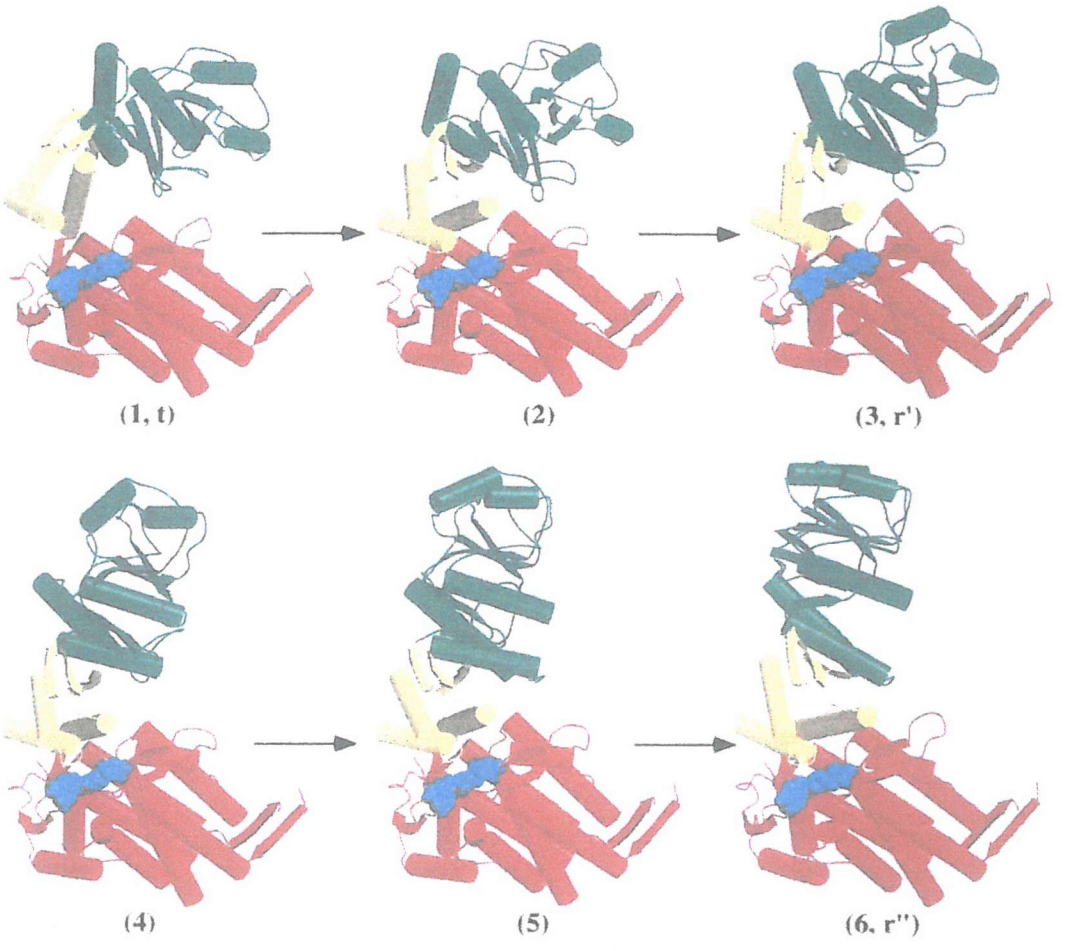


Fig 1.11.5 The conformational transition of a GroEL subunit.

The conformational transition is best modelled in two distinct stages; the first stage involves the binding of a nucleotide (t to r'), structures 1-3. Structures 4-6 correspond to the second stage involving GroES binding, ie the r' to r'' transition. Structures 1 (Boisvert *et al.*, 1996) and 6 (Xu *et al.*, 1997), labelled t and r'', respectively, are the x-ray structures. The other, intermediate, structures were obtained from the TMD simulation (Ma *et al.*, 2000).

The second level of allostery is between the two rings of GroEL (inter-ring negative cooperativity). The negative allostery is described by sequential transitions from the 't' state via the asymmetric 'tr' state to the (r'r'') symmetric state and arises from (i) the need to preserve the inter-ring interface, whereby a subunit in the trans-ring is forced to move away from the central axis by about 2^{o} (Xu *et al.*, 1997) and (ii) the need to avoid steric clashes that arise between equatorial domains in the two rings if they both undergo a conformational change (Ma *et al.*, 2000).

Additional support for the nested cooperativity model comes from the transient kinetic analysis of the ATP-induced allosteric transitions of GroEL (Yifrach and Horovitz, 1998, Cliff *et al.*, 1999). Cooperativity in ATP hydrolysis by GroEL, with respect to ATP, has also been measured in the absence and presence of different concentrations of non-folded α-lactalbumin (Yifrach and Horovitz, 1996), and is also in agreement with the nested cooperativity model.

Very little is known about the mechanism of inter-ring allostery for the type II chaperonins. The heterogeneity in their subunit composition and the fact that each subunit in one ring is in contact with only one subunit in the other ring, not two subunits as is the case with GroEL, suggests a mechanism of inter-ring (negative) allostery quite different from that of type I chaperonins (Llorca *et al.*, 1999).

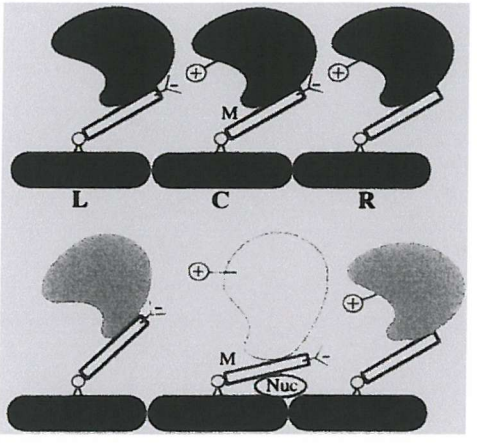


Fig 1.11.6 Schematic illustration of positive cooperativity in GroEL.

Nucleotide binding to subunit (C) causes it to undergo a conformational change from state 't' to state 'r', and both the left-hand (L) and right-hand (R) protomers become affected. The Arg197, Glu386 salt-bridge and the effect of the steric interactions are indicated. Helix M represents the intermediate domain. The lightened shading of the apical domains denotes their altered nucleotide binding affinity.

1.11.3 The Reaction Cycle of GroEL

Non-native substrate protein binds to the *trans*-ring of GroEL, at which time ADP and GroES are bound in the 'opposite' GroEL ring. Binding is facilitated by the hydrophobic surfaces of the apical domain lining the cavity in the GroEL ring. Subsequent ATP binding to the *cis*-ring leads to release of the ADP and GroES co-cpn, followed by binding of ATP and GroES to the *cis*-ring which results in the extensive conformational changes that ultimately enlarge the central cavity. This conformational change triggers the release of the substrate protein from the surface of the apical domain, which eventually effects substrate release. Hydrolysis of the ATP in the *cis*-ring weakens the interaction between GroES and the *cis*-ring, and binding of ATP to the *trans*-ring leads to the complete release of the GroES co-cpn and substrate protein (Rye *et al.*, 1999). One of the many recently offered models for the GroEL reaction cycle is depicted in Fig 1.11.7.

It has been proposed that the enclosed cavity functions as an 'Anfinsen cage', i.e. a protected chamber that isolates the polypeptide under conditions of infinite dilution and allows it to fold according to its thermodynamic potential (Hendershot *et al.*, 1996).

Horwich and colleagues (1996) have shown that under normal conditions the rate of hydrolysis of ATP, in the ring *trans* to the bound GroES, determines the rate of release of the GroES and subsequently the rate of dissociation of the substrate protein, with a half-life of ~15sec (Burston *et al.*, 1996). Confirmation of the existence of this temporal state comes from observations on the GroEL assisted folding of mitochondrial malate dehydrogenase (MDH). Trapping experiments demonstrated that the substrate retention time on the complex is only of the order of 20sec (Ranson *et al.*, 1997). This correlates well with both the rate of ATP turnover and the retention time of GroES in the complex, but is much shorter than the time required for the substrate to commit to the folded state.

There is also evidence to indicate that GroEL is able to unfold and refold misfolded proteins (Acton *et al.*, 1993, Nieba-Axmann *et al.*, 1997, Sparrer *et al.*, 1997, Todd *et al.*, 1996, Todd *et al.*, 1994). Clarke and co-workers (1995) studied the GroEL-assisted folding of mitochondrial MDH and showed that the chaperonin accelerated the dissociation of misfolded intermediates formed by reversible aggregation of an early, partially folded intermediate, through a cycle of binding and release, coupled to ATP hydrolysis. It has been proposed that the "unfoldase" activity of GroEL actually arises from its preferential affinity for the unfolded conformation (Weissman *et al.*, 1994). This would indicate that, according to the law of mass action, misfolded intermediates can be unfolded and, presented with a conducive opportunity, fold productively in the GroEL folding cavity.

Finally, experimental data show that GroEL prevents aggregation as well as providing a sequestered folding cage for the folding of variable-state substrates, and can also be said to be directly involved in the prevention of aggregation and misfolding of substrate polypeptides.

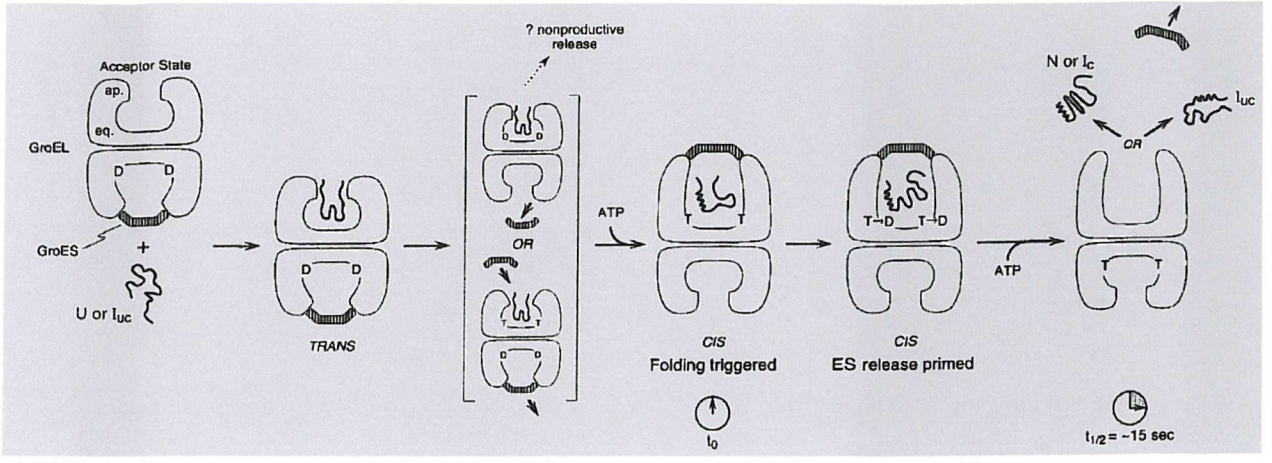


Fig 1.11.7 A Model for the GroEL-GroES-Mediated Folding reaction.

The asymmetric GroEL-GroES complex comprises; the polypeptide acceptor state *in vivo*, binds unfolded polypeptides (U) or kinetically trapped intermediates (I_{uc}), forming a *trans* ternary complex. GroES then binds in the presence of ATP; two possible pathways of GroES rearrangement, leading to the *cis* complex are shown. When GroES binds to the ring containing polypeptide in the presence of ATP it forms the folding-active *cis* intermediate and major conformational changes occur in the *cis* GroEL ring. Polypeptide release from the apical binding sites is triggered and folding commences (t_0). Hydrolysis of ATP in the *cis* ring weakens the GroEL-GroES interaction, priming GroES release while polypeptide folding continues. Binding of ATP in the *trans* ring ejects GroES from the *cis* ring, giving the polypeptide the opportunity to leave ($t_{1/2}$ = ~15s). The released polypeptide is either native (N) or committed to fold (I_c), or is in an uncommitted or kinetically trapped state (I_{uc}), which can rebind to the same or a different GroEL complex and undergo another cycle of folding. In the complexes, D designates ADP; T, ATP; and T→ D, denotes ATP hydrolysis.
(Bukau and Horwich,1998)

1.12 Group II Chaperonins

Group II chaperonins, such as chaperonin containing TCP-1 (CCT), or TriC, are found in archaea and in the eukaryotic cytosol (Kubota *et al.*, 1995). Some archaeal chaperonins, such as that of *Methanococcus jannaschii* consist of a single subunit (Bult *et al.*, 1996). However, the majority of archaeal chaperonins, such as the thermosomes from *Thermoplasma acidophilum*, *Pyrodictium brockii*, *Archaeoglobus fulgidus* (Phipps *et al.*, 1991) and *Thermococcus sp.* (Osipiuk and Joachimiak, 1997) consist of two different but related subunits, termed α and β, each possessing an eight-membered ring. Eukaryotic CCT (Kubota *et al.*, 1995) is also comprised of two eight-membered rings, each composed of eight different, but sequence-related subunits to form the sixteen-subunit complex.

Whereas group I chaperonins like GroEL assist in the folding of many other proteins, only a small number of proteins, mainly actin and tubulin, have been described as natural substrates of group II proteins, such as CCT. This specificity may well be related to the structural divergence of the eight CCT subunits (Llorca *et al.*, 1999).

Similarities in the heat-shock induction and the ATPase activation of lysed *Pyrodictium* cells were immediately ascribed to the similar chaperonin actions of the GroEL (hsp60/cpn60) family (Phipps *et al.*, 1991), leading to the inference that this protein is an archaebacterial chaperonin.

Subsequently, it was named a thermosome, emphasising its heat induction properties and its extreme thermostability. A subsequent report, describing a similar nine-fold symmetrical complex from *Sulfolobus shibatae* (TF55; for the thermophilic factor, ~55kDa subunits), which was able to bind unfolded proteins *in vitro*, corroborates the original hypothesis (Trent *et al.*, 1990).

E.coli GroEL (group I) (Braig *et al.*, 1994), and the thermosome, from *Thermoplasma acidophilum* (group II) (Ditzel *et al* ., 1998), both possess a similar structural organisation in their three domains (apical, intermediate and equatorial) within a given subunit (Waldmann *et al.*, 1995). Indeed, the equatorial domain is relatively conserved among all chaperonins. The greatest sequence divergence between the two subunits is found in the apical domains, which probably contain the substrate binding site (Kim *et al.*, 1994).

Whereas group I chaperonins, such as GroEL, act in unison with the co-chaperonin (co-cpn) GroES, the small 10kDa heptamer functioning as a lid to seal the chaperonin folding cavity, the group II chaperonins appear to possess a 'built-in' co-cpn in the form of a helical protrusion (Ditzel *et al.*, 1998, Klumpp *et al.*, 1997). Comparison of the thermosome amino acid sequence with those of the eukaryotic CCT subunit sequences demonstrates the presence of a long sequence insertion, indicative of a helical protrusion (Ditzel *et al.*, 1998), indicating that the structure of the CCT subunits is similar to that of the thermosome. Although the globular cores of the apical domains in both group I and group II chaperonins appear to be homologous, they nevertheless appear to function differently. In GroEL the hydrophobic apical domain, known to be the polypeptide-binding surface, is located between helices H8 and H9 (Fig 1.12.1),(Buckle *et al.*, 1997, Fenton and Horwich, 1994), whereas the corresponding region in the group II chaperonins is mostly hydrophilic.

Fig 1.12.1 Three-dimensional structure of mini-chaperone GroEL (Apical Domain).

The crystal structure of GroEL (191-376) [at 1.7 Å resolution]. Depicted is a 17 residue N-terminal tag, which interacts with the active site of a neighbouring molecule, is also depicted suggesting a model for the binding of long peptides around the rim of seven subunits in the cavity of the tetradecamer (Buckle *et* al., 1997). Helices H8 and H9 are coloured cyan. (adapted from: Chatellier *et al.*, 1999).

However, in group II chaperonins a hydrophobic surface area, with a degree of conservation that is consistent with a role in substrate binding, is located on the helical protrusions (Klumpp *et al.*, 1997, Nitsch *et al.*, 1998). These protrusions offer group II chaperonins the means to control access to the folding compartment without having recourse to a detachable GroES-like lid. The crystal structure of the complete thermosome from *Thermophilus acidophilum* (Ditzel *et al.*, 1998) confirmed this proposal, by showing that the protrusions project towards the central axis of the thermosome and collectively form a lid domain that occludes the central cavity by forming a β-barrel plug, see Fig 1.12.2.

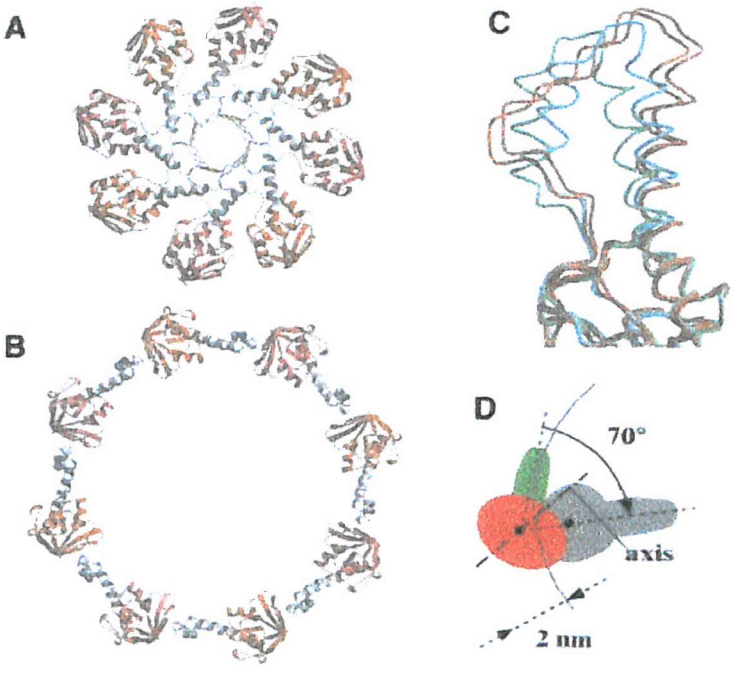


Fig 1.12.2 Structural Plasticity of the Group II Chaperonin *T.acidophilum*

(A) An ariel view of the closed thermosome demonstrating the iris-like arrangement of the apical domain protrusions in the crystal structure. (B) A model of the apical domain arrangement in the open state of the thermosome. (C) The structural variability of the helical protrusion in two different crystal forms of the isolated α-apical domain (orange and red) and in the α (green) and β (cyan) apical domains in the closed thermosome particle. (D) Mechanics of the open-to-closed transition of the apical domains. The radius of the apical zone shrinks by 2nm upon closure of the thermosome particle. (Gutsche *et al*; 1999).

Despite its similarity to bacterial chaperonins, the range of substrates of the eukaryotic cytosolic chaperonin TriC/CCT has been a matter of some debate. While some confirmed substrates, such as actin and tubulin, have been identified for TriC, little else is known about the natural substrates for this thermosome. It has been suggested that TriC is a specialised chaperonin that folds only a few cytoskeletal proteins (Hynes *et al.*, 1996). Given that most of the heterogeneity among TriC subunits resides in the putative substrate-binding domain (Kubota *et al.*, 1995, Gutsche *et al.*, 1999), it is possible that different subunits in the complex have evolved to recognise differing motifs in their respective substrate proteins. In addition, a structural analysis of the chaperonin-action complex, using immunoelectron microscopy supports the idea that the polypeptide interacts with specific subunits in the TriC chaperonin (Llorca *et al.*, 1999).

Comparative studies performed with TriC and GroEL demonstrated that both chaperonin subfamily types bind unfolded actin and tubulin efficiently, although it had been shown previously that both chaperonin types recognised different substrate binding regions (Melki and Cowan, 1994, Rommelaere *et al.*, 1999). These studies also demonstrated that only TriC was able to channel the folding of both substrates to their native states. A further difference in substrate recognition between GroEL and TriC was highlighted when malate dehydrogenase (a model substrate of GroEL) and rhodanese were not recognised at all by TriC. This has led to a proposal that GroEL only recognises the overall hydrophobicity of its targets, whereas precise recognition of the hydrophobic patterns was the *de rigour* for the TriC chaperonin (Rommelaere *et al.*, 1999).

There is now wide acceptance that TriC binds its substrates, both post-translationally (Eggers *et al.*, 1997) and co-translationally (Frydman and Hartl, 1996, Frydman *et al.*, 1992, Frydman *et al.*, 1994, Hartl, 1996). Currently, there are two models offered for the cellular mode of action for TriC, the "pathway model" and the "network model". The pathway model advocates that *de novo* protein folding takes place in a protected environment, created by the concommitent interaction between chaperonins and the translation machinery (Siegers *et al.*, 1999, Thulasiraman *et al.*, 1999).

The network model proposes that non-native proteins freely partition within the cytosol between the folding and degradation systems (Buchberger *et al.*, 1996, Farr *et al.*, 1997, Fenton and

Horwich, 1997). The pathway model implies that substrates released from the chaperonin *in vivo* are in a conformation committed to reach the native state. The network model implies that nonnative substrates may bind to the chaperonin repeatedly until they eventually reach the native state.

1.13 Chaperonin GroES Type Homologues

The absence of GroES homologues in both archaeal and eukaryotic chaperonins is intriguing. Regardless of the apical protrusions now known to exist, if no co-factor (co-cpn) exists, how can both the TriC and thermosome chaperonins prevent their respective substrates from leaving the central folding cavity prematurely? Given that the putative co-factors of β-tubulin, termed A and B (Mw 28 & 100kDa, respectively) described in earlier studies (Gao *et al.*, 1993) were shown not to interact directly with TriC (Melki *et al.*, 1996).

More recently new co-cpns, prefoldin (Vainberg *et al.*, 1998) also called GimC (Geissler *et al.*, 1998), implicated in the biogenesis of actin and tubulin in yeast, has been shown to bind unfolded actin and transfer it to TriC in a nucleotide-independent manner (Vainberg *et al.*, 1998), although a physical interaction between TriC and GimC/prefoldin has yet to be demonstrated. Disagreement exists as to the role of the GimC/prefoldin complex in protein substrate folding. One view point, held by Siegers and colleagues (1999) is that GimC/prefoldin acts in complex with TriC and eventually accompanies the near-native actin after its release from the chaperonin. In contrast, Hansen and his colleagues (1999) suggest that GimC/prefoldin binds nascent actin chains co-translationally and assists in their delivery to the TriC particle after the end of elongation.

An important observation supporting the role of GimC/prefoldin as a general co-cpn of group II chaperonins, is their co-related occurrence, i.e. they are absent from eubacteria but are conserved in the eukaryotes and archaea. A GimC/prefoldin-like hexameric complex from *Methanobacterium thermoautotrophicum* composed of two subunits in a 2:1 stoichiometry has been shown to possess chaperonin activity in that it prevents the aggregation of a variety of nonnative substrates (Leroux *et al.*, 1999).

It is evident that further work will produce additional caveats to the data already gained, and define more clearly the role of protein co-factors in group II chaperonin-substrate assembly pathways.

1.14 The Plastid Chaperonin

Although, there exist considerable conservation in the overall organisation of the chaperonin subunits, differences in structure have been identified amongst chaperonins from certain species. One such difference is seen in the plant chloroplasts, which possess the dual distinction of containing atypical chaperonin subunits and co-chaperonins. The chloroplast belongs to a family of developmentally inter-related organelles, collectively called plastids. In plants, chloroplasts are formed by the differentiation of a progenitor plastid, the proplastid. Proteins destined for the interior of the chloroplast are made as pre-cursors and the N-terminal 'transit peptides' that direct importation are removed once the protein is inside. This process requires, as a final event, the folding and assembly of these proteins into biologically active units. In all plastids examined to date (including chloroplasts and leucoplasts), chaperonins are observed to be abundant, constitutive proteins that proliferate in varying degrees, due to cell related stresses such as heat shock, or an increase in the cellular content of unfolded proteins (Georgopoulos and Ang, 1990, Reading, 1989). Although the majority of chloroplast proteins are synthesised in the cytosol and translocated into the plastids, a significant number, most notably the large subunit (L) of rubisco (ribulose 1,5-bisphosphate carboxylase oxygenase), are produced in the stroma.

Rubisco, directly or indirectly, plays a vital role in the metabolism of all cells, and consequently is one of the most studied plant enzymes. From studies on the biogenesis of rubisco in isolated chloroplasts (Barraclough and Ellis, 1980, Roy *et al.*, 1982), it is apparent that assembly of nascent 'L' subunit chains into the active rubisco holoenzyme occurs rapidly and spontaneously, but also involves an "intermediary" protein.

The most abundant form of rubisco is an oligomeric complex consisting of eight 'large' subunits (Mw ~50-52kDa) and eight 'small' subunits (Mr ~12-18kDa). This complex found in plant chloroplasts and most prokaryotic organisms is an example of a 'calvin cycle'enzyme (Hopkins, 1995). Each 'L' subunit (carrying an active site and residues from adjacent subunits necessary to gain activity) is encoded in the plastid genome and synthesised by plastid ribosomes. The 'S' subunits, required to confer full activity when in conjunction with the eight 'L' subunits, is encoded in the nuclear genome and imported into the plastid after synthesis in its precursor form via free-cytosolic ribosomes (Lennox and Ellis 1986, Branden *et al.*, 1987, Gatenby *et al.*, 1987). The use of isolated chloroplasts, and the subsequent 'native' PAGE analysis, in assembly studies of newly synthesised 'L' subunits into the oligomeric form of rubisco, highlighted the migration of another non-covalently bound chloroplast protein of lower mobility, of about 60kDa subunit mass (Barraclough and Ellis, 1980).

It has been shown that the majority of newly synthesised 'L' subunits cannot assemble into the holoenzyme without the involvement of the associated, non-covalently bound, protein (Ellis *et al.*, 1980, Barraclough and Ellis, 1980); these co-migratory proteins are not synthesised by the isolated chloroplasts, but encoded for by nuclear genes and synthesised by the cytosolic ribosome. These observations led Barraclough and Ellis to suggest that the binding of rubisco 'L' subunits to the co-migratory protein may well be an obligatory step in the rubisco folding process.

The co-migratory protein was originally named the LSU-BP (Large Subunit-Binding Protein) until it was observed that imported rubisco 'S' subunits also bound to this 'binding protein' (Ellis and van der Vies, 1988, Gatenby *et al.*, 1987). LSU-BP is now accepted to be a chaperone, and as a plastid chaperonin form is now commonly referred to as the ch-cpn60 (chloroplast-chaperonin 60kDa) (Hemmingsen *et al.*, 1988), with a stoichiometry of binding of one 'L' subunit to one ch-cpn60 oligomer having been demonstrated.

Further support for the involvement of the plastid chaperonins assembly process came from the demonstration of the inhibitory affects of their *in vitro*-synthesised 'L' subunits in oligomeric form, their ability to co-migrate with the pre-existing rubisco, and their induction upon addition

of antibodies raised against the plastid chaperonin in chloroplast extracts (Cannon *et al.*, 1986, Roy *et al.*, 1982).

Unlike mitochondrial and eubacterial chaperonins, which contain only a single homologous cpn60 subunit (Hemmingsen *et al.*, 1988, Reading *et al.*, 1989), purified preparations of the plastid ch-cpn60 (subsequently determined a Group I chaperonin) consist of approximately equal amounts of two divergent, 60kDa subunits. These subunits are designated the α and β subunits (Mw ~61 and 60kDa, respectively), which are no more similar to each other (~50% sequence homology) than they are to GroEL, each being about 50% homologous to GroEL (Hemmingsen and Ellis, 1986, Musgrove *et al.*, 1987, Martel *et al.*, 1990). On the basis of this sequence similarity it is apparent that the ch-cpn60 and the GroEL chaperonin (cpn60s) are homologous (Musgrove *et al.*, 1987).

Although ch-cpn60 has been observed to occur most likely as a tetradecamer (Mr ~800kDa) (Bonk *et al.*, 1996, Viitanen *et al.*, 1995), the polypeptide composition and subunit dispersal of the α and β subunits have yet to be determined. The observation that the α subunits of the *Brassica napus* ch-cpn60 fail to assemble into 14-mers in *E.coli*, unless co-expressed with the β subunits (Cloney *et al.*, 1992a, Cloney *et al.*, 1992b), suggests that both isoforms reside in the same molecule. However, in the presence of Mg-ATP, the β subunits form 14-mers in a highly cooperative reaction.

The quaternary structure of ch-cpn60, 'observed' by cryo-electron microscopy superficially resembles that of its GroEL homologue (Chen *et al.*, 1993, Pushkin *et al.*, 1982). Both molecules consist of two stacked rings of seven subunits each. Indeed, the chloroplast chaperonin particles resolved by this technique bear a striking resemblance to those observed for its GroEL homologue (Chen and Pushkin, respectively).

1.15 A Chloroplast Co-Chaperonin Similar to the Bacterial Cpn10 (GroES)

The release of target proteins bound to GroEL and their subsequent journey toward the native state, is heavily dependant upon the interactions with its co-chaperonin GroES, and MgATP. The posited role for GroES in the effective dissociation of the GroEL-target protein suggests the possibility that the ch-cpn60 may also require the co-operative assistance of a GroES co-factor equivalent to facilitate the protein folding and release processes. Indeed, such related GroES homologues have been isolated from pea and spinach chloroplasts (Bertsch *et al.*, 1992). Interestingly, this chloroplast co-cpn is a polypeptide of ~21kDa (designated ch-cpn21), which is twice the size of its homologous bacterial and mitochondrial cpn10 equivalents. Primary amino acid sequence analysis reveals that ch-cpn21 is a "double" cpn10 molecule that has arisen through either gene duplication or the fusion of two different ancestral cpn10 genes. Both halves of the mature spinach protein exhibit remarkable homology to the bacterial GroES co-cpn10 (Chandrasekhar *et al.*, 1986), and to bovine mt-cpn10 (mitochondrial) (Lubben *et al.*, 1990).

The higher plant ch-cpn10 (Bertsch *et al.*, 1992), now designated ch-cpn21, was initially identified through its ability to form stable, isolatable complexes with the bacterial GroEL chaperonin. The GroEL protein has been used to isolate unknown GroES homologues residing in crude chloroplast extracts (Chandrasekhar *et al.*, 1986; Viitanen *et al.*, 1990) and the protein identified was observed to be interchangeable with the GroES co-cpn, via various functional criteria (Lubben *et al.*, 1990, Bertsch *et al.*, 1992).

Despite the differences in their molecular masses it has recently become clear that certain structural and functional properties of these GroES homologues do not significantly differ from their bacterial counterparts. The ch-cpn21 consists of two GroES-like sequences that are fused "head-to-tail", via a peptide linker region (Bertsch *et al.*, 1992, Argos, 1990), to form a single protein.

Surprisingly, both "halves" of this co-chaperonin have been demonstrated to function autonomously in certain GroES-deficient mutants (Baneyx *et al.*, 1995), and are highly conserved in regions that are believed to be important for bacterial cpn10 (GroES) function. In addition, the two tandemly linked domains each contain a polypeptide segment analogous to the "mobile loop" region of GroES (Koonin and van der Vies, 1995), which is known to be critical for the connective interaction between GroES and GroEL. This binary organisation of the ch-cpn60 co-chaperonin may well reflect a necessary adaptation of plants, which occurred in response to their evolutionary acquisition of two divergent ch-cpn60 isoforms (Martel *et al.*, 1990).

As with other GroES homologues, ch-cpn21 is expected to be a heptameric toroid of ~Mw 150kDa mass. However, upon gel filtration the retention time of recombinant spinach ch-cpn21 protein is strictly concentration dependent. In dilute solutions ch-cpn21 dissociates into low molecular weight species that range in size from dimers to tetramers (Ryan *et al.*, 1995, Bertsch *et al.*, 1992, Koumoto *et al.*, 1999), whereas at higher protein concentrations its apparent molecular weight greatly exceeds that of the expected heptameric form (Baneyx *et al.*, 1995).

In vitro GroEL protein folding assays demonstrate that a five-fold molar excess of the purified spinach co-cpn21 homolog is required to achieve saturation with GroEL. Although this may be due to the differing nature of the chloroplast and GroEL systems involved, it could be argued that it reflects the tendency of ch-cpn21 to interact with itself *in vitro*. Regardless, one study has since shown that recombinant spinach ch-cpn21 is functionally active with its cognate partner (Viitanen *et al.*, 1995). In these experiments the two homologous ch-cpn's were able to assist in the *in vitro* refolding of prokaryotic rubisco, under non-permissive folding conditions. Although ATP hydrolysis was required for both of these reactions, the unique binary ch-cpn21 was not shown to be an obligate co-cpn for the ch-cpn60, apparently the plastid ch-cpn21 co-chaperonin inhibits the "uncoupled" ATPase activity of GroEL and can effectively substitute for bacterial GroES in the chaperonin-facilitated refolding of denatured rubisco (Baneyx *et al.*, 1995).
Electron micrographs of ch-cpn21 reveal a ring-like organisation that is similar to that of the *E.coli* co-cpn GroES (Chandrasekhar *et al.*, 1986), even though the two proteins have differences in subunit sizes of 21kDa and 10kDa, respectively.

In summary, domain fusion to form ch-cpn21 occurred early in (plant) evolution, as evidenced by its presence in all photosynthetic eukaryotes studied to date. This fusion has since proved to be highly conserved, which further suggests that the retention of the two-domain co-chaperonin structure may be advantageous in the chloroplast chaperonin system (McLachlan, 1987).

However, other advantages conferred by domain fusion could be of greater significance. For example, domains are autonomous cooperative folding units and it has been found that domain fusion enhances the rate of folding and can improve stability by reducing the entropy of the unfolded state. This occurs because the two chains are no longer independent of one another, thus reducing the translational and rotational entropy of the unfolded polypeptide (Liang *et al.*, 1993). Certainly, the ch-cpn21 co-chaperonin has proven to be more stable than its separated C-terminal domain (representing one half of the binary complex), when expressed in *E.coli*, whereas domain cleavage has been shown to reduce stability in other proteins, e.g. the C-terminal domain of γB-crystallin (Mayr *et al.*, 1994). One possible necessity for domain fusion may be that a mechanism is necessary for the correct association of dissimilar, but related, subunits (Tang *et al.*, 1978). This would reflect an evolutionary advantage in the synthesis of a co-chaperonin molecule exhibiting polarity, containing two distinct, but functional surfaces exhibiting modular behaviour. This would imply the existence of a protein folding mechanism that requires selectivity in the binding of co-chaperonins with the α and β isoforms of the ch-cpn60 being necessary to achieve optimum interactions and efficiency in the polypeptide discharge reaction.

1.16 Endosymbiosis of the Chloroplast Plastid Chaperonin

The chloroplast is one of several types of organelles collectively termed plastids, all are formed from the same precursor proplastid and all of which contains the same chloroplast DNA. It is thought that the chloroplast began as an ancestral cyanobacterium (a photosynthetic prokaryote) that became endocytosed by an ancestral eukaryotic cell, and replicated with the cytoplasm. It is hypothesised that a fraction of bacterial DNA, which encodes for proteins that evolved into chloroplast proteins translocated into the nucleus, which led to some of the chloroplast protein being imported into the organelle after their synthesis in the cytosol (Gray, 1992; 1993).

Mitochondria are also thought to have originated by a similar process whereby a prokaryotic cell capable of oxidative phosphorylation became the inner mitochondrial membrane. The outer membranes of the mitochondrial and chloroplast may well be evolutionary remnants of the membrane of the endocytic vesicle into which these prokaryotes were internalised (Gray, 1993, Margulis, 1993).

The biogenesis of a chloroplast is similar to that of the mitochondria. Some regions of chloroplast DNA are strikingly similar to those of the DNA in present-day bacteria (Ohyama *et al.*, 1988, Clegg *et al.*, 1994).

The term chaperonin reflects the evolutionary relationship now accepted to exist between the chaperones found in prokaryotes (i.e. *E.coli's* GroEL cpn60) and eukaryotes (i.e. the chloroplast plastid chaperonin ch-cpn60), (Hemmingsen *et al.*, 1988, Zeilstra-Ryalls *et al.*, 1991, Ellis and van der Vies, 1991). It appears that the last common eubacterial ancestor from which the mitochondrion originated was related to the cyanobacterial group (Viale and Arakaki, 1994). Regarding the cpn60 evolutionary tree, the branching of chloroplast homologues with respect to the cyanobacterial group supports the inferences posited from earlier studies regarding the endosymbiotic origin of chloroplasts from the purple bacteria (Margulis, 1970, Schwartz and Dayhoff, 1978, Gray and Doolittle, 1982, Gupta and Golding, 1993).
More specifically, the relationship between plant chloroplasts and cyanobacterial homologues, is evidenced by the number of shared sequence characteristics (Bertsch *et al.*, 1992). It is now

generally accepted that the chloroplasts of plants and algae originated as endosymbionts, with those organisms having plastids surrounded by two membranes (rhodophytes, chlorophytes, and land plants) evolving from a progenitor phagotrophic eukaryotic host plus cyanobacterium (Douglas *et al.*, 1991).

The existence of two distinct polypeptides of the chloroplast plastid chaperonin, ch-cpn60 (α and β), marks a noticeable departure from the composition of the bacterial and mitochondrial chaperones. The α and β chaperonin proteins are encoded for by the nuclear genes and imported into the chloroplast, after synthesis in the cytosol (Ellis and van der Vies, 1988, Hemmingsen, 1990). The branching of both the α and β subunits of ch-cpn60 sequences within the cyanobacterial/chloroplast cluster, suggests that both of these homologues evolved by a gene duplication event that took place in a common ancestor of the ch-cpn60 chaperonin. The identification of two different cpn60 homologues, found in a cyanobacterial species, supports this inference (Lehel *et al.*, 1993). A direct correlation now exists for the endosymbiotic origin of organelles, mitochondria and chloroplasts, from prokaryotic ancestors held accountable for the origin and distinct identity of the GroEL cpn60 and ch-cpn60 families (Hemmingsen *et al.*, 1988, Cheng *et al.*, 1989, Reading *et al.*, 1989, Martel *et al.*, 1990).

The precursor sequences for the β subunit, of the pea chloroplast *Pisum sativum* (ch-cpn60), comprising 12 introns demonstrate that the intron positions were determined by comparison with the α and β cDNA sequences of *Brassica napus* and finally in the genomic sequences of *Arabidopsis thaliana* (personal communication, Ellis, 2002, and reviewed by Hill and Hemmingsen, 2001); the α genomic sequence possesses no introns (personal communication, Ellis, 2002). This observation could attest to a eukaryotic evolutionary origin for the β subunit of the ch-cpn60 chaperonin and a prokaryotic origin for the α subunit.

Research Aims

The tetradecameric cage-like structure of the plastid cpn60 (chaperonin) appears to be an evolutionarily conserved form of most chaperonins and it is of special interest to understand how this unique structure is formed. It is of importance to determine the stoichiometric distribution of the α and β subunits of the plastid cpn60, which may be achievable by nearest-neighbour chemical crosslinking. Analysis of homologous models to the plastid α and β subunits reveals that cysteine-519 of the β subunit can crosslink to lysine-22 of an adjacent α or β subunit, by use of heterobifunctional reagent SPDP (N-Succinimidyl 3-(2- pyridyldithio) propionate). Cyanogen bromide digestion and N-terminal sequencing should identify the crosslinked neighbours. The identity of the nearest-neighbour subunit will distinguish between models involving α7, β7, or α+β heptameric rings of subunits.

Unlike the β subunit, the α subunit does not self-assemble into homo-oligomeric complexes, although it is thought to participate in mixed complexes with the β subunit, in the 'native' protein complex. Determination of the X-ray structure of the α subunit will enable understanding of its unusual properties. The crystallographic techniques to be used in determining the three-dimensional structure of these subunits will be M.R. (Molecular Replacement), and M.A.D. (Multiwavelength Anomalous Dispersion) through the bioincorporation of the analog Selenomethionine in place of the proteins natural methionines, allowing 'direct phase' determination from the acquired X-ray data set.

With regard to the ATPase-driven mechanism of the plastid cpn60 it would be useful to understand further the 'positive' and 'negative' cooperative states that exist within the 'intra-' and opposing 'inter-' rings of a tetradecameric complex. To help gain further insight into these two states, and also the possibility of individual subunits being subjected to allosteric constraints whilst contained within the ring complex, a series of ATPase activity assays will be carried out using the individual subunits (α and β) of the plastid cpn60.

In this way, the data acquired may help to further elucidate the subunit dispersal of the complex.

CHAPTER TWO

Homology Modelling of the Plastid Cpn60

2.1 Introduction

The growing database of protein sequences and three-dimensional structures has revealed many protein families. Members are primarily associated by sequence homology, but there are increasing numbers of proteins with structural homology where little or no sequence identity exists. Modelling the three-dimensional structure of proteins on the foundation of family characteristics continues to make a valuable contribution to understanding protein structure and function (Pearl and Taylor, 1987). The rapid expansion in genome sequencing coupled with the likelihood that nature works from a 'fold-library' of finite size suggests that modelling will grow in influence.

As described previously, the amino acid sequences of the plastid cpn60 α and β subunits share approximately 50% identity with the *E.coli* protein, GroEL. The crystal structure of GroEL was determined in 1994 (Braig *et al*; 1994) and that of the GroEL-ES complex in 1997 (Xu *et al*; 1997). Structures have also been described for nucleotide complexes (Boisvert, 1996), for isolated domains (Buckle *et al.*, Zahn and Fersht, 1997) and for simulated trajectories of allosteric transitions (Karplus, 2002). We are now in a strong position to analyse homology models of the plastid proteins to seek explanation for their properties.

GroEL comprises 547 amino acids in a single subunit. Fourteen of these subunits are arranged as two seven-membered rings, stacked back-to-back to form a cage-like complex, enclosing a cavity where folding events are believed to be initiated. The molecule has been called an 'Anfinsen-cage'.

The N- and C- termini of the subunit sequence are close together, spatially. From the N-terminus the polypeptide chain visits the equatorial, intermediate and apical domains in turn, returning via the intermediate domain to complete the equatorial domain structure.

Depicted in Figure 2.1.1 shows the equatorial domain dominated by helical secondary structure elements as is the intermediate domain. However, the core of the apical domain comprises two four-stranded antiparallel β-sheets, enclosed by several helices.

Fig 2.1.1 Components of the GroEL Subunit

Depicted are the separate domains and secondary structural elements of a single GroEL subunit (drawn using the MOLSCRIPT graphics program). The domains are colour coded for recognition purposes; apical domain (gold), intermediate domain (red), and the equatorial domain (green). Secondary structural elements are indicated by cylinders (α-helices, labelled A to R) or arrows (β-strands, numbered 1 to 19) and extended strands.

The oligomeric structure of the protein is stabilised by a modest repertoire of inter-subunit contacts. The equatorial domains make the most substantial contribution to these contacts, interacting with neighbours within and between heptameric rings of subunits. Intermediate domains interact with apical domains of adjacent subunits and apical domains also contact neighbouring apical domains. However, in total only 6.6% of the solvent accessible surface of each subunit is buried in the oligomer. This is consistent with the extensive reorganisations that take place during the GroEL folding cycle and the associated modulation of the ATPase activity and with the case of chemical disassembly (i.e. urea) of the complex. These inter-subunit interactions are of central interest to studies of the plastid cpn60 as the β subunit is known to self-assemble, albeit to a less stable and less active oligomeric form to that of GroEL, while the α subunit does not self-assemble at all.

2.2 Model Analysis

The amino acid sequences of the plastid cpn60 α and β subunits were pasted into the Swiss Modeller Package (EXPASY, Swiss Institute of Bioinformatics (SIB)). The program searches for sequence homology to a known three-dimensional structure, modifies the structure and returns an energy minimised structure of the subunit. In many protein families, sequence variation accumulates on the surface of the protein and the modelling task is trivial with respect to the more conserved hydrophobic core that tends to define secondary motifs and their disposition. Insertions and deletions can be more difficult to model, but usually reside on loops interconnecting major structural elements. With an oligomeric protein such as cpn60, variations in surface side-chains take on a more important perspective as they may well contribute to subunit interactions.

As expected the Swiss-modeller returned a plausible model of both plastid subunits, with the homology models based on the various solved crystallographic structures of the GroEL chaperonin. In order to assess the models a sequence alignment was performed with 'MALIGN' to highlight sequence differences between GroEL and the plastid α and β subunits, with respect to the known secondary structure elements of GroEL and the known regions of inter-subunit interactions of the oligomer.

This type of analysis might reveal why α subunits do not interact well with each other within or between heptameric rings. This method of investigation may further reveal beneficial β subunit interactions, which give an indication as to a likely mixed oligomer composition of the two divergent subunits.

The sequence alignment representing a single subunit of GroEL and the plastid cpn60 α and β subunits, was accomplished through using the MALIGN program (appendix 1), and showed six sites where significant insertion/deletion events occur at inter- and intra- subunit sites, and these are highlighted on the structure in Figure 2.2.1 (drawn using the MOLSCRIPT program). Changes at residues 11 and 12 cannot be considered, as there is no structure is currently available for this region.

2.3 Sequence-Difference Analysis of the Adjacent Subunit Stabilising Loop Region

Figure 2.3.1 (drawn using the MOLSCRIPT program) shows a further insertion/deletion event having occurred at residue 42, strand 2 of the equatorial domain, where the lysine residue of GroEL is substituted by serine in the β subunit and deleted in the α subunit. The Swiss Modeller has restructured the loop between strands 2 and 3 to accommodate the deletion in the α subunit. These changes are of considerable interest because this loop region is involved in intersubunit contacts within the heptamer that undergo considerable change during the 'T to R' (Tense to Relaxed, refer to chapter one, sections 1.11.1 and 1.11.2) transition.

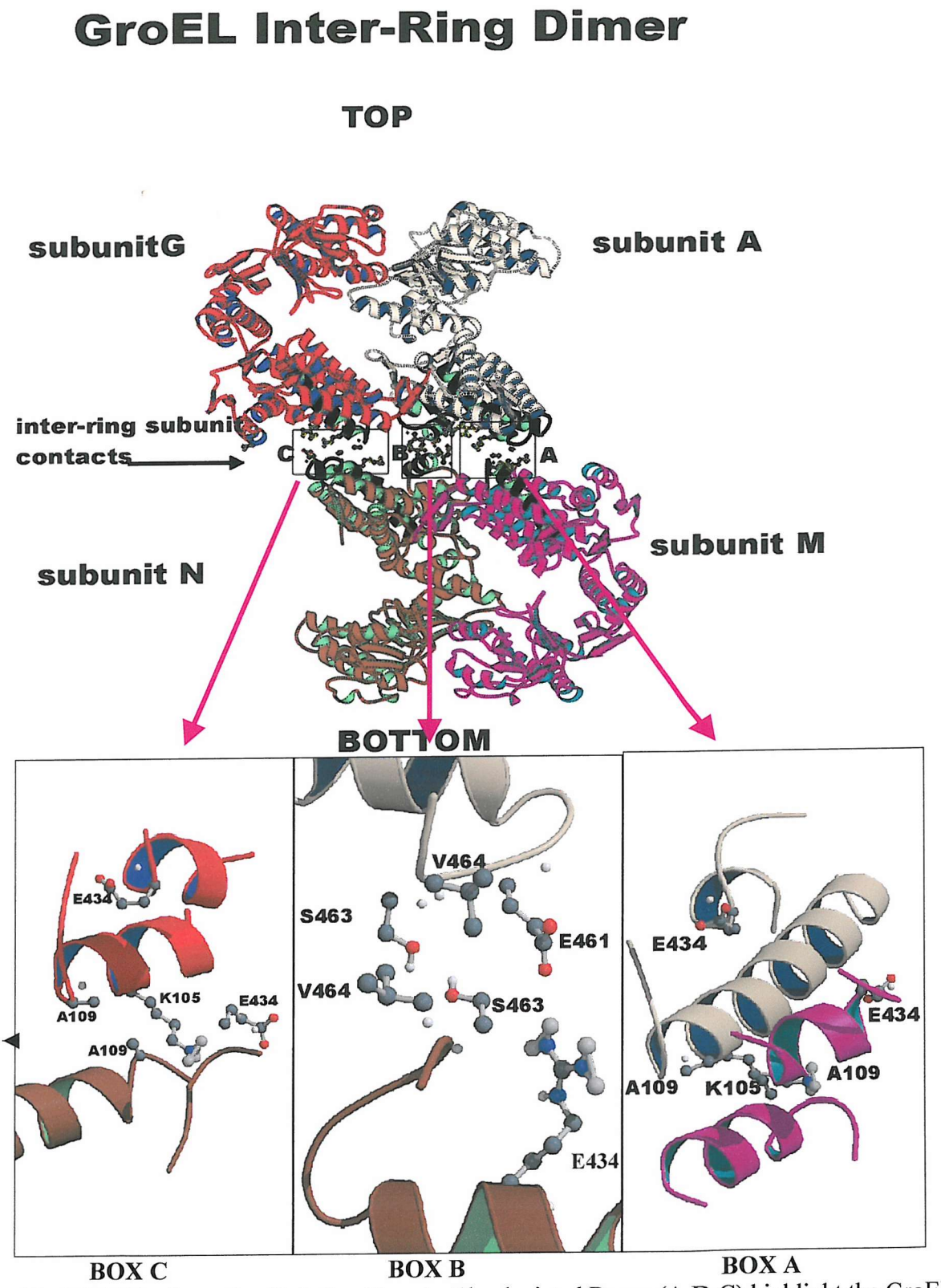


Fig 2.2.1 The Insertion/Deletion Events. The depicted Boxes (A,B,C) highlight the GroEL residues involved in the insertion/deletion events with the α and β subunits.

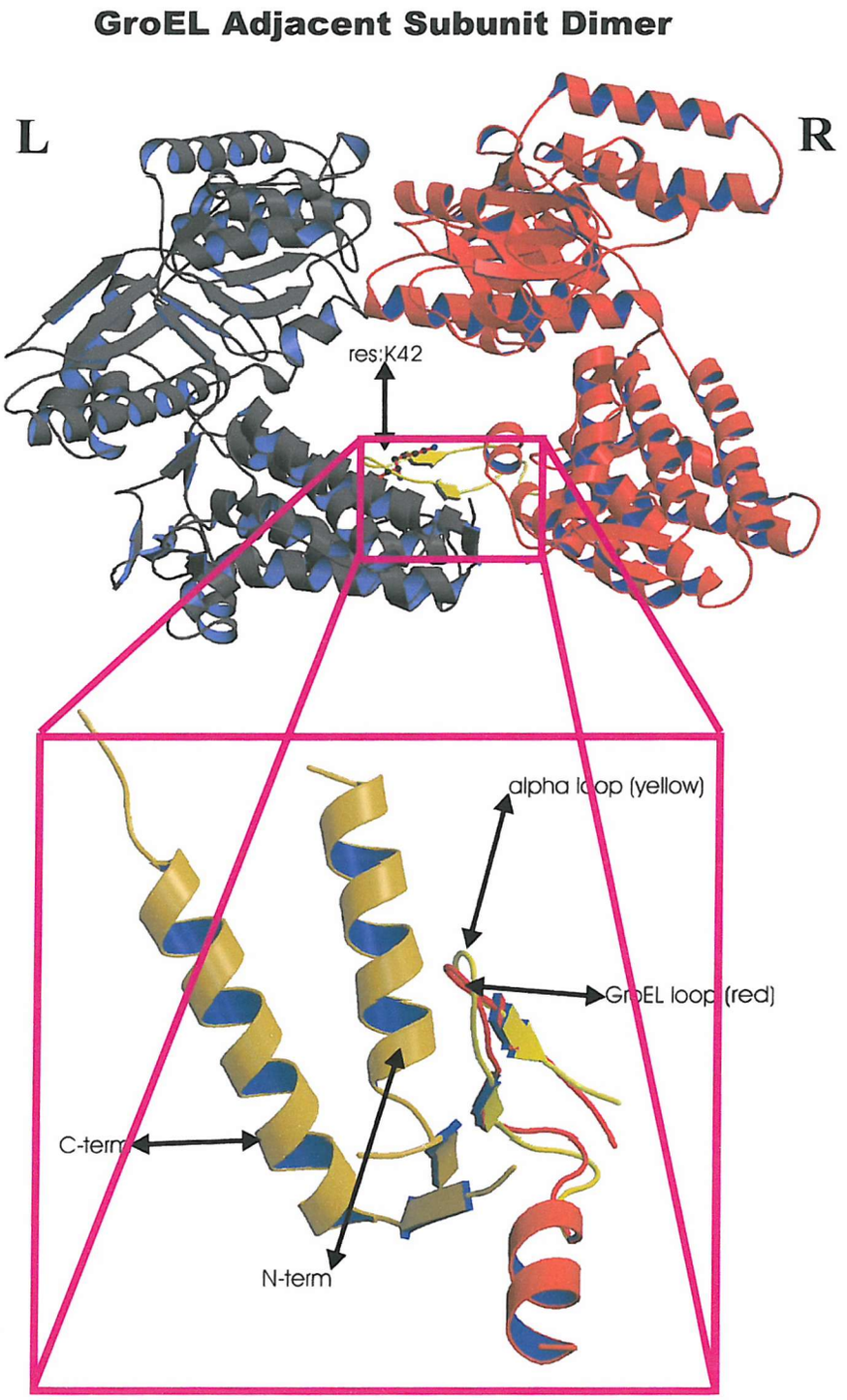


Fig 2.3.1 GroEL Lysine Residue 42 (loop region). The plastid cpn60 insertion is a serine(β subunit), and a deletion (α subunit).

The disposition of the GroEL loop region (residues 37 to 49) is shown in Fig 2.3.1. In GroEL residues 37 to 40 possess hydrogen bonds with residues 518 to 521 of the neighbouring subunit, providing a β-sheet stabilisation to the equatorial domain interaction. It is in this loop region that a greater significance can be placed upon the plastid cpn60α subunit deletion (residue 42), illustrated when one is confronted with the ion pairing known to exist in GroEL between the residues Lys 42 and Glu 59,which is seen to stabilise this loop region. This significance is illustrated further when one sees that the glutamine residue is conserved at position 59 for GroEL as well as for the plastid cpn60 β subunit, but is substituted for an alanine in the plastid cpn60 α subunit. The β subunit, although not significant for a deletion at the residue 42 position, does however lose the lysine for a serine insertion. This immediately questions the integrity and stability of the ion pairing that should exist in the β subunit between the 42 and 59 positions, due to the possibility of a required length deficit. In the same GroEL loop, the side-chain oxygen of Asp 41 hydrogen bonds with the Thr 522 of the adjacent subunit. By convention Asp 41 is from the left-handed partner and Thr 522 is from the right-handed partner, this would be apparent when viewing the interaction from within the cage. However, in both the α and β subunits the Thr 522 residue is substituted by a valine, deleting yet another stabilising interaction required to stabilise the heptamer.

2.4 Sequence-Difference Analysis for the Various Helices

Further deletions occurring at residue 144 on helix F of the β subunit intermediate domain are further compounded with insertions occurring at residue positions 433 and 434 in the α and β subunit. These insertions/deletions correspond to the loop interconnecting helices N (intermediate domain) and O (equatorial domain) near to the ATP binding site (Fig 2.1.1). Finally, there is a further insertion, proline, at position 478 in a loop interconnecting helix P and strand 16 of the equatorial domain (Fig 2.1.1).

To sum up this part of the sequence-modelling analysis, it is has been observed that all of the insertions and deletions occur in the equatorial domain or at the beginning of the intermediate domain. This data is in stark contrast to the observation that the greater overall sequence differences lie within the apical domains of the different proteins, depicted in Figure 2.4.1.

Fig 2.4.1 Apical Domain Sequences

Depicted are the residue alignments for the apical domains of the chaperonins GroEL (*E.coli*), and the α and β subunits for the chloroplast plastid chaperonin Ch-cpn60 (*Pisum sativum*), Res:191-376 (pink box). Due to the functional nature (substrate binding) of the apical domain, it is entirely reasonable to expect a greater range of residue variability when directly compared to their cohort domains, the intermediate and equatorial domains.

The most significant residue-event to be observed, between the relevant subunits of GroEL and the plastid α and β subunits, was the deletion which occurs at position 42 in the α subunit. This deletion must lead to an overall destabilisation of the intersubunit binding-contact in any α-containing oligomers. Although there are other sequence variations in this loop region, it is now strongly evident that the ionic interactions necessary to ensure nearest neighbour binding may have been seriously compromised for both the α and β subunits at residue 42.

Variations in the amino acid sequence have also been observed on the equatorial domain surfaces that interact between the opposing heptameric rings of the differing subunits. These contacts are concentrated in two zones, designated 'Left and Right' (as described above, section 2.3). The left-handed contact region in GroEL involves a polar interaction between Lys105 (side-chain nitrogen) of one ring and the Ala109 (main-chain oxygen) residue from the other. In the α subunit Leu105 cannot support this interaction. Furthermore, α-substituted Thr109 side chains must surely interact in a different manner in comparison to the Lys/Ala van der Waals contacts in GroEL. In the right-handed contact region, the ion pairing between Arg452 and Glu461 is removed completely in the α subunit model, this is entirely due to the Ala461 substitution.

However, perhaps the most telling, and critical, substitution in this region occurs at position 463. In both GroEL and the plastid β subunit, the conserved two-fold axis related Ser463 residues between the top and bottom subunits, exist very close together. However, in the plastid α subunit a Glu463 substitution immediately brings to bare an electrostatic repulsion at this position. The enormous significance of this very simple chemical fact, may be sufficient evidence enough to call into question the ability of the plastid cpn60 α subunit protein to electrostatically form an α-top/α-bottom complex arrangement.

2.5 Co-Chaperonin's cpn10 (GroES) and cpn21 (Plastid) Comparative Analysis

GroES homologues are also found to be present in chloroplasts. Interestingly, the chloroplast co-chaperonin is a polypeptide of about 21kDa, and is referred to as cpn21. Cpn 21 is twice the size of its bacterial and mitochondrial cousin's cpn10. In pea (*Pisum sativum*), chloroplast cpn21 was identified by its ability to both assist GroEL cpn60 in the ATP-dependent refolding of chemically denatured Rubisco, as well as possessing the ability to form a stable complex with the GroEL cpn60 complex in the presence of Mg.ATP (Bertsch *et al.*, 1992), indicating considerable functional conservation in the structure of the interactive chaperonin-co-chaperonin surfaces (Bertsch *et al.*, 1992, Burt and Leaver, 1994, van der Vies *et al.*, 1994).

Chloroplast cpn21 is comprised of two distinct cpn 10-like domains fused in tandem, via a peptide linker (Thr-Asp-Asp-Val-Lys-Asp, Baneyx *et al.*, 1995), to give a binary co-chaperonin structure (Bertsch *et al.*, 1992). Each of the two fused domains possesses several highly conserved amino acid residues that are found to be encoded in many GroES genes, suggesting that each domain could be independently functional whilst remaining in the fused state (Baneyx *et al.*, 1995). The two halves of the ch-cpn21 show ~40% amino acid identity to each other and are ~70% homologous if conservative changes are also considered (Bertsch *et al.*, 1992).

Surface regions of the chaperonin cpn60 apical domains are also involved in ligand binding and co-chaperonin binding during the folding cycle. Indeed, some of the same side chains are involved in both processes since the apical domains undergo a substantial reorientation when ligands are displaced into the cavity of the cpn60-cpn10 complex. The majority of these residues are found to be non-polar in character in GroEL and the corresponding α and β modelled counterparts.

However, a range of polar substitutions are observed in the plastid α subunit, at positions 204, 259, 263, 264; sites known to be involved in ligand binding. Polar substitutions are also found at residues 309, 314, 349 and are again sites which are known to fully participate in both the binding of polypeptide and co-chaperonin in the GroEL homologue (Fenton *et al.*, 1994). These substitutions may be associated with specificity for the foldee and /or novel cpn 10 interactions. Rubisco subunit folding and assembly could be segregated to chaperonin type but, thus far, there is no evidence for this. However, for the binding of GroES to its natural GroEL chaperonin there is available experimental data regarding the plastid homologue ((Bertsch *et al.*, 1992), also outlined in the introduction.

Homology modelling carried out for each of the cpn21 sequence repeats, based on a GroES template, provide plausible model's for a structural motif. However, without the experimental evidence the distribution of these motifs can only be inferred. It is not unreasonable to suppose that cpn21 motifs come together in two groups of seven, with cpn60 recognition loops necessarily displayed upon the two surfaces, Figures 2.5.1(a) and (b) (drawn using the MOLSCRIPT program). The organisation of these motifs and their cpn60 recognition loops experience a similar ambiguity in their organisation (rings of A motif and B motif or alternating ABAB, etc., motifs) to the plastids presence of two unique and distict α and β subunits (Hemmingsen *et al.*, 1988, Ellis and Hemmingsen, 1989, Martel *et al.*, 1990). That these two proteins are so divergent suggests that they have evolved to perform specialised functions (Martel *et al.*, 1990). It is, therefore, quite intriguing that the chloroplast should possess a double-domain co-chaperonin, which would appear to indicate a quite plausible hypothesis that the organisation of both the plastid co-chaperonin (cpn21) and its natural suitor (plastid cpn60) are, in some way, correlated. Depicted in Figure 2.5.2 is one such possible hypothesis, where the plastid cpn60 complex consists of an α/α heptameric ring in a complexed state with a β/β heptameric ring, conferring a fully active plastid cpn60 α and β complex conjoined by the domain-domain-liinked plastid cpn21 co-chaperonin. Domain fusion to form cpn21 was an early evolutionary event and the fusion has been highly conserved suggesting that the retention of the two-domain structure is advantageous in chloroplasts (Baneyx *et al.*, 1995).

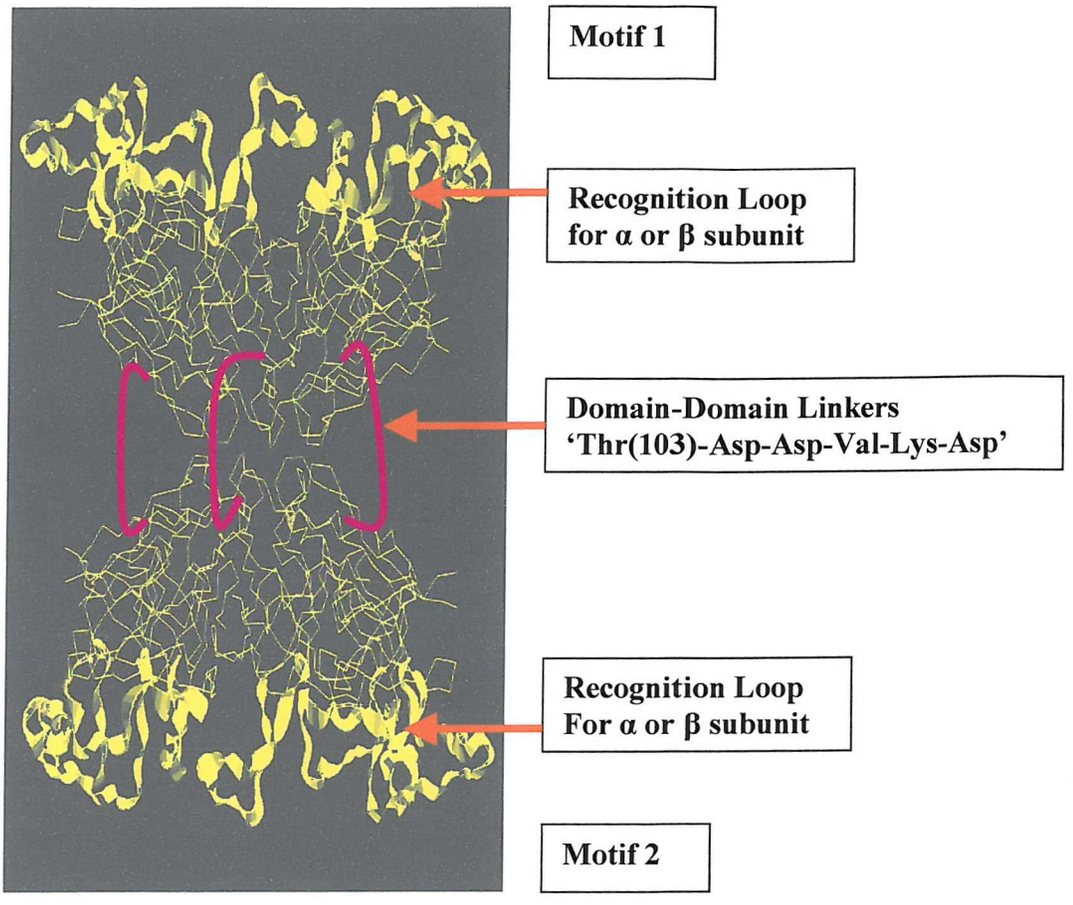


Fig 2.5.1(a) A Proposed Oligomeric Plastid cpn21 Double-Domain Model

Depicted above is the hypothetical structural distribution for the plastid cpn21. It is suggested that one such conformational possibility for the linear domain-linker-domain cpn21 is that one cpn10-domain (motif) could sit atop the other cpn10-domain, via the peptide linker, thereby conferring the ability to service either the hetero- (α/βα/β) or homo- (α/α/β/β) presented complex of the plastid cpn60. Highlighted in both motifs are the cpn60 apical domain recognition loops (Yellow ribbons) and the domain-domain linkers (Pink) allowing the tandemly linked domains the ability to invert against each other, becoming autonomous cooperative folding units, whilst retaining structural cooperative dependence upon one another viathedomain-domain-linker.

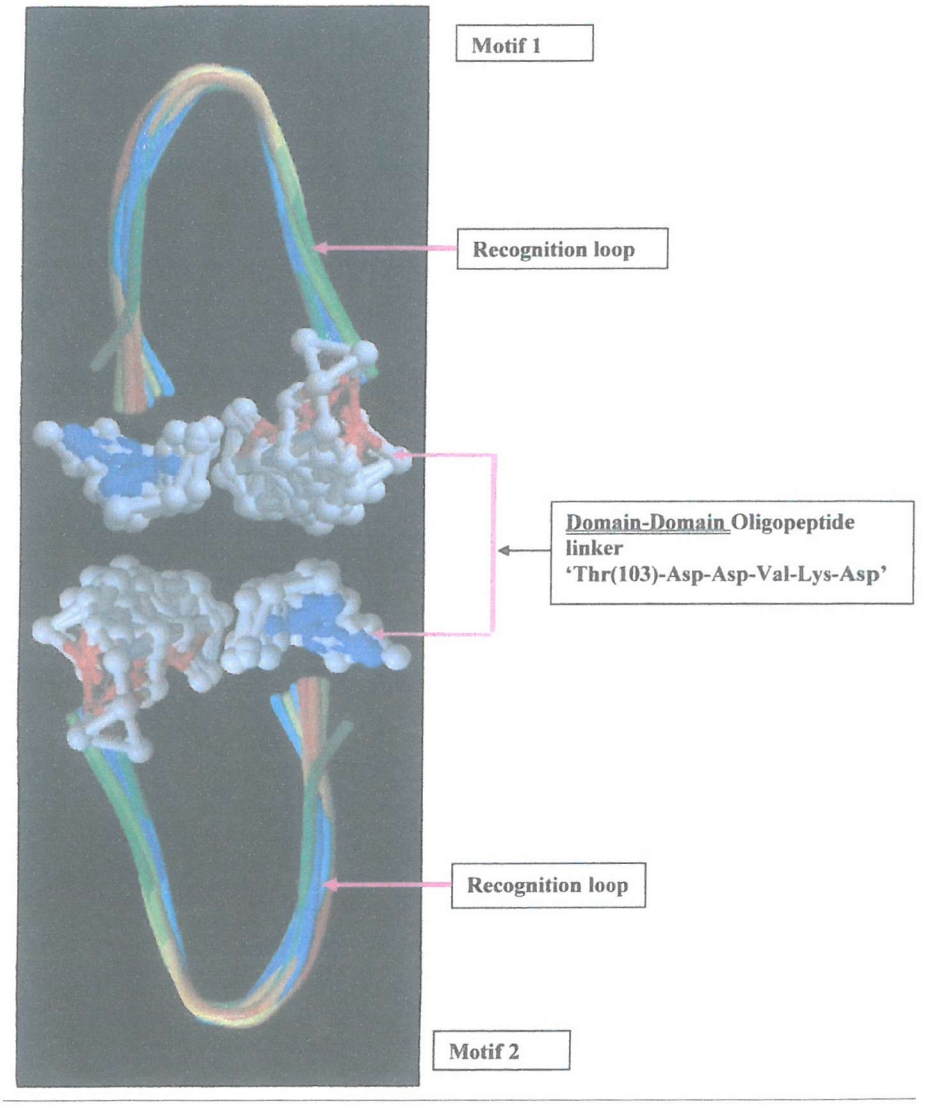


Fig 2.5.1(b) A Proposed oligomeric Inverted Domain-Domain-Linked cpn21

Depicted above is a modelled interpretation of just one of the plastid cpn21 subunits. The hypothesis for this model is the ability of each individual subunit to possess a significant structural flexibility, via the peptide linker, which in turn enhances the rate of folding as well as improving its stability by reducing the entropy of its unfolded state.

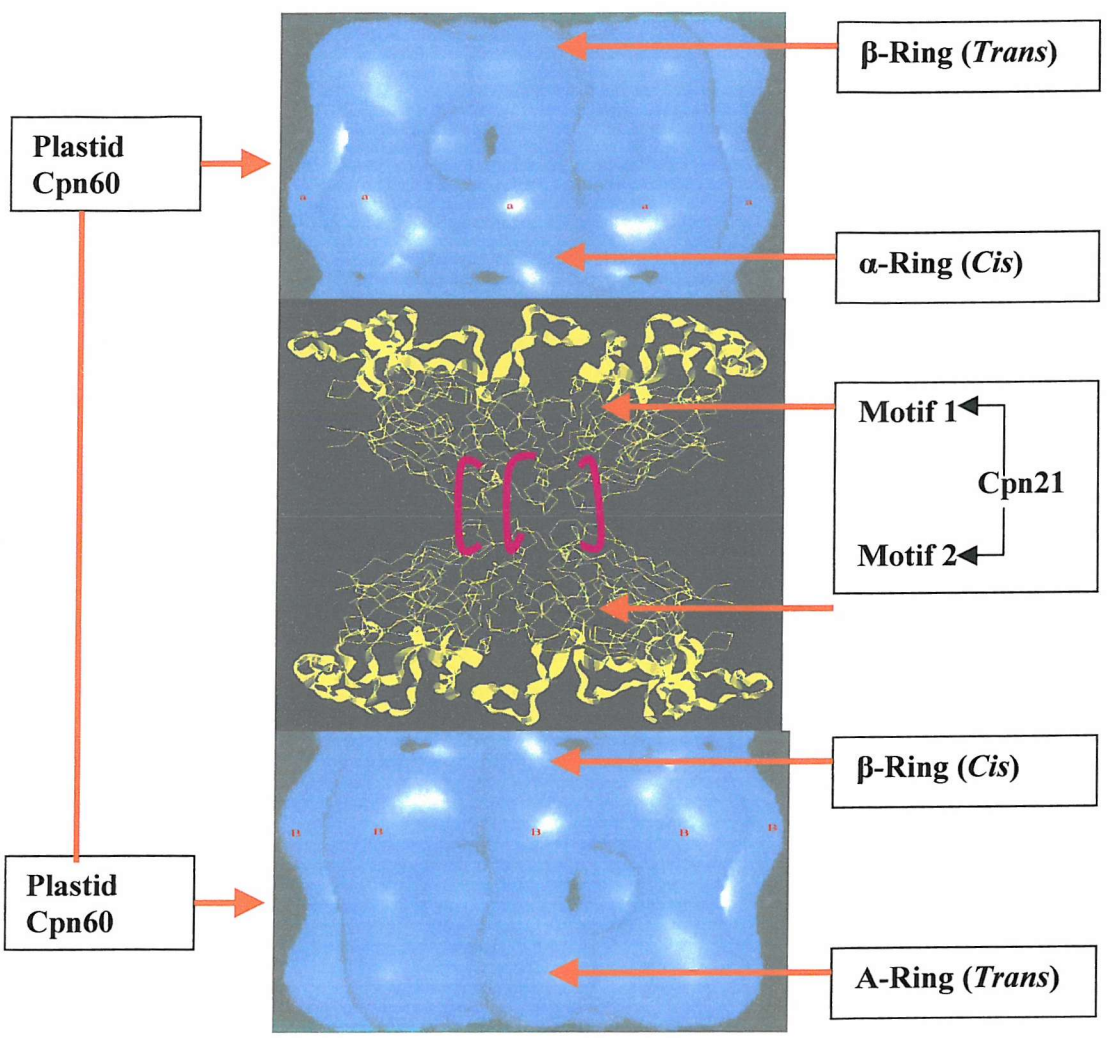


Fig 2.5.2 A Proposed Model for the Plastid cpn60 Protein Folding Mechanism

Depicted above is a suggested model strategy for domain fusion of cpn21. The co-chaperonin would exhibit polarity and contain two distinct but functional surfaces, exhibiting a modular behaviour. The implication of such activity would imply a protein folding mechanism that requires selectivity in the binding of co-chaperonins with the α and β forms of the plastid cpn60 to give optimum interactions and efficiency in the polypeptide discharge reaction.

2.6 Discussion

The model for cooperativity, based on GroEL switching of intersubunit salt bridge interactions, provided a general and prerequisite insight into the required allostery and cooperativity necessary to perform this set of modelling experiments.

The conformational changes promoted by ATP binding to GroEL have been crucially examined, using the structure of an ATPase mutant of the chaperonin, GroEL (D398A) complexed with ATP (Ranson *et al.*, 2001). By fitting GroEL subunit domains *en bloc* into a cryo-EM structure (10Å resolution) they were able to demonstrate that ATP binding changes the pattern of interactions throughout the double ring structure. Owing to tilts of the equatorial domains, the inter-ring interface is distorted, suggesting a pathway for propagating negative cooperativity to the opposite ring (see Chapter One and Chapter Seven).

These modelling experiments were carried out with the express intention of testing the structural binding-cohesiveness, via the chemical and ionic abilities, for both the plastid α and β proteins in a hyperthetical complex where both subunits may coexist or act in an independent manner. Experiments carried out using mutations at two sites in GroEL (Yifrach and Horovitz, 1994) go some way to supporting the findings of this set of modelling experiments. They observed that cooperativity in ATP binding by GroEL is mediated by the switching of salt bridges. They also showed that disruption of the intermediate-apical salt bridges present in unliganded GroEL by mutation R197A causes a large decrease in positive cooperativity. As described in section 2.4, sequence- difference analysis for the α and β subunits showed insertions/ deletions corresponding to the loop interconnecting helices N (intermediate domain) and O (equatorial domain) near to the ATP binding site (Fig 2.1.1). Direct comparison between these differences and those of Yirach and Horovitz's intermediate-apical domain disruption caused by the R197A mutation, shows that switching or distortion of salt bridge contacts between subunits can provide a structural framework for allosteric catastrophe, should the controlling elements necessary in illiciting the requisite conformational changes be either chemically incapable or not present.
Further, more significant potential destabilising events were observed at position 42, in the α subunit. The ionic interactions fundamental to ensuring adjacent subunit binding has now been called into question for both the α and β subunits at this residue position.

Further questions were raised regarding the ability of the plastid subunits to interact between opoosing rings at the surfaces of the opposing equatorial domains. GroEL sustains interaction at this important site via the polar interactions between Lys 105 of one ring and Ala 109 in the oopsing ring. Crucially, at the 105 position the α subunit comprises a leucine residue, incapable of supporting the polar interaction necessary with its counterpart residue at position 109 in the opposing equatorial domain.

However, the most strident inference to be drawn from this set of modelling experiments must surely be levelled at the α subunit glutamine substitution at residue 463 (Fig 2.2.1). It is the ineffectiveness of this α-glutamine to form an electrostatic attraction at this position that calls into question the ability of the plastid α subunit to sterically form an independent and stand-alone complex. This evidence ties in well with the known inability of the α subunit to form a complex unless in the presence of its β subunit cohort.

CHAPTER THREE

3.1 Ligation, Expression and Purification of the Ch-cpn60 α Subunit

During expression trials for the Ch-cpn60 α-subunit protein it became evident that a problem with the expression of the protein had arisen. Before any future experiments could be carried out using the α-protein, it was necessary to resolve the problem and regain the α-protein expression.

Likely causes for the loss of expression were assumed to be either, repetitive thawing of the plasmid (pET24a) containing the cDNA insert of the α subunit, which could manifest as cDNA degradation. A second likely cause was that the T7 promoter region within the pET24a vector may have developed a degradation problem. Another area of concern was the likelihood of point mutations having manifested in the α-cDNA insert sequence.

To rectify the problem and regain expression it was necessary to excise the cDNA insert from the old vector (pET24a) and ligate the excised insert into a new, and freshly tested, pET24a vector.

3.2 Methods and Materials

3.2.1 Microbiological techniques

3.2.2 Sterilisation

All pipette tips, media, stock solutions, and microfuge tubes were sterilised by autoclaving at 120°C for 90 minutes. Stocks of heat labile solutions such as antibiotics and IPTG were sterilised by filtration through disposable 0.22μm Millipore filters.

3.2.3 Bacterial strains

Strain	Genotype	Special Features
E.coli BL21(DE3)pLysS	F⁻ *ompT* $hsdS_B(r_B^- m_B^-)$*gal dcm* (DE3) pLysS (Cm^R)	BL21 *recA* mutant; stabilises tandem repeats. *Ion* and *ompT* protease deficient. pLysS carries gene for T7 lysozyme, a natural repressor, suppressing basal expression of target genes.

3.2.4 Cloning vectors

Vector	Fusion Tags		Special Features
	N-term	C-term	
pET-3a	T7.Tag Monoclonal Antibody purification	None	T7 controlled expression vector. Inducible with IPTG. Confers ampicillin resistance.
pET-24a	T7.Tag	His.Tag Metal chelation chromatography	As with pET3a, but confers Kanamycin resistance.

3.2.5 Culture media

Cultures were either grown in Luria broth or 2TY. All were supplemented, when necessary, by the aseptic addition of the relevant antibiotic at the following final concentrations; ampicillin 50μg/ml, chloramphenicol 35μg/ml, kanamycin 25μg/ml.

Media	Materials	g/litre
Luria Broth (L-broth)	Tryptone Yeast extract NaCl	10 5 5
2TY	Tryptone Yeast extract NaCl	16 10 5

Bacterial colonies, post transformation, were grown on plates of L-broth that had been solidified by the addition of 1.5% w/v of agar No.1.

3.2.6 DNA techniques

3.2.7 Restriction and digestion of $_p\alpha/_p\beta$ DNA

The plasmid vector pET24a contained both the α and β gene inserts, in separate vectors. Restiction enzymes *BamHI* and *NdeI* were used to cut out the respective inserts to re-introduce them back into fresh pET24a vectors carrying the C-terminal His.Tag fusion site, respectively.

3.2.8 Agarose gel electrophoresis

Submerged horizontal agarose mini-gels were used for the identification of, sizing, and purification of the respective DNA fragments. The percentage of agarose gel used varied between 0.7% and 1.0% w/v of agarose, depending on the respective fragment size being investigated. The agarose was dissolved in 50ml dH_2O and 1ml of 50x TAE buffer, in the microwave for 1 minute. It was allowed to cool to ~ 50°C at which time ethidium bromide was added at a concentration of 0.1µg/ml. The gel solution was then poured onto a prepared gel plate, with a seven lane comb in place, and allowed to set, after which the comb was removed and the plate placed into an eletrophoresis tank. The tank contained 1x TAE buffer. The DNA samples were prepared in 6x loading buffer and 10µl of sample was loaded into the wells. The gel was run at a constant voltage of 100mV. The fluorescent DNA bands were visualised with a UV transilluminator.

Reagents: 50x TAE (242g Tris.HCl, 57.1ml acetic acid, 200ml (0.25M) EDTA pH8.0/litre dH_2O)
Ethidium bromide (10mg/ml stock solution)
Agarose (Electrophoresis grade)
6x loading buffer (0.25% w/v Orange G)

3.2.9 Purification of DNA from agarose gels

Purification of DNA from the agarose gels was achieved by excision of the band of interest and subsequent purification using the Quiaquick mini-column spin kits supplied by Qiagen, in accordance with the manufacturer's instructions.

3.2.10 Plasmid DNA preparations

The pET vector plasmids were cleaved (*BamHI* and *NdeI*) and purified, in preparation for the α and β inserts, in the same manner as the inserts.

3.2.11 Ligation of vector and insert

"Sticky-end" ligations were performed in 10μl volumes using 1μl (~3 units) T4 DNA Ligase, 10-50ng of vector DNA (with a 1:3 ratio of vector-to-insert). The ligation mixture was placed at 4°C overnight. A ligation control, with no insert, was always included. Post-ligation, 4-5μl of each ligated mix was transformed, using heat shock protocols, into either *E.coli* BL21 (DE3) pLysS to check expression levels, or into *E.coli* B834 (DE3) pLysS *met*⁻ for the seleno-methionine crystallisation trials.

Reagents: 10x Ligation buffer (500mM Tris.HCl pH 7.8, 100mM $MgCl_2$, 200mM DTT, 10mM ATP, 0.5mg/ml BSA)

T4 DNA Ligase (3 units/μl)

3.2.12 Transformations

Transformations for the vector (pET-24a) into the required bacterial strains (*E.coli* BL21 (DE3) pLysS, were performed following the manufacturers (Novagen) recommendations, using the cold/heat shock method. The host bacterial strains were supplied in the competent state by Novagen.

All transformations were carried out on ice and the following procedure used throughout. 20µl aliquots of competent cells BL21(DE3)pLysS were pipetted into pre-chilled (ice) 1.9ml snap-cap polypropylene micro-centrifuge tubes. 1µl of purified plasmid DNA (pET24a) was added directly to the competent cells and pipetted gently to mix. The transformation mix was incubated on ice for a further 5 minutes. The tubes were then heated in a water bath for 30 seconds, at 42°C. The tubes were then returned to ice for a further 2 minutes. 80µl of supplied growth medium (SOC) was then added to the mix and those plasmids containing kanamycin resistance (pET-24a) were then shaken at 37°C (250rpm) for a further 30 minutes prior to plating. The cells were then spun down in a bench microfuge, at 1300rpm, for 1 minute and the supernatant removed. The pellet was resuspended in 200µl LB (Luria Broth) to concentrate the sample. Various quantities of these stock cells (5µl, 20µl, 100µl) were then spread onto LB-agar plates containing the respective selective antibiotic (Kanamycin at 25µg/ml for pET-24a) and then incubated at 37°C overnight. A positive control, using a supplied test plasmid, was used in conjunction with all transformations.

3.3 pET24a Expression Vector

The original construct, containing the pET24a vector (Novagen: Cat. No. 69772-3) and the ch-cpn60 cDNA α insert, was supplied by Professor Paul Viitanen (DuPont) in glycerol stock form (Dickson *et al.*, 2000).

pET24a (Novagen) is a transcription vector designed for expression from bacterial translation signals carried within a cloned insert. pET24a is a T7 promoter controlled vector with a 14-residue T7.TAG which affords monoclonal antibody purification, post-expression. The vector also carries an available 6-residue C-terminal HIS.TAG which allows for protein-specific purification, post-expression. Sites unique to the pET24a vector are depicted on the vector map in [Fig 3.3.1a] and vector cloning region sequences are shown in [Fig 3.3.1b].

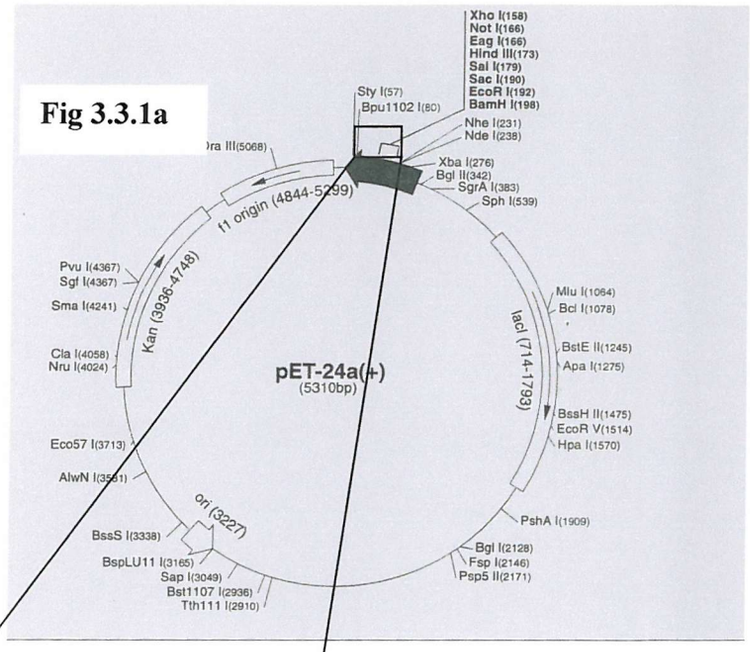
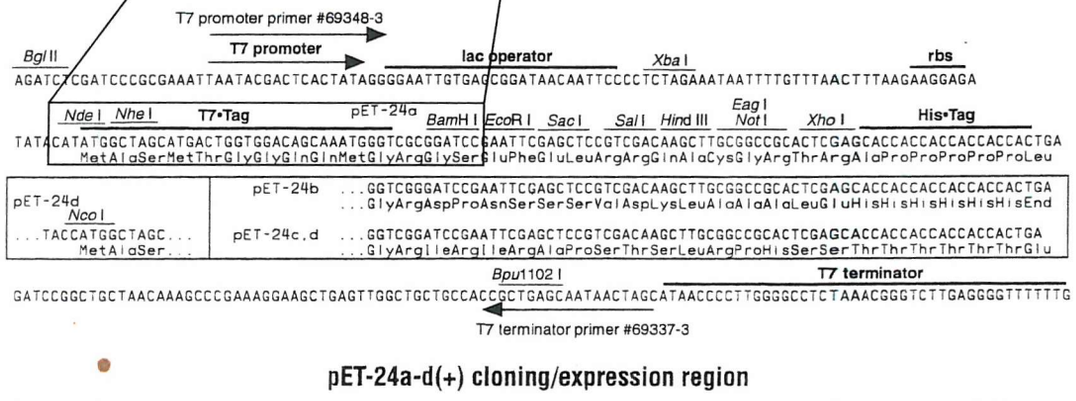


Fig 3.3.1a

Fig 3.3.1b



Figs 3.3.1(a and b) pET24a Vector Map and Cloning Sites

The highlighted regions between Figs 3.3.1a & 3.3.1b depict the cloning region used for the insertion of the ch-cpn60 α subunit cDNA sequence (Source: Novagen catalogue, 2000)

3.4 The making of the Ch-cpn60 Recombinant α-cDNA Insert

The plastid Ch-cpn60 α subunit precursor targeting sequences were removed and the modified genes then cloned downstream of the T7 promoter region of the pET24a vector (Dickson *et al.*, 2000). Two Polymerase Chain Reaction (PCR) primers were used to remove and amplify the targeting sequence.

Primer-1 (5'-CCTACCGCAACATATGGCCGCTAAAGATATCGCC-3') was hybridised to nucleotides 172-189 of the α gene insert, and introduced an *NdeI* restriction site (CA'TATG) and an initiator methionine residue at the transit peptide cleavage site (Dickson *et al.*, 2000).

Primer-2 (5'-CGCTCTAGAACTAGTGGATCC-3') was hybridised to the Bluescript KS^+ vector and introduced a *BamHI* restriction site (G'GATCC) downstream from the plastid α proteins own natural stop codon (TAG, present in the Primer 2 sequence). The PCR product was cut with *NdeI* and *BamHI* and then ligated into similarly digested pET24a vector. The resultant construct, pαCHEL, now encoded for the mature α ch-cpn60 and possessed the initiator Met at its N-terminus (Dickson *et al.*, 2000).

Introducing the *NdeI* and *BamHI* restriction sites into the oligonucleotide primers allowed easy incorporation of the α cDNA insert into the T7.TAG site of the pET24a vector (Fig. 3.3.1b). Retention of the α proteins natural stop codon (TAG) circumvented the unnecessary addition of an extra 18 amino acid tail (including the HIS.TAG) facility) that would otherwise have been expressed as part of the α protein product, had the pET24a vectors own stop codon (TGA) been utilised (Fig 3.3.1b).

3.5 Restriction/ Digestion of the Aberrant pET24a/pαCHEL construct

There were obvious advantages to be gained by using the same vector and restriction sites previously used in the original vector/insert construct. The *NdeI/BamHI* excision of the cDNA α insert would ensure the retention of the cDNA insert whilst negating the likliehood of incorporating any contaminated or degraded vector. Use of the same restriction sites would also offer the opportunity for ligation of the α cDNA insert back into the same cloning site (T7.TAG), within a new and potent pET24a vector (Novagen).

Before carrying out ligation of the α insert into the new vector, it was necessary and fundamental to excise the α insert from the old pET24a/pαCHEL construct to ensure its presence and potency (Fig 3.5.1).

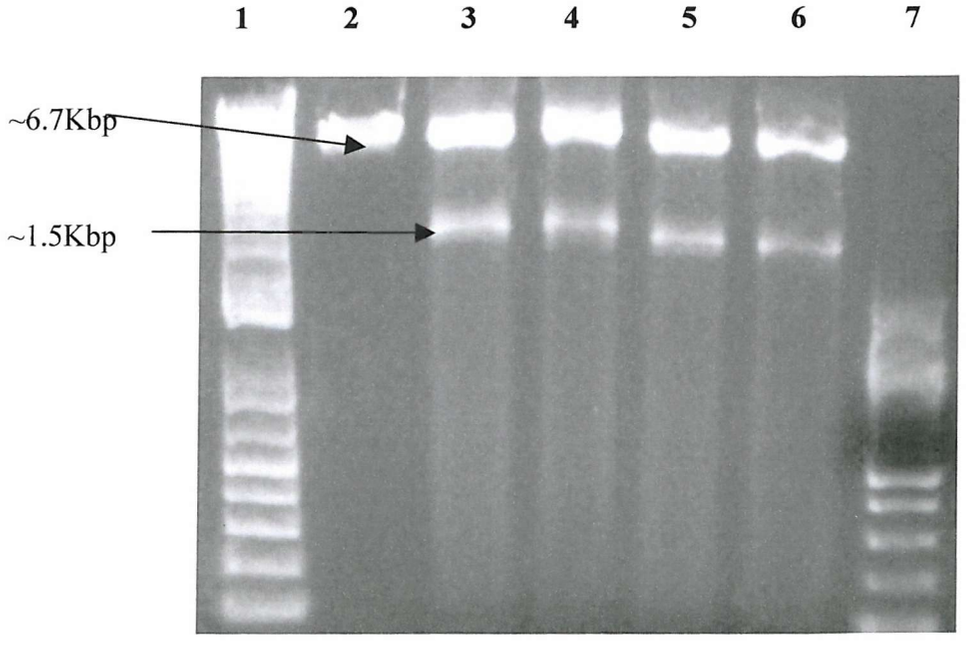


Fig 3.5.1 Restriction Digest (Nde1/BamH1) of the 'old' pET24a/pαCHEL

The vector size is 5236bp and the α insert size is 1500bp (Dickson *et al.*, 2000).

Lane 1: 1Kb marker; Lane 2: Uncut pET24a/pαCHEL; Lanes 3-6: Digested pET24a/pαCHEL (using Nde1/BamH1 restriction enzymes). There were x4 glycerol stocks (stored at -20°C) of the 'old' construct. The α insert is shown clearly at ~1500bp. Lane 7: 100bp marker.

3.6 Ligation and Transformation of the new pET24a/pα-insert construct

After achieving restriction/digestion of the α insert from the original construct, in parallel to the restriction/digestion of the new pET24a vector in preparation to receive the α insert, the next step was to purify each component in readiness for the ligation of the two products. Ligations were performed using T4 DNA ligase and a 1:3 ratio of vector-to-insert. Ligations were carried out overnight (o/n) at 4°C. Once ligation reactions had been performed, the ligation product (pET24a/pα insert) was then transformed into the *E.coli* bacterial host of choice, BL21 (DE3)pLysS, using the 'heat shock' transformation protocol recommended by the manufacturer (Novagen). The transformed cells were then spread, in varying dilutions, onto LB-Agar plates containing the selective antibiotic Kanamycin (25μg/ml) and incubated o/n at 37°C.

3.7 Expression of the pET24a/pαCHEL Construct

Selected colonies from the transformed plates were placed into separate vials containing 10ml fresh LB growth medium, and incubated at 37°C until mid-log growth phase had been obtained (O.D. $_{600nm}$ = 0.6). Prior to induction a 1ml sample was taken from the growth mixture to be retained as a pre-induction reference. A further 500μl was also extracted, to which 10% glycerol was added, this was stored at –20°C to be stored as plasmid stocks. IPTG (1mM/ml) was added to the remaining 8.5ml growth mixture to induce expression, and the cells were grown for a further 4hrs before harvesting. 1ml of the harvested cells was then spun down in a bench top centrifuge. The supernatant was discarded and the pellet was retained for expression analysis. A 4-12% SDS-PAGE (Nu-PAGE) analysis was carried out on both the pre-induction reference sample and the post-induction pellet. Figure 3.7.1 shows the results of the expression trials carried out on the constructed pET24a/pα insert.

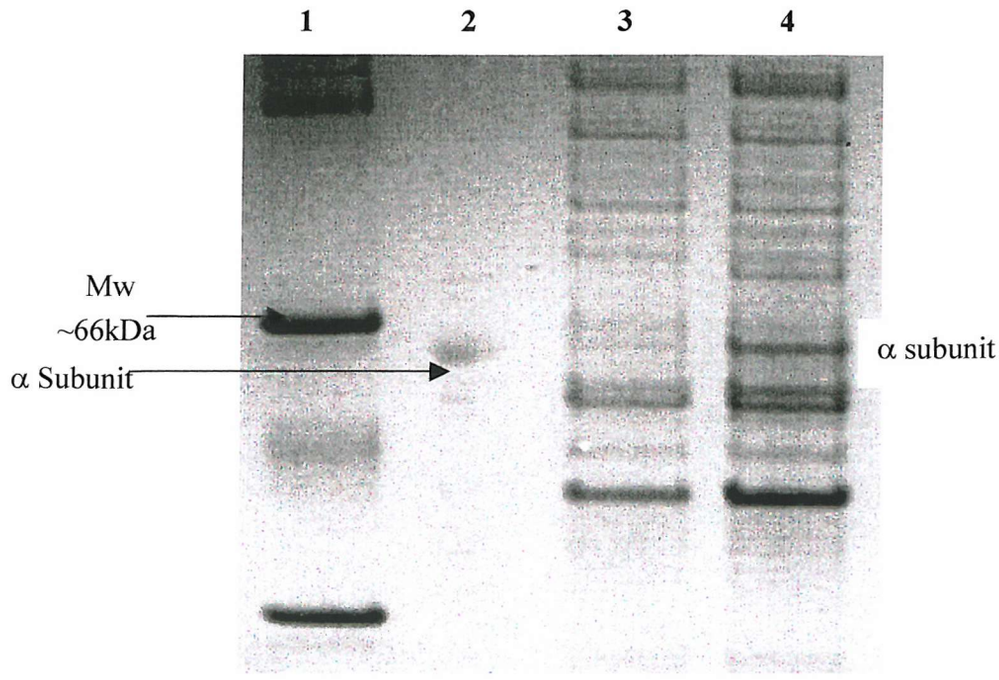


Fig 3.7.1 Expression Analysis of the Newly Constructed pET24a/pα Plasmid

The mass of the α monomer on a reduced SDS-PAGE is ~60kDa. Lane 1: Protein marker; Lane 2: Depicts a purified α protein, used as a mass marker for the expressed protein; Lane 3: Depicts the sample taken at pre-induction; Lane 4: Represents the pellet retained post-induction and after harvesting. The SDS-gel clearly indicates the expressed α protein in lane 4.

3.8 General Conclusions

By regaining the expression of the α protein, evidenced in Figure 3.7.1, it was apparent that initial assumptions, indicating a problem likely existing in the T7-promoter region of the old pET24a vector, could well have been substantiated. It can now be recognised that a potential problem essentially lies dormant and unobserved to the unaware researcher.
However, the fundamental recognition, and eventual solution, of this time-consuming problem should prove an essential aid to future experiments carried out with this particular protein

The expression of the cDNA α insert from the 'old' plasmid vector was successfully regained, after restriction and then ligation into a 'new' pET24a vector. Subsequently, experiments that required the α subunit could now be carried forward. Indeed, the expression problem initially came into focus due to the requirement for the protein to be competent in its expression. The experiment preferred was that of Multi-wavelength Anomalous Dispersion (MAD), designed to elicit the analog, Seleno-Methionine, bioincorporation into the plastid Ch-cpn60 α subunit protein in an attempt to solve the X-ray crystallographic structure (refer to Chapter Six).

CHAPTER FOUR

Chemical Crosslinking of the Plastid Ch-cpn60's Adjacent Subunits

4.1 Ch-Cpn60 (α and β) Subunit Stoichiometric Elucidation

Little is known about the dispersal of the ch-cpn60's two divergent subunits (α and β) other than a common acceptance that they exist in approximately equal proportions and likely exist in the same functional tetradecameric assembly (Dickson *et al.*, 2000). The α subunit is 587 amino acids in length, including a 47 amino acid transit peptide sequence (discarded once the membrane has been crossed); the β protomer is 595 amino acids long, and also includes a redundant 49 amino acid transit peptide once the protein is fully processed.

By visual comparison of the α and β protomer sequences (Fig 4.1), it can be observed that there is a significant difference in the number of cysteines contained within each subunit. The α subunit possessed one cysteine at residue-506, whereas the β subunit had five cysteine residues (218,418,506,507,519). The availability of these cysteines suggest the possibility of a chemical cross-linking experiment, with a potential for realising the dispersal of "nearest-neighbour" (adjacent) subunit identification. In this way it would be possible to determine either the homo- or hetero-oligomeric stoichiometry of the α and β subunits in relation to one another, possibly contained within a given heptameric ring of the tetradecamer. Fig 4.1.2 depicts possible α and β subunit dispersal scenario's within a given Ch-cpn60 complex.

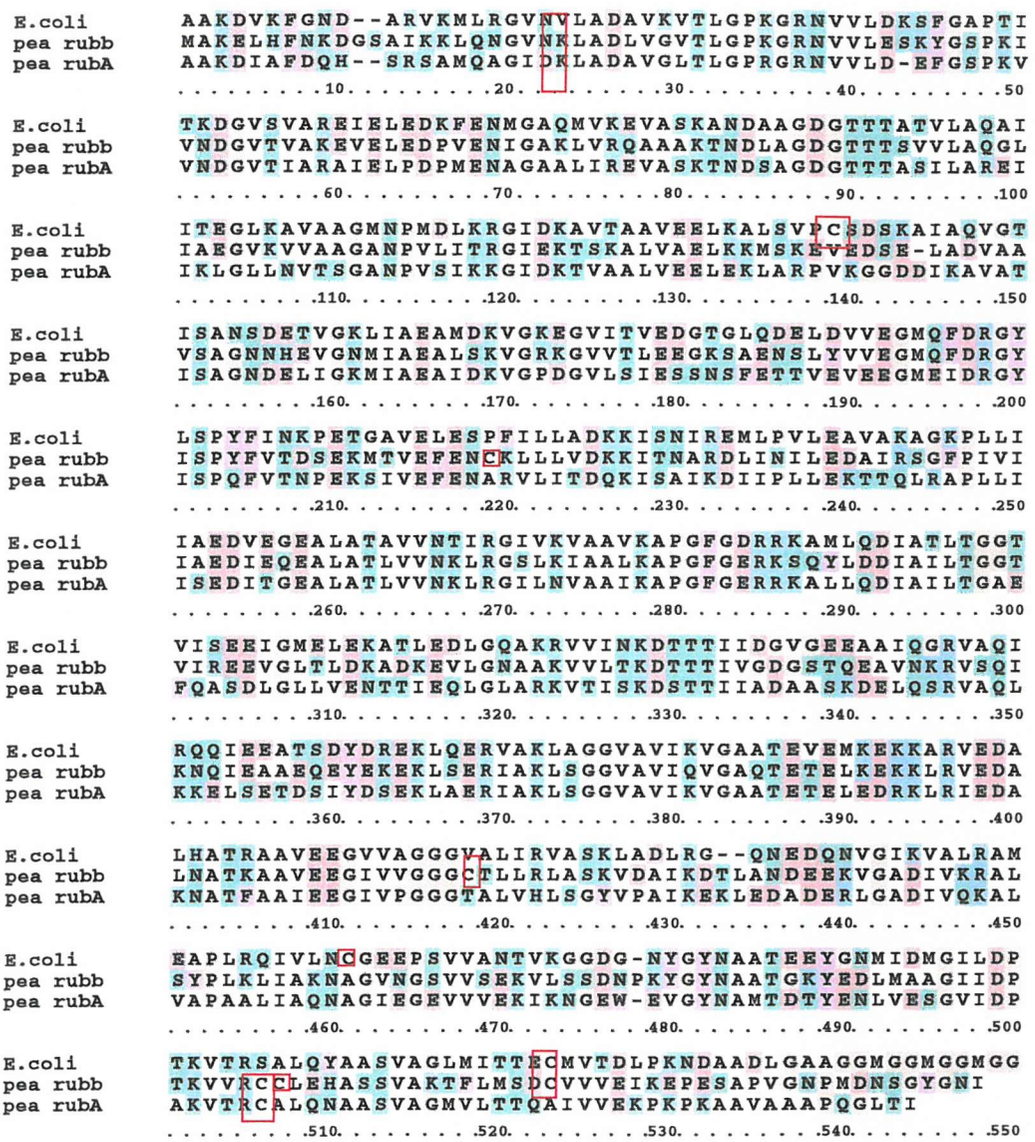


Fig 4.1.1 Sequence Display Alignment for the GroEL, and chloroplast Chaperonin α and β Protomers.

Cysteines of interest are at residues 519 (GroEL and β) and residues 506 (α and β). Cysteines are denoted by the colour: (yellow). Lysines of interest are at residues 22 (α and β). Lysines are denoted by the colour: (blue). Both cysteines and lysines of interest reside in the equatorial domains of all three protomers. (Source: http://www.biochem.ucl.ac.uk/bsm/dbbrowser/CINEMA2.02/index.html). *E.coli*=GroEL , pea rubb=rubisco β subunit, pea rubA=rubisco α subunit.

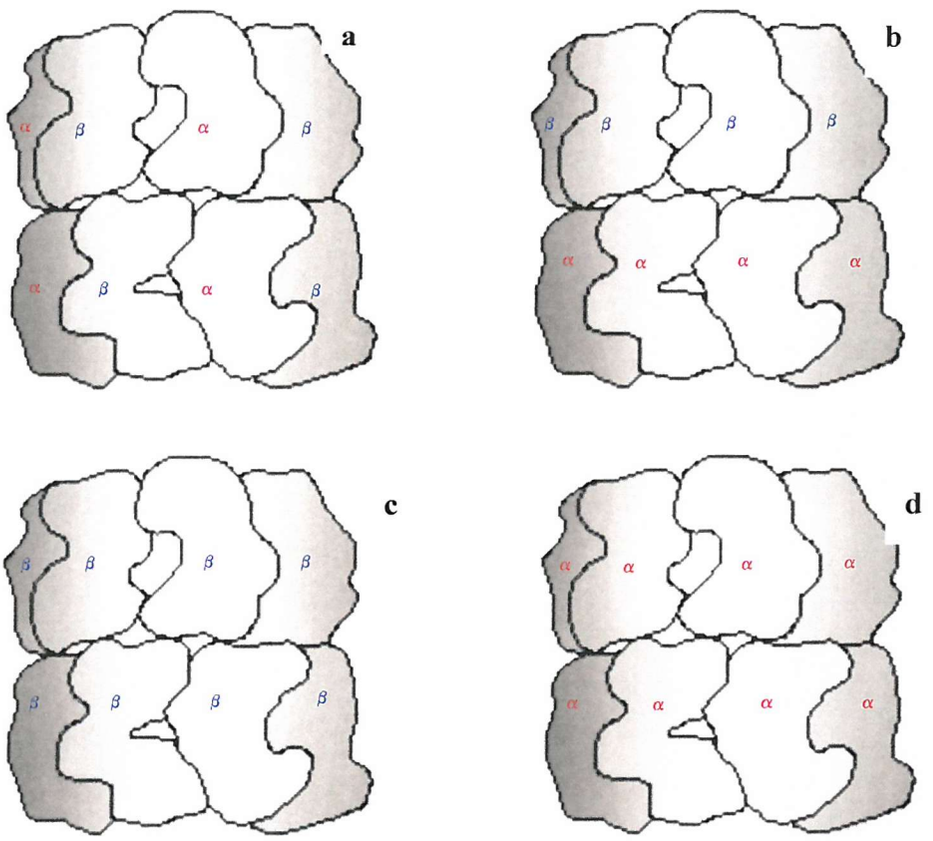


Fig 4.1.2 Possible Subunit Dispersal within the Ch-cpn60 Complex

The tetradecameric complex, a, depicts a possible heterogeneic dispersal of subunits within opposing rings of the complex. b depicts a possible homogeneic dispersal of subunits, again within opposing rings. c and d depicts a possible homogeneic dispersal of β and α subunits, respectively, in both rings of a complex.

4.2 Chemical Crosslinking of the GroE Chaperonin Complexes

Crosslinking of homo-oligomeric proteins with bifunctional reagents has been extensively used to obtain important information about the oligomeric state and symmetry of protein subunits within complexes (Davies and Stark, 1970, Hucho *et al.*, 1975, Sculley *et al.*, 1984, Tsfadia *et al.*, 1990).

Azem *et al.*, (1994), applied these same methods of crosslinking to $GroEL_{14}$ and $GroES_7$ oligomers, and were able to obtain important structural information. Using glutaraldehyde (GA) as the chemical bifunctional reagent they were able to investigate the structure of large oligomeric complexes such as $GroEL_{14}GroES_7$ (the asymmetric complex) and $GroEL_{14}(GroES_7)_2$ (the symmetric complex). In this way, crosslinking, followed by analytical denaturing electrophoresis, confirmed the number and arrangement of GroEL and GroES subunits within each individual oligomer, which confirmed previously known data from cryo-EM analysis (Hendrix 1979, Harris *et al.*, 1994).

Furthermore, crosslinking permitted a close examination of the effect of regulatory factors, such as nucleotides and free divalent cations, on the molecular structure of $GroEL_{14}$and $GroEL_{14}GroES_7$ (Azem *et al.*, 1994, Diamant *et al.*, 1995, Diamant *et al.*, 1995). The use of crosslinking experiments also permitted characterisation and quantitation of various chaperonin hetero-oligomeric complexes in $GroEL_{14}$, $GroEL_{14}GroES_7$, and $GroEL_{14}GroES_7$ in solution, and under conditions that also supported protein folding and ATP hydrolysis (Azem *et al.*, 1995).

4.3 Cysteine Acessibility

Paramount to the viability and success of this experiment would be the accessibility of the available cysteines to the cross-linking reagent of choice. A control protein was required to gauge the structural position and accessibility of any cysteines of the α and β subunits, and apply these comparative data to a known homologous structural model. Due to the high structural (~90%) and sequential (~50%) correlation that exists between the α and β subunits of the plastid ch-cpn60 and its bacterial homologue GroEL, it was decided to carry out cysteine accessibility trials on the GroEL tetradecamer. The amino acid sequence for GroEL revealed that it possessed three cysteine residues (139,461,523). The relevance of the initial observed differences in cysteine numbers became apparent when sequence alignment comparisons, between GroEL and the α and β subunits, were made (Fig 4.1).

4.4 The use of GroEL as a Trial Model

Use of GroEL as the homologous control model, would provide two types of data for analysis. Because of GroEL's likely high structural homology to the ch-cpn60 subunits, a DTNB colourimetric thiol quantitation trial could be used to indicate the likely accessibility of cysteine cross-linking sites in the α and/or β subunits, using GroEL as the trial model. Once cysteine accessibility data had been acquired for GroEL and the α and β subunits, sequence alignment between all three (Fig 4.1) could be used to give an indication of domain location, which could then be interpreted with the GroEL structural model (Braig *et* al., 1994) using the QUANTA modelling package. This would allow an accurate estimate of the location most likely to be a conducive crosslinking site between both the β/β subunits and/or the α/β subunits, dependent upon their dispersal within the complex.

4.5 Crosslinking of the β-14mer adjacent subunits

Before crosslinking adjacent subunits in the 'native' form of the *Pisum sativum* chaperonin Ch-cpn60$^{\alpha/\beta}$, trials were carried out on the β tetradecamer carrying both the site-specific cysteine and lysine residues, expressed from the cDNA form, carried in the pET24a plasmid vector/BL21 (DE3) pLysS strain.

4.6 Methods and Materials

4.6.1 SEC (gel filtration) purification of the β-14mer (post crosslinking)

Prior to the crosslinking trials, purification of the β-14mer was carried out. The buffers used were (20mM K_2HPO_4, pH 7.0, 5mM EDTA, 5mM β-mercaptoethanol). The reason for changing from the previously used buffer, Tris.HCl, to K_2HPO_4 was due to the ability of Tris.HCl to act as a primary amine for the crosslinking reagent used, an unwanted and uncontrollable reaction.

4.6.2 DTNB thiol colourimetric quantification of cysteine reactive residues

The degree of thiol accessibility was assessed using DTNB, Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid), which stoichiometrically yields the chromophore 5-mercapto-2-nitrobenzoic acid (410nm ~13,600 cm^{-1} M^{-1}) upon reaction with a thiol group.
GroEL and Ch-cpn60 (α and β) thiol quantification was performed using the following method:

Added to each 1ml quartz cuvette (Hellma synthetic far-UV quartz, type No: 104B-QG), with 1cm light path and light range 200nm-2500nm, was 900µl NH_4HCO_3, pH8.0, 50µl GroEL (at 112mg/ml) or 100µl α protein (at 1mg/ml) or 100µl β protein (at 1mg/ml), 50µl DTNB (at 10mM). Incubation was carried out for 5 minutes and at the same time the cuvette was placed in the spectrophotometer (Ultrospec 3000), λ412nm, and the absorbance reading for each protein taken.

4.6.3 Heterobifunctional crosslinking of adjacent β subunits

The heterobifunctional reagent SPDP (Succinimidyl 3-(2-pyridyldithio) propionate) was used to bridge the available lysine and cysteine residues. The crosslinking reaction was carried out in accordance with the manufacturers instructions (Sigma), but with the following modifications:

Each crosslinking reaction consisted of 320µg protein (β-14mer~200µM at 2mg/ml), 60µl SPDP (2mM per reaction, from 20mM stock dissolved in 500µl dimethyl sulphoxide, DMSO), buffer (20mM K_2HPO_4, pH7.0, 1mM EDTA) was added to make up a final reaction volume of 1.9ml in a 1.9ml eppendorf tube. Incubation time was determined at 1 hour at R/T (room temperature). The reaction mix was then transferred to a centricon centrifugal filter device (Amicon) YM-100 kDa membrane and centrifuged for ~30 minutes at 4°C in a Sigma (Howe) 3K15 centrifuge (using rotor number 12156) at 1000g, to concentrate the reaction products. The centrifuged retentate was then applied to the Superose 6 HR 10/30 column (Pharmacia), which had been equilibrated with the same buffer used in the crosslinking reaction mix, using the Pharmacia FPLC system. The appropriate samples relating to a crosslinked product, previously identified by SDS-PAGE (4-12%) analysis of all collected peak samples, were collected in 1ml fractions and then applied to the univap system to evaporate all solutions present. 70% TFA (trifluroacetic acid) was then added to the dried pellet, to avoid precipitation of the crosslinked products.

4.6.4 Cyanogen Bromide (CNBr) digest of the β-14mer crosslinked products

For this part of the protocol one large CNBr crystal was added to the crosslinked product (now in 70% TFA) and left o/n to the digest the protein down to peptides. The digest reaction mix was then placed in the univap once again to reduce the digest solution to a pellet form. The pellet was then resuspended in reaction buffer, plus 70% TFA.

4.6.5 Tricine Non-Reduced-PAGE and Western blot sequence preparation of the β_{14} Mer crosslinked product

The electrophoretic transfer technique was used to prepare the SDS-PAGE separated CNBr peptide products for protein sequencing.
NO-Reduced-PAGE (10-20% NuPage) of the CNBr digested peptide products was run in the usual way, but the relevant sample lanes were not stained prior to Western blotting. Blotting was always performed as soon as possible after peptide separation had been performed.

The blotting system used was the total immersion type, consisting of two outer Perspex plates perforated with 1cm diameter holes between which the blot sandwhich was assembled. Assembly was carried out in a tray containing CAPS buffer (10mM CAPS, 10x stock solution 22.1. (Sigma C-2632) pH 11.0). Two sheets of Whatman No.1 filter paper were cut to the size of the plates, as was the PVDF (polyvinyl difluoride) blotting membrane. The PVDF membrane was then thoroughly wetted with 100% methanol, before being placed on the Whatman filters, under CAPS buffer.

The tricine gel was then incubated in some CAPS buffer, for 5 minutes, before being placed on top of the PVDF membrane. Two extra sheets of Whatman filter paper were then placed onto the gel, before placing the final Perspex plate over the assembled system. The complete system was then slotted into the blotter, which was filled with CAPS buffer.

Blotting was carried out for 60 minutes at 200mA (constant). During blotting the buffer was cooled by passing it through a coil of tubing, buried in a bucket of ice. Under these conditions proteins with molecular weights up to 100kDa completely migrate out of the gel and onto the PVDF membrane.

After blotting, the membrane was withdrawn from the perspex packaged system and immersed in amido black stain (prepared in a 200ml volume as 0.1% (w/v) solution of Naphthol Blue Black (Sigma N-3393) in 10% methanol, 2% (v/v) acetic acid) for 10 minutes, whilst being continually agitated. The membrane was then transferred to destain and agitated for a further 20 minutes, with frequent changes of the destaining solution.

Bands of interest were then excised, using a clean scalpel and vortexed in a 1.9ml eppendorf tube containing deionised water, to remove any soluble impurities prior to sequencing.

Sequencing of the excised band of interest was carried elsewhere. Therefore, only the data from the protein sequencing will be mentioned.

4.6.6 Mass Spectroscopy Analysis

Samples were first desalted, using R2 and R3 reverse phase resin. 2µl of R2/R3 resin was applied to a micro bore needle. The resin was subsequently washed with 2x 4µl dH_2O. A 4µl sample was applied to the resin and washed with 3x dH_2O. The sample was eluted off the resin and into an electro-spray coated needle with 45% acetonitrile: 5% formic acid: 55% dH_2O. Between each desalting phase the sample was spun down using a bench top centrifuge.

Samples were then applied to the Time-of Flight Electro-spray Mass spectrometer by Neville Wright (Southampton University).

4.7 Cysteine Accessibility

Cysteine accessibility would be determined by the use of a common reagent for colourimetric thiol quantitation, DTNB(5,5-dithiobis-(2-nitrobenzoic acid) or Ellmans reagent (Ellman, 1959). The degree of thiolation and presence of residual thiols in a solution can be assessed using DTNB, which stoichiometrically yields the chromophore 5-mercapto-2-nitrobenzoic acid upon reaction with a thiol group (Σ_{410nm}–13600$cm^{-1}M^{-1}$).

4.8 SPDP Crosslinking Reagent

A number of methods presented themselves for the modification of functional groups in proteins. Many of them involved bifunctional reagents, i.e. reagents with two reactive groups that are capable of reacting with, and forming bridges between, the side chains of certain amino acids in the proteins. The cross-links or bridges may be either of the intra- or intermolecular type. The control of intra- versus intermolecular cross-linking is very difficult to achieve with homobifunctional reagents, where the two reactive groups are identical (Carlsson *et al.*, 1978). However, with a heterobifunctional reagent such as SPDP (N-Succinimidyl 3-(2-pyridyldithio) propionate), where the two reactive groups are directed toward different functional groups, one can usually conduct the coupling and the cross-linking in controllable sequential steps.

SPDP (Fig 4.8.1) contains one N-hydroxysuccinimide (NHS) ester moiety and one 2-pyridyl disulphide moiety. The hydroxysuccinimide ester reacts with primary amino groups which produce stable amide bonds, and the 2-pyridyl disulphide group reacts with aliphatic thiols to form aliphatic disulphides.

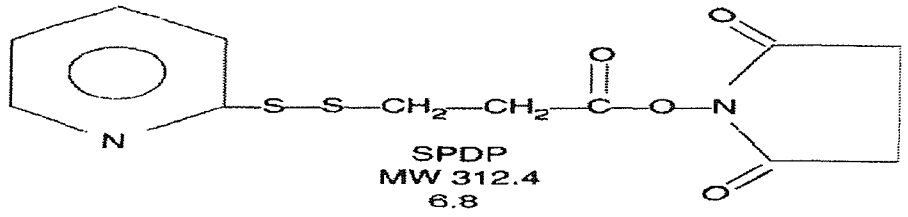


Fig 4.8.1 N-Succinimidyl 3-(2-pyridyldithio) propionate (SPDP).

During protein thiolation with SPDP, a 2-pyridyl disulphide group is introduced when any one of these reagents react with the primary amine group of the protein (Fig 4.8.2, Reaction 1). When conjugating an amine-containing protein with a sulfhydryl-containing protein using SPDP, the amine-containing protein is first conjugated to one of these cross-linkers via the NHS-amide group (Fig 4.8.2, Reaction 2a).

Both of these conditions proceed rapidly under very mild conditions, in aqueous media. It has also been observed that the N-hydroxysuccinimide ester moiety may also react with the thiol groups (Carlsson *et al.*, 1978), but much more slowly.

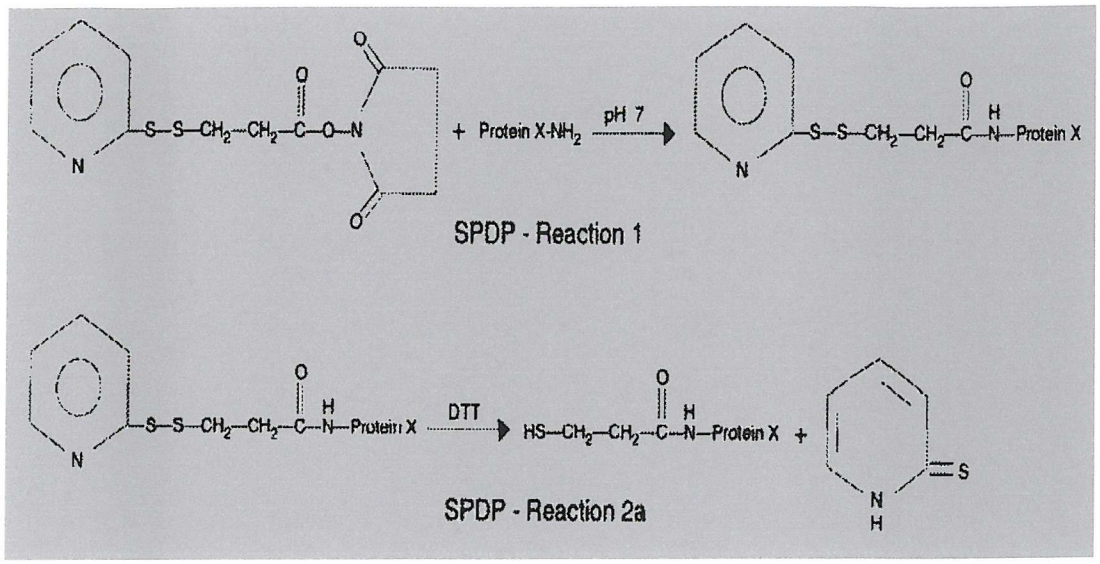


Fig 4.8.2 Reactivity of the Pyridyl Disulfide Moiety of SPDP

In addition to the appropriateness of chemical reactivity, SPDP also fulfilled the equally important criteria of size and length. The length of SPDP (6.8 Å) would ensure a greater control between the requisite intra-molecular crosslinked product and the unwanted inter-molecular products. SPDP was of the appropriate size to enter the 'windows', or apical domain-defined cavity entrance, and effect a reaction on the internal faces of the subunits of the full tetradecamer.

However, the use of SPDP as a cross-linking reagent comes at a price. The reagent reaction with the primary amines present in the protein, i.e. the lysines, would have the effect of 'blocking' the amino-terminals of any subunits involved in the cross-linked reaction. Therefore, it would prove further necessary to carry out a cyanogen bromide (CNBr) protein digest of the cross-linked products, freeing up the N-terminals, which would then allow the N-terminal sequencing of the dimerised product to be carried out, allowing elucidation of which two subunits had been cross-linked; whether they be αβ or αα/ββ (Fig 4.1.1).

4.9 Purification of the Ch-cpn60 β Subunit Protein

Due to the low quantities of 'native' Ch-cpn60α/β stocks available, all preliminary crosslinking experiments were carried out using the abundant plasmid stocks of the recombinant α and β subunit proteins, carried in the pET24a plasmid vector with the *E.coli* host BL21 (DE3)pLysS strain. Through this approach it would be possible to safely optimise the necessary conditions required for the successful completion of future crosslinking experiments involving the 'native' Ch-cpn60 complex.

Before the start of crosslinking trials, a sufficient supply of purified β_{14}mer was required. A major advantage in using the β_{14}mer complex was its known ability to form a 'stand alone' complex, as apart from the α subunit which could not form a complex unless in the presence of the β subunit (Dickson, *et al*; 2000). Further, initial analysis showed the β subunit to possess both of the residues required to acquire the desired crosslinking product.

Stocks of the Ch-cpn60 β_{14}mer were stored in partially purified form, having been applied to an ion exchange column. Further purification was carried out by size exclusion chromatography (SEC). Relevant peak fractions were pooled and then concentrated in a mini-centricon (AMICON) with a filter membrane cut-off of 100kDa (Mw of β_{14} ~800kDa). Analysis of the pooled fractions was carried out on a 4-12% SDS-PAGE (Novex). Fig 4.9.1 shows the gel filtration chromatogram representing the purification of the β_{14}mer and the analytical SDS-PAGE of the pooled fractions from the gel filtration protocol.

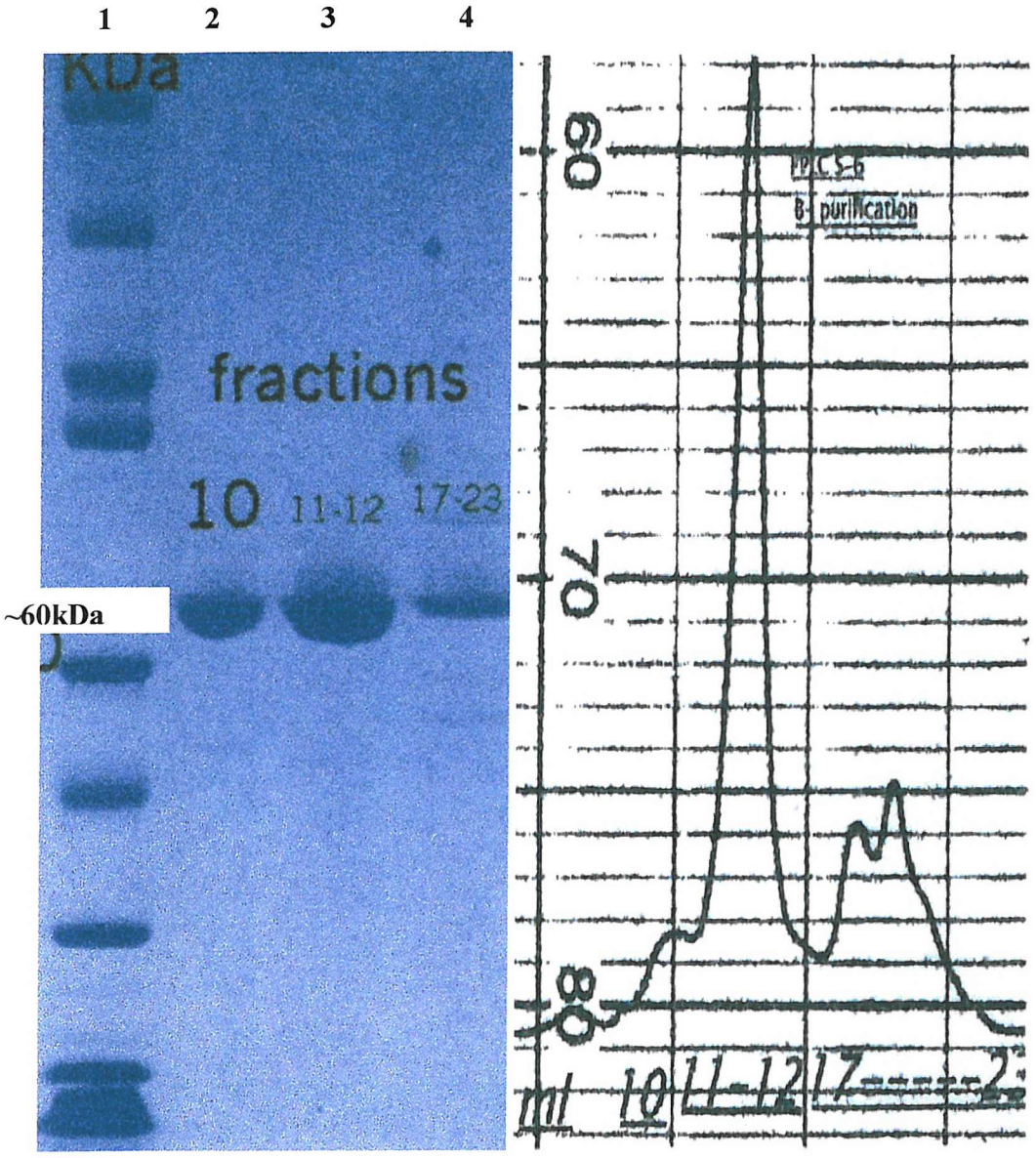


Fig 4.9.1 β_{14}mer Purification and Analysis

The Superose-6 gel filtration elution profile represents fractions collected in 1ml volumes. The purest form of β_{14}mer was at ~12ml, confirmed by 4-12% SDS-PAGE analysis. It was the pooled fractions at 11-12ml that were used for subsequent crosslinking experiments.

RESULTS

4.10 DTNB Analysis and the Ch-cpn60 Putative Crosslinking Sites

The DTNB data showed that all three cysteines in the GroEL protomer were accessible to a thiol reactive reagent. Data for the β subunit showed that only one cysteine was accessible and data for the α subunit showed no accessibility for the DTNB reagent. Using the sequence alignment map (Fig 4.1) it became evident that the accessible β subunit cysteine was sited at residue 519 (β Cys-519), as it was the only cysteine possessing homologous alignment with a reactive GroEL cysteine. Similarly, the only α subunit cysteine at residue 506 did not align with any GroEL cysteines but did align with the β cysteine at residue 506, which data showed not to be accessible to the reagent, allowing for the assumption that this particular cysteine was buried within the structural core of the subunit.

Because only the β subunit of the Ch-cpn60 complex possessed an accessible cysteine, this made the matter of 'nearest-neighbour' elucidation much simpler. It could now be assured that only subunits adjacent to a β subunit could be effectively crosslinked via the conjugated SPDP reagent.

Use of the GroEL homologue model would be used to find an appropriate crosslinking residue for the β Cys-519 residue. It was decided to link the β cysteine sulphydryl residue with a primary amine, in accordance with the reactive properties of the selected reagent SPDP. The GroEL model showed that at residue 22 a lysine was sterically available, in both the α and β subunits. That both subunits possessed a reactive lysine at residue 22 (equatorial domain) meant that no matter what the identity of the adjacent subunit was to the β subunit possessing the reactive cysteine at residue 519 it would possess a reactively available lysine at residue 22. Further analysis of the GroEL homologue showed that the distance between a β Cys-519 residue of one β subunit and an adjacent α/β lysine at residue 22 was ~5.2Å. The reactive distance of the SPDP reagent was 6.8Å.

The reactive properties between any given two subunits and the reagent were such that an optimum of control could be exerted over inter- versus intra-molecular crosslinking. It was an essential element of the experiment that intermolecular crosslinking between spatially separate complexes be avoided. The reactive distances of reagent and residues, combined with the location of the available reactive residues (both β Cys-519 and α/β Lys-22 are buried deep within the equatorial domain and close to their respective N-terminals, which curl up inside the complex and are not easily accessible to the outside), suggested a preferred reaction would occur between subunits rather than between complexes. Fig 4.10.1 depicts the Cα-trace of the GroEL protomer, showing the relevant β Cys-519 and α/β Lys-22 residues. It also depicts a close-up view of adjacent subunits highlighting the relative proximity of those residues.

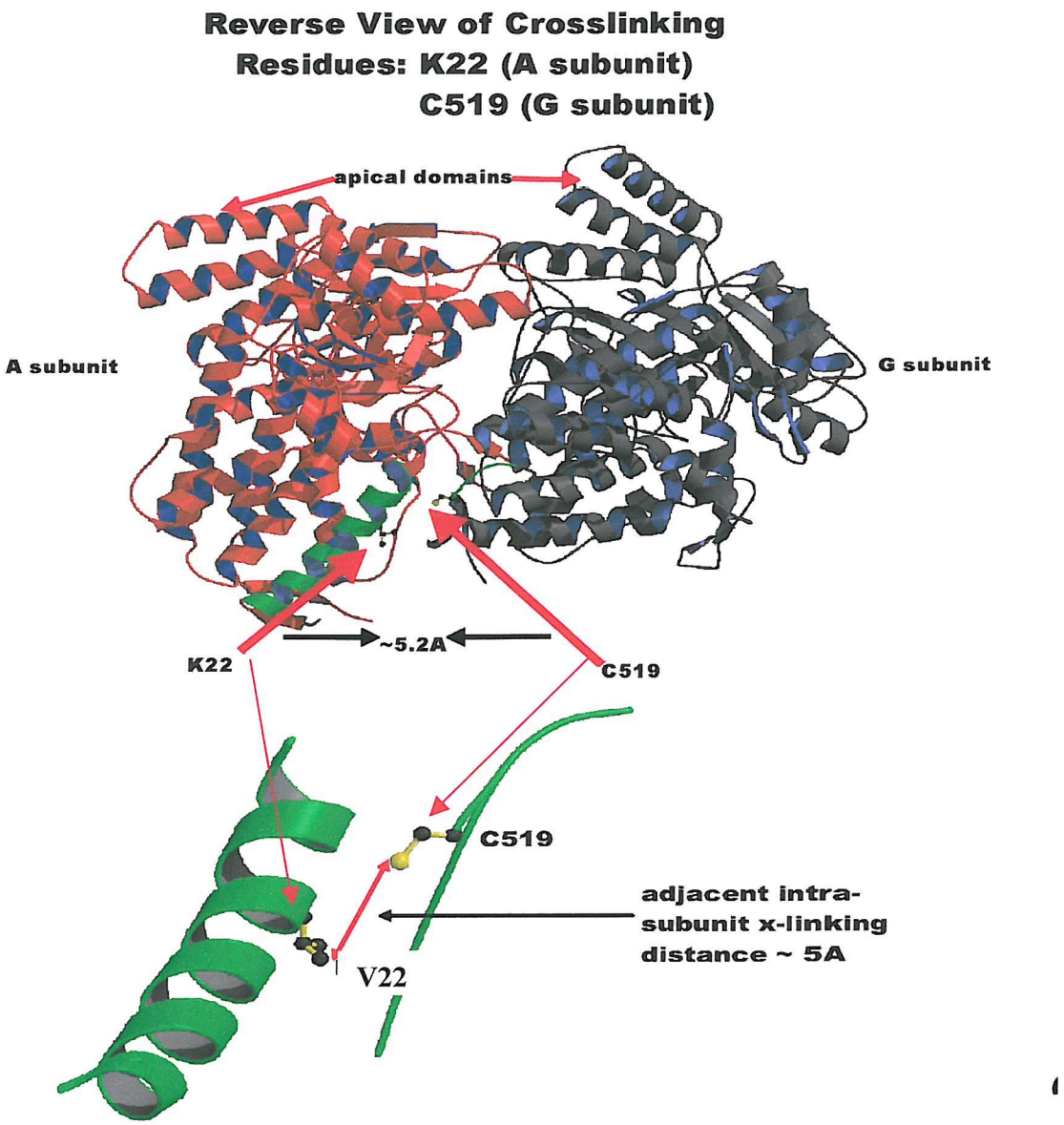


Fig 4.10.1 Cα-Trace of GroEL Protomer and Adjacent Subunit N-Terminals

Highlighted are the relevant cysteine and lysine residues expected to be the preferred crosslinking targets for likely conjugation with the SPDP reagent. Crosslinking occurring between β Cys-519 and Lys-22 residues (represented here by the GroEL Valine-22 (V22) residue) of the same β subunit was predicted to be structurally untenable, due to the distance involved. The picture (drawn using the MOLSCRIPT program) depicted on the left show the positions of the reagent-reactive residues within adjacent subunits. Subunit 1 depicts the position of the β Cys-519 residue and Subunit 2 depicts the position of an adjacent α/β subunit carrying the putative reactive lysine 22 residue.

4.10.1 Crosslinking Optimisation trials using β_{14}mer Complex and SPDP Reagent

The DTNB thiol quantitation experiments suggested that the β Cys-519 residue would most likely react with the SPDP reagent and initiate one half of the proposed heterobifunctional reaction. Structural modelling of the GroEL homologue suggested that lysine 22, resident in both the α and β subunits was the residue thought likely to complete the conjugate crosslinking reaction. In the initial crosslinking experiment the β_{14}mer complex was used to acquire fundamental analytical evidence before attempting any crosslinking experiments with the low stocks available, of the Ch-cpn60 'native' complex.

Several experimental factors required careful consideration before the crosslinking trials could proceed. Structural investigations of the possible crosslinking sites for adjacent subunits suggested that SPDP, with a reactive length of 6.8Å was sufficient to bridge the approximated distance between the relevant cysteine and lysine residues of adjacent subunits. Due to the small possibility of an intermolecular crosslinking event occurring, consideration was given to the optimum concentration ratios between β_{14}mer and the SPDP reagent, keeping in mind that sufficient protein concentration was required for subsequent analytical protocols.

4.10.2 Crosslinking of the α Subunit using the SPDP Reagent

To verify the negative data acquired from the DTNB experiments and the structural evidence gained from the GroEL structural model, suggesting that the α subunits were inaccessible to a thiol reactive reagent, an initial crosslinking experiment was carried out using the α subunit protein.

Due to the non-complex monomeric state of the α subunit there remained a possibility for the occurrence of anomalous intermolecular crosslinking between the α Cys-506 residue (possibly exposed in monomeric form) and accessible primary amines presented by other α subunits within conjugative distance of the SPDP reagent (~6.8Δ). The data of the trial showed no acquisition of a α/α crosslinking product having been formed, and subsequently acquired. The evidence of this negative result gave further assurances that future crosslinking experimental data, involving the β_{14}mer complex and the 'native' Ch-cpn60 form, would be representative of adjacent subunit dispersal occurring within a heptameric ring of the complex form.

4.10.3 SPDP crosslinking of the β_{14}Mer Complex Protein

Fig 4.10.3(a) depicts the elution profile of the reacted β_{14} product, run on the S-6 gel filtration column. Before gel filtration the reaction product was concentrated down from a reaction volume of 2ml to ~200µl in a Centricon-100kDa filter device (AMICON). After gel filtration selected fractions were pooled, in accordance with peak association, and once again concentrated in the Centricon-100kDa filter. These samples were then loaded and run on a 4-12% analytical SDS-PAGE (Nu-PAGE). Fig 4.10.3(b) depicts the gel which represents the peaks in order of elution off the S-6 filtration column. The gel presents evidence that 'intact' but probably tagged β_{14} complex eluted as a front shoulder to a main peak and at its known elution volume (Ve) of 12-13ml.

The conjoined peak (13-14ml) is thought to represent a reacted product with a mass approximating ~120kDa on the gel, which suggests a mass equivalent to the size associated with

a dimer of two conjoined subunits. Although contaminated by the presence of the front-shouldered peak (representing un-reacted β_{14}mer) it would be these fractions of reacted product that would be taken through subsequent protocols in preparation of N-terminal sequencing identification. The final set of pooled fractions, relating to peaks 17-18ml and 12-23ml showed nothing on the SDS-PAGE analysis and it was concluded that these peaks represented various salts used in the reaction as well as the SPDP reagent.

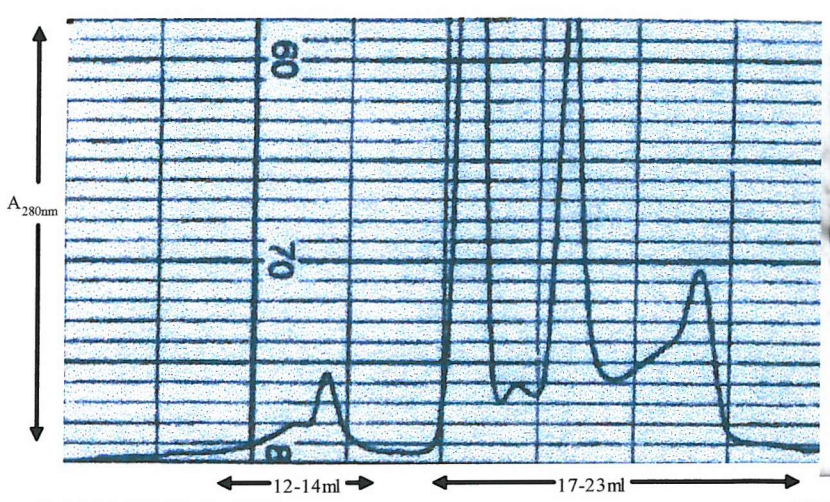


Fig 4.10.3(a)

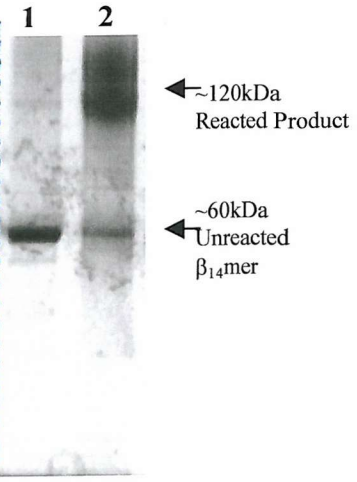


Fig 4.10.3(b)

Figs 4.10.3(a) and 4.10.3(b) β_{14} Crosslinked Elution Profile and Gel Analysis

Fig 4.10.3(a) depicts the elution profile obtained on gel filtration of 600µg of SPDP crosslinked β-cpn60 on Superose-6 (10x300mm) using 50mM phosphate buffer pH7.5. Intact β_{14}mer eluted at fractions 12ml in control experiments. Fig 4.10.3(b) depicts a 4-12% SDS-PAGE analysis of the pooled fractions. Lane 1: shows the 12ml fraction and represents intact β_{14}mer at ~60kDa. Lane 2: represents pooled fractions13-14ml and shows an amount of unreacted product at ~60kDa and an excess of crosslinked product at ~120kDa.

4.10.4 Cyanogen Bromide (CNBr) Peptide Fragmentation

Prior to N-terminal sequencing of the β_{14}mer crosslinked product, which would determine the sequence-related sites of the reacted product, removal of the N-terminal blockage acquired as a product of the SPDP reagents primary amine reactivity, was required. CNBr was selected as the best fragmentation agent for its residue cleavage sites and the size of the peptide fragments created by the cleavage.

CNBr is a chemical reagent commonly used for protein fragmentation on the C-terminal side of methionine residues, which produces a modification to the methionine into HomoSerine Lactone (HSL). The CNBr digest would produce an array of peptide fragments, one of which would be representative of the SPDP crosslinked product. The sequence of the Ch-cpn60 β subunit was used to create a CNBr fragment digest map (Fig 4.10.4) and from this map it was possible to determine a relative mass for a crosslinked product between the two respective residues (β cys-519 and β lys-22) of adjacent β subunits.

CNBr digestion of the selected S-6 filtration fraction (13-14ml) was then carried out. The reacted product was put into a 'univap' machine to evaporate off residual water. The product was then rehydrated in 70% Trifluoracetic Acid (TFA) producing a solvent environment which would avoid a previously observed precipitation event that occurred with the reacted product, in aqueous media. The peptide products produced by the CNBr digest were analysed on a 10-20% non-reducing tricine peptide gel (Nu-PAGE). Fig 4.10.5 depicts the result of the CNBr digest analysis. The gel shows very little product at the expected range of ~16kDa (see fig 4.10.4 for expected peptide sizes representative of crosslinked product), this could be attributed to the low starting concentration of protein as a necessary avoidance tactic against a spurious intermolecular crosslinking event, and a major reagent-induced disaggregation of 14-mer. However, the gel clearly shows two bands one at ~30kDa and the other at ~16kDa, both represented on the CNBr digest map.

mass	position	#MC	artif.modification(s)	peptide sequence
30086.253	260-540	0	HSL 540 30086.253	TVEFENCKLLLVDKKITNAR DLINILEDAIRSGFPIVIIA EDIEQEALATLVVNKLRGSL KIAALKAPGFGERKSQYLDD IAILTGGTVIREEVGLTLDK ADKEVLGNAAKVVLTKDTTT IVGDGSTQEAVNKRVSQIKN QIEAAEQEYEKEKLSERIAK LSGGVAVIQVGAQTETELKE KKLRVEDALNATKAAVEEGI VVGGGCTLLRLASKVDAIKD TLANDEEKVGADIVKRALSY PLKLIAKNAGVNGSVVSEKV LSSDNPKYGYNAATGKYEDL M
13826.692	50-183	0	HSL 183 13826.692	AKELHFNKDGSAIKKLQNGV NKLADLVGVTLGPKGRNVVL ESKYGSPKIVNDGVTVAKEV ELEDPVENIGAKLVRQAAAK TNDLAGDGTTTSVVLAQGLI AEGVKVVAAGANPVLITRGI EKTSKALVAELKKM
3315.779	210-241	0	HSL 241 3315.779	IAEALSKVGRKGVVTLEEGK SAENSLYVVEGM
2812.485	541-567	0	HSL 567 2812.485	AAGIIDPTKVVRCCLEHASS VAKTFLM
2624.207	184-209	0	HSL 209 2624.207	SKEVEDSELADVAAVSAGNN HEVGNM
2135.008	242-259	0	HSL 259 2135.008	QFDRGYISPYFVTDSEKM
2051.995	568-587	0	HSL 587 2051.995	SDCVVVEIKEPESAPVGNPM
839.353	588-595	0		DNSGYGNI

~ 16kDa

Fig 4.10.4 CNBr Digest Fragment Map of the β_{14}Mer Subunit Protein

The cleavage map depicts the expected combined peptide masses for an SPDP crosslinked product representing the β cys-519 and β lys-22 residues. The combined mass of the two peptides was expected to be ~16kDa, which would be evidence that a crosslinked product had occurred between adjacent subunits of a heptameric ring within the β complex. (Source: http://expasy.cbr.nrc.ca/cgi-bin/peptide-mass.pl)

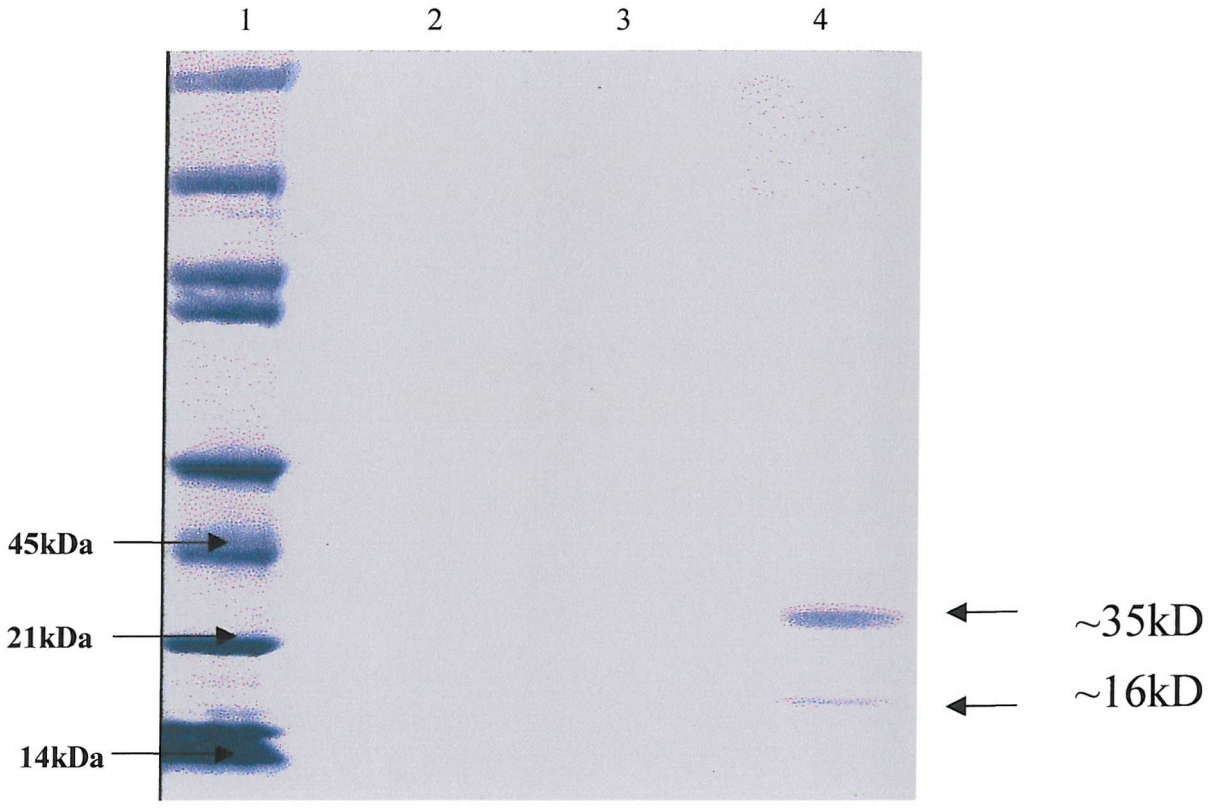


Fig 4.10.5 Non-reduced 10-20% Tricine Gel of the CNBr Digest for β_{14}

Lane 1: Mw marker, highlighted are the pertinent masses (45 & 21kDa). Lane 2 was left blank to avoid artefact interference with Lane 3, which contained unstained product in preparation for electroelution onto a PVDF membrane which would be used for N-terminal sequencing. Lane 4: depicts the product obtained from the CNBr digest and is one half of the product applied to lane 3. The peptide product at ~16kDa was representative of the mass range for the expected crosslinked product between the Cys-519 and Lys-22 residues of adjacent subunits.

4.10.5 Electroelution Transfer of the β_{14}Mer Crosslinked Product

The 'Western' transfer blot techniques was used to transfer the CNBr peptide fragment (~16kDa) from the tricine gel to a PolyVinyl DiFluoride (PVDF) membrane support in readiness for N-terminal protein sequencing. Amido black staining of the PVDF membrane was used to determine the presence and position of the respective peptide fragment at ~16kDa. The staining was too faint to depict a representative gel showing the presence of the peptide. However, it was discernible by eye and the respective band was excised from the membrane by scalpel and prepared for sequencing by vortexing the sample band in a 2ml eppendorf tube which contained deionised water, which removed any impurities.

Sequencing of the excised band of interest was done outside of this laboratory. Therefore only the data returned from the sequencing could be analysed and so discussed.

N-Terminal Sequencing Data

Predicted N-term sequence: AKELHFNK

Actual N-term sequence: AVEL –FNG

The predicted N-terminal sequence was taken from the CNBr digest map (Fig 4.10.4). This predicted sequence is representative of the N-terminal for the lysine 22 residue. The C-terminal end of the crosslinked product was expected to contain the cysteine product of the adjacent subunit, conjugated to the lysine via a disulphide bridge created by the SPDP reagent. The 'actual' sequence returned was for the first eight N-terminal residues, as requested. Comparisons between the 'predicted' and the 'actual' sequences were sufficiently close enough to suggest supportive evidence for the successful crosslinking of two adjacent subunits at the predicted residue sites of β Cys-519 and β Lys-22.

4.11 General Conclusions

Concerns about controlling the actual site of the crosslinking reaction were partially allayed with the results gained for the α subunit trials. The evidence suggested it was unlikely that the SPDP reagent could reach and therefore react with the α subunit cysteine at residue 506. This evidence agreed with the data gained from the DTNB experiments, which suggested that the α Cys-506 residue was inaccessible, which was further supported by the structural information gained from the GroEL homologue model. All of this evidence suggested that a good amount of identification and control could be exerted over the actual site of reaction.

Titration with DTNB is a standard method for thiol quantitation and was reliably used to demonstrate the possible accessibility of the cysteines possessed by both subunits. The DTNB data gained from the GroEL homologue and α and β subunits was sufficient evidence to predict that only the β cysteine 519 residue would be accessible to the SPDP reagent. This evidence, in conjunction with the structural model of the GroEL homologue, suggested an appropriate choice of reagent as well as a reactive partner for the cysteine residue of the β subunit.

The SPDP crosslinking experiments with the β_{14}mer highlighted areas both of interest and concern. The S-6 filtration chromatography was used to gain separation between reacted and non-reacted products.

Use of the CNBr digest protocol gave two distinct advantages. It would free up the N-terminal in preparation for the N-terminal sequencing of the crosslinked product, which allow sequential analysis of this product. The CNBr digest would also produce peptide fragments that would make for easier distinction of the expected crosslinked product (~16kDa). Analysis of the CNBr peptide gel showed two bands to be dominant over the remaining peptides and it was the band at 16kDa that provided a direct correlation to the mass predicted by the CNBr digest map as representing the crosslinking product required.

It can be strongly suggested, that the 16kDa peptide could well be representative of a crosslinked product that had occurred between β cys-519 of one subunit and β lys-22 of an adjacent subunit.

It is also reasonable to suggest that the opportunities afforded to the β cys-519 residue to react with any lysine other than the lysine at residue 22 would have been very low given the structural location of the cysteine. This observation was made with the use of the GroEL structural model which suggested that the location of the cysteine residue resided deep within the equatorial domain and close to the N-terminal, which itself forms within the hydrophobic 'cavity' of the complex and along side those of the adjacent subunit N-terminals. Furthermore, the reactive length of the SPDP reagent (6.8Å) suggests it would be highly improbable that sufficient reagent length was available for the reagents primary amine reactive tail to extrude from the cavity and react with a lysine of an opposing intermolecular complex.

The N-terminal sequencing of the selected crosslinked peptide product identifies only the first eight residues of the 16kDa peptide, which is to say that only the lysine 22 residue of the SPDP conjugated product is confirmed. It is the combined mass of the expected crosslinked product, predicted by the CNBr cleavage map, alluding to a successful linkage having occurred between the lysine 22 and cysteine 519 of adjacent subunits. It is the very fact that the lysine 22 residues were identified by N-terminal sequencing, and the ~16kDa mass of the peptide was comparable of the predicted mass, that is indicative of a controlled and predictive site-specific crosslinking event having occurred.

Therefore, evidence of the required crosslinked product is reliant upon the combined evidence of both the mass of the acquired peptide and the N-terminal sequencing. The ability to distinguish between the two possible crosslinked products, α/β, α/α, β/β, of adjacent subunits is paramount to the eventual success of these experiments. Therefore, a CNBr digest map of the α subunit (Fig 4.11.1) was made in order to predict the mass of a crosslinked linkage occurring between the α lysine 22 residue of one subunit and the β cysteine 519 residue of an adjacent β subunit. A correlation was drawn between the two CNBr maps and from this a predictive digest map (Fig 4.11.2) was made to depict possible crosslinked masses for α/β or β/β products.
There was no necessity to predict α/α products as this event could not be monitored, due the inaccessibility of the α subunits cysteine. If a β/β crosslinked product was the predominant species this event would be indicative evidence of homologous subunit dispersal, certainly within a heptameric ring of a given complex. This evidence could then be attributed to the α

subunit, again suggesting homologous subunit dispersal. The collective evidence of these experiments, coupled to the sequencing analysis, strongly suggests that the predicted residues cysteine 519 and lysine 22 of adjacent subunits were successfully crosslinked.

mass	position	#MC	artif.modification(s)		peptide sequence
31039.841	192-481	0	HSL: 481	31039.841	EIDRGYISPQFVTNPEKSIV EFENARVLITDQKISAIKDI IPLLEKTTQLRAPLLIISED ITGEALATLVVNKLRGILNV AAIKAPGFGERRKALLQDIA ILTGAEFQASDLGLLVENTT IEQLGLARKVTISKDSTTII ADAASKDELQSRVAQLKKEL SETDSIYDSEKLAERIAKLS GGVAVIKVGAATETELEDRK LRIEDAKNATFAAIEEGIVP GGGTALVHLSGYVPAIKEKL EDADERLGADIVQKALVAPA ALIAQNAGIEGEVVVEKIKN GEWEVGYNAM
9604.245	65-159	0	HSL: 159	9604.245	ENAGAALIREVASKTNDSAG DGTTTASILAREIIKLGLLN VTSGANPVSIKKGIDKTVAA LVEELEKLARPVKGGDDIKA VATISAGNDELIGKM
5036.717	16-64	0	HSL: 64	5036.717	QAGIIK LADAVGLTLGPRGR NVVLDEFGSPKVVNDGVTIA RAIELPDPM
3304.595	160-191	0	HSL: 191	3304.595	IAEAIDKVGPDGVLSIESSN SFETTVEVEEGM
3275.621	482-513	0	HSL: 513	3275.621	TDTYENLVESGVIDPAKVTR CALQNAASVAGM
2714.618	514-540	0			VLTTQAIVVEKPKPKAAVAA APQGLTI
1599.787	1-15	0	HSL: 15	1599.787	AAKDIAFDQHSRSAM

Fig 4.11.1 CNBr Digest Map for the α Subunit of the Ch-cpn60

Highlighted is the lysine 22 residue of the α subunit. The predicted peptide mass for this lysine 22 residue is ~5kDA. The predicted mass for the β Cys-519 peptide is ~2kDa, making a combined mass for both peptides of ~7kDa, if they were to occur as a crosslinked product. (Source: http://expasy.cbr.nrc.ca/cgi-bin/peptide-mass.pl)

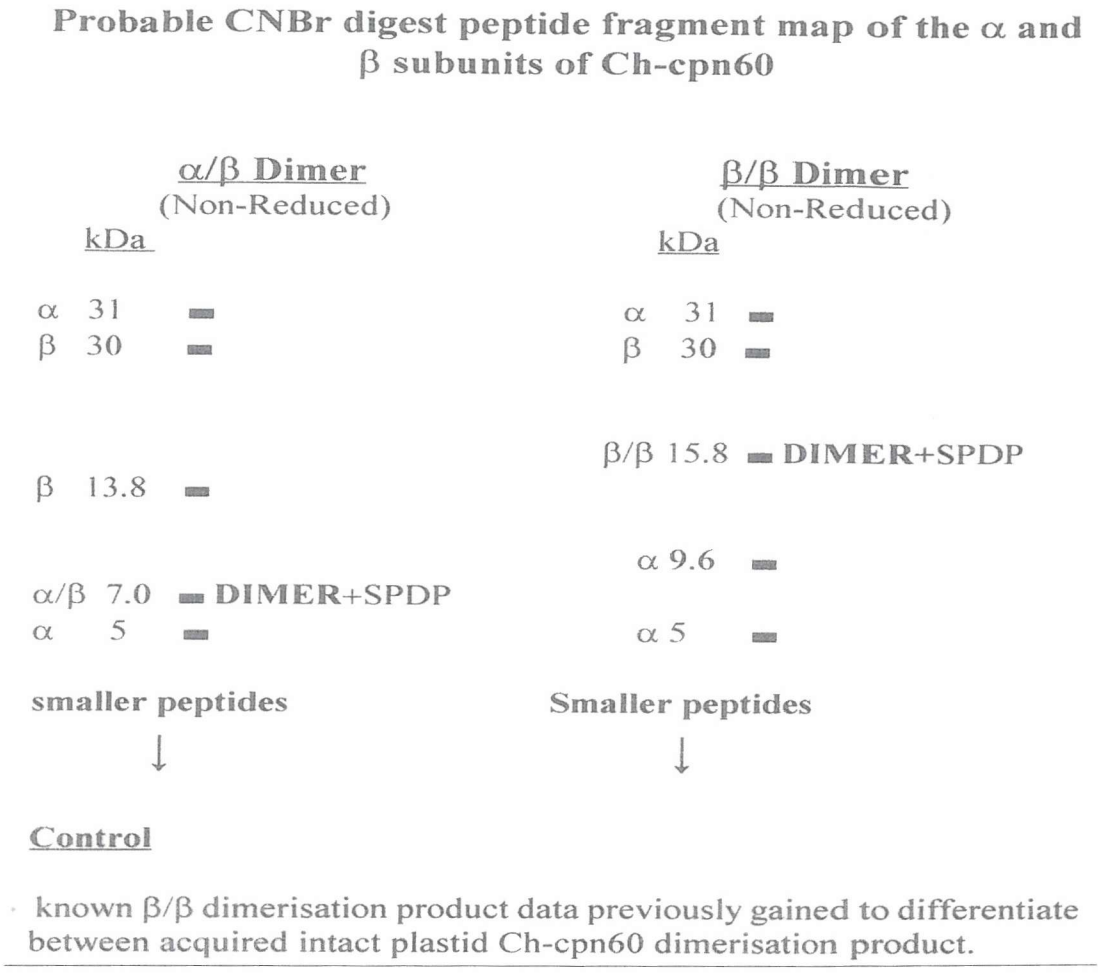


Fig 4.11.2 Representative Map Depicting Possible Crosslinked Product Masses

Highlighted are the possible dimerised products for a crosslinking reaction occurring between either the α/β subunits (~7kDa) or the β/β subunits (~16kDa), of adjacent subunits. The contrast in mass differences between the two possible products, which is greater than a two-fold, is sufficient enough to be suitable for identifying the origin of the lysines contained within the crosslinked product.

Finally, it is essential for one to understand not only the benefits to be gained from these experiments, but also the complexities that are ranged within them. Once the data for the identity of the adjacent subunits has been acquired it can only be ajudged as representative of which two subunits are adjacent to each other within one ring of the complex. Simply, this is because once a lysine 22 residue of one subunit (whether it be an α or β Lys22) has been reacted with the β-cysteine 519 of an adjacent subunit, within a heptameric complex, those subunits that were reacted would possess no residues with which a further crosslinking reaction couls occur. This would further suggest that the expected dominant species from a given set of crosslinking reactions is most likely found to be in a dimeric format. This evidence will afford an opportunity to determine whether the α and β subunits of the 'native' plastid Ch-cpn60 are dispersed throughout the complex in either a homo- or hetero-geneic format. Should a α/β subunit dispersal be the dominant crosslinked species then it could confidently be suggested that the α subunit is a necessary and integral part of the complex. If a β/β crosslinked species is dominant then it could suggest two alternatives to subunit dispersal, either there is a homogeneic dispersal of subunits within opposing rings of the same complex or that the α subunit may well exist as a separate entity to that of a β complex. Although these experiments offer some way forward in elucidating the dispersal of the Ch-cpn60's heterogeneic subunits, there would be further questions requiring an answer before subunit dispersal could be stated with any amount of certainty. However, this experimental design offers a way forward to resolving at least one of the fundamental questions with regard to the unknown subunit dispersal, ranged within theactive form of the 'native' plastic cpn60 complex. That is whether the 'native' complex is comprised of both divergent subunits (α and β) or if they exist as separate entities, whilst residing within the same space and possessing an inextricable dichotomy that requires the presence of both forms so as to ensure a symbiosis that imparts a full and active mechanism.

The next set of intended experiments will incorporate crosslinking a complex compiled of both the α and β subunits and the data gained from this set of experiments will be a precursor to carrying out crosslinking experiments with the 'native' Ch-cpn60 complex.

Chapter FIVE

Crystallisation and Structure Determination for the Ch-cpn60 α Subunit Protein

5.1 Introduction

Although evidence exists for the assembly (Cloney *et al.*, 1992) and functional characteristics (Dickson *et al.*, 2000, Viitanen *et al.*, 1995) of the chloroplast Ch-cpn60 protein from the *Pisum sativum* pea plant, the structural organization and dispersal of the divergent α and β subunits have yet to be fully elucidated. Both the Viitanen and Pushkin groups (Viitanen *et al.*, 1995, Pushkin *et al.*, 1983) have produced electron micrographic evidence that strongly suggests the quaternary structure of the Ch-cpn60 superficially resembles that of its Group I homologue GroEL (Viitanen *et al.*, 1995, Chen *et al.*, 1994, Pushkin *et al.*, 1983). Both molecules consisted of two stacked rings of seven subunits and as a consequence two distinct orientations of particles were evident in the electron micrograph data Fig 5.1.1.

E.coli GroEL chaperone is the most structurally studied and best understood of all the Group I chaperonins, with the structures of the GroEL oligomer, and its various complexes, having been determined by X-ray crystallographic methods (Braig *et al.*, 1994, Boisvert *et al.*, 1996, Xu *et al.*, 1997). The prediction that a high structural homology exists between GroEL and the plastid Ch-cpn60 presented an opportunity for the crystallographic study of the Ch-cpn60's α subunit. GroEL's structural data would provide the necessary phase information to carry out a 'molecular replacement' experiment on the diffracted X-ray data acquired on the Ch-cpn60 α subunit protein.

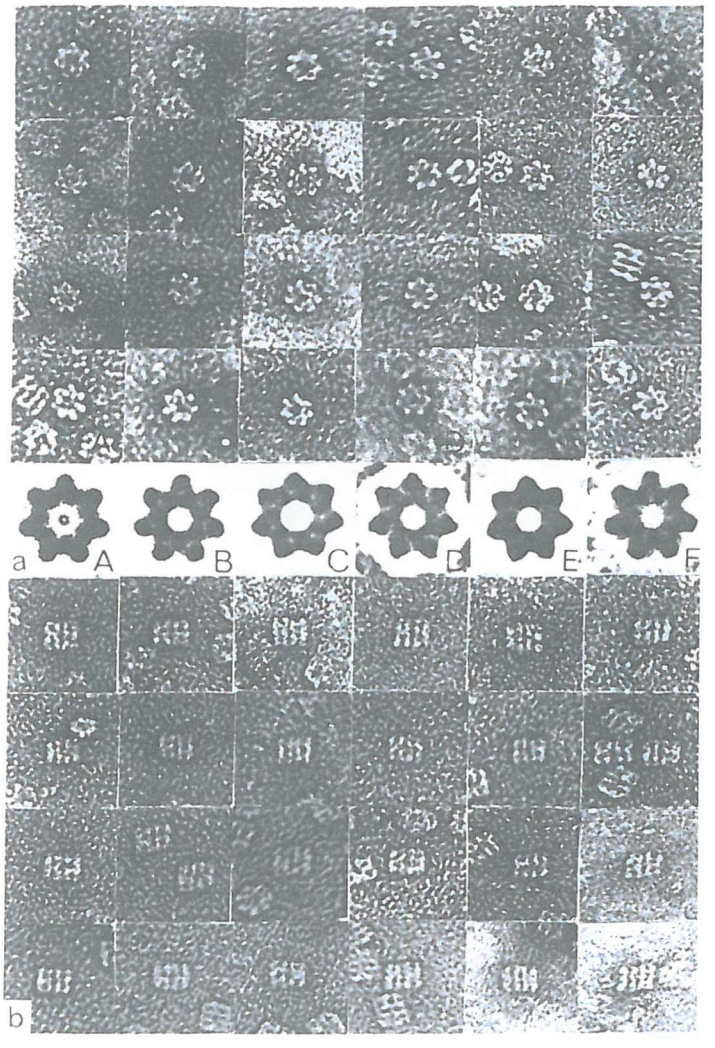


FIG 5.1.1 Cryo-Electron Micrograph Images of the Pea Leaf Plastid Ch-cpn60 Protein

Magnification x500000. (a) Round particles: images superimposed with 360°/7 turn angle can be seen in the lower row. Original micrographs of the particles are marked A,B,C,D,E and F. (b) Side projections of the particles, where four distinct stripes are observed (Pushkin *et al*, 1983).

5.2 Crystallography

Although the forces that contribute to protein structure and to the binding of ligands are well-understood, it is not possible to predict the three-dimensional structure of a protein from sequence information alone. X-ray crystallography currently provides the only way of determining the full three-dimensional structure of proteins at approaching atomic resolution. However, developments during the 1980's in Nuclear Magnetic Resonance (NMR) spectroscopy have made it possible to determine structures of some small proteins in solution. These NMR experiments can provide complete structural information, as well as providing information on the dynamics of the protein molecule. At the present time, only a few structures of proteins above 30 kDa molecular weight have been solved using the NMR technique. Ostensibly, because large molecules tumble slower, this results in faster relaxation and consequently broader lines in the NMR spectrum, the corresponding spectra show poor resolution and sensitivity and, in practice, it becomes very hard to determine structures from proteins above 40 kDa. However, recent advances in extending this size limit have been made with the introduction of novel NMR techniques and new biochemical approaches (Wider and Wuthrich, 1999).

X-ray crystallography is a technique used for directly determining the three-dimensional arrangement of atoms in protein molecules. This structural information provides a firm basis for elucidating the molecular mechanism of the biological reactions mediated by proteins. The protein of interest must first be obtained in crystalline form in preparation for the next step in the technique, X-ray diffraction. From the X-ray diffraction pattern of the protein crystal an electron density map of the protein can be derived, which can then be interpreted in terms of the positions of the atoms in the protein molecule, and finally a structure of the protein obtained.

X-ray Diffraction

The wavelengths of the X-rays used for protein crystallography (0.8 – 1.6Å typically) are close to the lengths of covalent bonds. When X-rays irradiate an object, X-ray waves are scattered by the electrons of the atoms in the molecule. In a crystal this scattering is amplified and concentrated in defined directions, producing discrete reflections. The total wave diffracted in a particular direction is the result of the superposition of the waves scattered in that direction by each of the electrons in that object, making the diffracted wave dependent on the spatial distribution of electrons in the object, Fig 5.2.1.

Bragg's Law

Bragg's Law was used to explain the interference pattern of X-rays scattered by crystals. X-ray diffraction spots are usually called reflections, produced by Bragg planes in the crystal. When the X-ray beam is reflected, the angle of incidence (the angle at which it strikes the plane of the crystal) is equal to the angle of reflection, and the beams hitting the plane of the crystal in phase also exit in phase. Bragg's Law is demonstrated in Fig 5.2.2.

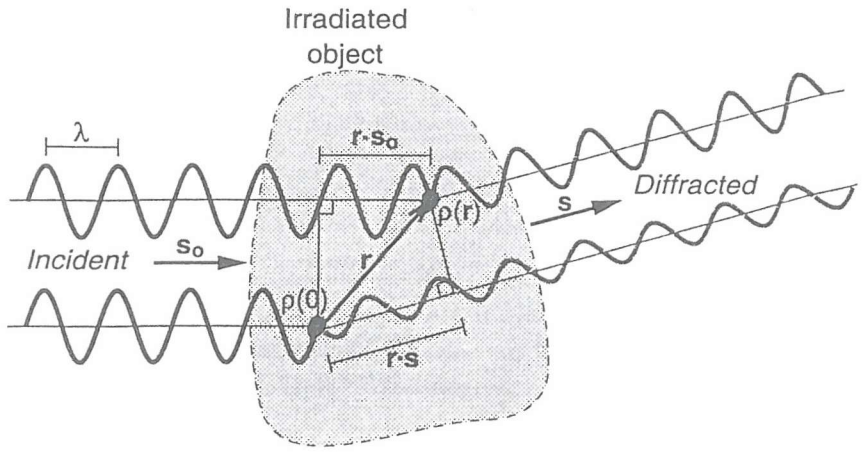


Fig 5.2.1 X-ray Scattering in the direction (s) by two volumes of electron density

The unit vectors $\mathbf{s}_0$ and **s** define the directions of travel of the incident and diffracted X-rays, respectively. The phase difference between the waves scattered by $\rho(0)$, located at the origin, and $\rho(\mathbf{r})$ is $\pi\mathbf{r}.\mathbf{S}$, where $\mathbf{S} = (\mathbf{s}-\mathbf{s}_0)/8$. A wave having amplitude A and phase M is conveniently represented as a complex exponential A exp(iM). Then, the total wave, diffracted by the entire object in the direction specified by **S**, can be represented by the integral $I\ \rho(\mathbf{r})\exp(2\pi i\mathbf{r}.\mathbf{S})d\mathbf{r}$, where the integration is over all infinitesimal volumes of electron density in the object.

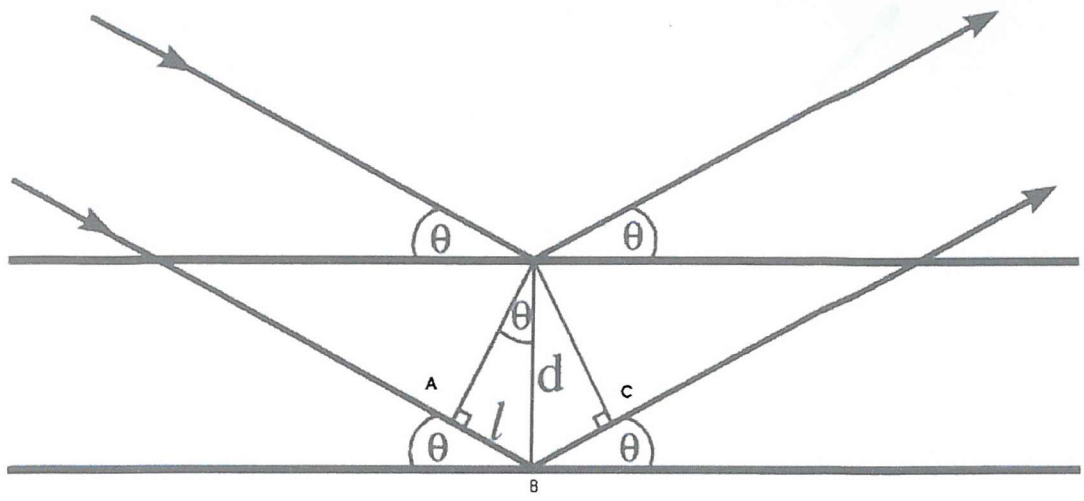


Fig 5.2.2 Bragg's Law using reflection geometry and applied trigonometry

The lower beam must travel the extra distance (AB + BC) to continue traveling parallel and adjacent to the top beam.

Simple trigonometry states that the distance **l** is equal to **d** sinθ. For the two rays to be diffracted in phase, twice **l** must be equal to the wavelength, giving the relationship:

$$\lambda = 2\ d \sin\theta$$

The two waves will be in phase if the pathlengths differ by any multiple of the wavelength, so Bragg's Law is usually expressed $n\lambda = 2\ d \sin\theta$. However, when this equation is applied to the data of a diffraction pattern, **d** is chosen in place of **n** which is then equal to one.

In Bragg's Law, as the angle increases, **d** must become smaller for the pathlength to remain equal to one wavelength. This is demonstrated by various rearragements of the common equation.

$$\sin\theta/1 = 1/(2\ d)$$

$$d = 1/(2 \sin\theta)$$

5.3 X-ray Diffraction by a Crystal

A crystal is composed of an object (motif) repeated translationally on a three-dimensional lattice. The periodicity of repeats in each of the three lattice directions is described by the unit cell, a parallelepiped, containing a single motif in the simplest example. The motif usually possesses internal symmetry, which inter-relates the asymmetric units that constitute the motif. There are 230 different ways to combine the allowed symmetry operations in a crystal. However, not all 230 space groups can occur in the protein crystal form. The reason for this is that the presence of mirror planes and inversion centres, found in non-biological crystals, would alter the asymmetry of the amino acids (the building blocks of proteins) from the normal condition of levo-rotatory to dextro-rotatory, left-to-right handed, a symmetry condition never found in proteins. The combination of symmetry elements present is described by the space group, of which only 65 are found to occur with protein crystals. The ultimate objective of this crystallographic process is to determine the atomic structure of the asymmetric unit, usually consisting of one protein molecule.

With a crystal, the contents of each unit cell diffract the X-rays identically, and the waves diffracted in specific directions with respect to the crystal lattice axes are "constructively" reinforced. However, the waves diffracted in all other directions are "destructively" cancelled. This constructive wave diffraction pattern of the X-rays appears on an image plate as a series of discrete spots (reflections) of varying intensities. In addition, symmetry within the unit cell of the crystal gives rise to symmetry in the diffraction pattern, represented by the discrete spots.

X-ray diffraction by a crystal can be conveniently represented by saying that for the crystal's real lattice there is a reciprocal lattice, indexed with the integers 'h, k and l', and it is this lattice which defines the permissible directions of the diffracted X-rays. The unit distances in the reciprocal lattice are related reciprocally to the unit distances in the crystal. Each reciprocal lattice point corresponds to one diffracted beam.

The reciprocal lattice is an imaginary concept for determining the direction of the diffracted beams, where if the crystal rotates the reciprocal lattice rotates with it. In an X-ray diffraction experiment the direction of the diffracted beams depends on two factors; the unit cell distances in

the crystal form (from which the unit cell distances in the reciprocal lattice are derived), and the X-ray wavelength.

5.4 The crystallographic Phase Problem Associated with Structural Determination

Each diffraction spot corresponds to a point in the reciprocal lattice and represents a wave, which possesses both amplitude and a phase (measured relative to a point of origin of the diffracted beams). Fig 5.4.1 shows two Bragg planes, representing how the phase and amplitude of the overall scattered wave arise from the individual scattered waves.

The vector, amplitude and phase, representing the overall scattering from a particular set of Bragg planes is termed the 'structure factor' (F). The structure factors, for the various points on the reciprocal lattice, correspond to the Fourier Transform of the electron density distribution within the unit cell of the crystal. A convenient property of the Fourier Transform is that it is reversible, whereby an inverse of the Fourier Transform applied to the structure factors represents the electron density distribution of the crystal. However, the major problem in X-ray crystallography is to determine the phase angles of the X-ray reflections. In order to calculate the electron density it is necessary to know both the amplitude and phase of each of a three-dimensional array of reflections. Amplitude information can be measured experimentally from an X-ray diffraction image, but phase information cannot be obtained directly. These are not represented in the measured intensity of the scattered wave and although it may be possible to measure the diffraction pattern and take square roots of the intensities thus giving the diffraction amplitudes, information relating to the phases has been lost during experimentation. This represents the 'phase problem'.

However, it is often possible to modify protein molecules in the crystal by the inclusion of electron-dense heavy atoms. The electron-dense heavy atoms in the derivative crystals alter significantly the strengths of reflections. Using the differences in the intensities of the native and

derivative crystals it is possible to determine the position of the heavy atoms within the repeating unit of the crystal. The data from more than one such heavy atom derivative are measured, and the quality of data is good, this information can be used to give an estimate of the phase of each reflection in the native crystals.

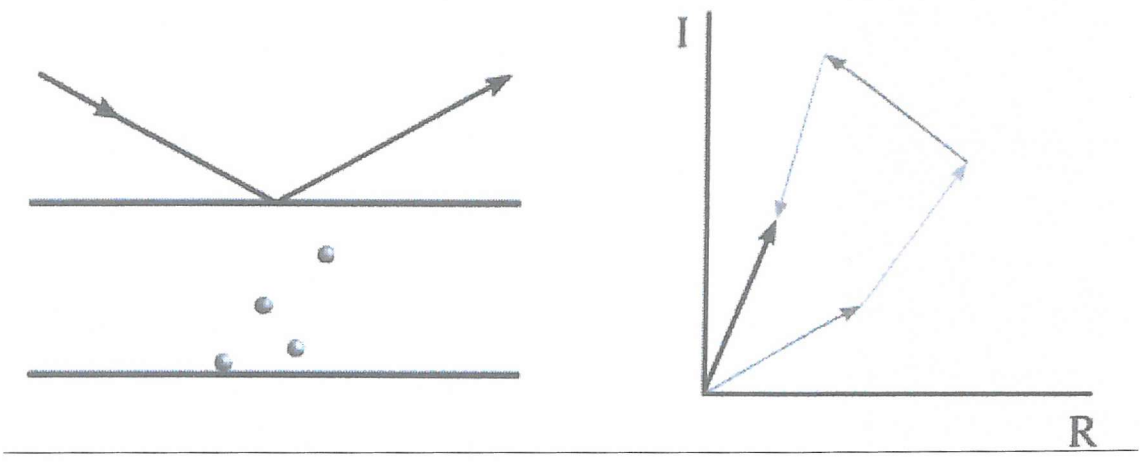


Fig 5.4.1 Bragg planes representing Phase and Amplitude combined into a Vector

The relative phase (from 0 to 360 degrees) depends on the relative distance of the atoms between the planes that define a phase angle of zero. The atoms, and their contributions to the scattering, are represented as vectors (Argand diagram, far right).

5.5 The Patterson Function

An understanding of the rotation and translation functions is obtained most easily by consideration of the Patterson Function (a self-convolution of the electron density), best described as a Fourier summation with the intensities as coefficients and no phase angles. Most traditional molecular replacement methods are based on the properties of the Patterson function.

The Patterson Function is important because it can be computed without phase information and the computed Patterson Function of a trial atomic model can be compared to an observed Patterson. When the model is oriented correctly and placed in the correct position in the unit cell, the two Pattersons should be similar. It was the Patterson Function that allowed the six-dimensional problem of a combined rotation and translation function to be divided into two constituent parts (Hoppe, 1957).

The Patterson map is a vector map (where vectors between atoms in the target structure appear as vectors from the origin to maxima in the Patterson map), with peaks at the positions of vectors between atoms in the unit cell. The vectors in the Patterson map can be divided into two categories; intramolecular vectors, which depend only on the orientation of the molecule, and not on its position in the cell, and these vectors are exploited in the rotation function and the intermolecular vectors, which depend both on the orientation of the molecule and on its position. Once the orientation is known these intermolecular vectors can then be exploited in the translation function.

The Patterson function can be expressed:

$$P_{(u,v,w)} = 1/V \sum_{h}\sum_{k}\sum_{l} |F_{hkl}|^2 e^{-2\pi(hu+kv+lw)}$$

where;

$P_{(u,v,w)}$ = the Fourier series (Patterson function). The coordinates (u,v,w) locate a point in a Patterson map, which is directly comparable to the (x,y,z) coordinates, which locate a point in an electron density map.

$\sum_{h}\sum_{k}\sum_{l}$= each term in the series is derived from one hkl summation.

(hu+kv+lw) =where h is in the u direction, k is in the v direction, and l is in the w direction.

5.6 Molecular Replacement

One way to solve the 'phase problem' is to possess an atomic model, from which estimates of the phases can be computed. Such a model can be obtained when you know the structure of a related protein. This method is based on the observation that proteins, homologous in their amino acid sequence, possess a very similar folding of their polypeptide chain. This level of sequential and structural homology (rule of thumb for molecular replacement is >40% sequence identity) is the first pre-requisite assigned to the intended use of molecular replacement techniques.

The next pre-requisite to carrying out a molecular replacement experiment is the need to place the model structure in the correct orientation (rotation) and position (translation) in the unit cell of the, as yet, unknown structure. The solution for these pre-requisites was put forward by Rossman and Blow, (1962). They stated that to orient a molecule it was necessary to specify three rotation angles and to place it in the unit cell, there is also a requirement for three translational parameters. This presents molecular replacement as a six-dimensional search. However, this is overcome by separating the molecular replacement experiment into two three-dimensional problems, with a rotation function computed to solve the three rotation angles, which can then be used to solve the three-dimensional translation angles within the cell. Fig 5.6.1 depicts the two stages (rotation and translation) required to obtain the preliminary information necessary to gain a molecular replacement solution.

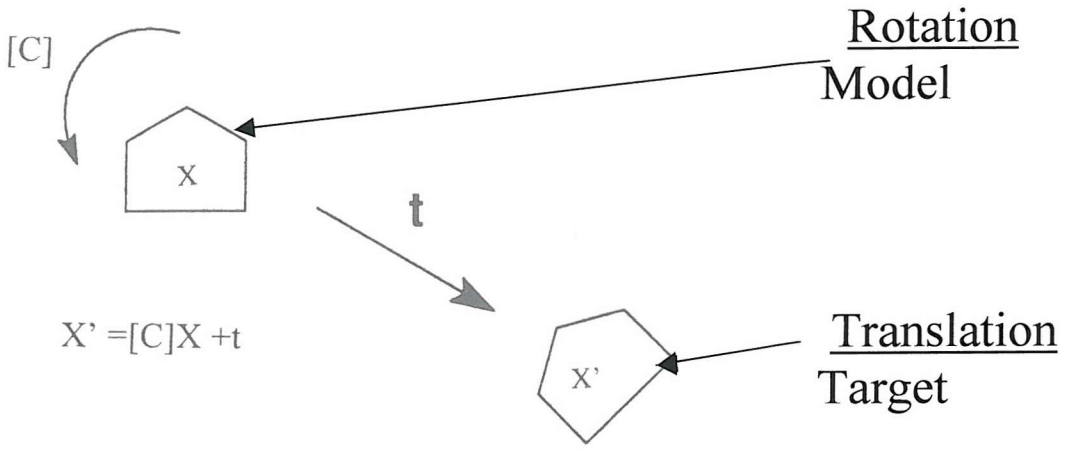


Fig 5.6.1 Using a Search Model in a Molecular Replacement Experiment

Gaining a molecular replacement solution for the 'target' structure involves determining a rotation matrix and then a translation vector to apply to the co-ordinates of the 'search' model. The diagram depicts the method of superimposition of the known structure (X) onto the target structure (X') in its unit cell. The transformation of X to X' is described by X' =[C]X +t where [C] is the rotation matrix that rotates coordinates X into a new orientation and t is the translation vector defining the position. The success of the procedure depends on determining 3 orientational parameters (rotation function) and 3 translational parameters (translation function).

5.6.1 The Molecular Replacement Method and the Rotation and Translation Functions

5.6.2 Rotation Function

The first step in molecular replacement (MR) is the 'rotation function' method. Phases from the structure factors of the 'search model' can be used as initial representative estimates for the phases of the 'target molecule'.

Rossman and Blow, (1962), expressed the rotation function, C, of the rotation, Ω, by the integral:

$$C(\Omega) = I_U P_1(r)\, \Re P_2(r)\, d\, r$$

Where P_1 is the target Patterson function, and YP_2 the rotated Patterson function, and $\Re P(r) = P(R^T.r)$, where R is the rotation matrix, x_2 the 'model' and x_1 the 'target' which are related by six rigid body parameters, 3 angles of 'rotation' defining the rotation matrix, R, and 3 'translation' components defining the translation vector, t. The integration performed over all points, r, is carried out in a spherical volume, U, centered on the origin.

Thus, it is possible to search for the best orientation, independently of location, by using the Patterson function. If the search model and target molecule possess high structural and sequential homology, and if they are oriented in the same way in unit cells of the same dimensions and symmetry, they should then derive very similar Patterson maps. This observation is suggestive that the Patterson map function presents a means whereby the determination for the best orientation of the model in the unit cell of the target protein, is a possibility.

It is this concommitent procedure of the Rossman and Blow method conjoined with the Patterson map function that may allow the attainment of an acceptable rotation solution, and that a single orientation solution may be carried forward in the search for the best position of the 'model' in relation to the real position of the 'protein'.

5.6.3 Translation Function

Once the orientation of the search model in an unknown cell has been obtained, the second stage of the Molecular Replacement method can proceed, namely the determination of its absolute position relative to the crystallographic space group symmetry elements. Only when all the molecules in the crystal cell are in the correct position can the 'calculated' Patterson cross-vectors superimpose to those of the 'observed' Patterson. Therefore, it can be stated that when the 'phasing', or 'search', model has been successfully superposed with that of the 'target' molecule, the intermolecular translation vectors (as opposed to the intramolecular vectors used to determine the orientation solution) are coincident and the local origin has been correctly defined with respect to the symmetry element.

Generally, the programs available to determine a translation solution work on the simple principle of 'step-wise' movements, over a resolution-constrained three-dimensional grid, which generate symmetry-equivalent molecules. From these data a measure of fit is acquired from each trial position between the 'observed' and 'calculated' structure factor amplitudes, F, or their equivalents in Patterson or real-space functions. However, although the translation functions are defined either in 'real' space or 'reciprocal' space, it has become the rignant format that the translation functions are transformed in reciprocal space, using structure factors.

Crowther and Blow; 'T' function (1967):

$$T(t) = I_v P_{obs}(u) P_{calc}(u,t) du$$

Where $P_{obs}(u)$ = Patterson map of the unknown structure

$P_{calc}(u,t)$ = calculated Patterson map of the phasing model, after a translation 't'

This can be expanded in reciprocal space to give:

$$T(t) = \sum_h I_{obs}(h)\, F_{M1} F^*_{M2}\, e^{-2\pi i h,t}$$

where I_{obs} = the observed intensities (from the target model)

F_{M1} = the Fourier transform of molecule 1 of the search model

F^*_{M2} = the complex conjugate Fourier transform of molecule 2 of the search model

h = the reciprocal space vector hkl

t = the translation vector between molecule M1 and molecule M2

Furthermore, due to the very nature of a translation function requiring to be processed in a 'step-wise' intermolecular fashion, and although related to the rotation process in the manner of being the second mechanism in the quest for a molecular replacement solution, the process was nevertheless independent of the need to process intramolecular vectors.

5.6.4 Assessing the Quality of a Gained Molecular Replacement Solution

5.6.5 The R-Factor

The atomic co-ordinates produced by protein crystallography represent an interpretation of the observed density as a fixed set of atomic parameters. It is important, therefore, to understand that the positional accuracy of the search model and the movements associated with the atomic positions should be taken fully into account, and critically evaluated. The resolution (expressed in Ångstrom units, Å, equal to 0.1nm) reflects the sampling of the electron density map that can be calculated from the data, that is the resolution determines the sharpness of the electron-density map and thus the amount of detail that is discernible.

A useful index of the accuracy of a model structure is the R-factor:

$$R = \Sigma(|F_o|-|F_c|) / \Sigma|F_o|$$

where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. Essentially, the lower the R-factor, the better the agreement between the observed and calculated structure factors. A typical R-value for a well-refined crystal structure at 2.0Å resolution is less than 0.20. Unrealistically, lower R-values can be obtained if bond lengths and angles are allowed to deviate from ideal values. It is also possible to improve an R-value by limiting the reflections used in its calculation. The mobility associated with an atom (its B value or temperature factor) is refined at the same time as its co-ordinates. The B-value is equal to $8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the mean-squared displacement of the atom from its equilibrium position, in Δ^2. The average B value for all the atoms in a protein is ~20, atoms with a B factor of greater than 79 possess a root mean square displacement of more than 1.0Å, and can safely be considered as disordered.

5.6.6 Common Molecular Replacement Packages

5.6.6.1 XPLOR

The XPLOR package (Brünger, 1992) is well known for its refinement algorithm which uses energy minimisation and molecular dynamics in conjunction with diffraction data to improve the coordinates of a protein structure. Molecular replacement is performed in Patterson space, using the real-space rotation function Patterson search method, originally implemented by Huber (1985).

The model Patterson maps are computed from structure factors |F| obtained by placing the model coordinates into an orthorhombic cell. The stipulated highest Patterson vectors gained from a given search radius range, are rotated using pseudo-orthogonal Eulerian angles (α, β, γ). The resultant Patterson map is then compared, computationally, with the Patterson map obtained from the target molecule. The search Patterson is then rotated in the three-dimensional coordinate system to find the orientation that best fits the target Patterson. Once the rotation function is complete, rigid-body Patterson correlation (PC) refinement is used in the capacity of a filter and if successful can explore, and discriminate between, the correct and incorrect orientations of the search model that may then justifiably improve the agreement between the target molecule and the search model. This is made possible by carrying out a minimisation against a target function defined by Brünger (1990), the search, or phasing, model is then optimised for each of the selected peaks of the rotation function. For each refined orientation, a correlation coefficient is computed and the orientation giving the highest correlation coefficient is chosen as the best orientation for the phasing model. The X-PLOR package then produces a sorted list of correlation coeffient results, derived from the observed and calculated squared normalised amplitudes gained from the target and search model, respectively, simplifying the final interpretation (Brünger, 1990).

However, at this stage the Patterson maps reflect the target molecule's orientation, but not its position.

The second, and final, stage of the X-PLOR molecular replacement package is to determine the best position for the correctly oriented search model in conjunction to the target molecule, within the unit cell. That is to say, that it is the search model that is once more used as the conduit means by which the correct position of the target may be obtained and thereby gain the initial phases required to obtain a correct molecular replacement solution for the target molecule. The translation search is performed over a three-dimensional grid that can be specified in fractional or orthogonal coordinates. Once again a correlation coefficient is employed, similar to an R-factor, which is obtained between the observed structure factor amplitudes $|F_{obs}|$ (target) and the calculated structure factor amplitudes $|F_{calc}|$ (model) to search for the best positional correlation between target and molecule, in order to best position the search model within the unit cell.

5.6.6.2 AMoRe (Automated Molecular Relacement)

Navaza (1994) wrote a molecular replacement program called AmoRe which incorporated several features, such as the exploration of many potential solutions and the already existent knowledge of the models. Instead of improving the search model in order to promote the correct peak to be positioned at the top of the rotation function peak list, the program strategy (Navaza, 1994) was to explore all the retained orientation solutions, in a fast and automated fashion. Instead of manipulating the coordinates, in order to calculate the structure factor amplitudes, during each orientation and rigid-body refinement of the search model in an iterative manner, as with previous program methods, Navaza determined that the AMoRe program would limit the use of atomic coordinates. Instead the atomic coordinates were to be used only once during a rotation function program, and that was to compute the Fourier coefficients that corresponded to the search models electron density. Subsequent structure factor calculations were to be performed by simple interpolation.

AMoRe consists of three main programs: ROTING, TRAING, and FITING. Also, there are two initialising program's, SORTING and TABLING. These two programs effectively sort the data into an internal format suitable for AMoRe, which then adjusts the model coordinates to the optimal position and subsequently prepares a table of continuous Fourier coefficients from the model coordinates, respectively.

The ROTING programs sole purpose is to calculate the rotation functions by using numerical integration, instead of the Crowther Fourier-Bessel expansions, in the radial variable (Crowther, 1972, Navaza, 1987). Further numerical instabilities are removed by the introduction of a new algorithm, which calculates the reduced rotation matrices (Navaza and Vernoslava, 1995), and as a consequence, improve accuracy in the results obtained, whilst further limitations were lifted concerning the "outer radius of integration/data resolution" ratios.

In practice, the structure factor amplitudes for the search model are calculated in a P1 box, the size parameters of which are dependent on the radius of integration, resolution of the data and

the dimensions of the search model. The rotation algorithm then computes the spherical harmonic coefficients for the search model and target Patterson distributions. Finally, before the rotation function program is initiated, a number of available parameters afford the user the option to modify certain criteria such as the input temperature factor, or selecting normalised structure factors (E (normalisation) values calculated by the CCP4 program ECALC), to allow an improved search.

The translation search algorithm in the AMoRe package implements both the Crowther and Blow overlapping function (1962) and a simplified reciprocal version of the full symmetry phased translation function implemented and developed by Bentley and Houdusse (1992). The correlation functions, using FFT's, are based on approximations of the intensities, a method devised by Harada *et al.*, (1981), and the exact computed overlaps of the intensities (Navaza and Vernoslova, 1995).

Finally, the FITING program performs fast rigid-body refinement, first implemented by Huber (1985), and assumes that the individual model Fourier transforms have been set to a common scale.

At the end of a fully automated AMoRe run, the output file expresses an ajudged correctness based on each translation function solution acquired. These data are depicted in an R-factor versus a correlation coefficient format, derived from the model and target structure factor data.

5.6.6.3 MOLREP

MOLREP, devised by Vagin and Teplyakov (1997) and subsequently CCP4-supported, is an automated program for molecular replacement which utilises effective new approaches in orientation and positional searching. Starting with the principal X-ray diffracted data, for the unknown molecule, and a relevant search model, the MOLREP program possesses the ability to go through all the requisite stages of molecular replacement, automatically.
Somewhat more pertinent, in difficult cases, is its ability to run and calculate the rotation, translation and rigid-body refinement programs in a separate and detached manner. This benefits the user by allowing for manual input of a set of parameters, necessary for the individual problem to hand. Structure factors acquired through the fast Fourier transformation of the model's electron density.

The rotational search is performed using the fast rotation function of Crowther (1972). However, to reduce the noise, the rotation function is calculated for three orthogonal orientations of the model and an average taken of the sum. The radius of the Patterson sphere is derived from the size of the search model and is usually twice the radius of gyration. However, when manually inputting the Patterson radius, or radius of integration, it should not exceed half of the unit cell dimension by more than 25%, the following of this simple rule deters any significant influence that would otherwise be exerted by the intermolecular vectors.

The translation function originates from the T2 function of Crowther and Blow (1967). The MOLREP package exhibits two main differences compared with other translation function programs. First, the translation function is calculated as the product of T_{jk} functions, rather than their sum, which presents the added benefit of increasing the contrast gained within the translation function. Secondly, the translation search is coupled to a packing function, which removes the false maxima corresponding to the interpenetrating molecules.

Both the translation function and packing function programs both use a similar fast Fourier technique. Both the translation and packing function programs allow for the incorporation of a second model, already placed in the cell. The orientation and position of the model obtained

from the rotation and subsequent translation function may be refined by the conventional rigid-body multidomain refinement, as incorporated in the MOLREP package. The minimised function is the ΔF or $|F_{obs}| - |F_{calc}|$. The best solution is indicated by the highest correlation coefficient. The refinement of the orientation of the model, prior to the translational search, is performed in space group P1, with cell dimensions chosen equal to the size of the model plus the radius of integration in each direction (Lifchitz, 1983). This allows for a certain neutrality and redundancy to coexist between model and target cell dimensions. Subsequently, this is reflected in the active minimisation of the number of intermolecular vectors that could be contributed by the search model.

5.7 Model Refinement

5.7.1 Introduction

In fitting an atomic model to a set of experimental data the model parameters are refined to best reproduce the measured structure factors. However, when working with macromolecules the introduction of additional parameters, such as B-factors, multiple conformations and solvent atoms, must be applied with extreme caution. The necessity for such reticence is due to the not uncommon pratice of working with search models whose atomic parameters exceeds the number of measured intensities obtained from the unknown molecule. In this situation there is a propensity for simply fitting the model to the noise in the collected data, rather than improving the fit of the model to the relevant set of data that has been acquired.

5.7.2 R-factor versus R-free Discrimination

The traditional indicator of how well the model fits the data is the R-factor, which measures the agreement between the $|F_{obs}|$ and $|F_{calc}|$. However, an arbitrarily gained low R-factor may well be obtained by the overuse of too many parameters, applied to the model, which may well lead to a functionally misleading model. This presents a problem in deciding which model parameters may be adequately determined from a noisy data set.

The R-free (Brünger, 1992) is a statistical technique that allows for a discrimination to be made between the signal-to-noise ratio, when modelling. The R-free, or cross-validation, approach requires the omission of a reasonable number of randomly selected reflections, approximately 1000 or 5-10% of the available data that are omitted from the starting set of reflections and retained as an independent set of 'test' data. In this way an assessment can be objectively made of any modification carried out to the model during a building cycle.

In principle, the R-free statistic measures the agreement between the atomic search model and the diffraction data of the unknown molecule, via a "test" set of unbiased reflections randomly

chosen and which are not actively involved in any refinement cycles. In this way, if a cycle of refinement results in a lowering of the R-factor, but a maintained or subsequent rise in the R-free set, then one can safely assume that over-fitting of the model to the noise of the data has occurred. Conversely, if both the R-factor and the R-free independently present improved results then it can be implied that the additional parameters, used in the preceeding building cycle, have provided a meaningful fit to the data.

The R-free is defined in the same way as the R-factor, but is calculated over the reflections in the free set alone, where:

R-factor

$$R = \Sigma_{allh} ||F_{obs}| - |F_{calc}|| / \Sigma_{allh} |F_{obs}|$$

and

R-free

$$R = \sum_{h,\ free\text{-}set} ||F_{obs}| - |F_{calc}|| / \sum_{h,\ free\text{-}set} |F_{obs}|$$

5.7.3 Refinement

Refinement and model building represents the last step in the solving and attainment of a molecular structure.

The refinement processes consists of fitting the structure factors of an atomic model $|F_{calc}|$ to the diffraction data $|F_{obs}|$. The refinement process can be greatly enhanced by incorporating chemical information such as bond length restraints and angles to typically observed values (Hendrickson, 1985). However, the refinement of a structure can come face-to-face with the problem of 'multiple minima', where the search model can become trapped in the many 'local' minima of the target molecule. This can have a contradictory effect on the placement of individual atoms in respect to the peaks and troughs of energy minimisation, where the search model atoms may well be placed in an incorrect 'local' position due to the lower energy effects being applied, rather than being given the opportunity to be 'globally' placed in the correct position, via the application of a higher energy minimisation. Molecular dynamics methods, coupled with simulated annealing techniques, greatly improves the possibility of gaining the global minimum (minimising the disagreement between the current density function and an ideal one that fulfils all physical requirements), whilst reducing the need for 'user' intervention (Brünger, 1997).

It is at this stage that the introduction of the R-free value has been instrumental in allowing an objectiveness to be applied, when assessing the validity of a refined structure and in so doing compensates for the possible error of overfitting of the model data to the observed data.

5.7.4 Least-Squares versus Simulated Annealing Refinement Methods

The initial model gained for the protein structure is often not good enough to allow a fully automated refinement process to take place. Therefore, semi-automated methods are available to help improve the agreement between the calculated and observed structure factor amplitudes.

Constrained, or rigid-body, refinement is a widely used technique. That is, when the geometry of a group of atoms is accurately known and with such confidence that the particular group would not be subjected to significant modifications, then this technique allows the entire group of atoms to be treated as a rigid entity. This method presents the advantage of requiring a considerably reduced number of parameters in order to perpetuate the refinement process. For each rigid group of n atoms, the number of independent positional parameters reduces from $3n$ to 6 namely (x_0,y_0,z_0), and ω,ψ,ϕ for the group position and orientation. To keep the number of required parameters small, thermal parameters are usually considered as isotropic and a single thermal parameter, for the whole group, is usually assumed.

Another function of the least-squares method is that of applied restraints. Whereby, restraints may be imposed to some functions of the parameters being used, permitting only realistic deviations of their values to be applied, from the fixed standard ones. Restraints applied to parameters often include set interatomic distances, fixed atomic coordinates along a given polar direction, which in turn fixes the centre of gravity of the molecule, thereby determining the origin. Several other types of restraints can be also be imposed concerning van der Waals distances, planarity of groups, bond lengths and torsion angles, thermal motion, and positional and thermal parameters related by non-crystallographic symmetry (NCS). The effect of these applied restraints is to add contributions to the elements of the normal matrix. In this way, highly correlated parameters can be found, albeit that the parameters are being deliberately correlated by the very restraints imposed.

However, it warrants only small standard deviations to make 'restraints' become equivalent to a 'constraints' method. The application of either constraints and/or restraints in a least-squares

process usually leads to larger residuals than those derived by unconstrained refinement methods.

5.7.5 Least-Squares (Reciprocal-Space Refinement)

In least-squares optimisation methods (Hendrickson, 1985, Sussman *et al.*, 1977), the observed structure factor amplitudes have fixed values and the parameters are varied, such that the calculated values approach the observations as closely as possible in the refinement process. Because the computation requires a direct comparative analysis between calculated and observed structure factor amplitudes (reciprocal space), these methods are duly referred to as 'reciprocal-space refinement'. The essence of the least-squares refinement method is to select atom positions that minimise the squares of differences between corresponding calculated and observed structure factor amplitudes, so that the goal is to seek the defined differences between observed and calculated F's, for each reflection *hkl* as $(|F_{obs}| - |F_{calc}|)_{hkl}$ and thereby minimise the function Φ,

$$\Phi = \sum_{hkl} w_{hkl} (|F_{obs}| - |F_{calc}|)^2 \; hkl$$

where:

Φ = the sum of the squares of differences between observed and calculated amplitudes.

hkl = the sum taken over all reflections, currently in use.

w_{hkl} = the term attributed to weighted differences, a number that depends on the reliability of the corresponding measured intensity. Simply, the weight should be $1/(\sigma_{hkl})^2$, where σ is the standard deviation from multiple measurements of $|F_{obs}|$.

As previously discussed, the use of constraints and restraints can increase the probability of finding the global minimum. This is essential, due to the complicated function of Φ exhibiting a propensity for many local minima. Imposing only a basic least-squares method will only

encourage the minimum energy movements of each atom that is nearest the starting point. Therefore, when using the least-squares method, it is imperative that the starting model parameters be as close to the global minimum as possible, in order that a best fit agreement between observed and calculated structure factors can be acquired. The flip-side to this imperative is that the refinement can converge into an incorrect local minimum. An important concept is the 'radius of convergence', which is the largest possible movement that an atom can undergo from its correct position. This has the potential to adversely decrease the radius of convergence, which in turn increases the opportunity for local minimisation to occur. It is the Converse of this inclusion of lower resolution data, initially in the refinement, which affords an opportunity for the model to graduate to the global minima, ensuring the correctness of atom placement.

5.7.6 Simulated Annealing (Molecular Dynamics)

The least-squares method only has the potential for the function to follow a downhill gradient to find its minimum value, therefore the refinement process may become trapped in a local minimum, instead of achieving correct atom placement.

Molecular dynamics (Brünger, 1987, Brünger and Nilges, 1993), or simulated annealing, provides the potential to overcome these energy barriers. This is achieved by enabling the movement of the atoms, within the refinement function, to flow uphill against the natural energy gradient by way of increasing the potential energy available to the function, and by so doing increasing the opportunity for correct atom placement. The significance of this increased potential is to improve the R-factor value when compared to the least-squares optimisation function. The likelihood of going 'uphill' is determined by a control parameter referred to as the "temperature" at which the simulation is done, whereby the higher the temperature, the more likely it is that the optimisation will overcome the higher energy barriers. By starting at a higher temperature (usually between 2000°K – 3000°K), and to then subsequently simulate the cooling of the molecule at a rate which is slow enough to ensure that system reaches equilibrium at each stage of the annealing schedule. Thus, simulated annealing has the potential benefit to increase the radius of convergence.

In effect, when the initial model is in its early form, simulated annealing can draw groups of atoms away from their final position. This corresponds with the potential energy minimum being placed in those final positions, without the need for manual intervention. This process lies in stark contrast to the limitations imposed on the least-squares method due to its inability to overcome the energy gradient.

5.7.7 Model Building

Once the phase angles (gained from the search models phases) have been estimated for the unknown structure, in this instance via the molecular replacement method and/or refinement, the calculation of an electron density map is the next step in solving the protein structure.

The solving of an *Ab initio* protein structure comes about through the interpretation of two combined techniques, electron density mapping and real-space refinement. Manual refinement is where the 'mark-one eyeball' is used in the visual building and placement of atoms and residues via available computer graphics packages. In contrast, in the real-space refinement method, electron density maps are used to derive inproved atomic positions for the model by minimising the difference between the observed and calculated electron density values. Generally however, it is accepted that both techniques go together.

One particular danger identified with the molecular replacement method is that of phase bias, ever present where use of the search model phases the root to an initially gained solution. This is where the phases from the model dominate the Fourier series, which could culminate in the Fourier series containing amplitudes purely from the intensity data and phases purely from the model. However, model bias can be compensated for with the introduction of electron density maps using various difference Fourier syntheses, in which the amplitude of each term is of the form ($|n|F_{obs}|$ - $(n-1)|F_{calc}|$), where 'n' is usually 2, which reduces overall model influence by subtracting the calculated structure factor amplitudes ($|F_{calc}|$) from some multiple of the observed amplitudes ($|F_{obs}|$). An easy difference map to interpret is the $2|F_o|$ - $|F_c|$ map. In this map the model bias is still reduced, however not to the extent presented when using the $|F_o| - |F_c|$ map, which tends to show density for missing atoms as positive peaks, and wrongly placed atoms as negative peaks.

With each successive map, where the phases of the new molecular model are combined with the observed structure factor amplitudes, new molecular features appearing in the map are added to

the model as, and when, they become discernible. In this way, errors can be corrected, allowing the structure to be brought to full completion.

5.7.8 Types of Electron Density Map

1 "$|F_o|$ map"

$$\rho (x\ y\ z) = 1/V\Sigma_{hkl} \mid F_o (hkl) \mid \exp [-2\pi i(hx+ ky + lz) + i\phi_{(hkl)}]$$

Phases may come from MR (Molecular Replacement), MIR (Multiple Isomorphous Replacement), density modification or from a model. It produces an image of the molecular structure. However, this map is the most prone to model phase bias.

2 "$F_o - F_c$ map" (Difference map)

$$\rho (xyz) = 1/V\Sigma_{hkl} (\mid F_o \mid - \mid F_c \mid) \exp [-2\pi i(hx+ ky + lz) + i\phi_{c(hkl)}]$$

$|F_c|$, ϕ_c come from the model. This map shows the difference between the currently-modelled structure and the information in the observed structure factors. Positive peaks indicate unaccounted for electron density, whereas negative peaks indicate wrongly placed atoms. These features appear at half-height. In this way the difference map can become a more objective tool toward overcoming phase bias.

3 "$2|F_o| - |F_c|$ map"

$$\rho (xyz) = 1/V\Sigma_{hkl} (2 \mid F_o \mid - \mid F_c \mid) \exp [-2\pi i(hx+ ky + lz) + i\phi_{c(hkl)}]$$

This map can be regarded as the sum of electron density of the model and of the difference map. Besides showing the electron density of the model phases, it also depicts the differences between the target and model (at full height). This map is considered to be one of the best for model building.

5.8 Crystallisation of Proteins

The identification of conditions appropriate for protein crystal growth is largely a multi-dimensional search conducted by trial and error, often representing the major obstacle in a crystallographic experiment. Crystals are grown from a supersaturated solution of the protein. The supersaturation state is usually achieved by lowering the protein solubility through the addition of a precipitating agent (most commonly a salt), an organic solvent, or polyethylene glycol. Optimisation between protein and precipitant concentrations are the two conditions that are sought in any searching screen protocol, Fig 5.8.1 depicts the relationship between protein and precipitant concentrations in a crystallization experiment.

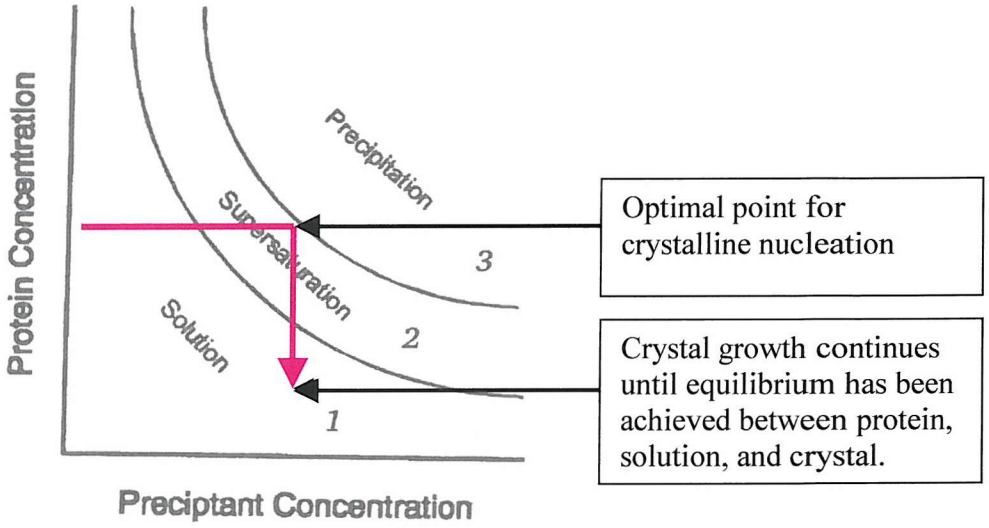


Fig 5.8.1 Two-Dimensional solubility diagram between protein and precipitant

The protein precipitant solution can exist in three distinct phases:

(1) Non-saturated, the protein solution is thermodynamically stable; (2) Supersaturated, the protein solution is kinetically stable but thermodynamically unstable; to reach equilibrium the protein concentration must reduce. If nucleation sites are present or added, precipitation of the protein to either amorphous or crystalline precipitates allows the solution to reach equilibrium; (3) Precipitation zone, the solution is both thermodynamically and kinetically unstable and a precipitate will form spontaneously. The red line indicates the course of a hanging drop experiment (see later).

Other factors that affect protein solubility include pH, temperature, and the presence of specific ligands and metal ions. Protein crystals are typically ~50% solvent, which fills extended channels between protein molecules. These channels allow small ligands (such as substrates, inhibitors, or heavy atom reagents) to be diffused into the crystal.

To achieve crystalline nucleation (a metastable state allowing crystal growth) the optimized conditions allow the solution to move into the zone of supersaturation. It is this move toward the supersaturation state that increases the concentration of the protein and the likelihood of protein precipitation. It is generally understood that it is on the border between the supersaturation zone and the precipitation zone that suitable nuclei for crystal growth are generated.

A common first approach to the problems that beset any crystallization experiment is the use of the "sparse matrix screens" provided for example by Hampton Research Laboratories (Cudney *et al.*, 1994). These screens (I and II) allow a wide range of conditions, incorporating salts and precipitants, to be tested.

5.9 Crystallisation of the Plastid Ch-cpn60 α Subunit Protein

All of the crystal screens related to the growth of the α protein were conducted by the vapour diffusion, hanging drop method (Engel, *et al*; 1995). Vapour diffusion has become the most successful method used for producing diffraction quality crystals. The method requires that a drop of protein solution, suspended from a cover slip, is mixed with precipitant taken from the mother liquor volume present in the same well. This solution is slowly dehydrated in a sealed well, by equilibration with the reservoir of mother liquor, which is at a higher precipitant concentration.

The Ch-cpn60 α and β proteins were first screened for optimum crystallization conditions, using the Hampton sparse matrix screens, I and II, by Dr Maria Hidalgo (Southampton University). Each vapour diffusion hanging drop condition contained 1ml of the particular precipitant in the well and on the cover slip 2μl of this mother liquor was mixed with 2μl of protein (either α or β) at 10mg/ml and suspended over the well.

Two crystals of the Ch-cpn60 α protein were grown to an X-ray diffractable size (size unknown, due to being grown before I arrived) in well 28 of crystal screen II of the Hampton sparse matrix. These crystals were acquired only after ~24 months of growth. Other wells in the same screen did not produce crystals. It was decided to perform a 'Factorial' crystal growth design (Carter and Yin, 1994) to try and optimize the growing conditions in the hope of reducing the growing time need to obtain suitable α protein crystals.

5.10 Optimising Growth screens for the Ch-cpn60 α Subunit Protein

The precipitant conditions for the successful crystal growth of the α protein, in well 28 of crystal screen II, were 1.6M sodium citrate, pH6.5. A subsequent crystal screen was tailored around these conditions in the hope of gaining an optimum to the growth and time of growth associated with the crystal form. The changes introduced were; pH range (6.2-6.8), and a variation on the protein-to-precipitant ratio mix, Table 5.10.1 depicts the growth matrix used in this experiment. Crystal growth was seen after 2 months, in Quadrant 2, Well A4 (hanging drop conditions: 1.6M sodium citrate (pH 6.8), 10% glycerol, 3:1µl of protein:precipitant at 10mg/ml). However, the crystals were not of diffractable size (~20µm) and consequently were left for a further growth period, Fig 5.10.1 shows a photograph of the crystals.

To date, more than 18 months after initial growth, there has been no appreciable increase in size of these crystals. These crystals could now serve as a 'seed' source for further crystallization experiments. This will entail crushing the crystals and introducing them into a hanging drop mixture, in the hope that they will form nuclei from which nucleation may occur.

The next *ad hoc* factorial screen was based around the partial success obtained with the second trial (seed crystals), which produced crystals in Well A4. Condition adjustments were; pH remained unchanged (pH 6.8), concentration range of the sodium citrate was changed to 2M and 2.5M with the additional introduction of either 10% glycerol or 10% ethylene glycol, hanging drop protein:precipitant ratios were also varied. Table 5.10.2 depicts the matrix used for this experiment.

quadrant 1		pH 6.2	pH 6.2	pH 6.8	pH 6.8	quadrant 2
		1	2	3	4	
A	Na_3Citr	3:1	2:2	2:2	3:1'crystals'	+
B	Na_3Citr	3:1	2:2	2:2	3:1	-
C	NaH_2Citr	3:1	2:2	2:2	3:1	-
D	NaH_2Citr	3:1	2:2	2:2	3:1	+
quadrant 3		[HighProtein] 10mg/ml	[LowProtein] 10mg/ml	[LowProtein] 10mg/ml	[HighProtein] 10mg/ml	quadrant 4

Table 5.10.1 Optimisation (1) Screen for the Ch-cpn60 α Subunit Protein

Depicted is the first optimization growth matrix for the α protein. The + and – for rows A and B indicate addition or no addition of 10% glycerol. The + and – for rows C and D indicate addition or no addition of ethylene glycol. Protein:precipitant ratios are highlighted by a green background. Crystals grew in Quadrant 2, Well A4 in conditions; sodium citrate 1.6M (pH 6.8), 10% glycerol, and a protein:precipitant ratio of 3:1µl (protein concentration 10mg/ml).

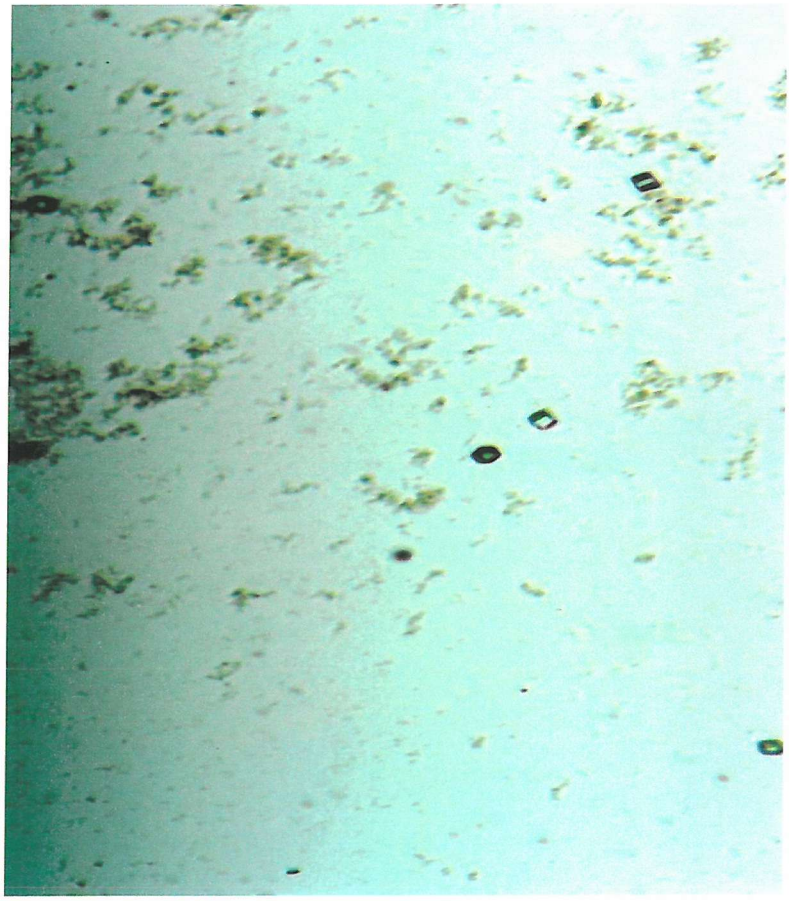


Fig 5.10.1 α Protein 'Seed" Crystals grown from First Optimizing Screen

The crystals depicted in the photograph were grown under the matrix conditions depicted in Table 5.10.1. These crystals were ~20μm diameter (magnification 2.5), too small for X-ray diffraction but possibly suitable for a 'seeding' experiment with further additions of the α protein.

quadrant 1		pH 6.5	pH 6.5	pH 6.8	pH 6.8	quadrant 2
		1	2	3	4	
A	Na_3Citr	2:2	3:1	2:2	3:1	+Gly
B	Na_3Citr	2:2	3:1	2:2	3:1	+E/G
C	Na_3Citr	2:2	3:1	2:2 2xcrystals	3:1	+Gly
D	Na_3Citr	2:2	3:1	2:2	3:1	+E/G
quadrant 3		Low [Protein] 10mg/ml	High [Protein] 10mg/ml	Low [Protein] 10mg/ml	High [Protein] 10mg/ml	quadrant 4

Table 5.10.2 Optimisation (2) Screen for the Ch-cpn60 α Subunit Protein

Depicted is the second optimized growth matrix for the α protein. **Gly** = 10% Glycol; **E/G** = 10% Ethylene Glycol. Protein:precipitant ratios are highlighted by a green background. Two crystals were grown (~60μm Diameter) in Quadrant 4, Well C3 in conditions; sodium citrate 2.5M (pH 6.8), 10% glycerol, and a protein:precipitant ratio of 2:2μl (protein concentration 10mg/ml).

Two crystals were grown in Quadrant 4, Well C3, and after ~7 months although small, these crystals were of sufficient size (~60μm) to attempt to gain X-ray diffraction data. Fig 5.10.2 shows a photograph of both crystals.

Fig 5.10.2 <u>Ch-cpn60 α Protein Crystals grown from Second Optimising screen</u>

The crystals depicted in the photograph were grown under the matrix conditions described in Table 5.10.2. These crystals were ~60μm diameter (at 2.5 magnification) and were judged of suitable size for X-ray diffraction.

5.11 Data Collection for the Ch-cpn60 α Protein

The first two crystals of the ch-cpn60 α protein, grown by Dr. Hidalgo, were sent to the European Synchrotron Research Facility (ESRF) based in Grenoble, France for X-ray data collection. Unfortunately, only one of the crystals survived X-ray beam irradiation. Data collection at the ESRF facility was collected by Dr. Alun Coker (Southampton University).

Data was collected on Beam Line ID29 for 240 images, crystal-to-detector distance was 250mm, $\lambda = 0.98$Å, ϕ start = 0^{o} and $\Delta\phi = 0.5^{o}$ oscillations, resolution was ~2.7Å. Fig 5.11.1 depicts one of the images collected.

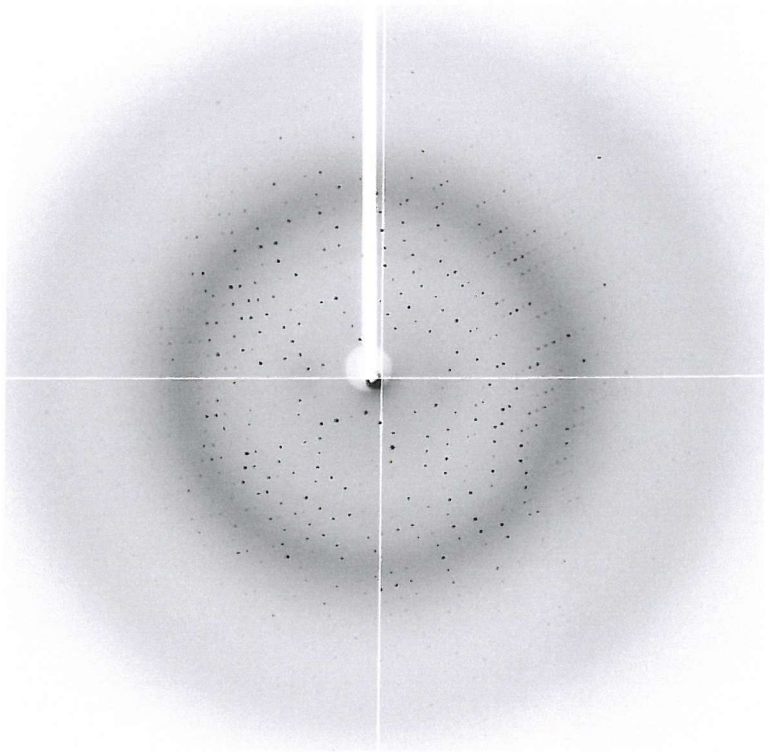


Fig 5.11.1 X-ray Diffraction Image of the α Protein Crystal

X-ray diffraction data for the α crystal was collected at ESRF, Grenoble, France. Processing on-site X-ray diffraction data resolved to ~2.7Å. Data processing was to suggest the high symmetry rhombohedral space group of R32.

5.12 Data Processing of the Ch-cpn60 α Subunit Protein

The acquired α crystal data was processed using the MOSFLM program. The MOSFLM suite of programs is designed to facilitate processing of rotation data collected either on image plate or film. Originally developed by Wonacott and Arndt (1977), the program has since been extensively developed (Leslie, 1992; 1994). The basic procedure for data processing is independent of the type of detector, film or image plate, being used. MOSFLM performs the actual integration of the reflection intensities. It generates the reflection list, reads the digitised image, integrates the spots and writes the intensities and standard deviations into a file. The program has the additional capability of refining the crystal orientation and cell parameters during data processing, using the intensities of partially recorded reflections. The MOSFLM program also gives an indication of the highest symmetry space group possible, by means of a lowest penalty order profile listing. The composite log file for these data can be found in Appendix 2.

After the data has been processed through MOSFLM it is scaled using SCALA from the CCP4 suite (Collaborative Computing Project Number 4, 1994). This program scales together multiple observations of reflections and merges these multiple observations into an average intensity. The merging algorithm analyses the data for outliers, and gives detailed analyses. It generates a weighted mean of the observations of the same reflection, after rejecting the outliers. The SCALA program completes three passes through the acquired data:

1. A scaling pass, whereby an initial estimate of the scales are obtained, followed by a refinement of the scaling parameters.
2. An analysis pass is carried out to analyse any discrepancies and accordingly adjust the standard deviation (SD) estimates.
3. A final pass is then used to apply scales, analyse agreement and write the output file, usually with merged intensities, but alternatively as a copy of the input file with evaluated scales appended to each observation.

SCALA also gives the completeness of data, the multiplicity of reflections (indicating the amount of times a reflection was counted in the data collection, the higher the multiplicity, the better the data), and finally the R-Factor and R-Merge, as well as the I/σ (I) (standard deviation of spot intensity) data. R-merge is where, after scaling the level of agreement among different frames are acquired. Through averaged, scaled intensities for all observations of one reflection, the $|F_{obs}|$'s values are derived. The corresponding $|F_{calc}|$'s values are derived from scaled intensities for individual observations of the same reflection.

Finally, the program TRUNCATE (CCP4) was used to convert the intensities of the spots, which was the data gained from the SCALA program, into F values (structure factor amplitudes). The TRUNCATE progam (French and Wilson, 1978) reads the reflection data file output by the SCALA program and, in turn, outputs an MTZ reflection data file containing the mean amplitudes and the original intensities, based on Bayesian statistics. The structure factors are calculated using the prior knowledge of Wilson's distributions for acentric and centric data (calculated in shells of reciprocal space in a first pass through the data) and the mean intensity and standard deviation values. The Wilson plot is a part of the program which attempts to calculate an absolute scale and temperature factor for a set of observed intensities, based on the number of amino acids in the asymmetric unit. The Wilson plot will deviate from a straight line from about 3.0Å - 4.0Å. Although all the points on the Wilson plot are plotted the scales are only determined from the resolution range determined by the user. Tables 5.12.1 and 5.12.2 depict the combined data processing statistics acquired for the α protein crystal, using all of the above programs.

All programs used and approaches tried were done so with the full intention of obtaining a molecular replacement solution for the α subunit protein, using the three space groups (SG) indicated in the MOSFLM auto-indexed highest-order penalty listing (Table 5.12.1). The following resultant data encompasses the direction taken with R32. However, it should be also accepted that the protocols and approaches applied to the R32 experiments were equally applied to the space groups R3 and C2.

5.12.1 Highest Order Symmetry

Auto-indexing of the X-ray data set, using the MOSFLM program, produced a list of possible Laue space groups, sorted on a penalty index. Table 5.12.1 shows the top order penalty index presented for the data set. It is common practice to seek the highest symmetry group on which to carry out a post-refinement of the data set. With a penalty of only 12, the rhombohedral space group R32 was selected as the space group most likely to represent the α subunit's order of orientation and position in the crystal, along with any non-crystallographic symmetry (NCS) or crystallographic symmetry, relating to the protein in the asymmetric unit. It transpired that the space groups R32 (penalty 12), R3 (penalty 12), and C2 (penalty 9) would all be used in an attempt to acquire a rotation solution.

No	PENALTY	LATT	a	b	b	alpha	beta	gamma	Possible spacegroups
14	127	oP	80.18	81.08	81.23	96.3	84.2	96.0	P222,P2221,P21212,P212121
13	105	hR	119.83	108.18	134.66	90.5	94.4	116.2	R3,R32
12	102	oC	107.92	119.83	81.23	98.2	90.4	89.4	C222,C2221
11	96	mP	80.18	81.23	81.08	96.3	96.0	84.2	P2,P21
10	96	mP	80.18	81.08	81.23	96.3	84.2	96.0	P2,P21
9	95	oC	108.33	120.87	80.18	82.1	90.1	89.9	C222,C2221
8	94	mP	81.08	80.18	81.23	84.2	96.3	96.0	P2,P21
7	41	mC	119.83	107.92	81.23	90.4	98.2	89.4	C2
6	36	mC	120.87	108.33	80.18	89.9	97.9	89.9	C2
5	31	aP	80.18	81.08	81.23	96.3	84.2	96.0	P1
4	12	hR	107.92	108.18	154.01	90.7	89.3	119.8	R3,R32
3	9	mC	119.83	107.92	81.23	89.6	98.2	90.6	C2
2	2	mC	120.87	108.33	80.18	90.1	97.9	90.1	C2
1	0	aP	80.18	81.08	81.23	83.7	84.2	84.0	P1

Table 5.12.1 Auto-indexed Penalty List for the α Protein Data Set

R32 is 4th in the penalty list, suggesting a high likelihood that a solution for this space group was a possibility. The highest symmetry space group suggested from the MOSFLM data was R32 (4th in rank order with a penalty of 12). The unit cell dimensions were a = b = 108Å, c = 154Å, γ = 120°, determined through the REFIX algorithm, as part of the MOSFLM program.

Number of reflections	**150851**
Total number of unique reflections	**8806**
Resolution	**2.7Å**
Completeness	**99.9%**
Multiplicity	**5.2**
R_{merge}	**0.103**
Average I/σ (I)	**4.1**
Solvent content	**47%**
Asymmetric Unit	**~51% of one subunit (putative proteindegradation; refer to section 5.14)**

Table 5.12.2 SCALA and TRUNCATE Statistics for the α Protein in Space Group R32

The data for this table was acquired using the SCALA and TRUNCATE programs (CCP4). The composite log files for these data is in Appendix 2.

5.13 The Trigonal Crystal System of R32

The highest symmetry space group suggested from the MOSFLM data was R32 (4th in rank order with a penalty of 12, Table 5.12.1). Whereby, the unit cell dimensions were a = b = 108Å, c = 154Å, $\alpha = \beta = 90^{o}$, $\gamma = 120^{o}$. This system can be treated either with hexagonal axes or with rhombohedral axes (Fig 5.13.1). According to Drenth's (1994) 'Seven Crystal System' tables, in order to be classed a rhombohedral crystal system, the conditions imposed on the cell geometry are;

$a = b = c$; $\alpha = \beta = \gamma$. Conversely, to fulfil the conditions for the hexagonal class, the imposed conditions are: $a = b \neq c$; $\alpha = \beta = 90^{o}$; $\gamma = 120^{o}$.

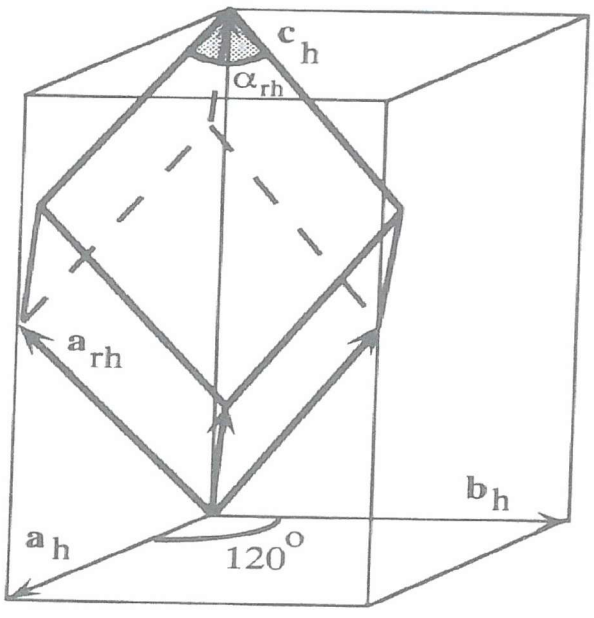


Fig 5.13.1 Rhombohedral cell and its corresponding hexagonal cell

In the hexagonal unit cell (a_h) and (b_h) are equal in length and have an angle of 120^{o} to each other. (c_h) is perpendicular to the ab-plane and differs in length from (a_h) and (b_h). The rhombohedral cell (a_{rh} b_{rh} c_{rh})has three equal axes at angles not necessarily 90^{o} with each other. (Source: Drenth, Principles of Protein X-ray Crystallography. 1994).

5.14 Putative Proteolytic Cleavage of the α Subunit in Crystalline Form

From the TRUNCATE program, and using the Matthews constant (1968) (V_m=2.63Å^3/Da), the solvent content was calculated to be ~47%, based on the assumption that 300 of the subunits 526 residues were present in the asymmetric unit. By carrying out several TRUNCATE runs, initially all of the α subunit residues (526) were used, then residues were detracted from the full compliment in suitable increments, until it was concluded that only ~53% of a full α subunit could be accounted for in the R32 space group (relating to ~30kDa Mw of subunit present in the asymmetric unit). The knowledge that there are 18 symmetry equivalents (international tables for x-ray crystallography, page 261, 1969), 18 asymmetric units, to the R32 unit cell further indicated it was unlikely that the asymmetric unit of the cell could accommodate a full α subunit (~58 - 60kDa MW).

The following formula was used to acquire the solvent content of the R32 asymmetric unit cell:

Solvent content = 1- (n Υ_p Mw/N_A V)

Where:

n = Number of molecules in unit cell (symmetry equivalents) R32 (155) = 6 (international tables for x-ray crystallography, page 261, 1969)

Υ_p = partial specific volume of protein (~ 0.737 units for most proteins)

Mw = molecular weight of protein (estimated for asymmetric unit) ~30kDa

N_A = Avogadro's number ($6.02x10^{23}$)

V = Volume of unit cell in cm^3 (∴ Parallelogram vol of 93Åx154Åx108Åxsin120^o = $150000x10^{-24}$Å^3) = $1.5x10^{-18}$cm^3

This observation suggested that it would be prudent to carry out SDS-PAGE on the crystal screen wells that originally produced the crystals, from which the X-ray diffracted data had eventually been collected and processed. In this way it could be possible to verify whether the α subunit had undergone some form of proteolytic cleavage, whilst still in its soluble state and prior to the crystallisation event having taken place.

An SDS-PAGE 4-12% (Nu-PAGE), followed by a silver staining of the gel, was run with the remaining product formed in four of the original wells (screen II, 1.6M Sodium Citrate at pH 6.5) which had produced the α subunit crystals (screens were laid by Dr Maria Hidalgo, Southampton University). The results of the SDS-PAGE Fig 5.14.1, show putative cleavage having occurred within the α subunit proteins laid down for crystal growth.

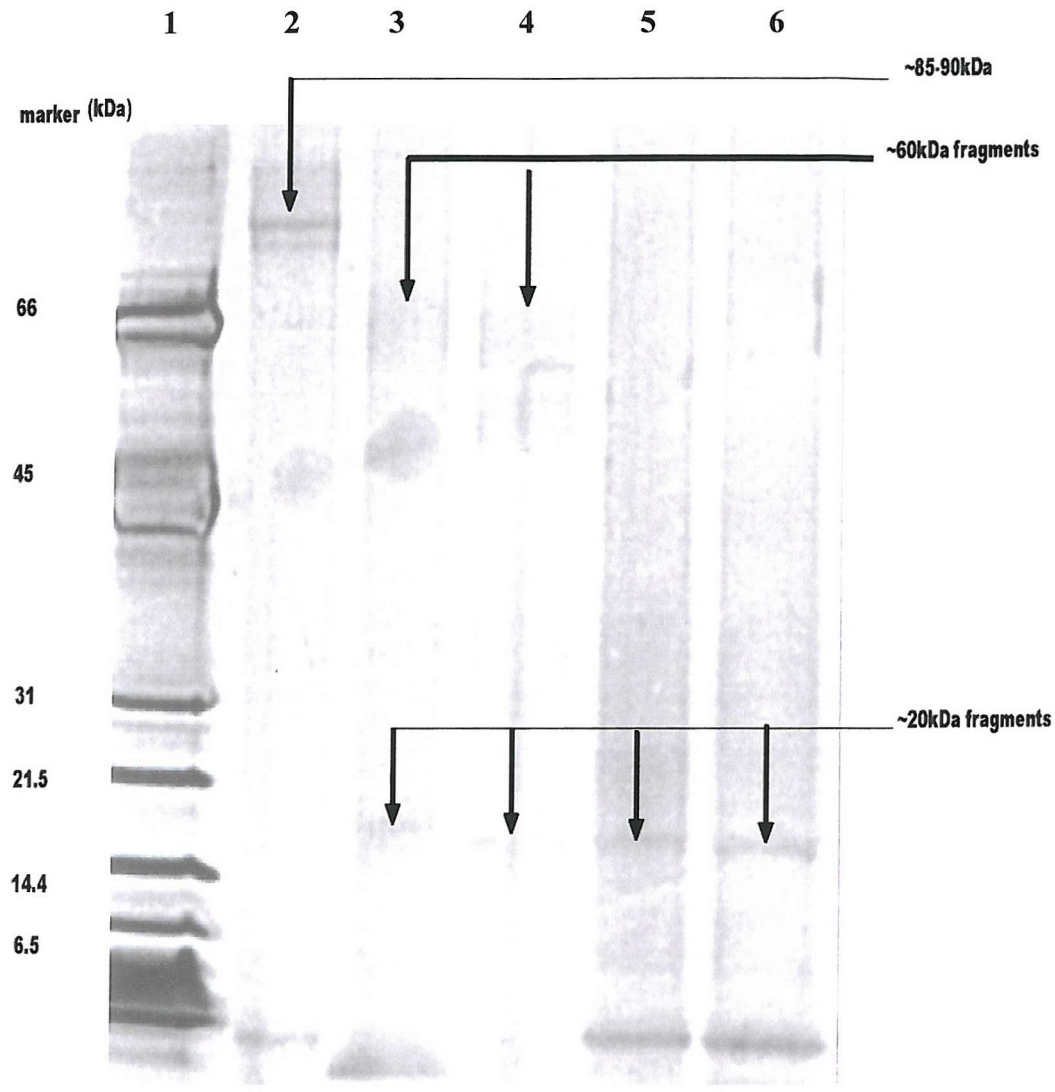


Fig 5.14.1 SDS-PAGE (4-12%) and Silver Staining of the α subunit.

Lane 1 = protein marker (PROTEAN® II), Lanes 2 to 6 = screen solutions from various wells taken from the original crystal screen. Lane 2, shows a product ~20-25kDa bigger than the α subunit. Lanes (3and4) show heavier bands that could be representative of the α subunits expected mass (~60kDa), but also show bands at the ~20kDa Mw range. Lanes 5and6 show lower fragmented bands, ~20kDa (greatly below the expected α subunit Mw of ~58-60kDa), which could be indicative of some form of cleavage having occurred during crystallisation.

5.15 Molecular Replacement Analysis for the Ch-cpn60 α Protein

Now that the collected X-ray data for the α protein had been processed it was possible to carry out a molecular replacement experiment in an attempt to find a solution for the protein and eventually attempt to resolve the structure.

The processed data would be representative of the 'target' molecule in a molecular replacement experiment, and the 'search' model phases to be used would come from the structurally and sequentially homologous *E.coli* GroEL chaperonin.

The Type I chaperone GroEL is constructed with fourteen homologous subunits, and each subunit possessed ~50% sequence identity to the plastid Ch-cpn60 α protein, as well as an estimated ~90% structural similarity. This amount of homology suggested GroEL as the preferred choice of 'search' model for the molecular replacement experiment for acquiring the initial phases that could then be applied to the α protein, presented as the 'target' molecule.

In practice, the intra-molecular vector-set (short vectors) are used to estimate the relative orientation of the structures. The longer inter-molecular vectors are used to determine the relative position of the structure. The molecular replacement method dictates that a rotation solution, between model and target, must be solved before attempting to gain a translational solution.

5.16 Self Rotation Functions (POLARRFN)

The self-rotation, or POLARRFN (fast polar rotation) function was written by Kabsch and is part of the CCP4 program suite (1994). It produces sections of a constant rotation angle (Kappa) for different axis directions, defined by omega (ω), representing the angle from the pole, and phi (φ) representing the angle around the equator. The polar angles are mapped to the surface of a sphere and displayed in spherical polar coordinates, not Eulerian angles. Primary to its function is when searching for noncrystallographic (NCS) symmetry, where POLARRFN has proved readily convenient to work with in spherical polar rotations. The advantage of this function stems from the commonness found in rotations of kappa = 180° and 120°, which subsequently allows the search to be restricted to a fixed value for kappa.

The self-rotation function is achieved when two Patterson maps are superposed upon one another. A search is then commenced by rotating the Pattersons, corresponding to interatomic vectors, over one another in the search for peaks, which whenever the two maps superpose, the rotation function will be high. These peaks can then be correlated to either NCS and/or the presence of crystallographic symmetry.

The self-rotation function carried out using the R32 unit cell dimensions, as depicted in Figs 5.16.1(a) and 5.16.1(b), clearly showed the existence of a 3-fold crystallographic symmetry (Kappa = 120°), and a 2-fold crystallographic symmetry (Kappa = 180°), looking down the c-axis. Furthermore, it was evident that no non-crystallographic symmetry existed in the asymmetric unit.

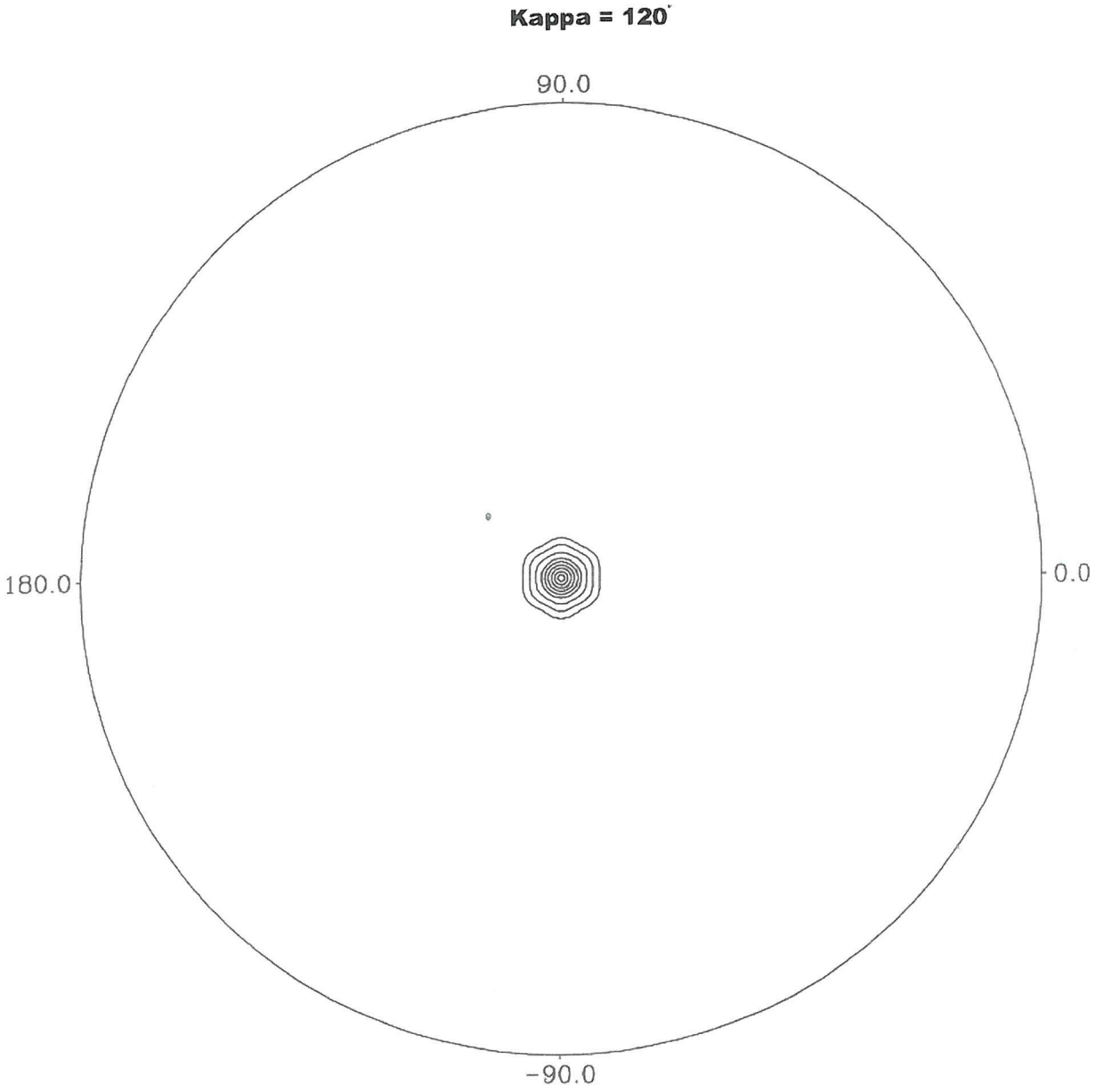


Fig 5.16.1(a) Self-rotation function of the α subunit protein. Kappa = 120°

The function was carried out at a resolution of 2.8Å and a radius of integration of 45Å.

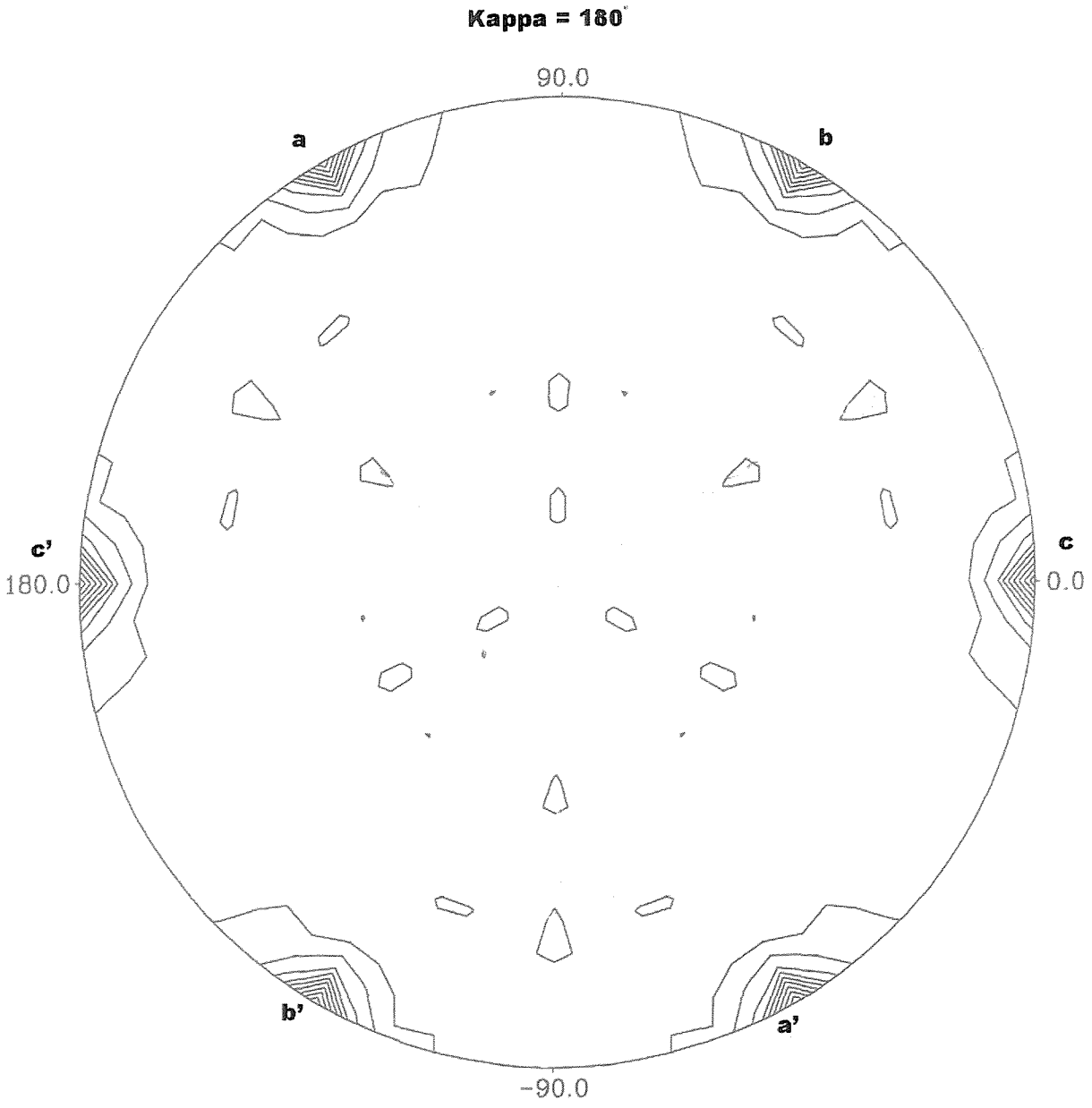


Fig 5.16.1(b) Self-rotation function of the α subunit protein. Kappa = 180°

Self-rotation function of the α subunit as determined from the program POLARRFN [Kabsch, 1993]. (a) Peaks on this section (κ = 120°) show the position of crystallographic 3-fold axes. (b) Peaks on this section (κ = 180°) show the position of crystallographic 2-fold axes. These 2-fold axis are perpendicular to the 3-fold axis.

The function was carried out at a resolution of 2.8Å and a radius of integration of 45Å.

5.17 Whole Model Rotation Functions

Although a question mark now remained over the integrity of the α subunit in crystalline form, it was still essential to start the molecular replacement experiments using a 'whole' model search protocol. In part, this was due to being unable to answer whether the protein had been cleaved or not, and if so which part/parts of the structure remained intact. Therefore, by using an initial whole model search it may be possible to detect any structural variability in domain integrity.

The model used for the initial search was the GroEL heptamer, solved to 2.8Å, accession number 10EL, (Braig *et al.*, 1994) of which only data relating to a single subunit was used.

However, notwithstanding exhaustive use of the programs currently available to help acquire molecular replacement rotation solutions, MOLREP, CCP4 Suite's AMoRe rotation program, CNS, and XPLOR, none gave any discernible or likelihood solution peaks above the mean.

5.18 Equatorial Domain Rotation Functions

Due to the lack of a rotation solution being found by using the whole search model. It was decided to refine the rotation search parameters by using only the equatorial domain (residues 2-135 and 410-526) of the Braig GroEL search model. It may have been possible to gain a rotation solution for the α subunit in that part of its structure. Although only a part of the whole subunit, the equatorial domain represents the largest of the three domains that subsequently comprise a whole subunit. It was concluded that the α subunit may be oriented in a non-conformed position relevant to the whole subunit orientation of the Braig model being used as a search model, due in part to the possibility of cleavage activity effecting a structural aberration.

By breaking the model down into its constituent parts (domains) it may afford a rotation solution to be gained for at least part of the subunit. Once a rotation solution for this domain had been gained, the angles of the equatorial solution could be further used as reference points in the search for the orientation of the remaining domains in the molecule.

This tactic proved fruitless in the search for a rotation solution, where once more no discernible peaks attested to a rotation solution having been acquired.

5.19 Apical Domain Rotation Functions

The approaches applied to the equatorial domain of the Braig model were also applied to the apical domain of the same model (residues 191-376). This approach proved no more successful than the previous approach.

Four separate programs, each specific to gaining a rotation solution, were utilized to their maximum effect. After exhaustive searches had been carried out for one space group, the whole process was repeated for the other two space groups, leaving little to chance should a rotation solution have been obtainable.

5.20 Systematic Residue Deletions of the Braig Search Model

The next approach was to use systematic deletions. This approach provides the possibility of acquiring a rotation hit by systematically deleting a given number of residues and repeating the rotation search after each deletion series (McCree, 1993). If parts of the model that contribute to a rotation hit are likely to be homologous to the unknown structure, a drop in the peak below the known baseline peak for the model will occur. However, if a portion of the model is interfering with the rotation search, then by removing this portion a peak will obtained higher above that of the mean baseline peak.

The first step applied to this approach was to systematically remove all of the loop regions associated with the model and gain as closely related homologous structure to the α subunit as possible (Pro33-Asp52, Ala176-Glu191, Tyr199-Val213, Ala475-Ala493). Rotation function data were systematically checked following each deletion for any significant improvements set against previously acquired results. The next approach incorporated systematic deletion of 3 successive residues, starting from the N-terminal of the equatorial domain (due to slight differences in tail length that exist between the GroEL model and the α subunit), and 3 residues from the C-terminal and N-terminal ends of the apical domain, consisting of a maximum of 25 residues per terminal (N-term and C-term). Any more than the deletion of 25 residues per terminal would have been counter-productive.

Each deletion was run through a rotation function analysis, using the AMoRe program (CCP4 suite). A peak analysis table was determined, representing the significance (positive or negative against the baseline) for each deletion. Each deletion gained-peak that was significantly above the base line of the undeleted model rotation solution were then combined into one PDB file, making a composite search model, and run through each of the other rotation programs.

However, this approach would restrict the short vector search parameters, of an already shortened version of the Braig model. This concern was manifest in the poorest showing of rotation peak differences acquired, when compared to the previously applied approaches.

5.21 Cross-Correlation Trials

The final approach used in the search for a rotation solution was to make use of an equatorial and apical domain, but from different models. The equatorial domain was taken from the Braig structure and the apical domain from the Xu structural model, accession number 1AON (Xu *et al.*, 1997). These domain models were used as separate entities in separate rotation function trials.

The apical domain of the plastid α subunit is capable of undergoing major allosteric rearrangements, which may account for the inability of the Braig models apical domain residue coordinates to find a rotation solution. It is likely that the 'target' molecule's apical domain is in a wholly different conformational state when compared to the apical domain of the Braig model. By using the two domains (apical and equatorial), supplied by differing sources and presenting different apical conformations, in separate rotation functions it may be possible to acquire angle correlations between the two domains. The idea was that when angles from a separate equatorial (Braig) rotation function analysis closely correlated in comparison to those gained for the apical domain (Xu), also run separately, the PDB files relating to those individual solutions could be amalgamated by way of angle orientation and then compliled into a composite PDB file, using the QUANTA modelling package to bring the separate domains together. This new composite PDB file could then be run in a new rotation function experiment to ascertain an orientation profile relative to the two domains.

Subsequent trials with this approach did show correlated angles had been acquired between the two domains, from independent rotation functions. Although modeling (Quanta Modeling Package) of the composite PDB file proved a success both spatially and in orientation, when compared with the Braig/Xu models, when the composite PDB file was put through a rotation function it did not provide a relevant significant peak that could be remotely considered as correlating to a feasible solution.

The approaches applied to gain a rotation solution for the α subunit protein, using modified and unmodified versions of the Braig and Xu structures, required eight months of application and were finally deemed unsuccessful.

Table 5.21.1 depicts a compilation of the rotation and translation function data that was acquired during the various attempts made in order to gain a molecular replacement solution for the plastid Ch-cpn60$^{\alpha}$ subunit protein. Also represented in Table 5.21.1, are the assortment of molecular replacement programs that were used in this set of experiments.

Program	**Space Group**	**Function** **Rot** **Tran**	**Rotation Function peaks height**	**RF/σ (Peak Diffs)**
CNS	R32	•	7.66 7.47	Nil
XPLOR	R32	•	1.2 0.8	Nil
MOLREP	R32	• •	6523 6489	4.25 4.22
AMORE	R32	•	67.6 67.8	25.8 25.1
MOLREP	R32	•	4506 4461	3.93 3.89

Table 5.21.1 depicts the various experiments carried out to gain a molecular replacement solution for the plastid Ch-cpn60$^{\alpha}$ subunit. Although only the relevant peaks shown relate to R32, this trend of disappointing and negative data carried through to both R3 and C2. At this stage it was evident that the opportunities available of finding a solution were all but depleted. Possible failure was due in part to the ambiguity of identifying the space group, but fundamentally in all probability due to the possibility of proteolytic cleavage having occurred within the protein, prior to crystallisation. Peak heights show only the top two peaks from a mass of raw data.

5.22 Putative Molecular Replacement Solution gained for the Ch-cpn60 α Subunit Protein

Eventually, after an exhaustive perusal of all available models which consisted of phases suitable to be borrowed in the search, a putative, but weak, molecular replacement solution was acquired for the α subunit protein.

The tertiary structure of the plastid Ch-cpn60 α subunit was solved using the molecular replacement method incorporated into the MOLREP computer program (Vagin and Teplyakov, 1997). The refined crystal structure of the GroEL protein, consisting only of the apical domain, solved to 1.7Å (Buckle *et al.*, 1997) was used as the probe structure throughout these final molecular replacement experiments. Figure 5.22.1(a) depicts a molscript-generated structure of the Buckle, Zahn and Fersht apical domain, Figure 5.22.1(b) depicts a molscript-generated structure of the plastid α subunits apical domain, the plausible model structure of which was gained via the Swiss Modelling Package on the EXPASY web site. Figure 5.22.1(c) depicts both models superposed with one another to show the differences of structural homology that exist between them. Due to fact that the plastid cpn60α model was structured on the available GroEL solved structures, the goodness of structural fit is not so surprising. However, the plastid cpn60 model does clearly show slight structural differences to occur at the sites of sequence differences between the two chaperonins.

The new search model introduced to the molecular replacement experiments consisted only of the apical domain, solved to 1.7Å, from the *E.coli* chaperonin GroEL. The atomic coordinates were placed in the protein data bank, December 1996 (accession number 1KID: P06139, Buckle*et al.*, 1997) and consisted of the apical residues 191 – 376, mutated with Ala 262 replaced with Leu and Ile 267 replaced with a methionine

The α subunit apical domain is the only domain of the three subunit domains (apical, intermediate and equatorial) which runs as an uninterrupted piece of the linear sequence, whereas the other two domains are composed of residues top-and-tailing the N-terminal and C-terminal sides of the apical sequence (Fig 5.22.2).

When directly compared to its GroEL homologue, it has been observed that the α subunit apical domain possesses less sequentially conserved regions than its two cohort domains. In part, this could be attributed as a direct consequence of the substrate binding role that the apical domain plays, as a part of the chaperonin mechanism, compared to the main role of its cohorts (the intermediate and equatorial domains) in driving the ATPase mechanism necessary for the chaperonins function whilst also sustaining structural integrity both in the monomeric form and as a functioning complex.

Therefore, it was decided to change these nonidentical residues, contained within the search model (1KID) pdb file, to alanine residues. In this way, the constraints forced on the search would endeavour to favour an overall main-chain structural search, which in turn would give a greater likelihood of a molecular replacement solution being found.

Fig 5.22.1(a) The GroEL Apical Domain

Depicted is a MOLSCRIPT generated picture of the GroEL apical domain, resolved to 1.7Å by Buckle, Zahn and Fersht, 1996.

Fig 5.22.1(b) The Plastid cpn60 α subunit Apical Domain

Depicted is a MOLSCRIPT generated picture of the α subunit apical domain, the coordinates of which were produced using the Swiss Modeller Package (EXPASY).

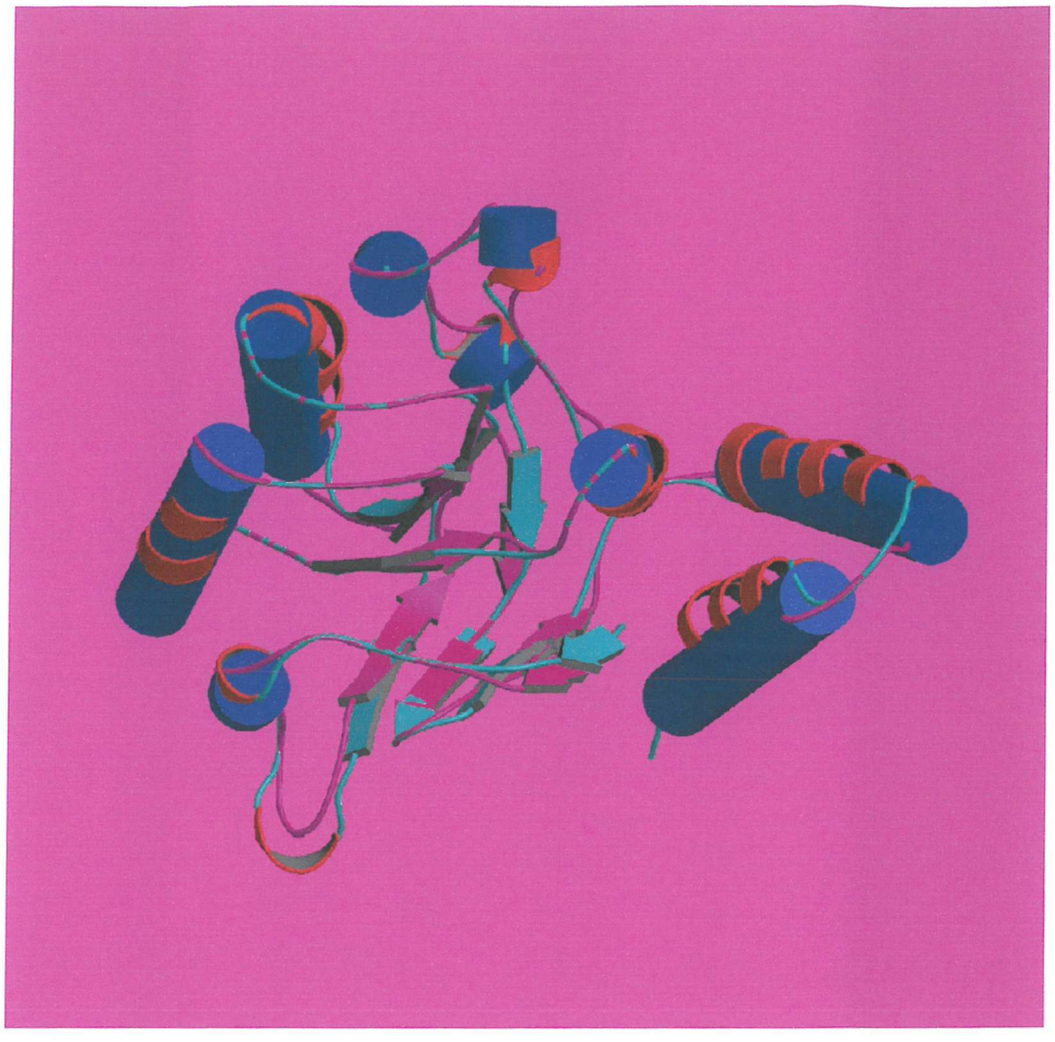


Fig 5.22.1(c) Superposed GroEL and α Apical Domains.
The α subunit is indicated by red helices is shown superposed onto the Buckle *et al.*, 1.7Å search model indicated by the blue cylinders. Evident differences of good structural alignment are apparent within the opposing β-strands and extended strands. This is due to the sequence difference's that exist between the two chaperonins, which was taken into account by the modelling package.

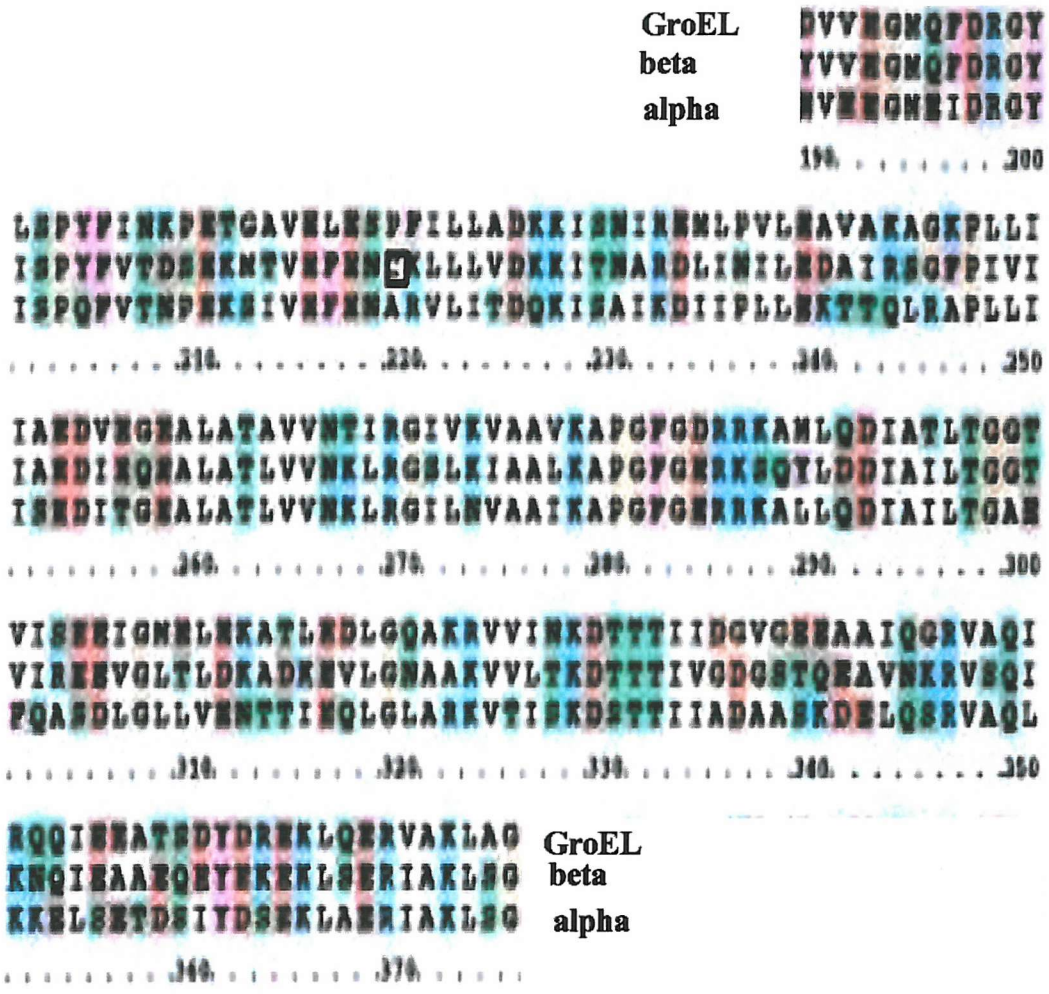


Fig 5.22.2 Apical Domain Sequences

Depicted are the residue alignments for the apical domains of the chaperonins GroEL (*E.coli*), and the α and β subunits for the chloroplast plastid chaperonin Ch-cpn60 (*Pisum sativum*). Due to the very nature of the apical domain, in that it is the substrate binding domain, it is reasonable to expect a greater range of residue variability when directly compared to its cohort intermediate and equatorial domains.

5.23 Cross Rotation Search

The rotation search was carried out using the Patterson search procedures in MOLREP. The rotation search was performed using reflections within the resolution range 3Å to 40Å, searching with a 'radius of integration' of 45Å. The twenty highest Patterson vectors in the range 3Å - 40Å were selected and rotated using the pseudoorthogonal Eulerian angles, (α, β, γ). MOLREP produced a sorted list of the correlation results, and the search yielded two significant solutions (α = 80.76°, β = 92.31°, γ = 215.22° and α = 77.77°, β = 85.10°, γ = 217.38°), Table 5.23.1.

Peak Number	α (°)	β (°)	γ (°)	Rf/σ
1	80.76	92.31	215.22	17.17
2	77.77	85.10	217.38	13.20
3	99.06	90.59	239.00	11.81
4	96.57	95.63	234.58	11.06
5	85.70	90.32	352.96	10.49

Table 5.23.1 The five highest peaks obtained from the cross rotation calculations performed between 3Å to 40Å.

5.24 Translation search

A translation search as implemented in the MOLREP program was used to find the molecular position of the now oriented search model in the α protein's unit cell. It was the highest peak from the cross rotation function, in conjunction with the translation search, that went on to produce a single position which emerged at X = 0.978, Y = 0.843, Z = 0.185 with a correlation coefficient (CC) of 22.1%.

Depicted in Table 5.24.1 are the orientation and positional coordinates for the gained molecular replacement solution of the Ch-cpn60 α subunit protein, using the apical domain of *E.coli*'s GroEL Group I chaperonin (Buckle *et al.*, 1996).

Function/ Peak No.	α (°)	β (°)	γ (°)	X	Y	Z	C.C.
Rot 1	80.76	92.31	215.22				
Tran 5				0.978	0.843	0.185	22.1

Table 5.24.1 The highest rotation and translation peaks.
The top peak, gained in the orientation search, was also chosen by the MOLREP program, to give the best position in the R32 space group.

The crystal packing of the solution was visually tested using the MOLPACK packing function program (Wang *et al.*, 1991). This showed that a sensible packing solution had been acquired and that no asymmetric subunit inter-contact clashes were apparent (Fig 5.24.1).

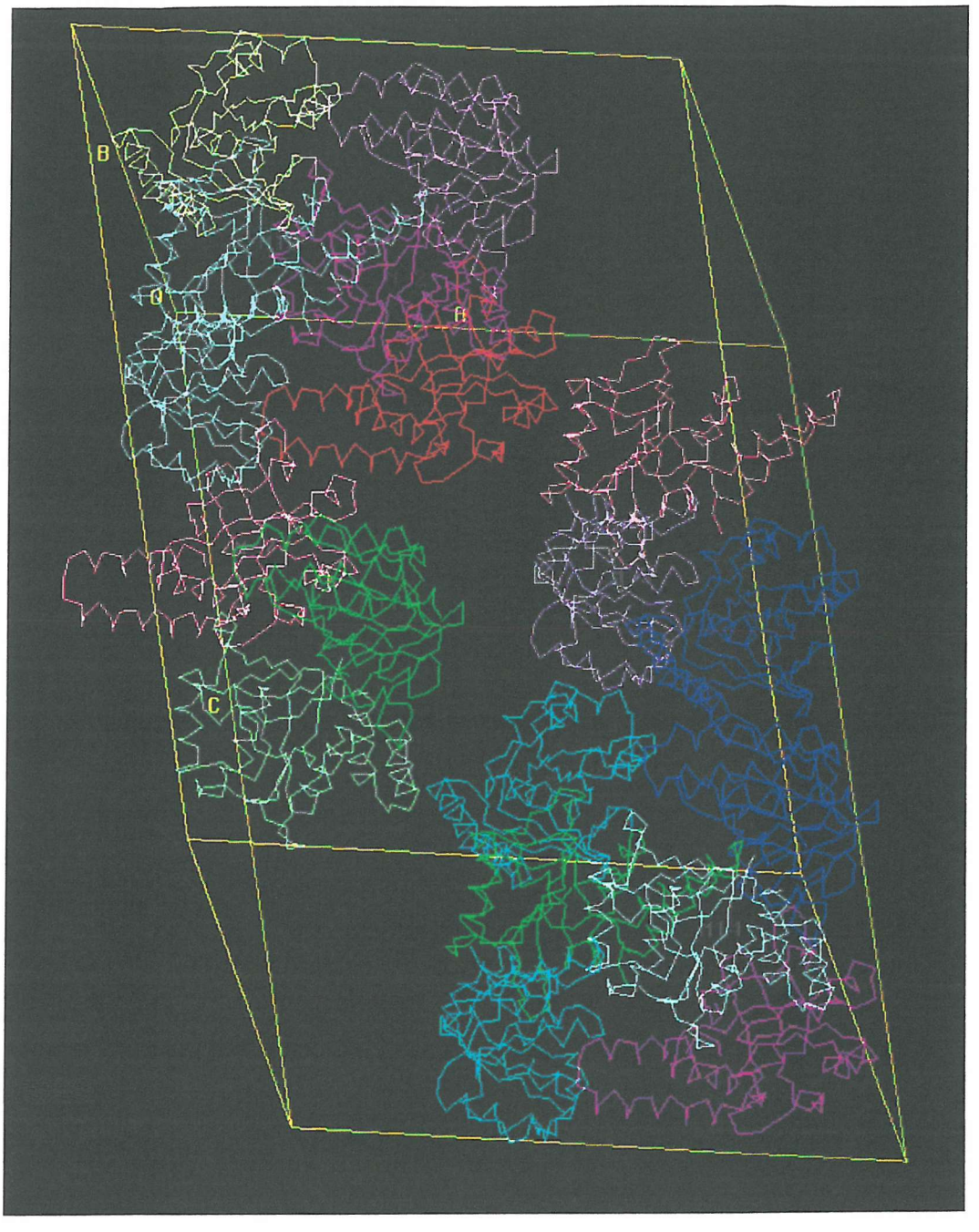


Fig 5.24.1 The crystal packing solution for the plastid chaperonins α subunit apical domain, gained from the *E.coli* GroEL subunits apical domain, used as the search model (MOLPACK; Wang *et al.*, 1991). The α subunit protein crystallises in the trigonal space group R32 (hexagonal, triple primitive, showing 18 symmetry equivalants), (a = 108Å, b = 108Å, c = 154Å, α = β =90°, γ = 120°). With approximately 47% solvent content, this allowed for approximately 30kDa of the ~58kDa subunit to be present in any one asymmetric unit.

5.25 Preliminary Structure Refinement of the Plastid Ch-cpn60 α Subunit Protein

The crystallographic refinement of the search models positions and orientations was carried out using a simulated annealing, slow-cooling, protocol (Brünger *et al.*, 1987, 1990), and using a resolution range of 2.77Å - 6Å, with a test set of reflections (6% of data) used for the R_{free} (non-biased) calculations. The model was subjected to an initial round of simulated annealing refinement (slow cool from 2000 to 300° K, in 25 K incremental decreases).

The newly acquired coordinates were then used to calculate a $2|F_o| - |F_c|$ electron density map, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes.

All model building was done on an SGI Workstation (Silicon Graphics) with the general purpose macromolecular modelling software package **O** (Kjeldgaard and Jones, 1995). Careful examination of the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ (a difference electron density map) maps were carried out at each refinement step. All residues were inspected on the graphics system at several stages throughout the refinement process. After each round of iterative refinement and model adjustment, as well as the R-factor and R_{free} scores, results were scrutinised against each newly acquired electron density map. In this way, the hope of immediately identifying and thereby eradicating the possibility of any progressive model bias influencing the refinement process was addressed.

5.25.1 Progress of Refinement

Table 5.25.1 shows the various stages of the refinement progress made after each model building experiment.

Stage	Number of reflections	R-factor	R-free	RMS deviations
1	1197	0.507	0.606	0.239
		MODEL	**REFINEMENT**	
2	1446	0.476	0.567	0.177
		MODEL	**REFINEMENT**	
3	1296	0.405	0.526	0.136
		MODEL	**REFINEMENT**	
4	1296	0.388	0.524	0.102
		MODEL	**REFINEMENT**	

Table 5.25.1 Shows key stages of refinement building for the α subunit protein.
The numbered stages (1,2,3,4) signify the simulated annealing (slow cool) events that were used to monitor the effects of any model bias that may have encroached upon the building process. The unnumbered stages, set between the simulated annealing events, signify the model building events that took place using $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron density maps.

In order to demonstrate the validity of the putative molecular solution it was necessary to present evidence that could stand alone from any density, which would be indicative of phase bias toward the search model, which was the conduit used in gaining the putative solution during the MOLREP protocols. The intention with the following pictures is to highlight regions within the electron density that present visual evidence of density that alludes to side-chain density that could not be filled with residues from the search model, due to the lack of sequence homology between the model and the molecule.

Initial examination of the maps revealed well defined $|F_o| - F_c|$ density for side-chain residues, in the apical domain, other than the density attributable to the *E.coli* 1.7Å apical search model. The maps also revealed a density for a partial α-helix, outside of the phase and structure factor amplitude influence of the search model used. This was encouraging as it intimated the validity of the molecular replacement solution, whilst further indicating that the calculated maps were not completely dominated by model bias.

Figure 4.25.1 shows an alanine at residue 229, which in the *E.*coli apical domain is a serine residue, and in the plastid Ch-cpn60 α subunit apical domain is a lysine residue. It is clear from the density surrounding the A229 residue that a larger residue belongs to this region, and indeed larger than the *E.coli* search model's serine residue at this position. Figure 5.25.2 depicts the inserted mutation of the A229 side-chain residue to a lysine and can be interpreted as fundamental in providing the evidence required to substantiate the acquisition of a correct solution.

Figure 5.25.3 depicts a proline residue (P200), which is correct for the plastid Ch-cpn60 α subunit protein. However, in the search model used (*E.coli* GroEL chaperonin, apical domain) the identified residue for this region is a lysine (K200). The electron density map showed a larger, rounder form of density that implied a better fit would be acquired for the insertion of a proline residue, rather than that of a lysine residue. These observations were once again paramount in proving the possibility of a gained molecular replacement solution for the target molecule involved, the α subunit protein.

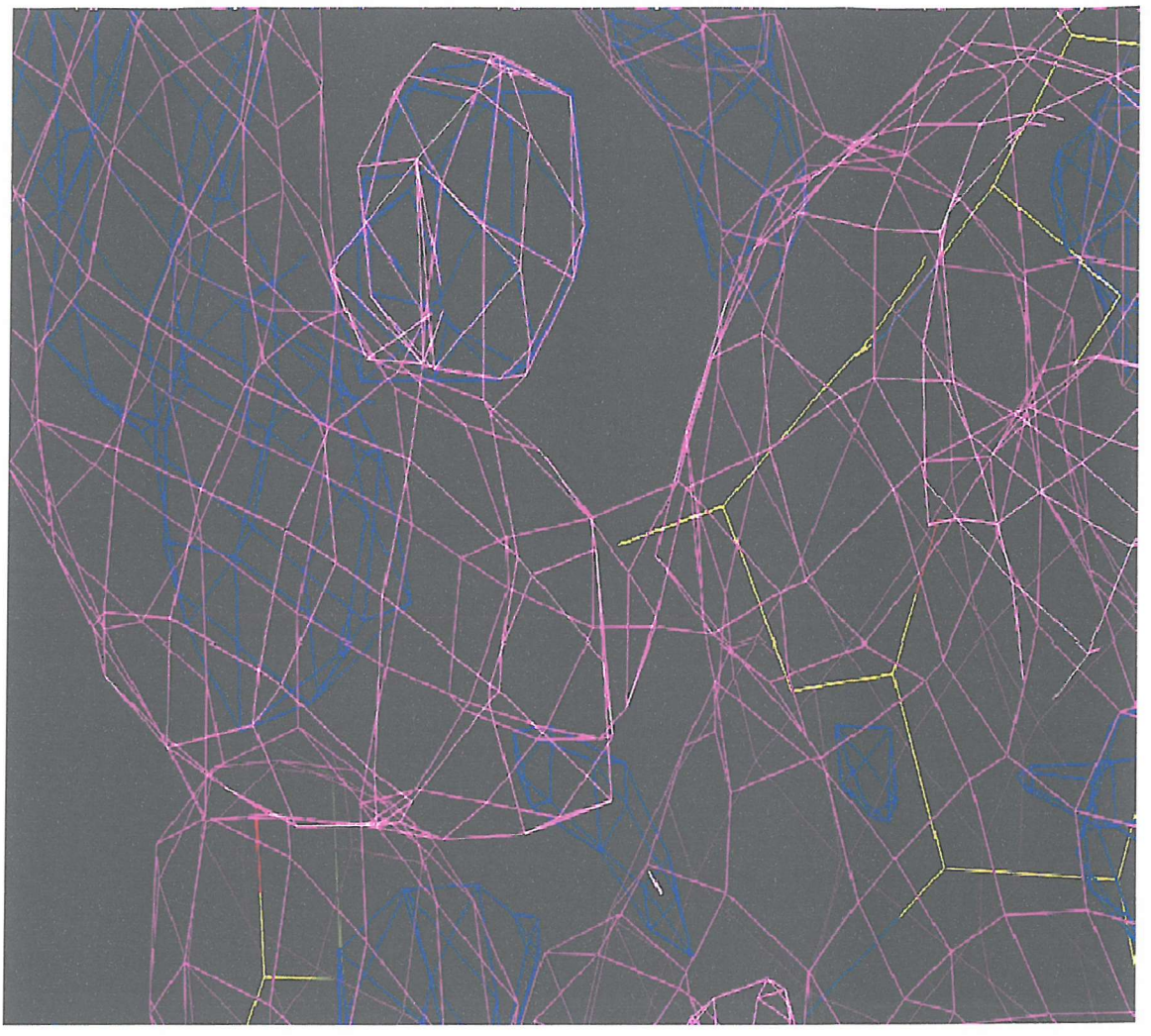


Figure 5.25.1 Good evidence that a correct molecular replacement solution for the α subunit protein was acquired. The residue (A229) depicted in this picture and contoured at 1 σ, taken from the modelling graphics package O (Kjeldgaard and Jones, 1995), is from the poly-alanine apical domain search model used to gain the putative molecular replacement solution for the α subunit protein. Note the excess electron density available in the map necessary to insert the side-chain residue relevant to the α subunit protein.

Figure 5.25.2 Depicted is the mutated insertion of the residue lysine 229 (K229), contoured at 1 σ.

In conjunction with figure 5.25.1, this picture intimates the correct placement of the residue relevant (K229) to the α subunit protein. In so doing, this inserted residue, and the many more applicable residues eventually inserted, that replaced the poly-alanine state of the search model, can be deemed to present good evidence of the acquired solution.

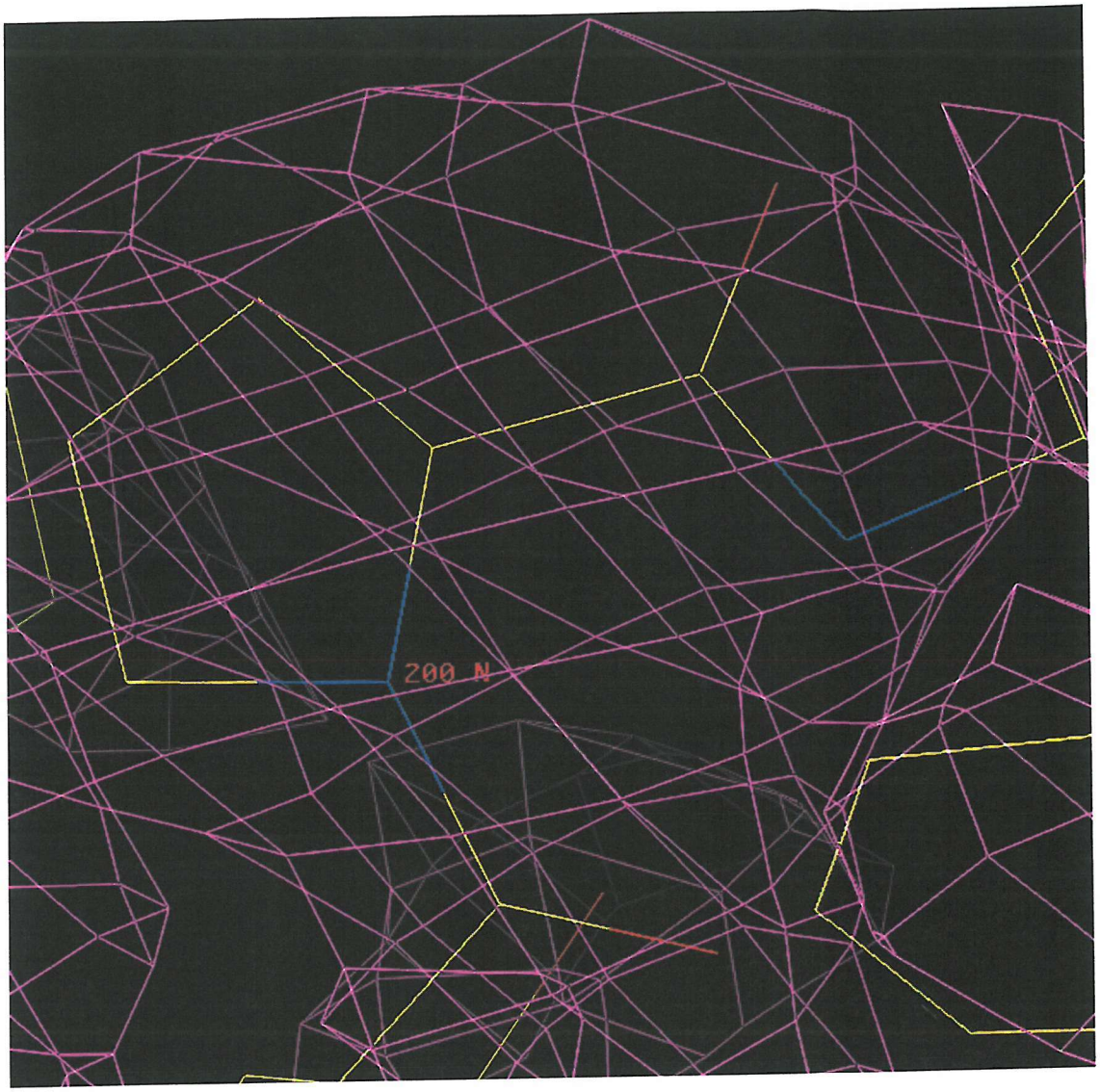


Fig 5.25.3 Depicted is the insertion of the residue proline 200 (P200), contoured at 1 σ. This picture shows good electron density coverage for the bulbous proline residue.

5.26 General Discussion

The plastid Ch-cpn60 α subunit protein crystallisation condition screening was initially performed using the Crystal Screen I and II kits (Hampton Research, USA). Crystallisation was performed using the hanging-drop vapour-diffusion method. However, it was not until some two years later that two crystals were grown and reached dimensions significant enough to be determined as diffraction-quality in form and size. Following further optimisation trials of the growth conditions, highlighted by the attainment of the first crystals, small crystal growth appeared within nine months of the new screens. However, although appearing to be rhombohedral in shape, and therefore resembling the first successful crystal acquisition, the crystals were subsequently monitored for a further year in the hope of gaining better diffraction-quality crystals. However, the anticipated crystal growth enlargement did not occur. The screens showed two size-forms of the crystals acquired. The first were decided to be only large enough to be used as 'seeding' crystals for further experiments, and the second type were harvested for an X-ray diffraction experiment. These second, slightly larger crystals diffracted to give data to 4Å.

Observation of the second optimisation growth screens also showed the protein drops to contain a rather thick film of denatured protein and precipitate, suggesting that the crystallisation solution may have provided a further purification step to the crystallisation process, which could be seen as indicative of the nature of the protein being handled. Quite simply, the function of the chaperonin is to bind small peptides at any given opportunity and so a certain amount of heterogeneity of the protein preparation and purification could have been expected to occur.

Data collection and subsequent processing for the initial diffracted data (from the crystals grown by Dr Hidalgo, Southampton University) showed the crystallisation of the alpha subunit had occurred in the space group R32 and resolved to 2.8Å. This was to be repeated for the second crystal growth experiment, which resolved to 4Å, again in R32.

However, completed processing of the data quickly showed that it was not possible that a complete subunit of the α protein had been crystallised, in all likelihood due to putative proteolytic cleavage having occurred during crystallisation of the α subunit protein. It was estimated that the diffracted data represented only approximately 50% of the full subunit, which equated to some 30-33kDa of the total protein Mw of ~58kDa.

It quickly became apparent that the acquisition of a solution for the α subunit protein, using the whole GroEL subunit homologue as a search model, faced what was to become insurmountable problems, certainly using the molecular replacement method. No discernible peak solutions were obtained with the use of any of the available programs, space groups, or models. The approach of using separate search model domains added to, rather than detracted from the problems encountered. Initial data gained for the cross-correlational experiment gave hope for this approach, which had been supported by, and implemented via, the modeling of the composite PDB file with respect to the two search model domains (apical and equatorial). However, disappointment again followed with the resultant rotation function analysis of the composite model, which showed once more that no feasible solution had been acquired.

It was not until the introduction of the last search model to be used in the molecular replacement trials, the apical domain fragment from the *E.coli* chaperonin GroEL, solved to1.7Å (Buckle *et al.,* 1997), consisting of the apical residues 191 – 376, that a reasonably provocative rotation solution was observed. It was the acquisition of this rotation solution that led to a translation solution being acquired, which when depicted in the MOLPACK program, showed no apparent clashing between molecules, within the unit cell. When it is considered that some 18 symmetry equivalent molecules are resident in R32, to detect no clashing was no mean feat of the putative solution.

However, it was not until the introduction of electron density maps, calculated with $2|F_o| - |F_c|$ coefficients, that a putative molecular replacement solution was clearly revealed. From this juncture a preliminary refinement and model building experiment could be attempted. Several successive and iterative rounds of refinement (via simulated annealing) and model building (via the O program) were carried out with the express intention of improving both the R factor and the interpretation of the model between refinement rounds. The putative degradation of the subunit was to become the major hurdle to fully solving the structure as well as interpreting the density maps.

Had there been the presence of a full alpha subunit, it is likely that greater advances with the refinement and modelling processes would have occurred sooner, as well as taking the R'factor to a reasonable level of acceptance. However, if cleavage of the subunit had occurred, and it is likely evident that this event did occur, then it was not likely apparent to which areas of which domains that cleavage would have occurred. This in turn, would have had a proportional effect on the ability of both the molecular replacement programs used, as well as the models used, which would be manifest in the 'actual' structural presentation of the protein, as opposed to the 'natural' form of the alpha subunit. As for the appearance of the electron density maps, when interpreting the electron density outside of the influence exerted by the search model, this presented a scenario of difficulties. At this juncture, one was forced to consider not only if the pertinent density being interpreted was representative of non-model bias (due to being outside the bounds of the search model), but also to which domain, if any, outside the apical domain that this particular piece of tangible density belonged to.

From the evidence presented in this chapter it is likely that some form of molecular replacement solution has been gained for the plastid Ch-cpn60 α subunit protein. That said, it would be hard to assuage my conscience that a full and resolute solution has been acquired, sufficient to fully interpret and solve the structure. Finally, matters were made more difficult by the protein eventually crystallising in such a convoluted space group as R32. It is felt that the requisite growth conditions required for the protein to reach mature, diffraction-quality, crystals precludes any future attempt at using molecular replacement as a preferred method of choice for solving the structure of this, perhaps too easily cleaved, protein.

CHAPTER SIX

6.1 Multi-Wavelength Anomalous Dispersion Experiment for the Ch-cpn60 α Protein

The Multi-wavelength Anomalous Dispersion (MAD) method was initially introduced to solve the phase problem that besets crystallography. The review by Fourme and Hendrickson, (1990), called attention to the phenomenon of anomalous scattering and the important role it could play in crystallography.

The MAD approach to macromolecular structure determination has potential advantages for both accuracy and convenience in phase evaluation. Isomorphism is exact with this method, presenting the possibility of an algebraically exact analysis; relative scattering power and phasing power increase with the scattering angle, and all required diffraction data can be measured from a single crystal.

Cowie and Cohen, (1957) in their work on the mechanistic biosynthesis of active proteins containing selenium substituted sulphur, demonstrated that methionine could be replaced by selenomethionine (SeMet) in *E.coli*. This approach to SeMet analog substitution was eventually fostered by the MAD method (Hendrickson *et al.*, 1990), in a bid to conquer the phase problem. Selenium replacing sulphur in the methionine residues became recognized as a suitable resonance for MAD experiments, which require synchrotron radiation with precise beam wavelength control. The X-ray absorption edge of selenium (λ = 0.98Å) is easily accessible with synchrotron radiation. At this wavelength radiation damage is limited and X-ray absorption, by the specimen, is low.

Since selenium is similar to sulphur in its chemical, physical properties and electronegativity ($18e^-$), it is reasonable to expect that the native protein and its fully substituted SeMet analogs exhibit similar properties (Budisa *et* al., 1995, Doublie, 1997). This presents a positive step forward for the production of derivative crystals that are wholly isomorphous with native crystals.

The proposed SeMet bioincorporation into the protein represents a general approach to the MAD method, since several requirements are fulfilled: (i) convenience of the labeling procedure, (ii) high isomorphism with the parent molecule, (iii) stability of the amino acid analog, (iv) phasing power, with relative abundance and mobility of the target sites (Hendrickson *et al.*, 1990).

The α subunit protein offered itself as a suitable candidate to the MAD method, possessing a suitable number of native methionines (6 per subunit). The aim of this experiment would be to solve the X-ray crystallographic structure of the plastid Ch-cpn60 α protein.

The procedures for engineering and crystallizing selenomethionyl proteins are relatively straight forward and can be divided into 4 steps; expression, cell growth, purification, and crystallization (Budisa *et al.*, 1995).

6.2 Methods and Materials

6.2.1 Microbiological techniques

6.2.2 Sterilisation

All pipette tips, media, stock solutions, and microfuge tubes were sterilised by autoclaving at 120°C for 90 minutes. Stocks of heat labile solutions such as antibiotics and IPTG were sterilised by filtration through disposable 0.45μm Millipore filters.

6.2.3 Bacterial strains

Strain	Genotype	Special Features
B834 (DE3) pLysS	F⁻*ompT* *hsdS*$_B$($r_B^- m_B^-$)*gal* *dcm* *met* (DE3) pLysS (Cm^R)	Met auxotroph allowing high specific activity labelling of target protein with selenomethionine for crystallography.

6.2.4 Cloning vectors

Vector	Fusion Tags N-term C-term	Special Features
pET-24a	T7.Tag His.Tag Metal chelation chromatography	As with pET3a, but confers Kanamycin resistance.

6.2.5 Culture media

Cultures were grown in Luria Broth, 2TY and/or NMM (New Minimal Media). All were supplemented, when necessary, by the aseptic addition of the relevant antibiotic at the following final concentrations; ampicillin 50μg/ml, chloramphenicol 35μg/ml, kanamycin 25μg/ml.

Media	Materials	g/litre
Luria Broth (L-broth)	Tryptone Yeast extract NaCl	10 5 5
2TY	Tryptone Yeast extract NaCl	16 10 5

Media	Materials	mM/L	mg/L	μg/L
NMM	Ammonium sulphate	7.5		
	NaCl	8.5		
	KH_2PO_4	55		
	K_2HPO_4	100		
	Magnesium sulphate	1		
	D-Glucose	20		
	Ca^{2+} ions		1	
	Fe^{2+} ions		1	
	Cu^{2+} ions			1
	Mn^{2+} ions			1
	Zn^{2+} ions			1
	MoO_4^{2-} ions			1
	Thiamine		10	
	Biotin		10	
	18 $_L$-amino acid mix ($_L$met-) in		50	
	NaCl	137		
	KCl	2.5		
	Na_2HPO_4	10		
	KH_2PO_4 (pH7)	1.76		
	Kanamycin	25⇐μg/ml		
	Chloramphenicol			34μg/ml

Bacterial colonies, post transformation, were grown on plates of L-broth that had been solidified by the addition of 1.5% w/v of agar No.1.

6.2.6 Transformations

Transformations for the vector (pET24a), into the required bacterial strain *E.coli* B834 (DE3)pLysS *met* ⁻), were performed following the manufacturers (Novagen) recommendations, and using the cold/heat shock method. The host bacterial strain was supplied in the competent state by Novagen.

All transformations were carried out on ice and the following procedure used throughout. 20µl aliquots of the competent cells (B834(DE3)pLysS) were pipetted into pre-chilled (ice) 1.9ml snap-cap polypropylene micro-centrifuge tubes. 1µl of purified plasmid DNA (pET24a) was added directly to the competent cells and pipetted gently to mix. The transformation mix was incubated on ice for a further 5 minutes. The tubes were then heated in a water bath, for 30 seconds at 42°C. The tubes were then returned to ice for a further 2 minutes. 80µl of supplied growth medium (SOC) was then added to the mix and those plasmids containing kanamycin resistance (pET-24a) were the shaken at 37°C (250rpm) for a further 30 minutes prior to plating. The cells were then centrifuged in a bench microfuge, at 1300rpm, for 1 minute and the supernatant removed. The pellet was resuspended in 200µl LB (Luria Broth) to concentrate the sample. Various quantities of these stock cells (5µl, 20µl, 100µl) were then spread onto LB-agar plates containing the respective selective antibiotic (Kanamycin at 25µg/ml for pET-24a) and then incubated at 37°C overnight. A positive control, using a supplied test plasmid, was used in conjunction with all transformations.

6.2.7 Trial expression and optimisation of the α subunit

The analog Seleno-methionine is naturally toxic to cells. To gain eventual incorporation into the protein of choice, the cells must undergo an initial phase of growth in non-harsh conditions. Therefore, very low levels of LB containing L-methionine are initially introduced into the NMM growth media, favouring normal cell growth conditions. Equally important is the optimisation of these initial LB concentrations whereby, enough $_{L}$-methionine is being added to drive the translation phase of expression, but not so much that the introduction of the seleno-methionine is not out-competed for the methionine sites available in the fully expressed α subunit. Optimisation trials to ascertain the correct amount of LB to use, in conjunction with the NMM and incorporated seleno-met, were carried out.

Depicted below are several small-scale growth media trials (A-F, plus control)

NMM Variant	Basic NMM mix	+30mM stock $_{DL}$-Seleno-Met	+water	100% LB added
A	3.11ml	0.1ml	6.74ml	-
B	3.11ml	0.1ml	5.79ml	0.95ml
C	3.11ml	0.1ml	4.79ml	1.95ml
D	3.11ml	-	6.84ml	-
E	3.11ml	-	5.89ml	0.95ml
F	3.11ml	-	4.89ml	1.95ml
Control	-	-	-	9.95ml

All trials included 50μl of O/N (over night) starter culture. $_{DL}$-Selenomethionine was prepared commercially by Sigma (S-3875).

The effect of various concentrations of $_{L}$-methionine (incorporated in the LB media) was examined in the presence and absence of $_{DL}$-seleno-methionine, to determine a limiting concentration of LB which could be used to ensure initial cell growth, later to become depleted, thereby allowing the eventual incorporation of the selenomethionine into the cellular protein product.

To each 9.95ml growth medium, 50µl of overnight cells grown in 100% LB was added, making a final growth volume of 10ml, which also adjusted the LB levels in each reaction to either 0.5%, 10%, or 20%. These samples were then incubated at 37°C (250rpm). Samples were induced with IPTG to a final concentration of 1mM at $O.D._{600nm}$ 0.6-0.9, and cells were harvested 4 hours post-induction, except tube A where harvest was 8 hours post-induction and tube D where no IPTG was added. Harvested cells were mixed 1:1 with SDS-PAGE 2x sample buffer and the α subunit was visualised using 4-12% Nu-Page SDS-PAGE Bis-Tris gels (Invitrogen) run at 190mV.

6.2.8 Large-scale expression / purification of seleno-met incorporated α subunit

A 10ml o/n culture of the *E.coli* B834(DE3)pLysS cells, containing the α subunit construct, was grown up in 100% LB medium at 37°C (250rpm). 4ml (plus 4ml LB) of the o/n culture was added to 794ml of NMM containing 0.3mM $_{DL}$-selenomethionine. Cells were then incubated again at 37°C (250rpm) and at $O.D._{600nm}$ 0.6 induction using IPTG (final concentration 1mM/ml) was carried out. Cell harvest was at 8 hours post-induction by centrifugation at 7227xg, at 4°C for 30 minutes. Cell pellets were resuspended in degassed (minimising selenium oxidation) lysis buffer (50mM potassium phosphate pH7.5, 50mM EDTA, 5mM fresh PMSF, and 5mM β-mercaptoethanol) and sonicated (15 seconds on, 30 seconds off, x 12 cycles). Cell debris was spun down at 7740xg for 30 minutes at 4°C and the supernatant removed and retained (representing the soluble portion of the selono-met incorporated α subunit). The pellet form was also retained (representing the insoluble portion of the seleno-methionine incorporation).

6.2.9 Q-Sepharose anion exchange purification

Ion exchange chromatography is capable of separating molecules that have only slight differences in charge to give a high-resolution separation. This technique is most suited for the capture or intermediate steps in purification. The separation is based on the reversible interaction between a charged molecule and an oppositely charged chromatographic medium.

The particular ion exchange column used for the initial purification of the α protein was the XK26 'HiLoad 26/10 Q Sepharose Fast Flow' anion exchange column (Amersham Pharmacia Biotech). The dominant application of this column is the ability for fast separations of crude samples (100ml loads were used), whilst being ideal for scale up. It has a working pH range of 2-12, with a recommended working flow rate of up to 18ml/min.

6.2.10 Ion exchange purification for the Seleno-Met incorporated α protein

The supernatant, from the lysis and sonication, was loaded onto a pre-washed, pre equilibrated Q-Sepharose anion exchange column (XK26/10 Amersham Pharmacia), using the Pharmacia FPLC (Fast Protein Liquid Chromatography) system. Anion exchange was performed using degassed anion exchange buffer, with a NaCl (sodium chloride) gradient of 0-0.5M, 500ml (no salt buffer) and 500ml (NaCl buffer). The flow rate was 5ml/min and 10ml fractions were collected. Samples from prominent eluted column fractions were resolved on a 4-12% NuPage SDS-PAGE for initial analysis of the success of the experiment. Seleno-Met/α protein samples were then put into an aqueous buffer solution, containing 10mM DTT (Dithiothreitol) added to prevent oxidation of the selenomethionine, and the samples were then prepared for further purification and separation, using SEC (Size Exclusion Chromatography or Gel filtration).

6.2.11 SEC (gel filtration) purification of the Seleno-Met incorporated α subunit

SEC was performed on the Superose 6 HR 10/30 column (Pharmacia), which separates proteins in the 5000 – $5x10^6$ Da range, using the Pharmacia FPLC sysytem. After the column had been equilibrated with the requisite buffer (20mM Tris.HCl pH 7.5, 5mM EDTA, 5mM β-Mercaptoethanol), samples of 100μl were applied to the column at a flow rate of 0.2ml/min. Samples were collected in 1ml fractions and the appropriate fraction peaks resolved and analysed on a 4-12% NuPage SDS-PAGE.

6.3 Choice of Bacterial Host

The host bacterial strain used to carry the α protein was necessarily required to be auxotrophic for methionine (Met^-), as the growth medium to be used would already have selenomethionine incorporated. In this experiment the host strain of choice was *E.coli's* B834 (DE3)pLysS (Novagen catalogue #:69042-3), which was auxotrophic for methionine. The designation (DE3) indicates that the host is a lysogen of DE3, carrying a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. Such strains are suitable for production of protein from target genes (the α cDNA) cloned in pET vectors. pLysS is the designation given to hosts carrying pET compatible plasmids that encode T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase. These strains are used to suppress basal expression of T7 RNA polymerase prior to induction and thus stabilize pET recombinants encoding target proteins that affect cell growth and viability. The tight control of background expression and efficient control of the cloned α gene insert was necessary. This was imperative, due to previous evidence showing the existence of high background expression, where the labeled protein was found to be contaminated by the parent protein as well (Budisa *et al.*, 1997)

The experimental approach taken to facilitate the uptake of the SeMet analog, replacing the native methionines, was to carry out an intial growth phase of the bacterial host (B834) in a minimal media, deficient of methionine but with a limited source of sulphur in the form of Luria Broth. The exact composition of growth medium influences expression levels and so trials were carried out to determine the optimum levels of new minimal media (NMM), to use (Budisa *et al.*, 1995).

6.4 Transformation and Expression Trials into *E.coli's* B834 Met$^-$

Transformation of the pET24a/α gene construct into the B834 (DE3) pLysS host was carried out in accordance with the supplier's recommendations. The transformed product was placed on LB-agar plates and incubated at o/n 37°C. Selected colonies were grown in fresh LB medium to check the expression of the α protein in the new *E.coli* host cell. Evidence for the success of the transformation and expression was acquired by SDS-PAGE analysis. Trials were next carried out to optimize growing conditions for the incorporation of the SeMet analog. The large-scale expression and purification of the SeMet incorporated α protein was carried out in a final growth volume of one litre. Again expression evidence was sought by taking 1ml of the expression culture which was analysed on a 4-12% SDS-PAGE (Nu-PAGE), Fig 6.4.1 shows the SDS-PAGE analysis for pre- and post induction of the SeMet incorporated α protein.

6.5 Purification of the SeMet Incorporated α Protein

Purification of the SeMet incorporated α protein should be similar to that for the native protein, with the exception of the protein having become a more hydrophobic and less soluble form (Doublie, 1997). To avoid oxidation of the SeMet into its seleno-oxide equivalent (Yang *et al.*, 1990), all buffers used throughout the purification process were degassed. In addition, dithiothreitol (DTT) (5-20mM) and the chelating agent ethylene diamine tetraacetic acid (EDTA) (5-50mM) were used (Scopes, 1982).

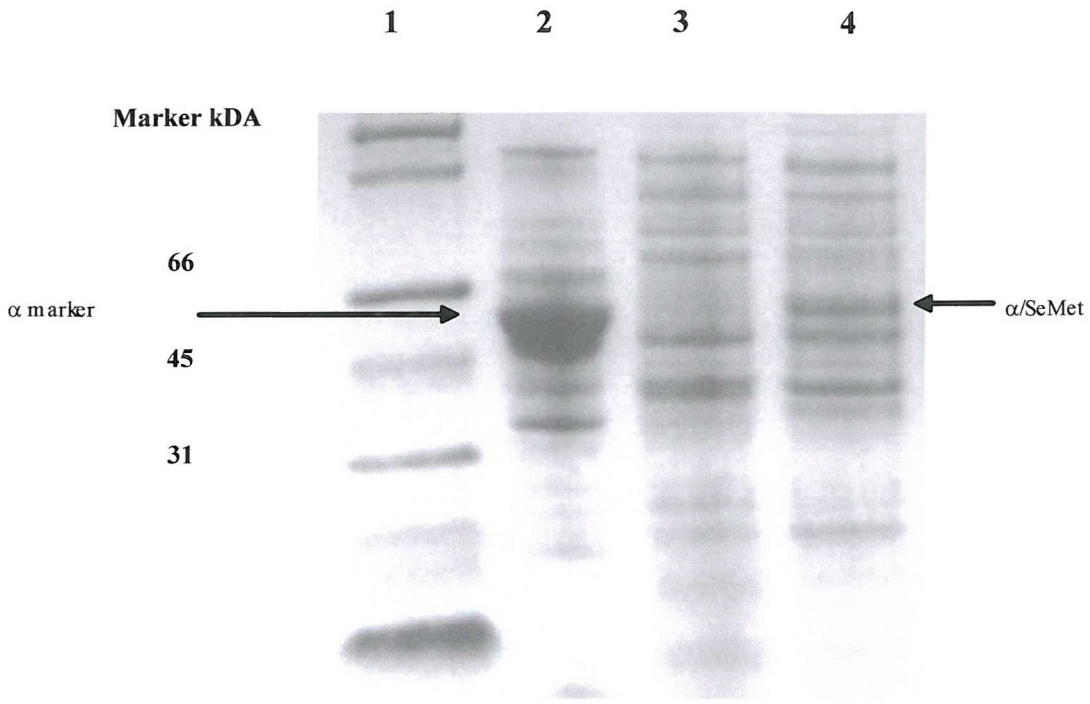


Fig 6.4.1 SDS-PAGE Expression Analysis for the SeMet Incorporated α Protein

Lane 1: Protein marker. Lane 2: α protein used as a reference marker. Lane 3: depicts a pre-induction sample, which shows no expression of the SeMet incorporated α protein. Lane 4: shows good evidence that the α protein has been expressed in a bacterial host environment lacking in natural methionines (B834) and rich with the SeMet analog. This would suggest that incorporation of the SeMet analog into the α protein has been successful.

6.6 Cell Lysis

The 1litre culture, containing the α-SeMet reaction mix, was centrifuged at 14000rpm for 30 mins at 4°C, in a 1ltr container. The resultant pellet was retained and the wet cell mass was 3.22g/litre. The pellet was then resolubilised in 20ml of degassed RIPA (lysis) buffer:

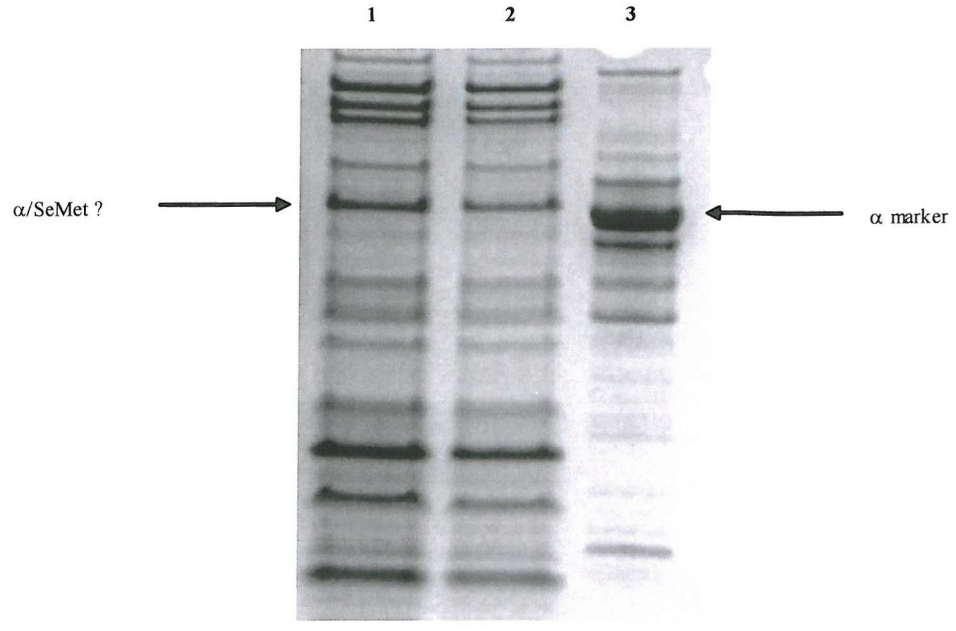
0.1% SDS (Sodium Dodecyl Sulphate)
1.0% Triton x-100 detergent
1.0% Sodium Deoxy Cholate
25mM Tris.HCl pH 7.5
150mM NaCl
1mM PMSF
25mM EDTA

To gain maximum cell lysis the resolubilised pellet was then sonicated; 15 seconds on: 30 seconds off; x 12 cycles.

6.7 Ion Exchange Purification

The supernatant obtained from the lysis/sonication phase was then loaded onto a pre-equilibrated Q-Sepharose anion exchange (IEX) column (XK26/10 Amersham Pharmacia).

10ml fractions were collected, pooled and then concentrated in an Amicon filter (30kDa membrane). Analysis of the resultant fractions was carried out on a 4-12% SDS-PAGE. The results of the gel analysis are shown in Fig 6.7.1.



6.7.1 SDS-PAGE Analysis of IEX Fractions for the SeMet Incorporated α Protein

Lanes 1 and 2: depict two of the pooled fractions from the same peak eluted at between 30-40ml on the ion exchange run. The elution point of these fractions directly relate to the known elution point for the α protein, when run on ion exchange, from previous runs carried out by Dr Hidalgo (Southampton University). Lane 3: α protein reference marker.

There appears to be mobility retardation for the ion exchange products against the α marker. However, this may be due to either overloading of the marker or the addition of the selenomethionine analogs may affect the charge of the protein in some way.

6.8 SEC (Filtration) Purification of the SeMet Incorporated α Protein

The samples were further purified using size exclusion chromatography (SEC). Resultant fractions were pooled according to their peak distribution profile and concentrated in Centricons (30kDa membranes) to a final volume of ~ 2000-400µl. These fractions were subsequently analysed on a 4-12% SDS-PAGE, Fig 6.8.1 depicts the results.

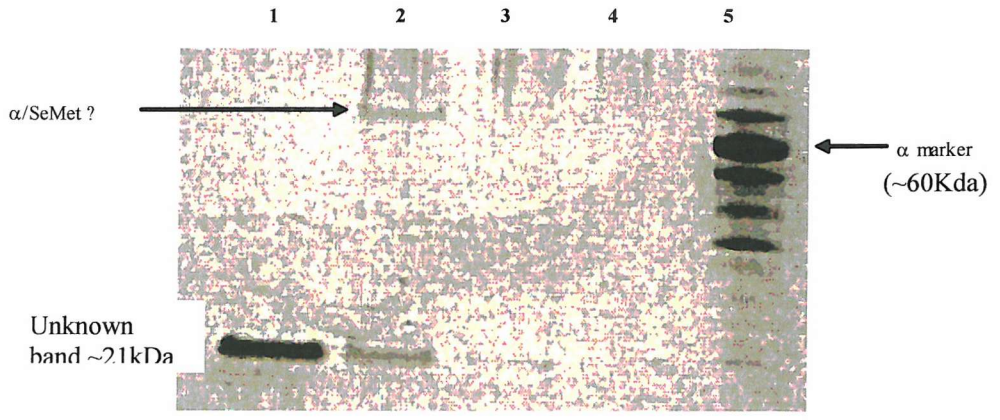


Fig 6.8.1 SDS-PAGE Analysis of SEC Fractions for the SeMet Incorporated α Protein

Lanes 1-4 represents the various eluted peaks (and two unknown bands at ~21kDa, lanes 1 and 2). Lane 2: depicts the α protein, which eluted at the appropriate point of ~13ml off the column. Again there is mobility retardation for the protein when aligned with the α subunit (reference) marker in Lane: 5. Most notable, at this juncture, was the heavy loss of α/SeMet product in comparison to the gained product post-ion-exchange (Fig 6.7.1). Further, it is apparent that a lower Mw protein (lane:1) has been expressed in abundance.

6.9 Analysis of the SeMet Analog Incorporation into the α Protein

Analysis as to the success of selenomethionine substitution for the native methionines of the α protein was carried out by time-of-flight electrospray-mass spectroscopy (TOF MS-ES). This method allows the differences in mass, presented by each SeMet analog (~47Da), to be accurately measured by MS-ES.

The calculated molecular mass (average mass) for recombinant α protein is 56911.14Da. The mass measurements acquired from MS-ES for the α subunit prior to SeMet incorporation was 56857.05± 0.44Da, giving a mass difference of -54.09± 0.44Da between estimated and MS-ES masses, Fig 6.9.1 shows the ES-MS profile for the native α protein.

The mass difference between methionine and selenomethionine is 46.89Da and so for all six methionine residues of the α protein to be SeMet incorporated an overall mass increase of 281.34Da would be expected for the protein.

The sample representing the SeMet incorporated α protein was then applied the MS-ES for analysis, Fig 6.9.2 depicts the MS-ES profile obtained.

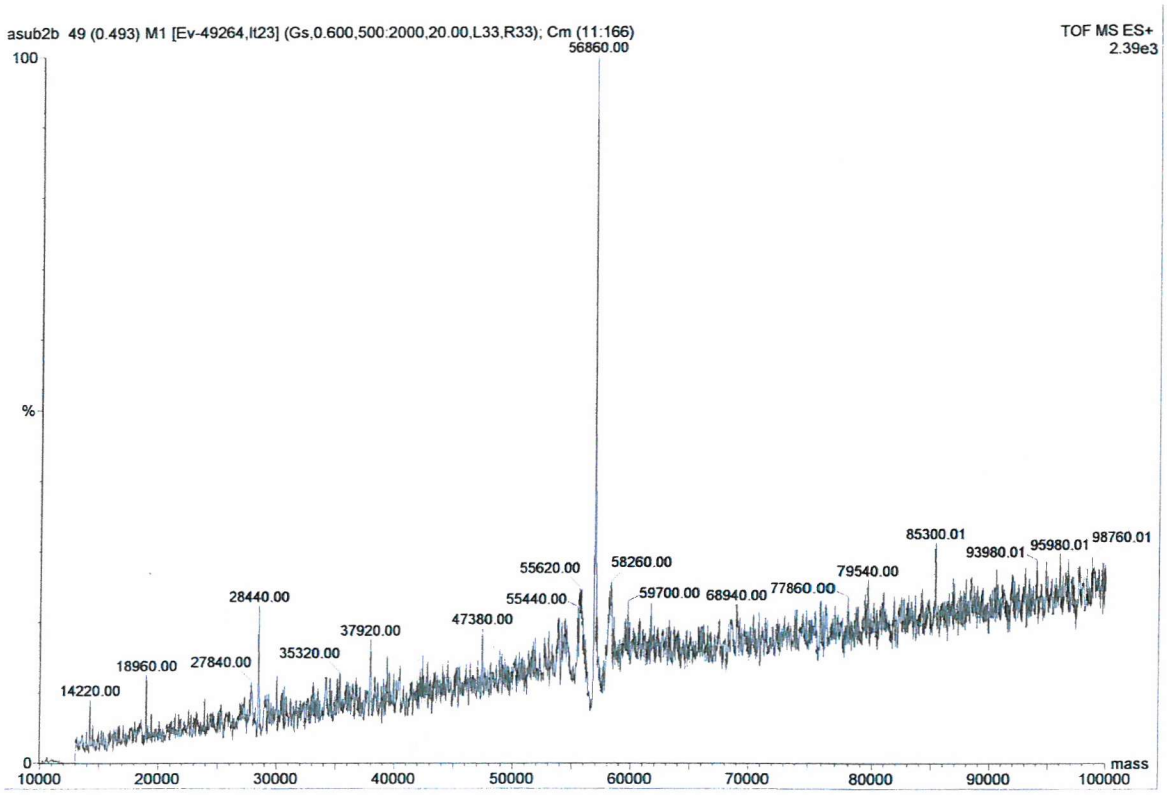


Fig 6.9.1 MS-ES Profile for the α Protein (Native)

This profile represents the mass of the α protein (56860Da) without the incorporation of the selenomethionine. This profile would be used as a reference against which the SeMet incorporation trial could be estimated.

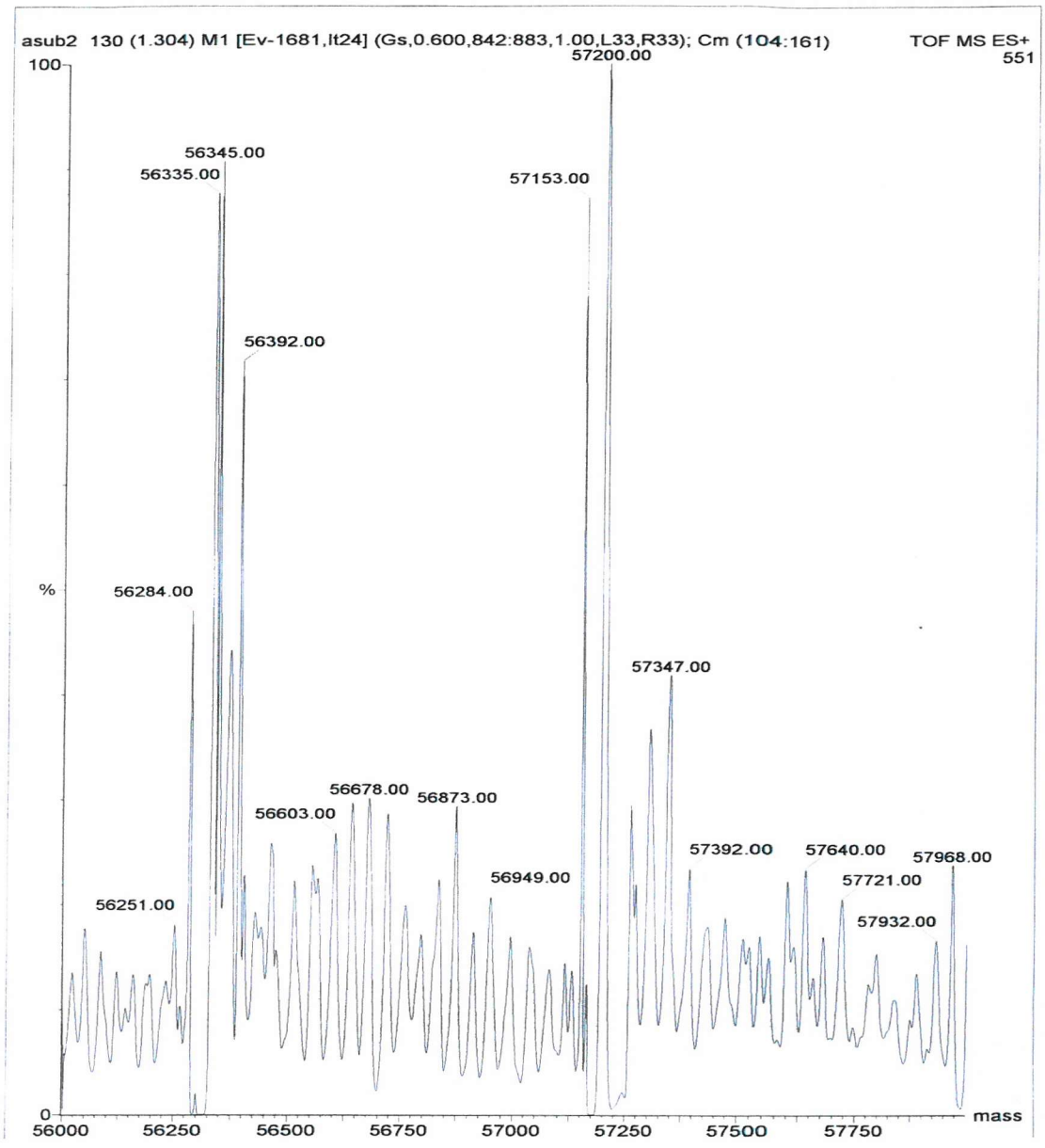


Fig 6.9.2 MS-ES Profile for the SeMet Incorporated α Protein

This profile represents the α protein with putative SeMet incorporated substitution of the natural methionines. However, there is no definitive evidence regarding SeMet incorporation with respect to the expected mass range (56860Da), as opposed to the mass range (57200Da) obtained from this profile.

6.10 General Conclusions

To initiate this trial it was first necessary to transform the plasmid (pET24a), carrying the cDNA form of the α protein (previously reconstituted and expressed, refer to Chapter Two), into a methionine auxotrophic bacterial host (B834). There was strong evidence that this initial part of the experiment was carried out successfully (Fig 6.4.1).

The mass spectrometry (MS-ES) profiles (Figs 6.9.1 and 6.9.2), between the α native protein mass and the expected mass increase for the α-SeMet protein, were inconclusive. In consequence, and since growth time for any α protein crystal (suitably large enough for X-ray diffraction) is known to extend over a substantial time period, it would not be prudent to lay a crystal screen on the strength of the evidence presented, unwisely raising the hope that a future MAD experiment would either be successful or assured.

The lysis buffer, which contained two strong detergents (SDS and Triton T-100), proved inappropriate for this experiment. The Triton X-100 carries a sufficient charge down the mass spectroscopy (28kDa) that, when present in abundance, would likely mask the profile of the SeMet incorporated α protein.

Notwithstanding the questions and concerns raised by the detergent issue, further concerns were raised over the evident and unexplained low yield of the final α/SeMet purified product, which was not commensurate with the initial expression profiles acquired (Fig 2.6.1, Chapter Two and Figs 6.4.1 and 6.7.1). However, there was strong evidence produced for the presence of expressed SeMet incorporation at the final conclusion of the experiment (Fig 6.8.1). Throughout the purification process there was evidence presented in the gels relating to the mobility retardation of the SeMet incorporated α protein, which could have represented a cause for some concern, had strong evidence for the α proteins expression not been presented on the SDS-PAGE in Fig 6.8.1.

Therefore, it is likely that on the evidence presented here, future experiments would be encouraged to use a cell lysis protocol that did not include the presence of strong, detrimentally charged, detergents at such high concentrations.

CHAPTER SEVEN

Steady-State Kinetic Analysis of the Nucleotide Induced Allosteric Transitions of the Plastid Ch-cpn60$^{\alpha/\beta}$ Chaperonin

7.1 Introduction

Protein folding in the presence of the chaperonin GroEL is thought to proceed by a cycle of substrate binding, encapsulation by GroES, followed by release from the cavity (Weissman *et al.*, 1995,1996, Burston *et al.*, 1996, Corrales and Fersht, 1996, Rye *et al.*, 1997). Each opposing heptamer of the ring complex acts alternately as a protein binding unit, with the affinity for non-native polypeptide regulated allosterically by nucleotide binding (Badcoe *et al.*, 1991, Jackson *et al.*, 1993, Staniforth *et al.*, 1994, Todd *et al.*, 1994). The ADP-bound state possesses a much higher affinity for the binding of the polypeptide than does the ATP-bound state. When GroES is bound, the protein-binding residues are set apart from the cavity surface (Roseman *et al.*, 1996, Xu *et al.*, 1997). Release of the encapsulated substrate protein and the co-cpn GroES from a ring is dependent on ATP hydrolysis on that ring (*in cis*) and subsequent ATP binding to the other ring (*in trans*) (Rye *et al.*, 1997).

GroEL has a high affinity for nucleotide (Gray and Fersht, 1991, Jackson *et al.*, 1993, Yifrach and Horovitz, 1994), as well as for its co-chaperonin GroES if nucleotide is bound (Jackson *et al.*, 1993, Todd *et al.*, 1994, Hayer-Hartl *et al.*, 1995). When substrate is bound this forms a *trans* complex with respect to GroES (Smith and Fisher, 1995) and when in conjunction with ATP binding in *trans*, accelerates the release of GroES and substrate, thus allowing the cyclic encapsulation process to start again (Rye *et al.*, 1999).

A titration of GroEL with ATP suggested that seven ATP's are hydrolysed in one turnover (Burston *et al.*, 1995), and when coupled with the observation of negative cooperativity between the rings (Yifrach and Horovitz, 1994), suggested that hydrolysis of ATP occurs on each ring alternately, further suggesting that one molecule of ATP is sequestered per subunit.

The GroEL chaperonin provides us with an interesting example of a macromolecular assembly that shows unusual allosteric properties caused by the interactions between its constituent subunits (Yifrach and Horovitz, 2001).

The allosteric control of the protein-binding sites by nucleotide and GroES binding is supported by the cooperativity exhibited in ATP binding. Positive cooperativity in ATP binding exists within the rings (Gray and Fersht, 1991, Bochkareva *et al.*, 1992, Jackson *et al.*, 1993, Kovalenko *et al.*, 1994, Todd *et al.*, 1994), and recent studies have revealed that the intra-ring allosteric transitions in GroEL are concerted (Shiseki *et al.*, 2001). The positive cooperativity in ATP binding can be approximated by the Monod, Wyman, and Changeau (MWC) model of cooperativity (Monod *et al.*, 1965). This model would suggest that each ring is considered a cooperative unit that can exist in either the T-state (Tense) or R-state (Relaxed) (Gray and Fersht, 1991, Yifrach and Horovitz, 1994. The negative inter-ring cooperativity is described as an inhibition of the ATPase function when ATP is bound to the second ring and results in an asymmetric oligomer where the two rings of the protein adopt different conformations (Burston *et al.*, 1995). It was these observations of positive and negative inter-ring cooperativity which led to the development of the 'Nested' model of cooperativity (Yifrach and Horovitz, 1994, 1995).

Conformational changes brought about by nucleotide binding and hydrolysis necessarily induce a time and space realignment of the chaperonin structure, which can be considered a mechanism for the ultimate control of its functional cycle. Given this context, it can be deduced that inter-ring allosterism is a critical factor when considering the continuous cycling of GroEL rings between protein accepting and releasing states, which still remains vague in comparison to our understanding of the intra-ring cooperativity states (Sigler *et al.*, 1998).

Directly, it can be stated that it is the very displacement of the equatorial domains, of opposing rings, which most likely provides the structural basis for the development of the signal that eventually couples ATP binding to the trans-ring when coupled to GroES release state from the cis-ring (Rye *et al.*, 1997), regardless of the state of stasis (Xu *et al.*, 1997), or conformational change (Ranson *et al.*, 2001), existent at the inter-ring interface.

7.2 Investigating the Allosteric Cooperativity of the Ch-cpn60 α and β Proteins

It was with the express intention of investigating the allosteric cooperativity, and transitional, states that exist for the plastid Ch-cpn60 protein (specifically the ATPase activity for the monomeric form of the α and β subunits), that a set of ATP-regenerating activity assays were performed. Through steady-state kinetic analysis it was likely that information could be gained, regarding the allosteric constraints that may exist upon the complex as a whole and the subunit monomeric form. Further, from these analyses it should be possible to answer such questions as the activity rate differences that may exist between the β_{14} mer complex and the β_m (subunit) form; when their activities are compared does the β subunit, within the complex, have more or less activity when compared to the activity of the β_m in stand alone monomeric format? The available ATPase activity data for the GroEL homologue should provide a stable scaffold upon which to compare the activity data.

A further set of questions could also be put forward, regarding a further interest in the ATPase activity of the monomeric α subunit form when compared to the β14 mer and βm data. Subject to the ATP activity data gained, analysis could likely indicate any differences that may exist between the allosteric constraints exerted upon the two divergent subunits, were they to exist within the same rings. Further to this, it is possible that determination could be made as to whether the two divergent subunits could or could not actually belong in the same ring?

7.3 The Steady-State ATPase Activity Method

The usual way of measuring the rate of an enzyme reaction is to monitor the change in concentration of one the substrates involved in, or one of the products produced by, the reaction.

The coupled enzyme ATPase assay is based on the conversion of phosophoenolpyruvate (PEP) to pyruvate, by pyruvate kinase (PK) coupled to the conversion of pyruvate to lactate by lactate dehydrogenase (LDH), (Kreuzer and Jongeneel, 1983). The latter step requires NADH which is oxidized to NAD^+. NADH absorbs strongly at 340nm but NAD^+ does not, enabling the utilization of NADH to be followed by monitoring absorbance at 340nm. The decrease in OD_{340nm} can be converted into ATPase activity where 1 molecule of NADH oxidized to NAD^+ corresponds to the production of 1 molecule of ADP by the adenosine triphosphotase (ATPase) hydrolysis reaction. Fig 7.3.1 depicts the typical pathway followed by this reaction.

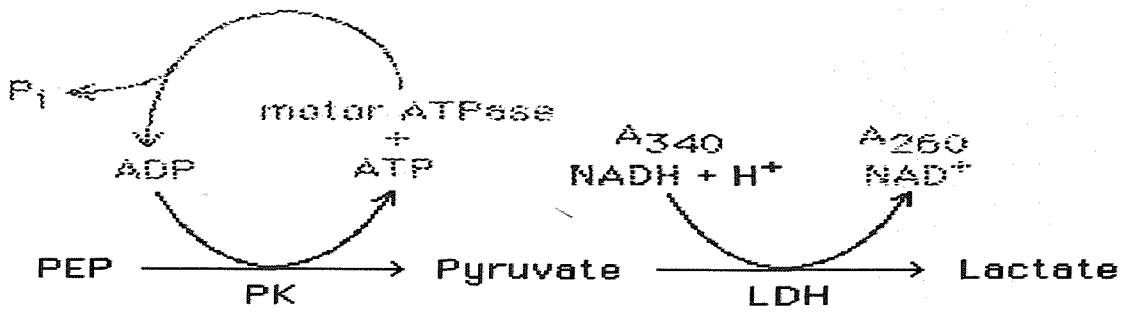


Fig 7.3.1 The ATPase Activity Pathway

7.4 Materials and Methods

Purification and aquisiton of GroEL, β_{14}, and α_m (monomer) was carried out and analysis to check the purity of these protein forms was achieved by 4-12% SDS-PAGE, as previously described.

The acquisition of βm was achieved by the addition of potassium thiocyanate (KSCN) at a 4:1 ratio of KSCN to protein.

7.4.1 Bicinchoninic Acid Protein Concentration Assay

Accuracy of protein concentration was essential to this experiment and therefore the Bicinchoninic Acid (BCA) protein concentration determination assay was used. The BCA assay is more sensitive than the Bradford Lowry methods and produces more stable, and so less variable, concentration data to work from.

Materials:

The reagents for the assay were purchased from Sigma.

Bicinchoninic Acid Solution (B 9643)

Copper Sulphate Pentahydrate solution (C 2284)

1.0mg/ml Bovine Serum Albumin

Pharmacia biotech ultrospec 3000 uv/visible Spectrophotometer

1ml Quartz cuvettes

10 ml Running buffer

- 20mM Tris (pH 7.5)
- 50mM EDTA
- 5mM Mercaptoethanol

The BCA working reagent protocol was as follows; mix 16ml of Bicinchoninic Acid solution with 0.32ml Copper Sulphate Pentahydrate solution, giving a 50:1 16.32ml working reagent.

To prepare protein standards of different known concentrations between 200-1000μg/ml dilute the 1mg/ml Bovine Serum Albumin appropriately. Each dilution should be diluted with the same buffer that the protein to be determined is in as this takes into account any interference that may be due to the buffer. e.g the chelating agent EDTA. In this case the buffer that the chaperonin has been kept in is the running buffer used on the column during protein purification.

For the assay

(1) Add 100μl of the protein standards of BSA (concentration between 200-1000μg/ml) into separate test tubes according to the chart below.

(2) For the blank add 100μl of the running buffer.

(3) Dilute the unknown protein sample appropriately so that the diluted protein concentration falls in the range 200-1000μg and add 100μl of this to the correct test tube. You need to know roughly the concentration of your sample.

(4) To each test tube add 2ml of the BCA working reagent. (Table 7.4.1)

(5) Vortex each sample to ensure correctly mixed.

(6) Incubate at 37°C for 30 minutes and then allow to it cool down to room temperature.

(7) Transfer 1ml of each solution to separate cuvettes and measure the absorbance using spectrophotometer at 562nm over a minute and record average. Colour development continues slowly after cooling to room temperature but if the tubes are read 10minutes between each other there should be no significant error.

(8) Plot a standard curve of BSA protein standard concentrations (μg/ml) against net absorbance at 562nm (Abs) to determine unknown protein concentration. Take into account dilution factors of unknown protein sample.

Tube no.	Sample (ml)	[BSA] Protein standard (mg/ml)	BCA Working Reagent (ml)	Role
1	0.1	0	2	Blank
2	0.1	200	2	Standard
3	0.1	400	2	Standard
4	0.1	600	2	Standard
5	0.1	800	2	Standard
6	0.1	1000	2	Standard
7	0.1	(unknown)	2	Unknown (1 in 10 dilution)
8	0.1	(unknown)	2	Unknown (1 in 5 dilution)

Table 7.4.1 Depicted in the above table are the constituent reagent and protein mixes necessary to obtain the BCA standard protein concentration plot.

7.4.2 ATPase Activity Assays

Materials:

1.0mM Phosphoenolpyruvate (PEP)

0.2mM NADH

3-5 units of Pyruvate kinase (PK)

3-5 units of Lactate dehydrogenase (LDH)

5-50μg/ml GroEL

0.5-2.0mM ATP

100ml Reaction buffer containing;

50mM Tris (pH 7.5)

4mM MgCl

4mM KCl

To make reaction buffer add 0.606g of Tris to 100ml of distilled water, adjust to pH 7.5 as required by the addition of HCl drop by drop. Now add 0.038g of MgCl and 0.030g of KCl and stir until dissolved.

The enzymes come in the same solution in an approximate 3:2 ratio, (LDH:PK).The possible amounts of enzymes to be added are;

3μl gives (3.9units of LDH) (2.78units of PK)

4μl gives (5.2units of LDH) (3.71units of PK)

5μl gives (6.5units of LDH) (4.64units of PK)

Method of assay determination:

(1) Determine the protein concentration of GroEL or other chaperonin using the BCA-1 assay.

(2) Due to the small quantities of compounds used in this assay higher concentration stock solutions have to be made and appropriately diluted. For example, a 100mM PEP stock solution is made by adding 0.019g to 1ml of distilled water (or reaction buffer for more accurate results). Adding 100µl of this to a further 900µl gives a stock solution of a 10-fold dilution. Finally the addition of 100µl of this solution to the 1ml assay solution will give the required 1.0mM PEP concentration. This method should be followed to make stock solutions of NADH (2.0mM stock) and ATP (10mM stock).

(3) Into a 1.5ml eppendorph tube add 4µl of LDH/PK, 100µl of PEP from 10mM stock (giving a concentration of 1.0mM), 100µl of NADH from 2.0mM stock (giving a concentration of 0.2mM), 100µl of ATP from a 10mM stock (giving a concentration of 1.0mM) and 671µl of reaction buffer.

(4) Mix thoroughly.

(5) Set up the spectrophotometer to absorb at a wavelength of 340nm, with a delay after the sample has been added of sixty seconds to allow the reaction to reach a steady state and to record every second for 300 seconds.

(6) Add the1ml of reaction solution into a cuvette.

(7) Blank the spectrophotometer. i.e. with a solution containing a replicated mix.

(8) Add 25µl of the 800µg/ml GroEL solution to the cuvette.

(9) Mix using parafilm and inverting a few times. Careful not to produce bubbles as these will effect the absorbance reading

(10) Place in spectrophotometer and leave for 300 seconds collecting data.

(11) Repeat three times and take mean average of data.

(12) Repeat the experiment varying ATP concentrations and GroEL concentrations keeping all other variables constant and adjusting the amounts of solutions appropriately.

N.B It was an integral part of the assay to keep the temperature constant i.e. at 25 degrees Celsius.

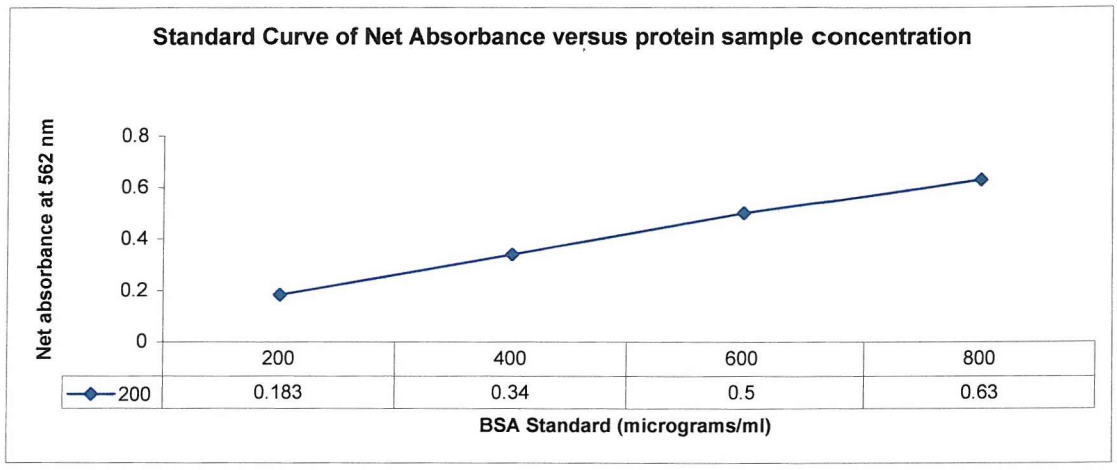
In the presence of GroEL ATP hydrolysis should occur and increase the concentration of ADP, allowing reaction 1 to proceed, and being coupled to reaction 2 will drive the oxidation of NADH to NAD+, therefore a decrease in absorbance should be seen.

Calculating the ATPase Rate:
(micromoles of ATPhydrolyzed/min/microgram of chaperone) from the change in absorbance at 340nm/min requires the use of the Beer Lambert law. ($A = \varepsilon CL$) where, A is the absorbance change/min, ε is the molar extinction coefficient for NADH (6220), and L is the path length.

7.5 Results

7.5.1 Protein Concentrations

The first stage of the experiment was involved with the acquisition of the pure forms of GroEL, β14, β_{mon}, and α_{mon} (monomers were acquired via 2M KSCN reduction). Protein concentration for the various protein states was determined by the use of the BCA Assay, and the protein used to obtain a standard reference curve was Bovine Serum Albumin (BSA), Fig 7.5.1 shows the BSA standard curve profile and Table 7.5.1 shows the concentrations for the relevant proteins and their variable states.



Protein	Dilution	Net Absorbance at 562nm (average 5 mins)	Protein Conc. (mg/ml)
None	none	0	0
Fridge(1) Beta-14mer	1 in 10	0.342	4
Freezer(2) Beta-14mer	1 in 10	0.22	2.4
GroEL	none	0.955	11.6
β monomer	none	0.245	0.29
α monomer	1 in 20	0.35	8.8

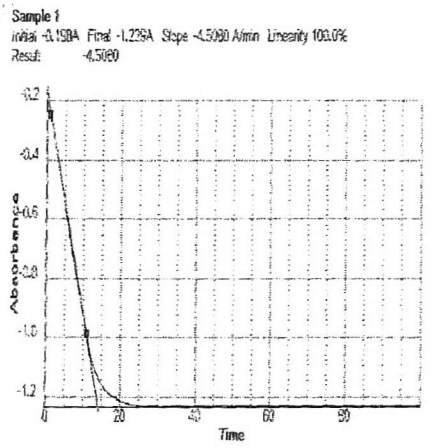
Figure 7.5.1 and Table 7.5.1 BCA Standard Graph and Protein concentrations

Fig 7.5.1 depicts the BCA standard reference graph obtained using BSA. Table 7.5.1 depicts the concentrations obtained for the various proteins used in the experiment.

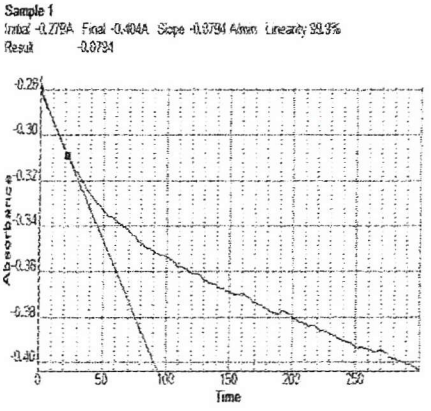
7.5.2 ATPase Activity Assays

The ATPase assays were carried out under identical conditions and using the same photospectrometer, Figure 7.5.2 depicts a selection of some of the assay profiles gained for the GroEL:ATP activity assays. Figure 7.5.3 depicts the comparative data gained for the ATPase activity assays carried out on GroEL, β_{14}mer, β_{mon} and α_{mon} (the monomeric subunit form of the proteins was acquired using 2M KSCN) at various ATP concentrations (1 to 4mM). Figure 7.5.4 depicts comparative data analyses made between the ATPase activity data for GroEL and the activity data for β_{14}mer. A comparison of the data obtained from Figure 7.5.3 and 7.5.4 made it possible to elicit a predicted ATPase activity rate for the β_m, which was then compared with the activity data obtained from the β_m form *in vitro*. From the comparative analysis of the predicted and *in vitro* β monomeric activity assays, the evidence was suitably indicative to determine whether or not allosteric constraints were being applied to the individual subunit whilst part of the ring complex.

(1)



(2)



(3)

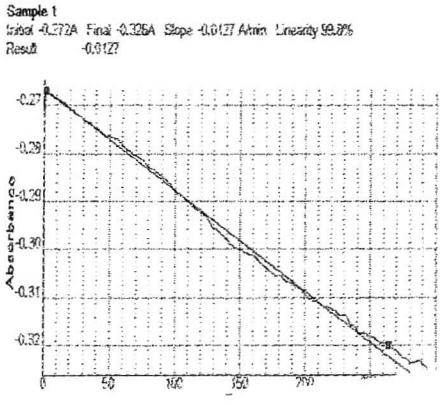


Figure 7.5.2 Activity Assay Profiles for GroEL

ATPase activity assays for GroEL (1) ADP, (2) 2mM ATP, 20μg GroEL, (3) 2mM ATP, 20μg GroEL, 4mM potassium.

[ATP] mM	Average Rate (Abs change/min)	Rate (micromoles/min/microgram)	
0	0		0
0.5	0.0255		0.205
1	0.049		0.394
1.5	0.069		0.555
2	0.087		0.7
2.5	0.0934		0.751
3	0.1164		0.936
3.5	0.0945	Negative Cooperativity ?	0.76
4	0.09		0.723

Table 7.5.2a GroEL (14mer) ATPase Activity Data

[ATP] mM	Average Rate (Abs change/min)	Rate (micromoles/min/microgram)	
0	0		0
0.5	0.0162		0.13
1	0.04518		0.363
1.5	0.0595		0.478
2	0.079		0.635
2.5	0.0988		0.794
3	0.1092		0.878
3.5	0.0954	Negative Cooperativity ?	0.767
4	0.0920		0.720

Table 7.5.2b β14mer ATPase Activity Data

Tables 7.5.1 and 7.5.2 depict the activity data acquired for the GroEL complex form and the β complex form. Quite evident is the drop in activity for both proteins at 3.5mM ATP, suggesting negative cooperativity is occurring. Whereby, the abundant presence of ATP can no longer be utilised by the chaperonins opposing-ring mechanism due to the blocking effect set up by the now replete *cis*-ring. All data are an accumulated average for each set of experiments, which were repeated three times per individual experiment.

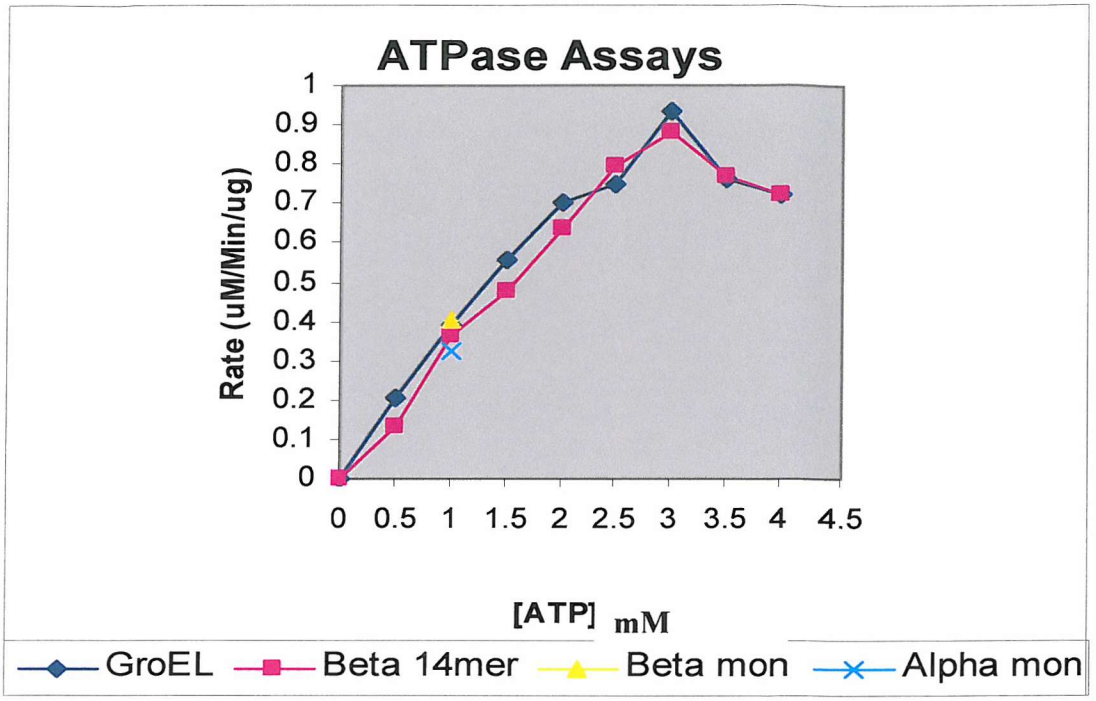


Figure 7.5.3 ATPase Activity assays for GroEL, β_{14mer}, β_{mon} and α_{mon}

Depicted is a graph representing the data presented in Tables: 7.5.1, 7.5.2, 7.5.3 and 7.5.4. Interestingly, at 3.5mM ATP the activity rate for both GroEL and β_{14} decreases, indicating –ve cooperativity occurrence between rings. Also depicted are the 'actual' activity rates for both the α_{mon} and $\beta_{mon,}$ due to both being in monomeric format no cooperativity was expected and therefore only a 1mM ATP assay was used. All data are an accumulated average for each set of experiments, which were repeated three times per individual experiment.

[ATP] mM	Average Rate (Abs change/min)	Rate (micromoles/min/microgram)
1	0.0503	0.404

Table 7.5.3 Data for β_{mon} ATPase Activity Assays

[ATP] mM	Average Rate (Abs change/min)	Rate (micromoles/min/microgram)
1	0.0386	0.322

Table 7.5.4 Data for α_{mon} ATPase Activity Assays

Depicted are the activity assay data for the β_m and the α_m. Predictions for complex activity were made for each subunit from the data gained on their single subunit activity. This would suggest that very little, if any, allosteric constraints are applied to the β subunit (monomer) whilst committed as a constituent part of a full complex and when compared to the GroEL assays (Fig 7.5.3). However, activity for the α subunit (monomer), although only 0.082 μMoles/Min/μg lower in activity than the β subunit, when directly compared to its β counterpart in Figure 7.5.3, shows a drastic reduction in activity sufficient to suggest that problems would be encountered when in the complex state, (see also Chapter Two). All data are an accumulated average for each set of experiments, which were repeated three times per individual experiment.

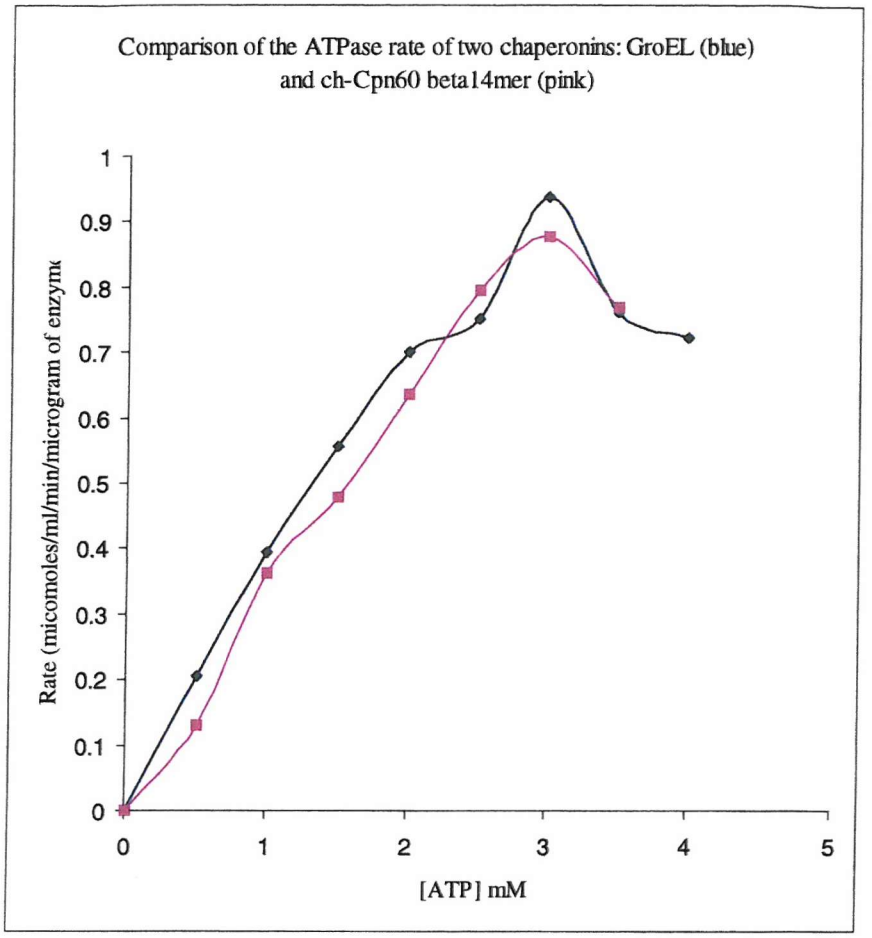


Fig 7.5.4 ATPase Activity Comparison Profile for GroEL$_{14}$ and β$_{14}$ Complexes

Taken from the activity data for both proteins (Tables 7.5.1 and 7.5.2), it was possible to present a graph profile depicting the curve of the activity. There are two prominent features suggested by this profile. The first is that the activity for both proteins is similar, and the second prominence to the profile is the decrease in activity for both proteins at ~3.5mM ATP. This would suggest that at this point the *cis* rings of both proteins possess a full compliment of ATP molecules and until they have been hydrolysed the opposing *trans* ring is incapable of ATP molecule uptake. This phenomona is suggestive for the occurrence of negative cooperativity between rings. All data are an accumulated average for each set of experiments, which repeated three times per individual experiment.

7.5.3 General Conclusions

The chaperonin GroEL drives its protein-folding cycle by cooperatively binding one molecule ATP to each subunit of one of its rings (*cis*). Priming that ring to become folding-active upon GroES binding, while simultaneously discharging the previous folding chamber from the opposite ring (*trans*). The effects of this ATP binding to a GroEL-GroES-ADP complex, suggest structural models for the ATP-induced reduction in affinity for polypeptide and for cooperativity.

The cooperative binding of ATP to a GroEL ring allows rapid binding of GroES to that ring, converting it from a polypeptide-accepting to a folding-active state. At the same time, ATP binding allosterically discharges the ligands from the opposite ring.

Positively cooperative binding of ATP to one of the two GroEL rings promotes a conformational change in the chaperonin and is a prerequisite for binding of the cochaperonin GroES to that ring (Chandrasekhar *et al.*, 1986; Gray and Fersht, 1991, Jackson *et al.*, 1993, Todd *et al.*, 1994). Negative cooperativity with respect to ATP binding results in asymmetric complexes, with the two rings adopting different conformations (Yifrach and Horovitz, 1995, Burston *et al.*, 1995).

It is clear that ATP binding plays a pivotal role in chaperonin function, but the structural consequences of ATP binding remain poorly understood. Even more problematic, it has been observed that nonhydrolyzable ATP analogs can neither promote productive folding of stringent substrates nor drive dissociation of folding-active complexes (Rye *et al.*, 1997), suggesting that the analogs cannot mimic the stereochemical action of ATP on GroEL.

7.5.3.1 ATPase Activity Assays for GroEL

Measurements of initial rates of ATP hydrolysis at increasing concentrations of nucleotide have shown that GroEL undergoes two different allosteric transitions that can be explained assuming a nested allosteric model (Yifrach and Horovitz, 1995) (Table 7.5.1). The first transition takes place at ATP concentrations lower than 3mM, and reflects positive cooperativity in ATP binding and hydrolysis within one ring. At the higher nucleotide concentrations (ATP 3.5mM, reporting a drop in absorbance A_{340nm} = 0.76), a second level of allostery was suggested, caused by negative inter-ring cooperativity, where the opposing (*trans*) ring could take up ATP in an independent manner due to the allosteric restraints put on the respective ring. This effect has been previously observed by Yifrach and Horovitz (Yifrach *et al*; 1996, Yifrach *et al*; 2000, Horovitz, 1993, Yifrach *et al.*, 1996), who fitted their observed data to a model of 'nested' cooperativity using the (WMC) and Koshland, Nemethy, and Filmer (KNF) models, (Monod *et al.*, 1965, Koshland *et al.*, 1996).

7.5.3.2 B_{14}Mer ATPase Activity

The β_{14}mer ATPase activity data showed very similar activity rates to those acquired for GroEL. This suggested a similar correlation could exist for the two levels of allosteric transition that occurs to effect positive and negative cooperativity (Table 7.5.2 and Fig 7.5.3). These data further suggest that β_{14} is capable of providing an active folding environment in isolation of the α subunit and at a similar rate to that of GroEL.

7.5.3.3 ATPase Acivity Data for the β_m and α_m Subunits

The activity rates for the β_m (Table 7.5.3) suggested that when compared to the activity rates of the β_{14} complex, there was little difference between the activity rates of the β subunit when part of the complex and the β_m as a 'stand-alone' subunit. This would suggest that very little allosteric constraint is being applied to the individual subunit when part of the ring complex. This result would further suggest that each individual subunit, when part of a ring, contributes to the overall ATPase activity of the complex as a single entity and only requires the presence of its neighbouring subunits to confer an overall allostery to the active complex. That is to say, each subunit within a ring is capable of acting in an individual manner when in the presence of nucleotide. The choreographed allostery occurring within rings as well as in opposing rings is the sum of the individual parts coming together to project a symmetrical complex conferred by the mechanism of the individual subunits. Although no one subunit controls the complex, no complex can accomplish a folding operation without its presence.

Table 7.5.4 shows the data obtained for the activity rate of the α subunit. This data shows a lower activity rate (0.082) in comparison to that for obtained the β subunit (Table 7.5.3). This result agrees with current opinion that the α subunit requires the presence, and therefore, the ATPase activity of the β subunit to effect an active form of the complex.

Future recommended work in this area of research should include ATPase activity assays for the GroEL subunit and the activity assays for the α and β subunits complexed *in vitro*. The data from these future experiments should allow for further comparative analyses and allow further ATPase activity comparative analyses to be made between the GroEL monomer and the β monomer, this with a view to extending the current knowledge on whether GroEL subunit has full ATPase activity outside of its complex.

References

Acton S, Resnick D, Freeman M, Ekkel Y, Ashkenas J, Krieger M. (1993) J. Biol. Chem. **268**(5) 3530-3537.

Ang D, Liberek K, Skowyra D, Zylicz M, Georgopoulos C. (1991) *J. Biol. Chem.* **266**(36) 24233-24236.

Argos P. (1990) *J. Mol. Biol.* **211**(4) 943-958.

Anfinsen C.B. (1973) *Science* **181**(96), 223-230.

Azem, A., Kessel, M., Goloubinoff, P. (1994) *Science* **265**(5172) 653-656.

Azem A, Diamant S, Kessel M, Weiss C, Goloubinoff P. (1995) *Proc. Natl. Acad. Sci.* USA **92**(26) 12021-12025.

Badcoe IG, Smith CJ, Wood S, Halsall DJ, Holbrook JJ, Lund P, Clarke AR. (1991) *Biochemistry* **30**(38) 9195-9200.

Baneyx F, Bertsch U, Kalbach CE, van der Vies SM, Soll J, Gatenby AA. (1995) *J. Biol. Chem.* **270**(18) 10695-10702.

Baneyx F, Gatenby AA. (1992) *J. Biol. Chem.* **267** 11637-11644.

Bardwell JC, Craig EA. (1987) *Proc. Natl. Acad. Sci.* USA **84**(15) 5177-5181.

Barraclough R, Ellis RJ. (1980) *Biochim Biophys. Acta.* **608**(1) 19-31.

Battistoni A, Carri MT, Steinkuhler C, Rotilio G. (1993) *FEBS Lett.* **322**(1) 6-9.

Beckmann RP, Mizzen LE, Welch WJ. (1990) *Science* **248**(4957) 850-854.

Bentley GA, Houdusse A. (1992) *Acta. Cryst.* **A48** 322-328.

Bertsch U, Soll J, Seetharam R, Viitanen PV. (1992) *Proc. Natl. Sci.* USA **89**(18) 8696-8700.

Bochkareva ES, Lissin NM, Flynn GC, Rothman JE, Girshovich AS. (1992) *J. Biol. Chem.* **267**(10) 6796-6800.

Boisvert DC, Wang J, Otwinowski Z, Horwich AL, Sigler PB. (1996) *Nat. Struct. Biol.* **3**(2) 170-177.

Bonk M, Tadros M, Vandekerckhove J, Al-Babili S, Beyer P. (1996) *Plant Physiol.* **111**(3) 931-939.

Bose S, Weikl T, Bugl H, Buchner J. (1996) **274**(5293) 1715-1717.

Bowden GA, Georgiou G. (1990) *J. Biol. Chem.* **265**(28) 16760-16766.

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB. (1994) *Nature* **371**(6498) 578-586.

Brünger AT, Kuriyan J, Karplus M. (1987) *Science* **235** 458-460.

Brünger AT. (1990) *Acta. Cryst.* **A46** 46-57.

Brünger AT. (1992) *Nature* **355** 472-475.

Brünger AT, Nilges M. (1993) *Quarterly Rev. Biophys.* **26** 49-125.

Brünger AT. (1997) *Methods Enzymol.* **277** 366-396.

Brunshier R, Danner M, Seckler R. (1993) *J. Biol. Chem.* **268** 2767-2772.

Bryngelson JD, Wolynes PG. (1987) **84**(21) 7524-7528.

Buchberger A, Schroder H, Hesterkamp T, Schonfeld HJ, Bukau B. (1996) *J. Mol. Biol.* **261**(3) 328-333.

Buchner J, Schmidt M, Fuchs M, Jaenicke R, Rudolph R, Scmid FX, Kiefhaber T. (1991) *Biochemistry* **30**(6) 1586-1591.

Buchner J, Bose S, Mayr C, Jakob U. (1998*) Methods Enzymol.* **290** 409-418.

Buckle AM, Zahn R, Fersht AR. (1997) *Proc. Natl. Sci.* USA **94**(8) 3571-3575.

Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. (1995) *Eur. J. Biochem.* **230**(2) 788-796.

Budisa N, Karnbrock W, Steinbacher S, Humm A, Prade L, Neuefeind T, Moroder L, Huber R. (1997) *J. Mol. Biol.* **270**(4) 616-623.

Bukau B, Horwich HL. (1998) *Cell* **92**(3) 351-366.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC. (1996) *Science* **273**(5278) 1058-1073.

Branden CI, Schneider G, Lindqvist Y, Andersson I, Knight S, Lorimer G. (1987) Cold Spring Harb. Symp. Quant. Biol. **52** 491-498.

Burston SG, Clarke AR. (1995) *Essays Biochem.* **29** 125-136.

Burston SG, Ranson NA, Clarke AR. (1995) *J. Mol. Biol.* **249** 138-152.

Burston SG, Weissman JS, Farr GW, Fenton WA, Horwich AL. (1996) *Nature* **383**(6595) 96-99.

Burt WJ, Leaver CJ. (1994) *FEBS Lett.* **339**(1-2) 139-141.

Cannon S, Wang P, Roy H. (1986) *J. Cell Biol.* **103**(4) 1327-1335.

Carlsson J, Drevin H, Axen R. (1978) *Biochem. J.* **173**(3) 723-737.

Carter CW Jr, Yin Y. (1994) *Acta. Cryst.* **D50** 572-590.

Chandrasekhar GN, Tilly K, Woolford C, Hendrix R, Georgopoulos C. (1986) *J. Biol. Chem.* **261**(26) 124141-12419.

Chen S, Roseman AM, Hunter AS, Wood SP, Burston SG, Ranson NA, Clarke AR, Saibil HR. (1994) *Nature* **371**(6494) 261-264.

Chen S, Roseman AM, Hunter AS, Wood SP, Burston SG, Ranson NA, Clarke AR, Saibil HR. (1994) *Nature* **371**(6494) 614-619.

Cheng MY, Hartl FU, Martin J, Pollock RA, Kalousek F, Neupert W, Hallberg EM, Hallberg RL, Horwich AL. (1989) *Nature* **337**(6208) 620-625.

Clegg MT, Gaut BS, Learn GH Jr, Morton BR. (1994) *Nature* **91**(15) 6795-6801.

Cliff MJ, Kad NM, Hay N, Lund PA, Webb MR, Burston SG, Clarke AR. (1999) *J. Mol. Biol.* **293**(3) 667-684.

Cloney LP, Wu HB, Hemmingsen SM. (1992a) *J. Biol. Chem* **267**(32) 23327-23332.

Cloney LP, Bekkaoui DR, Wood MG, Hemmingsen SM. (1992b) *J. Biol. Chem.* **267**(32) 23333-23336.

Collaborative Computational Project, Number 4, (1994) *Acta. Crsyt.* **D50**, 760-763.

Corrales FJ, Fersht AR. (1995) *Proc. Natl. Acad. Sci.* USA **92**(12) 5326-5330.

Corrales FJ, Fersht AR. (1996) *Proc. Natl. Acad. Sci.* USA **93**(9) 4509-4512.

Craig EA, Weismann JS, Horwich AL. (1994) *Cell* **78**(3) 365-372.

Crowther RA. (1972) The Molecular Replacement Method 173-178, Ed. Rossmann MG., New York, Gordon and Breach.

Crowther RA, Blow DM. (1967) *Acta. Cryst.* **23** 544-548.

Cudney B, Patel S. (1994) *Am. Biotechnol. Lab.* **12**(7) 42.

Davies GE, Stark GR. (1970) *Proc. Natl. Acad. Sci.* USA **66**(3) 651-656.

Diamant S, Azem A, Weiss C, Goloubinoff P. (1995) Biochemistry **34**(1) 273-277.

Diamant S, Azem A, Weiss C, Goloubinoff P. (1995) *J. Biol. Chem* **270**(47) 28387-28391.

Dickson R, Weiss C, Howard RJ, Alldrick SP, Ellis RJ, Lorimer G, Azem A, Viitanen PV. (2000) *J. Biol. Chem.* **275**(16) 11829-11835.

Dill KA, Phillips AT, Rosen JB. (1997) *J. Comput. Biol.* **4**(3) 227-239.

Dill KA, Chan N. (1997) *Nature Struct. Biol.* **4** 10-19.

Ditzel L, Lowe J, Stock D, Stetter KO, Huber H, Huber R, Steinbacher S. (1998) *Cell* **93**(1) 125-138.

Doublie S. (1997) *Methods Enzymol.* **276** 523-530.

Drenth J. (1994) Principles of Protein Crystallography, Springer.

Eggers DK, Welch WJ, Hansen WJ. (1997) *Mol. Biol. Cell* **8**(8) 1559-1573.

Ellis RJ, Robinson C, van der Vies SM, Kirwin PM. (1988) *Biochem. Soc. Trans.* **16**(5) 703-704.

Ellis RJ, Hemmingsen SM. (1989) *Trends Biochem. Sci.* **14**(8) 339-342.

Ellis RJ, van der Vies SM. (1991) *Annu. Rev. Biochem.* **60** 321-347.

Ellis RJ. (1996) The CHAPERONINS. Academic Press, Inc. San Diego, California.

Ellis RJ, Hartl FU. (1996) *FASEB J.* **10**(1) 20-26.

Ellis RJ. (1997) *Curr. Biol.* **7**(9) R531-533.

Ellman GL. (1959) *Arch. Biochem. Biophys.* **82** 70-77.

Engel A, Hayer-Hartl MK, Goldie KN, Pfeifer G, Hegerl R, Müller S, da Silva ACR, Baumeister W, Hartl FU. (1995) *Science* **269** 832-836.

Epstein CJ, Goldberger RF, Anfinsen CB. (1963) Cold Spring Harb. Symp. *Quant. Biol.* **28** 439.

Escher A, Szalay AA. (1993) *Mol. Gen. Genet.* **238**(1-2) 65-73.

Farr GW, Scharl EC, Schumacher RJ, Sondek S, Horwich AL. (1997) *Cell* **89**(6) 927-937.

Fayet O, Louarn JM, Georgopoulos C. (1986) *Mol. Gen. Genet.* **202**(3) 435-445.

Fenton WA, Horwich AL. (1997) *Protein Sci.* **6**(4) 743-760.

Fenton WA, Kashi Y, Furtak K, Horwich AL. (1994) *Nature* **371**(6498) 614-619.

Fisher MT, Yuan X. (1994) *J. Biol. Chem.* **269** 29598-29601.

Fong G, Bridger WA. (1992) *Biochemistry* **31**(24) 5661-5664.

Fourme R, Hendrickson WA. (1990) *Synchrotron Radiation and Biophysics* (Hasnain, SS., Ed.) Ellis Harwood, Chichester, pp. 156-175.

Fowler HW, Fowler FG. (1976) The Concise Oxord Dictionary. Ed. Sykes JB. Clarendon Press, Oxford.

Freeman BC, Morimoto RI. (1996) *EMBO. J.***15**(12) 2969-2979.

French S, Wilson KS. (1978) *Acta. Cryst.* **A34** 517-525.

Frydman J, Nimmesgern E, Erdjument-Bromage H, Wall JS, Tempst P, Hartl FU. (1992) *EMBO J.* **11**(13) 4767-4778.

Frydman J, Nimmesgern E, Ohtsuka K, Hartl FU. (1994) *Nature* **370**(6485) 111-117.

Frydman J, Hartl FU. (1996) *Science* **272**(5267) 1497-1502.

Gao Y, Vainberg IE, Chow RL, Cowan NJ. (1993) *Mol. Cell Biol.* **13**(4) 2478-2485.

Gassler CS, Buchberger A, Laufen T, Mayer MP, Schroder H. (1998) *Proc. Natl. Acad. Sci.* USA **95**(26) 15229-15234.

Gatenby AA, van der Vies SM, Rothstein SJ. (1987) *Eur. J. Biochem.* **168**(1) 227-231.

Gatenby AA, Ellis RJ. (1990) *Annu. Rev. Cell Biol.* **6** 125-149.

Georgopoulos C, Ang D. (1990) *Semin. Cell Biol.* 1(1) 19-25.

Geissler S, Siegers K, Scheibel E. (1998) *EMBO J.* **17**(4) 952-966.

Go N. (1994) *Annu. Rev. Biophys. Bioeng.* **12** 183-210.

Goloubinoff P, Christeller JT, Gatenby AA, Lorimer GH. (1989) *Nature* **342**(6252) 884-889.

Gomer CJ, Ryter SW, Ferrario A, Rucker N, Wong S, Fisher AM. (1996) *Cancer Res.* **56**(10) 2355-2360.

Gragerove A, Nudler E, Komissarova N, Gaitannaris GA, Gottesman ME, Nikiforov V. (1992) *Proc. Natl. Acad. Sci.* USA **89**(21) 10341-10344.

Gray TE, Fersht AR. (1991) *FEBS Lett.* **292**(1-2) 254-258.

Gray TE, Fersht AR. (1993) *J. Mol. Biol.* **232**(4) 1197-1207.

Grimm R, Donaldson GK, van der Vies SM, Schafer E, Gatenby AA. (1993) *J. Biol. Chem.* **268**(7) 5220-5226.

Gray MW, Doolittle WF. (1982) *Micobiol. Rev.* **46**(1) 1-42.

Gray MW. (1992) *Int. Rev. Cytol.* **141** 233-357.

Gray MW. (1993) Curr. Opin. Genet. Dev. **3**(6) 884-890

Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R. (1997) *Nature* **386**(6624) 463-471.

Gupta RS, Golding GB. (1993) *J. Mol. Evol.* **37**(6) 573-582.

Hanein D, Matlack KE, Jungnickel B, Plath K, Kalies KU, Miller KR, Rapoport TA. (1996) *Cell* **87**(4) 721-732.

Hansen JE, Gafni A. (1993) *J. Biol. Chem.* **268**(29) 21632-21636.

Harada Y, Lifchitz A, Berthou J, Jolles P. (1981) *Acta. Cryst.* 398-406.

Harris JR, Pluckthun A, Zahn R. (1994) *J. Struct. Biol.* **112**(3) 216-230.

Harrison CJ, Hayer-Hartl M, Di Liberto M, Hartl FU, Kuriyan J. (1997) *Science* **276**(5311) 431-435.

Hartl FU, Hlodan R, Langer T. (1994) *Trends. Biochem.* Sci. **19**(1) 20-25.

Hartl FU, Martin J. (1995) *Curr. Opinion Struct.* Biol. **5** 92-102.

Hartl FU. (1996) *Nature* **381**(6583) 571-579.

Hayer-Hartl MK, Ewbank JJ, Creighton TE, Hartl FU. (1994) *EMBO J.* **13** 3192-3202.

Hayer-Hartl MK, Martin J, Hartl FU. (1995) *Science* **269**(5225) 836-841.

Hayer-Hartl MK, Weber F, Hartl FU. (1996) *EMBO J.* **15**(22) 6111-6121.

Hendershot L, Wei J, Gaut J, Melnick J, Aviel S, Argon Y. (1996) *Proc. Natl. Acad. Sci.* USA **93**(11) 5269-5274.

Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ. (1988) *Nature* **333**(6171) 330-334.

Hemmingsen SM. (1990) *Semin. Cell Biol.* **1**(1) 47-54.

Hendrick JP, Hartl FU. (1993) *Annu. Rev. Biochem.* **62** 349-384.

Hendrick JP, Langer T, Davis TA, Hartl FU, Weidmann M. (1993) *Proc. Natl. Acad. Sci.* USA **90**(21) 10216-10220.

Hendrickson WA. (1985) *Methods Enzymol.* **115** 252-270.

Hendrickson WA, Horton JR, LeMaster DM. (1990) *EMBO J.* **9**(5) 1665-1672.

Hendrix RW. (1979) *J. Mol. Biol.* **129**(3) 375-392.

Hill JE, Hemmingsen SM. (2001) *Cell Stress Chaperones* **6**(3) 190-200.

Holl-Neugebaur B, Rudolph R, Schmidt M, Buchner J. (1991) **30**(50) 11609-11614.

Hopkins WG. (1995) Introduction to PLANT PHYSIOLOGY. John Wiley & Sons, Inc.

Hoppe W. (1957) *Acta. C*ryst. **10** 751.

Horovitz A. (1993) *J. Mol. Biol.* **231** 58-64.

Horwich AL, Low KB, Fenton WA, Hirshfield IN, Furtak K. (1993) *Cell* **74**(5) 909-917.

Huber R. (1985) Experiences with the application of Patterson search techniques 58-61, Molecular replacement, (Machin PA. comp.). Proceedings of the Daresbury Study Weekend, S.E.R.C., Daresbury, U.K.

Hucho F, Mullner H, Sund H. (1975) *Eur. J. Biochem.* **59**(1) 79-87.

Hynes G, Celis JE, Lewis VA, Carne A, U S, Lauridsen JB, Willison KR. (1996) *Electrophoresis* **17**(11) 1720-1727.

Jäätelä M. (1990) *Scand. J. Immol.* **31**(2) 175-182.

Jackson GS, Staniforth RA, Halsall DJ, Atkinson T, Holbrook JJ, Clarke AR, Burston SG. (1993) *Biochemistry* **32** 2554-2563.

Jaenicke R. (1998) *Biol. Chem.* **379**(3) 237-243.

Jakob U, Meyer I, Bugl H, Andre S, Bardwell JC, Buchner J. (1995) *J. Biol. Chem.* **270**(24) 14412-14419.

Jakob U, Lilie H, Meyer I, Buchner J. (1995) J. Biol. Chem. **270**(13) 7288-7294.

Jakob U. (1996) *Front Biosci.* 1 d309-317.

Kabsch W. (1993) *J. Appl. Cryst.* **21** 916-924.

Karplus M. (1997) *Fold Des.* **2**(4) S69-75.

Karplus M. (2002) *Acc. Chem. Res.* **35**(6) 321-323.

Kaufmann SH. (1990) *Immunol. Today* **11**(4) 129-136.

Kaufmann SH, Schoel B, van Embden JD, Koga T, Wand-Wurttenberger A, Munk ME, Steinhoff U. (1991) *Immunol. Rev.* **121** 67-90.

Kaufmann SH. (1994) *Int. Arch. Allergy Immunol.* **103**(4) 317-322.

Karzai AW, McMacken R.(1996) *J. Biol. Chem.* **271**(19) 11236-11246.

Kim PS, Baldwin RL. (1990) *Annu. Rev. Biochem.* **59** 631-660.

Kim S, Willison KR, Horwich AL. (1994) *Trends Biochem. Sci.* **19**(12) 543-548.

Kjeldgaard K, Jones TA. (1995) O Manual. BMC. Uppsala University, Sweden.

Klumpp M, Baumeister W, Essen LO. (1997) *Cell* **91**(2) 263-270.

Koonin EV, van der Vies SM. (1995) Trends Biochem. Sci. **20**(1) 14-15.

Koshland DE Jr, Nemethy G, Filmer D. (1966) *Biochemistry* **5**(1) 365-385.

Koumoto Y, Shimada T, Kondo M, Takao T, Shimonishi Y, Hara-Nishimura I, Nishimura M. (1999) *Plant J.* **17**(5) 467-477.

Kovalenko O, Yifrach O, Horovitz A. (1994) *Biochemistry* **33**(50) 14974-14978.

Kreuzer KN, Jongeneel CV. (1983) *Methods Enzymol.* **100** 144-160.

Kubota H, Hynes G, Willison K. (1995b) *Eur. J. Biochem.* **230** 3-16.

Kumamoto CA, Beckwith J. (1985) *J. Bacteriol.* **163**(1) 267-274.

Laminet AA, Ziegelhoffer T, Georgopoulos C, Pluckthun A. (1990) *EMBO J.* **9**(7) 2315-2319.

Langer T, Pfeifer G, Martin J, Baumeister W, Hartl FU. (1992) *EMBO J.* **11**(13) 4757-4765.

Laskey RA, Honda BM, Mills AD, Morris NR, Wyllie AH, Mertz JE, De Roberts EM, Gurdon JB. (1978) Cold Spring Harb. *Symp. Quant. Biol.* **Pt 1** 171-178.

Lee SC, Olins PO. (1992) *J. Biol. Chem.* **267**(5) 2849-2852.

Lehel C, Los D, Wada H, Gyorgyei J, Horvath I, Kovacs E, Murata N, Vigh L. (1993) *J. Biol. Chem.* **268**(3) 1799-1804.

Lennox CR, Ellis RJ. (1986) *Biochem. Soc. Trans.* **14**(1) 9-11.

Leopold PE, Montal M, Onuchic JN. (1992) *Proc. Natl. Acad. Sci.* USA **89**(18) 8721-8725.

Leroux MR, Fandrich M, Klunker D, Siegers K, Lupas AN, Brown JR, Scheibel E, Dobson CM, Hartl FU. (1999) *EMBO J.* **18**(23) 6730-6743.

Leslie AGW. (1992) Joint CCP4 and ESF-EACMB *Newsletter on Protein Crystallography* **No 26**, Daresbury Laboratory, Warrington, U.K.

Leslie AGW. (1994) Mosflm User Guide.

Levinthal C. (1966) *Sci. Am.* **214**(6) 42-52.

Liang H, Sandberg WS, Terwilliger TC. (1993) *Proc. Natl. Acad. Sci.* USA **90**(15) 7010-7014.

Lifchitz A. (1983) Acta. Cryst. **A39** 130-139

Llorca O, Perez-Perez J, Carrascosa JL, Galan A, Muga A, Valpuesta JM. (1997) *J. Biol. Chem.* **272**(52) 32952-32932.

Llorca O, McCormack EA, Hynes G, Grantham J, Cordell J, Carrascosa JL, Willison KR, Fernandez JJ, Valpuesta JM. (1999) *Nature* **402**(6762) 693-696.

Lubben TH, Gatenby AA, Donaldson GK, Lorimer GH, Viitanen PV. (1990) *Proc. Natl. Acad. Sci.* USA 87(19) 7683-7687.

Luby-Phelps K, Hori M, Phelps JM, Won D. (1995) *J. Biol. Chem.* **270**(37) 21532-21538.

Ma J, Sigler PB, Xu Z, Karplus M. (2000) *J. Mol. Biol.* **302** 303-313.

Makino Y, Taguchi H, Yoshida M. (1993) FEBS Lett. **336**(2) 363-367.

Margulis L. (1970) *J. Theor. Biol.* **26**(2) 337-342.

Margulis L. (1993) *Biosystems* **31**(2-3) 121-125.

Martel R, Cloney LP, Pelcher LE, Hemmingsen SM. (1990) *Gene* **94**(2) 181-187.

Martin J, Langer T, Boteva R. (1991) *Nature* **352** 36-42.

Maruya M, Sameshima M, Nemoto T, Yahara I. (1999) *J. Mol. Biol.* **285**(3) 903-907.

Mathews BW. (1968) *J. Mol. Biol.* **33** 491-497.

Mayer MP, Bukau B. (1998) *Biol. Chem.* **379**(3) 2610268.

Mayhew M, da Silva AC, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU. (1996) *Nature* **379**(6564) 420-426.

Mayr EM, Jaenicke R, Glochshuber R. (1994) *J. Mol. Biol.* **235**(1) 84-88.

McCarty JS, Buchberger A, Reinstein J, Bukau B. (1995) *J. Mol. Biol.* **249**(1) 126-137.

McCree DE. (1994) *J. Struct. Biol.* **125** 156-165.

McLachlan AD. (1987) Cold Spring Harb. Symp. *Quant. Biol.* **52** 411-420.

Melki R, Rommelaere H, Leguy R, Vandekerckhove J, Ampe C. (1996) *Biochemistry* **35**(32) 10422-10435.

Mendoza JA, Lorimer GH, Horowitz PM. (1991) *J. Biol. Chem.* **266**(26) 16973-16976.

Mendoza JA, Rogers E, Lorimer GH, Horowitz PM. (1991) *J. Biol. Chem.* **266** 13044-13049.

Mendoza JA, Demeler B, Horowitz PM. (1994) *J. Biol. Chem.* **269**(4) 2447-2451.

Miller AD, Maghlaoui K, Albnese G, Kleinjan DA, Smith C. (1993) *Biochem. J.* **291** 139-144.

Monod J, Wyman J, Changeux JP. (1965) *J. Mol. Biol.* **12** 88-118.

Musgrove JE, Johnson RA, Ellis RJ. (1987) *Eur. J. Biochem.* **163**(3) 529-534.

Nathan DF, Vos MH, Lindquist S. (1997) *Proc. Natl. Acad. Sci.* USA **94**(24) 12949-12956.

Navasa J. (1987) *Acta. Cryst.* **A43** 645-653.

Navasa J. (1994) *Acta. Cryst.* **A50** 157-163.

Navasa J, Vernoslava E. (1995) *Acta. Cryst.* **A51** 445-449.

Neiba-Axmann SE, Ottiger M, Wuthrich K, Pluckthun A. (1997) *J. Mol. Biol.* **271**(5) 803-818.

Nelson RJ, Heschl M, Craig EA. (1992) *Genetics* **131** 277-285.

Nitsch M, Walz J, Typke D, Klumpp M, Essen LO, Baumeister W. (1998) *Nature Struct. Biol.* **5**(10) 855-857.

Noll A, RoggenKamp A, Heesemann J, Autenrieth IB.(1994) *Infect. Immun.* **62**(7) 2784-2791.

Ohyama K, Kohchi T, Sano T, Yamada Y. (1988) *Trends Biochem. Sci.* **13**(1) 19-22.

Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. (1995) *Proc. Natl. Acad. Sci.* USA **92**(8) 3626-3630.

Osipiuk J, Joachimiak A. (1997) *Biochim. Biophys. Acta.* **1353**(3) 253-265.

Ostermann J, Horwich AL, Neupert W, Hartl FU. (1989) *Nature* (London) **341** 125-130.

Parsell DA, Kowal AS, Singer MA, Lindquist S. (1994) *Nature* **372**(6505) 475-478.

Pear LH, Taylor WR. (1987) *Nature* 329(6137) 351-354.

Pelham HR. (1989) *EMBO J.* **8**(11) 3171-3176.

Peralta D, Hartman DJ, Hoogenraad NJ, Hoj PB. (1994) *FEBS Lett.* **339**(1-2) 45-49.

Phipps BM, Hoffmann A, Stetter KO, Baumeister W. (1991) *EMBO J.* **10**(7) 1711-1722.

Pratt WB, Toft DO. (1997) *Endocr. Rev.* **18**(3) 306-360.

Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. (1997) *Cell* **90**(1) 65-75.

Pushkin AV, Tsuprun VL, Solov'eva NA, Shubin VV, Evstigneeva ZG. (1983) *Biokhimiia* **48**(9) 1441-1446.

Ranson NA, Burston SG, Clarke AR. (1997) J. Mol. Biol. **266**(4) 656-664.

Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Horwich AL, Saibil HR (2001) *Cell* **107** 869-879.

Reading DS, Hallberg RL, Myers AM. (1989) *Nature* **337**(6208) 655-659.

Rommelaere H, De Neve M, Melki R, Vandekerckhove J, Ampe C. (1999) Biochemistry **38**(11) 3246-3257.

Roseman AM, Chen S, White H, Braig K, Saibil SR. (1996) *Cell* **87** 241-251.

Rossman MG, Blow DM. (1962) *Acta. Cryst.* **15** 24-31.

Roy H, Bloom M, Milos P, Monroe M. (1982) *J. Cell Biol.* **94**(1) 20-27.

Rudolph R, Lilie H. (1996) *FASEB J.* **10**(1) 49-56.

Rutherford SL, Lindquist S. (1998) *Nature* **392**(6709) 336-342.

Ryan MT, Naylor DJ, Hoogenraad NJ, Hoj PB. (1995) *J. Biol. Chem.* **270**(37) 22037-22043.

Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, Horwich AL. (1999) *Cell* **97**(3) 325-338.

Scheraga HA. (1992) Protein Sci. **1**(5) 691-693

Schroder H, Langer T, Hartl FU, Bukau B. (1993) *EMBO J.* **12**(11) 4137-4144.

Schumacher RJ, Hansen WJ, Freeman BC, Alnemri E, Litwack G, Toft DO. (1996) *Biochemistry* **35**(47) 14889-14898.

Schwartz RM, Dayhoff MO. (1978) *Science* **199**(4327) 395-403.

Schwartz DH, Lucas AH, Chaisson RE. (1991) *N. Engl. J. Med.* **325**(26) 1837-1842.

Scopes RK, Stoter A. (1982) *Methods Enzymol.* **90** Pt E479-490.

Sculley TB, Spelsberg TC, Pearson GR. (1984) *Intervirology* **22**(4) 191-200.

Shiseki K, Murai N, Mtojima F, Hisabori T, Yoshida M, Taguchi H. (2001) *J. Biol. Chem.* **276**(14) 11335-11338.

Siegers K, Waldmann T, Leroux MR, Grein K, Shevchenko A, Scheibel E, Hartl FU. (1999) *EMBO J.* **18**(1) 75-84.

Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL. (1998) *Annu. Rev. Biochem.* **67** 581-608.

Smith KE, Fisher MT. (1995) *J. Biol. Chem.* **270**(37) 21517-21523.

Sparrer H, Rutkat K, Buchner J. (1997) *Proc. Natl. acad.* Sci. USA **94**(4) 1096-1100.

Staniforth RA, Burston SG, Atkinson T, Clarke AR. (1994) *Biochem. J.* **300** Pt 3 651-658.

Steinhoff MC, Auerbach BS, Nelson KE, Vlahov D, Becker RL, Graham NM, Szabo A, Korszun R, Hartl FU, Flanagan J. (1996) *EMBO J.* **15**(2) 408-417.

Stuart RA, Cyr DM, Neupert W. (1994) *Experientia* **50**(11-12) 1002-1011.

Suh WC, Lu CZ, Gross CA. (1999) *J. Biol. Chem.* **274**(43) 30534-30539.

Sussman JL, Holbrook SR, Church GM, Kim SH. (1977) *Acta. Cryst.* **A33** 800-804.

Taguchi H, Motohashi K, Yoshida M. (1994) *Tanpakushitsu Kakusan Koso.* **39**(7) 1233-1240.

Tang J, James MN, Hsu IN, Jenkins JA, Blundell TL. (1978) *Nature* **271**(5646) 618-621.

Theyssen H, Schuster HP, Packschies L, Bukau B, Reinstein J. (1996) *J. Mol. Biol.* **263**(5) 657-670.

Thomas JG, Baneyx F. (1996) *J. Biol. Chem.* 271(19) 11141-11147.

Thulasiraman V, Yang CF, Frydman J. (1999) **18**(1) 85-95.

Tian G, Vainberg IE, Tap WD, Lewis SA, Cowan NJ. (1995) *J. Biol. Chem.* **270**(41) 23910-23913.

Todd MJ, Viitanen PV, Lorimer GH. (1994) *Science* **265**(5172) 659-666.

Todd MJ, Walke S, Lorimer G, Truscott K, Scopes RK. (1995) *Biochemistry* **34**(45) 14932-14941.

Török Z, Vigh L, Goloubinoff P. (1996) *J. Biol. Chem.* **271** 16180-16186.

Trent JD, Osipiuk J, Pinkau T. (1990) *J. Bacteriol.* **172**(3) 1478-1484.

Tsfadia Y, Shaked I, Daniel E. (1990) *Eur. J. Biochem.* **193**(1) 25-29.

Vagin A, Teplyakov A. (1997) *J. Appl. Cryst.* **30** 1022-1025.

Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ. (1998) *Cell* **93**(5) 863-873.

Van der Veis SM, Viitanen PV, Gatenby AA, Lorimer GH, Jaenicke R. (1992) *Biochemistry* **31** 3635-3644.

Van der Veis SM, Gatenby AA, Georgopoulos C. (1994) *Nature* 368(6472) 654-656.

Viale AM, Araaki AK. (1994) *FEBS Lett.* **341**(2-3) 146-151.

Viitanen PV, Lubben TH, Reed J, Goloubinoff P, O'Keefe DP, Lorimer GH. (1990) *Biochemistry* **29**(24) 5665-5671.

Viitanen PV, Donaldson GK, Lorimer GH, Lubben TH, Gatenby AA. (1991) *Biochemistry* **30**(40) 9716-9723.

Viitanen PV, Gatenby AA, Lorimer GH. (1992) *Protein Sci.* **1**(3) 363-369.

Viitanen PV, Schmidt M, Buchner J, Suzuki T, Vierling E, Dickson R, Lorimer GH, Gatenby AA, Soll J. (1995) *J. Biol. Chem.* **270**(30) 18158-18164.

Waldmann T, Nimmesgern E, Nitsch M, Peters J, Pfeifer G, Muller S, Kellermann J, Engel A, Hartl FU, Baumeister W. (1995) *Eur. J. Biochem.* **227**(3) 848-856.

Waldmann T, Lupas A, Kellermann J, Peters J, Baumeister W. (1995) *Biol. Chem.* Hoppe Seyler **376**(2) 119-126.

Wang D, Driessen HPC, Tickle IJ. (1991) *J. Mol. Graph.* **9** 38 50-52.

Weiss JB, Ray PH, Bassford PJ Jr. (1988) *Proc. Natl. Acad. Sci.* USA **85**(23) 8978-8982.

Weissman JS, Kashi Y, Fenton WA, Horwich AL. (1994) *Cell* **78**(4) 693-702.

Weissman JS, Hohl CM, Kovalenko O, Kashi Y, Chen S, Braig K, Saibil HR, Fenton WA, Horwich AL. (1995) *Cell* **83**(4) 577-587.

Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL. (1996) *Cell* **84**(3) 481-490.

Wickner S, Gottesman S, Skowyra D, Hoskins J, McKenney K, Maurizi MR. (1994) *Proc. Natl. Acad. Sci.* USA **91**(25) 12218-12222.

Wider G, Wuthrich K. (1999) *Curr. Opin. Struct. Biol.* **9**(5) 594-601.

Wiech H, Buchner J, Zimmermann R, Jakob U. (1992) *Nature* **358**(6382) 169-170.

Wonacott AJ, Arndt UW. (1977) North Holland Publishing Company, Amsterdam.

Xu Z, Horwich AL, Sigler PB. (1997) *Nature* **388**(6644) 741-750.

Yang W, Hendrickson WA, Kalman ET, Crouch RJ. (1990) *J. Biol. Chem.* 265(23) 13553-13559.

Yifrach O, Horovitz A. (1994) *J. Mol. Biol.* **243**(3) 397-401.

Yifrach O, Horovitz A. (1995) *Biochemistry* **34**(16) 5303-5308.

Yifrach O, Horovitz A. (1996) *J. Mol. Biol.* **255**(3) 356-361.

Yifrach O, Horovitz A. (1998) *Biochemistry* **37**(20) 7083-7088.

Yifrach O, Horovitz A. (2000) *Proc. Natl. Acad. Sci.* USA **97**(4) 1521-1524.

Zahn R, Buckle AM, Perrett S, Johnson CM, Corrales FJ, Golbik R, Fersht AR. (1996) *Proc. Natl. Acad. Sci.* USA **93**(26) 15024-15029.

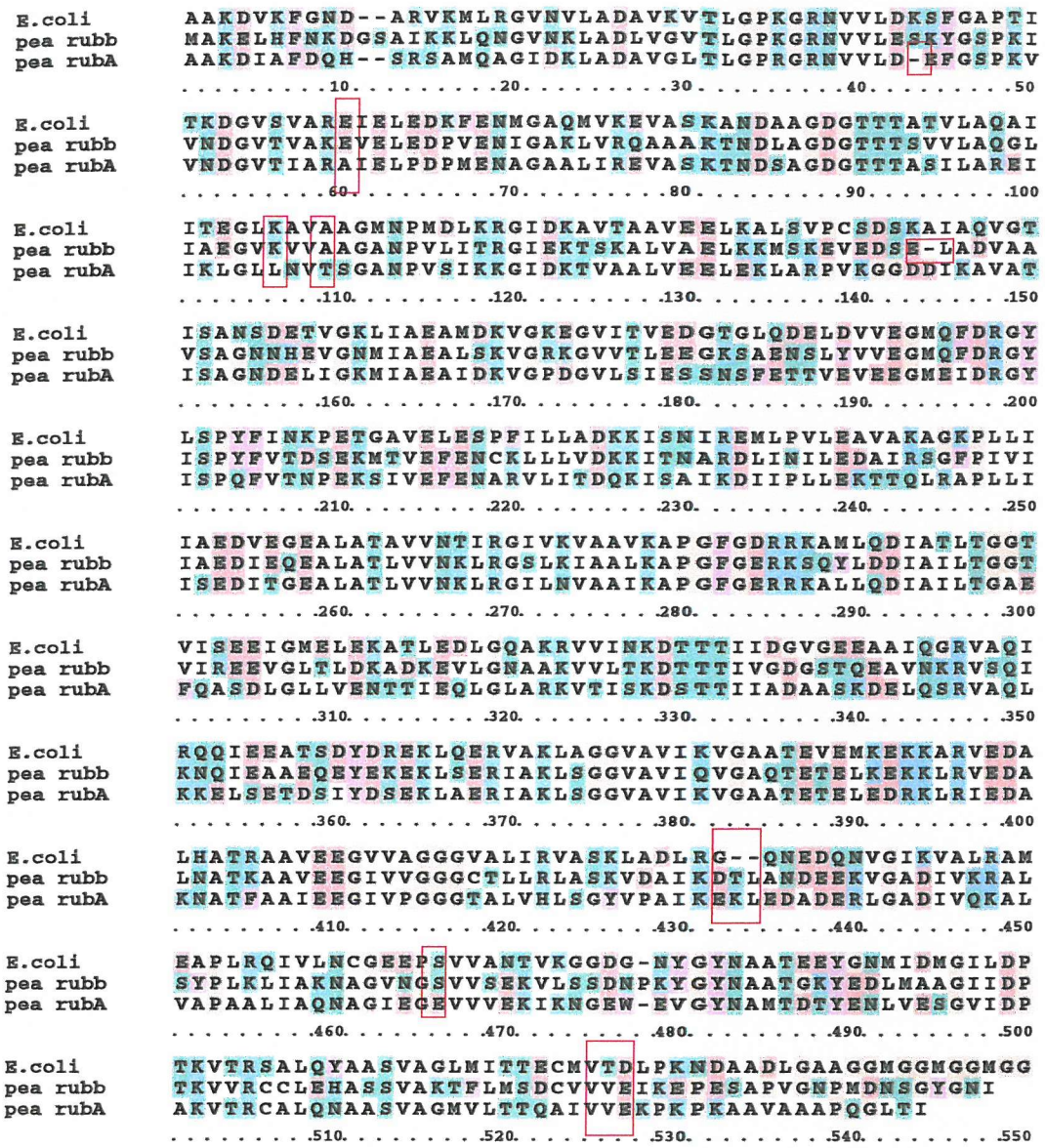
Zeilstra_Ryalls J, Fayet O, Georgopoulos C,(1996) *Annu. Rev. Microbiol.* **45** 301-325.

Zheng X, Rosenber LE, Kalousek F, Fenton WA. (1996) *J. Biol. Chem.* **268**(10) 7489-7493.

Ziemienowicz A, Skowyra D, Zeilstra-Ryalls J, Fayet O, Georgopoulos C, Zylicz M. (1993) *J. Biol. Chem.* **268**(34) 25425-25431.

Zimmerman SB, Trach SO. (1991) *J. Mol. Biol.* **222**(3) 599-620.

APPENDIX 1



Depicted in the red boxes are the deletion/insertion sites relevant to the GroEL and plastid α and βsubunits. For a complete alignment explanation for the above depictions, see Chapter Two.

The sequence alignment picture was acquired through the MALIGN program on the EXPASY website.

APPENDIX 2

```
 ###########################################################
  ###########################################################
  ###########################################################
  ### CCP PROGRAM SUITE: Scala        VERSION 4.0:        ##
  ###########################################################
  User: terry  Run date:  2/27/01  Run time:18:52:19
  Please reference: Collaborative Computational Project, Number 4. 1994.
  "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. D50, 760-763.

  as well as any specific reference in the program write-up.
                  ******************************************
                  *    SCALA - continuous scaling program  *
                  *                                        *
                  *    Version: 2.7.2                      *
                  *    Date   : 5/1/2000                   *
                  *                                        *
                  *   Phil Evans, MRC LMB, Cambridge, UK   *
                  *   pre@mrc-lmb.cam.ac.uk                *
                  *                                        *
                  ******************************************

 * Number of Columns =  16

  * Number of Reflections = 142247

  * Missing value set to NaN in input mtz  file

  * Number of Batches = 240

  * Column Labels :

 H K L M/ISYM BATCH I SIGI IPR SIGIPR FRACTIONCALC XDET YDET ROT WIDTH LP MPART

  * Column Types :

 H H H Y B J Q J Q R R R R R R

  * Cell Dimensions :

    108.311  108.311  154.682   90.000   90.000  120.000

 *  Resolution Range :

      0.00050     0.12997      (   44.721 -     2.774 A )

  * Sort Order :

      1     2     3     4     5

  * Space group = R32  (number   155)

 :Multiplicity v Resolution :N:2,9:
 : Rmeas, Rsym & PCV v Resolution :N:2,12,13,14,15,16:    $$
   N 1/resol^2 Dmin    Nmeas    Nref   Ncent   %poss Cm%poss Mlplcty AnomCmpl AnomFrc   Rmeas  Rmeas0  (Rsym)    PCV     PCV0
$
 $$
   1  0.013    8.77     1834     323      95    99.4    99.4     5.7     0.0     0.0   0.139   0.139   0.127   0.164   0.164
   2  0.026    6.20     3487     541     100   100.0    99.8     6.4     0.0     0.0   0.132   0.132   0.121   0.153   0.153
   3  0.039    5.06     4776     692      95   100.0    99.9     6.9     0.0     0.0   0.125   0.125   0.116   0.152   0.152
   4  0.052    4.39     5711     825     104   100.0    99.9     6.9     0.0     0.0   0.113   0.113   0.105   0.134   0.134
   5  0.065    3.92     6411     907      97   100.0    99.9     7.1     0.0     0.0   0.113   0.113   0.105   0.141   0.141
   6  0.078    3.58     7068    1005      98   100.0   100.0     7.0     0.0     0.0   0.169   0.169   0.156   0.215   0.215
   7  0.091    3.32     7693    1084      99   100.0   100.0     7.1     0.0     0.0   0.203   0.203   0.188   0.253   0.253
   8  0.104    3.10     8189    1147     102   100.0   100.0     7.1     0.0     0.0   0.256   0.256   0.238   0.319   0.319
   9  0.117    2.92     8862    1243      91   100.0   100.0     7.1     0.0     0.0   0.356   0.356   0.330   0.445   0.445
  10  0.130    2.77     8313    1252      95    96.2    99.4     6.6     0.0     0.0   0.509   0.509   0.469   0.633   0.633
 $$
"><b>For inline graphs use a Java browser</b></applet>

           Overall     62344    9019     976    99.4    99.4     6.9     0.0     0.0   0.151   0.151   0.140   0.184   0.184
                       Nmeas    Nref   Ncent   %poss Cm%poss Mlplcty AnomCmpl AnomFrc   Rmeas  Rmeas0  (Rsym)    PCV     PCV0

    Imax    Rfac  Rfull  Ranom  Nanom    Av_I    SIGMA  I/Sigma      sd   Nmeas   Nref   Ncent  FRCBIAS  Nbias  Abs.BIAS
 $$  $$
     943.   0.679  0.346  0.000      0     373.    419.6      0.9   477.4   26729   3895     466   -0.145  11675     -51.8
    1967.   0.260  0.138  0.000      0    1394.    513.1      2.7   476.4    9859   1391     113   -0.085   4178    -117.3
    3090.   0.202  0.113  0.000      0    2493.   1425.3      1.7   516.8    5944    833      66   -0.095   2679    -232.9
    4331.   0.169  0.086  0.000      0    3685.    836.9      4.4   587.7    4055    573      56   -0.029   1607    -107.9
    5718.   0.151  0.069  0.000      0    5025.   1336.7      3.8   665.1    2923    417      32   -0.079   1194    -393.1
    7290.   0.130  0.087  0.000      0    6559.   1173.3      5.6   767.2    2235    322      32   -0.093    930    -598.5
    9106.   0.131  0.078  0.000      0    8260.   1414.6      5.8   875.8    1799    260      16   -0.106    773    -857.4
   11253.   0.117  0.089  0.000      0   10268.   2675.0      3.8  1024.1    1515    218      20   -0.105    646   -1052.3
   13881.   0.108  0.066  0.000      0   12707.   1819.0      7.0  1235.9    1371    199      21   -0.099    548   -1245.6
   17269.   0.105  0.068  0.000      0   15543.   2160.9      7.2  1488.3    1342    194      28   -0.112    599   -1730.0
   22044.   0.105  0.068  0.000      0   19553.   2756.3      7.1  1793.8    1331    199      29   -0.133    556   -2571.5
   30208.   0.093  0.061  0.000      0   25572.   3081.5      8.3  2299.2    1122    162      19   -0.158    444   -4055.8
  218967.   0.099  0.059  0.000      0   52362.   7537.2      6.9  4419.1    2072    309      50   -0.183    785   -8869.1
 $$
"><b>For inline graphs use a Java browser</b></applet>

 Overall   0.140  0.085  0.000      0    5052.   1765.8      2.9   762.7   62297   8972     948   -0.131  26614     -607.7
            Rfac  Rfull  Ranom  Nanom    Av_I    SIGMA  I/Sigma      sd   Nmeas   Nref   Ncent  FRCBIAS  Nbias  Abs.BIAS
```

Selected SCALA (CCP4 suite) data for the Ch-cpn60 α subunit in space group R32.