CHARACTERISATION OF THE J DOMAIN AMINO ACID RESIDUES IMPORTANT FOR THE INTERACTION OF DNAJ – LIKE PROTEINS WITH HSP70 CHAPERONES

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

RHODES UNIVERSITY

FRITHA HENNESSY

January 2004

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy in Rhodes University. It has not been submitted before for any degree or examination in any other university.

_____

_____day of _____2004

ABSTRACT

The 70 kDa heat shock proteins (Hsp70s) are vital for normal protein folding, as they stabilise the unfolded state of nascent polypeptides, allowing these sufficient time to attain a correct tertiary structure. Hsp70s are aided by the DnaJ-like family of proteins, which interact with Hsp70s in order to enhance the chaperone activity of these proteins. DnaJ-like proteins contain a J domain, a seventy amino acid domain consisting of four α-helices, which is defined by the presence of an invariant tripeptide of histidine, proline and aspartic acid (HPD motif). This motif is key to the interaction between DnaJ-like proteins and Hsp70s. This thesis has focused on determining the presence of other conserved residues in the J domain and their role in mediating the interaction of DnaJ-like proteins with partner Hsp70s. DnaJ-like proteins from *Agrobacterium tumefaciens* RUOR were isolated and used as a model system. *A. tumefaciens* DnaJ (*Agt* DnaJ) was able to replace the lack of *E. coli* DnaJ in an *E. coli* null mutant strain, however, additional *A. tumefaciens* DnaJ-like proteins *Agt* DjC1/DjlA, *Agt* DjC2 and *Agt* DjC5 were unable to complement for the lack of *E. coli* DnaJ. Replacement of the *Agt* DnaJ J domain with J domains from these proteins resulted in non-functional chimeric proteins, despite some sequence conservation. The kinetics of the basal specific ATPase activity of *Agt* DnaK, and its ability to have this activity stimulated by *Agt* DnaJ and *Agt* DnaJ-H33Q were also investigated. Stimulation of the ATPase activity by *Agt* DnaJ ranged between 1.5 to 2 fold, but *Agt* DnaJ-H33Q was unable to stimulate the basal ATPase activity. Conserved amino acids in the J domain were identified *in silico*, and these residues were substituted in the J domain of *Agt* DnaJ. The ability of these derivative proteins to replace *E. coli* DnaJ was investigated. Alterations in the HPD motif gave rise to proteins unable to complement for lack of *E. coli* DnaJ, consistent with literature. *Agt* DnaJ-R26A was unable to replace *E. coli* DnaJ suggesting that Arg^{26} could be key to the interaction with partner Hsp70s. *Agt* DnaJ-D59A was unable to replace *E. coli* DnaJ; substitutions in Asp^{59} have not previously been shown to impact on the function of DnaJ. Substituting Arg^{63} in *Agt* DnaJ abrogated the levels of complementation. Substitution of several structural residues was also found to disrupt the *in vivo* function of *Agt* DnaJ suggesting that the maintenance of the structural integrity of the J domain was important for function. This study has identified a number of residues critical to the structure and function of the J domain of *Agt* DnaJ, and potentially of general importance as molecular determinants for DnaJ-Hsp70 interaction.

This thesis is dedicated to my grandparents

Andrea and Desmond Utton

and

Barbara and James Hennessy

ACKNOWLEDGMENTS

The following people deserve recognition for the roles they have played in my life over the course of the last four years.

At a professional level, the following people and entities deserved thanks.

Firstly, my supervisor Prof Greg Blatch, for providing a fascinating project and a wonderful environment in which to do this work in, as well as being a fantastic supervisor from whom I have learnt a vast amount, as well as allowing me the freedom to take this project in differing directions.

The National Research Foundation of South Africa, Volkswagen and the Wellcome Trust supplied monetary support.

Drs Mike Cheetham, Paul Chapple and Simon Wilkins kindly donated materials and reagents, as well as making me feel very welcome in the lab in London during the time I spent there.

Thanks are also due to Prof Rosemary Dorrington, Dr Carol Hartley and Ms Mesbaah Jiwaji, who donated *A. tumefaciens* RUOR genomic DNA and cells, and to Dr Olivier Deloche, who donated the *E. coli* OD259 strain and control plasmids.

At a personal level, the following people deserve thanks.

My parents, who have provided much support (monetary and otherwise), without which I would not have been able to survive the last four years. Thank you for your love, support and all you have done for me. My father also deserves thanks for helping with the parodies found in Chapter 1. The rest of my family, who have provided much support to me over the course of this project.

The members of the Chaperone Lab in the Department of Biochemistry, Microbiology and Biotechnology at Rhodes University, past and present: Delia, Adrienne, Tayo, Tonderayi,

Mike, Vicky, Linda, William, Addmore, Will and Caryn. We have learnt a lot together. Thank you for all the memories.

The honorary members of the Chaperone Lab have provided support and friendship during the tough, and not-so-tough times. Thanks firstly go to Delia, Henry, Mike, Vuyo, Will and Stewart for being such fantastic friends; to Doug, for everything; to Gwyn for her friendship and for always being there with a smile when I needed one; to Michelle, Lester and Shalene who have taught me so much, and for providing shoulders to cry on; to Alex and Nicky, who have provided the choral background; and finally, to Adrienne and Neil for the long distance support!

Thanks are also due to my father and Alexandra Holland for proof reading this manuscript. Thanks go to Gareth Huckle for helping with the printing of this manuscript.

> *"Never since have the Ainur made any music like to this music, though it has been said that a greater still shall be made before Ilúvatar by the choirs of the Ainur and the Children of Ilúvatar after the end of days. Then the themes of Ilúvatar shall be played aright, and take Being in the moment of their utterance, for all shall then understand fully his intent in their part, and each shall know the comprehension of each, and Ilúvatar shall give to their thoughts the secret fire, being well pleased."*
>
> J.R.R. Tolkien, Ainulindalë, The Silmarillion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

α	Alpha
β	Beta
λ	Lambda
μg	Microgram(s)
μl	Microlitre(s)
μM	Micromolar
μmol	Micromole(s)
°C	Degrees Centigrade
A	Absorbance
A. thaliana	*Arabidopsis thaliana*
A. tumefaciens	*Agrobacterium tumefaciens*
ADP	Adenosine diphosphate
Agt DnaJ	*A. tumefaciens* DnaJ
Agt DnaK	*A. tumefaciens* DnaK
Amp^R	Ampicillin resistance
ANJ1	*Atriplex numularia* DnaJ1
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
B. burgdorferi	*Borrelia burgdorferi*
B. japonicum	*Bradyrhizobium japonicum*
B. melitensis	*Brucella melitensis*
B. taurus	*Bos taurus*
Bag-1	Bcl2-associated athanogene 1
BAP	BiP associated protein
BiP	Binding protein
BLAST	Basic Local Alignment Search Tool
bp	Base pair(s)
BSA	Bovine serum albumin
CbpA	Curved DNA binding protein
csp	Cysteine string protein
CTAB	Hexadecyltrimethyl-ammonium bromide
D. melanogaster	*Drosophila melanogaster*
DjlA	DnaJ-like A
DjlB	DnaJ-like B
DjlC	DnaJ-like C

DNA	Deoxyribonucleic acid
DnaJ	Prokaryotic Hsp40
DnaK	Prokaryotic Hsp70
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetra-acetic acid
ER	Endoplasmic reticulum
ERdj5/JPD1	Endoplasmic reticulum DnaJ protein 5 / J protein domain 1
Fes1p	Factor exchange for Ssa1p
g	gram
G/F region	Glycine-phenylalanine rich region
H. sapiens	*Homo sapiens*
Hdj1	Human DnaJ-like protein 1
Hdj2	Human DnaJ-like protein 2
Hip	Hsc70 interacting protein
Hop	Hsp70 – Hsp90 organising protein
HPD motif	Histidine, proline and aspartic acid motif
His Tag	6 x Histidine Tag
Hsc	Heat shock cognate protein
Hsc20	Hsc of 20 kDa
Hsc66	Hsc of 66 kDa
Hsc70	Hsc of 70 kDa
HscC	Heat shock cognate protein C
Hsj1	*Homo sapiens* DnaJ protein 1
Hsp	Heat shock protein
Hsp40	Hsp of 40 kDa
Hsp60	Hsp of 60 kDa
Hsp70	Hsp of 70 kDa
Hsp90	Hsp of 90 kDa
IDT	Integrated DNA Technologies
IPTG	Isopropyl βD-thiogalactoside
kDa	Kilo Dalton
Ldj2	Leek DnaJ2
M	Molar
M. loti	*Mesorhizobium loti*
M. musculus	*Mus musculus*
mg	Milligram
ml	Millilitre
MSA	Multiple sequence alignment

NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
P. dentrificans	*Psuedomonas dentrificans*
P. falciparum	*Plasmodium falciparum*
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDI	Protein disulphide isomerase
pDJA1	pig DnaJ-like protein A1
Pi	Inorganic phosphate
RESA	Ring infected erythrocyte surface antigen
RMSD	Root mean square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
S. cerevisiae	*Saccharomyces cerevisiae*
S. meliloti	*Sinorhizobium meliloti*
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sHsp	Small Hsp
Sls1p	*S. cerevisiae* lethal mutations with the 7S RNA mutation
SPDBV	Swiss Protein Data Bank Viewer
SPR	Surface plasmon resonance
T Ag	T antigen
TBE buffer	Tris Borate EDTA buffer
TBS	Tris buffered saline
TBST	TBS containing Tween 20
TE buffer	Tris-EDTA buffer
TEMED	N,N,N’N’-tetramethylethylenediamine
TPR	Tetratricopeptide repeat
Tris	Tris-2-amino-2-(hydroxymethyl)-1,3,propandiol
URL	Uniform Resource Location
w/v	Weight per volume
Ydj1	Yeast DnaJ
YT agar	Yeast tryptone agar
YT broth	Yeast tryptone broth

The IUPAC-IUBMB three and one letter codes for amino acids were used, and single letter codes were used for nucleotides.

Chapter 1:
Literature Review

1 LITERATURE REVIEW

1.1 THE PROTEIN FOLDING PROBLEM

"To fold or not to fold? That is the question: -
Whether 'tis nobler in the cell to allow,
The twists and turns of outrageous folding,
Or to take arms against a sea of chaperones,
But by opposing them, end."[1]

The attainment of the correct tertiary structure of a primary amino acid sequence requires a complex mechanism, the precise nature of which is still being elucidated. The tertiary structure of any given polypeptide sequence is dependent upon its primary amino acid structure. This was the central tenet of structural biology initially formulated by Anfinsen following work on the refolding characteristics of the small protein ribonuclease.

1.1.1 The folding of ribonuclease

Anfinsen and co-workers demonstrated that the sole requirement for a protein to fold correctly was its primary amino acid sequence (Anfinsen, 1973). They showed that the protein upon refolding could regain its native conformation, without any other factors being present which could contribute to the re-attainment of the correct tertiary conformation. However, under those conditions ribonuclease took about 20 minutes to fold correctly, which was significantly longer than the length of time required in the cell. It was postulated that the reason for this discrepancy was due to the presence of disulphide bonds in the ribonuclease, and the increased time required for refolding was due to the incorrect formation of several of these bonds, which would need to be broken before the correct bonds could be formed (Anfinsen *et al.*, 1961; Haber and Anfinsen, 1962). Incorrect formation of these bonds prevented the ribonuclease from attaining full activity. The addition of a protein disulphide isomerase (Anfinsen, 1973) or sulphydryl containing substances (Haber and Anfinsen, 1962) to the refolding ribonuclease increased the speed of folding to native rates. The conclusion

[1] - With apologies to William Shakespeare. The original quote can be found in *Hamlet*, Act III, Scene I.

from this work, and a conclusion that has become a standard dogma in structural biology today, was termed the principle of self-assembly. It stated that the sole requirement for the correct folding of proteins, from their primary amino acid sequence to their correct tertiary and quaternary structures, was their amino acid sequence (Goldberger *et al.*, 1963).

Ribonuclease is a relatively small protein, and the number of potential conformations that it could adopt on its way to a correctly folded conformation will therefore be substantially less than for a large multimeric protein. This concept was summarised by Levinthal in 1968 (Levinthal, 1968; Zwanzig *et al.*, 1992). The Levinthal paradox states that if every amino acid in a polypeptide of 100 amino acids in length can adopt one of ten different conformations, then the number of different possible conformations that the polypeptide chain can attain will be 10^{100}. The time that it would take for the chain to attain every possible conformation would be approximately the length of the life of the universe, and as such is not feasible. Hence, the question that needs to be answered is how do polypeptide chains achieve their native tertiary and quaternary states in a cell, in the limited time frame that they do? A limited bias against energetically unfavourable structures would be one method (Zwanzig *et al.*, 1992). It was known that folding steps for a protein both *in vivo* and *in vitro* were identical (Ruddon and Bedows, 1997). However, the conditions in most *in vitro* experiments are fundamentally different from the physiological conditions in the cell (reviewed in Ruddon and Bedows, 1997). Therefore there must be a way that the cell can overcome the possibility of protein aggregation, which was the predominant outcome of protein misfolding, and allow for the synthesis of correctly folded protein. An analysis of all the possible answers is beyond the scope of this work, and instead this review will focus on a group of proteins that have evolved to play a significant role in aiding nascent polypeptides to attain their native conformation.

1.1.2 Effects of Cellular Stress on Cells

Some of the initial experimental work performed on eukaryotic systems focused on the common fruit fly, *Drosophila*. This was in part due to its relatively short life span (two weeks), which allowed for ease of performing experimental work over several generations. Initial work revealed a specific pattern of puffing on the polytene chromosomes seen in the salivary glands of *Drosophila* larvae. These polytene chromosomes form due to repeated

chromosomal replication, without the dividing of the chromosomes at the end of replication. This gives rise to extra large chromosomes, with a specific pattern of bands and inter-bands. Incorporation of P^{32} (Gross, 1957) and tritium labelled uridine (Pelling, 1959) was shown to occur around these puffs, predominately into RNA. These regions were transcribed at a higher rate than the rest of the chromosome; the genes encoded in this region were generally the essential housekeeping genes. Part of the work that was performed entailed looking at the effect of heat shock on *Drosophila bucksii* larvae. Larvae had their temperature inadvertently but serendipitously altered from 25°C to 30°C for 30 minutes and after this heat shock it was seen that the pattern of the puffing shifted dramatically (Ritossa, 1962; Ritossa, 1996). This implied that the transcription of the house keeping genes was shut down, and a new set of mRNAs was being transcribed, and hence a new set of proteins was being produced (Pauli *et al.*, 1992). The class of proteins that were isolated, corresponding to the gene products of the regions that showed increased transcriptional activity, were given the name heat shock proteins (Hsp) (Tissières *et al.*, 1974; Georgopoulos and Welch, 1993). The response of the organism to the heat shock involved a coordinated genetically defined response (Morimoto *et al.*, 1997). Later work showed that the increased production of these proteins also resulted from other forms of cellular stress (Feige and Mollenhauer, 1992). Consequently, it appeared that these Hsps were produced in response to the higher amounts of abnormally folded proteins that form during periods of cell stress (reviewed in Georgopoulos and Welch, 1993). The Hsps were named according to their apparent molecular mass when resolved using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Georgopoulos and Welch, 1993). These include proteins of 60 kDa (Hsp60), 90 kDa (Hsp90), 70 kDa (Hsp70) and 40 kDa (Hsp40), as well as the small heat shock proteins (sHsp).

Homologous proteins in various organisms that corresponded to constitutively produced Hsps have also been identified (Ingolia and Craig, 1982; Bardwell and Craig, 1984). These proteins were termed heat shock cognate proteins (Hsc). They are important in terms of the stabilisation of newly synthesised proteins; hence, their primary cellular role was to mediate the folding of these newly synthesised proteins during the normal cellular cycle. Many Hsp and Hsc homologues were initially identified performing another function. An example of this was the 70 kDa Hsc (Hsc70) protein called BiP (Binding Protein), which was initially identified as binding to the heavy chains of immunoglobulins (Haas and Wabl, 1983).

1.2 CHAPERONES AND CO-CHAPERONES

> *"If chaperones be the food of life, fold on;*
> *Give us excess of them, that sufficing,*
> *The aggregation may decrease, and so fail."*[2]

Heat shock proteins play an important role in the prevention of protein aggregation, and in the formation of the correct tertiary and quaternary structures of many nascent polypeptides. As such, they are members of the group of proteins that have come to be termed molecular chaperones (Ellis and van der Vies, 1991). This term was first proposed by Ron Laskey and co-workers (Laskey *et al.*, 1978). They were investigating an acidic, soluble, nuclear protein called nucleoplasmin. This protein mediated the *in vitro* assembly of nucleosomes and DNA. The addition of DNA to histone proteins at normal ionic strength caused precipitation of the histones, and was an example of a failure of the principle of self-assembly. However, addition of a molar excess of nucleoplasmin to the histones before the addition of DNA allowed for the formation of soluble nucleosome cores on the addition of DNA. In other words, nucleoplasmin chaperoned the formation of nucleosomes. This term was then expanded on by Ellis (Ellis, 1987; Ellis, 1990). He proposed

> *"...the term 'molecular chaperones' to describe a class of cellular proteins whose function is to ensure that the folding of certain other polypeptide chains and their assembly into oligomeric structures occur correctly"* (Ellis, 1987).

1.2.1 Heat Shock Proteins as molecular chaperones

Heat shock proteins can be produced either constitutively or under stress conditions such as heat shock (Ellis and van der Vies, 1991). Cellular stress can be caused by several factors such as heat shock, presence of heavy metals, oxygen or nutrient deprivation, and cold shock. Some chaperones were induced by several of these factors, whereas others such as the *E. coli* protein Hsc66 were induced by only one, in this case cold shock (Lelivelt and Kawula, 1995). The heat shock proteins produced during periods of cellular stress were vital in preventing

[2] With further apologies to William Shakespeare. The original quote is found in *Twelfth Night*, Act I, Scene I.

protein aggregation, and allowing for the rapid return of the cellular machinery to normal functioning after the removal of the stress. Constitutively produced chaperones play a role in the normal functioning of the cell, particularly with respect to co-translational protein folding (Federov and Baldwin, 1997).

1.2.2 Cellular requirements for molecular chaperones

Newly synthesised polypeptides can reach approximately 10 μM concentration in the cytoplasm and 50 μM in the rough endoplasmic reticulum (ER) (Flynn *et al.*, 1989) which could put pressure on cellular systems to ensure that unfolded proteins do not interact non-specifically and aggregate. However, it is still unclear what proportion of these proteins require molecular chaperones to fold. Pulse chase experiments, in combination with immunoprecipitation, indicate that Hsc70 will associate temporarily with newly synthesised polypeptides. These will range in size, but will be larger than 20 kDa (Thulasiraman *et al.*, 1999). The proportion of polypeptides that associated with Hsc70 was between 15 and 20 %, and it appeared to be a transient interaction as minimal amounts of polypeptide remained bound after 30 minutes (Thulasiraman *et al.*, 1999). By contrast work performed using the eukaryotic Hsp60 showed a far more specific peptide interaction, with peptides between 30 and 60 kDa being bound (Thulasiraman *et al.*, 1999). Similar results were seen when the role of the prokaryotic protein DnaK was examined. Between 9 and 18 % of newly synthesised polypeptides associate with DnaK (Deurling *et al.*, 1999). As seen for Hsc70, there was a large variability in the length of polypeptides that can associate with DnaK (Deurling *et al.*, 1999). Nascent polypeptide chains generally bound for a longer period of time in the eukaryotic system to Hsc70, than such chains in the polypeptide system do to DnaK (Deurling *et al.*, 1999; Thulasiraman *et al.*, 1999). Hence, there does appear to be a necessity for a certain proportion of nascent chains to associate with an Hsp70 in order to fold correctly.

1.3 HSP70 AND HSC70

"One Protein to find them all,
One Protein to bring them,
One Protein to hold them all,
And in the cytosol, fold them."[3]

Hsp70s are one class of molecular chaperones that recognize unfolded proteins, binding and stabilising them, thereby giving them enough time to fold correctly. They also have a variety of functions related to this, such as the stabilisation of precursor proteins, the aiding in the translocation of newly synthesized proteins, the rearrangement of protein oligomers, and the protection of the cell against the effects of cellular stress (Becker and Craig, 1994). Hsp70s have also been implicated in protein degradation, such as through the lysosomal pathway (Agarraberes and Dice, 2001), the specific degradation of the heat shock transcription factor σ^{32} in *E. coli* (Tatsuta *et al.*, 2000), and *via* the ubiquitination pathway in eukaryotes for example in the degradation of polyglutamine repeat containing proteins (reviewed in Sakahira *et al.*, 2002). These functions are all dependent on the Hsp70 keeping polypeptides in an extended conformation, and stabilising the exposed hydrophobic regions of the target polypeptide (Brodsky, 1996). The most investigated Hsp70 is the *Escherichia coli* protein DnaK. Its co-chaperone protein is DnaJ. DnaK is a homologue of Hsp70 present in *Drosophila melanogaster,* and DnaJ is related to Hsp40. They were initially identified as being necessary for bacteriophage λ DNA replication, hence the Dna nomenclature (Georgopoulos and Welch, 1993). Hsp70s are found in all prokaryotes and eukaryotes, as well as in some, but not all of the archaea (Conway De Macario and Macario, 1994; Macario *et al.*, 1999). Some methanogens, extreme halophiles and extreme thermophiles do not appear to synthesize, or even contain a genomic copy of, a 70 kDa heat shock protein; for example the genomic sequence for *Methanococcus jannaschii* appeared not to contain a genomic copy of an Hsp70 gene (Bult *et al.*, 1996; reviewed in Macario *et al.*, 1999).

There are two distinct classes of Hsp70s: the constitutive form (Heat shock cognate protein 70, Hsc70) (Ingolia and Craig, 1982) and the stress inducible form (Hsp70). There appear to

[3] With apologies to J.R.R. Tolkien. The original quote can be found in *The Lord of the Rings.*

be minimal differences between the two in terms of function. The cognate form will act as a molecular chaperone under normal cellular conditions, whereas the heat shock protein will function under stress conditions.

1.3.1 Domain Structure of Hsp70

All Hsp70s are composed of three regions, namely a highly conserved approximately 44 kDa ATPase domain, a conserved approximately 15 kDa substrate binding region, and a less conserved approximately 10 kDa C terminal region. The structures of the ATPase domain and the substrate binding region have been determined and are shown below in Figure 1.1.

1.3.1.1 ATPase domain

The ATPase domains of bovine Hsc70 (Flaherty *et al.*, 1990), *E. coli* DnaK (in conjunction with the nucleotide exchange factor GrpE) (Harrison *et al.*, 1997), bovine Hsc70 (in conjunction with the nucleotide exchange factor Bag-1) (Sondermann *et al.*, 2001) and human Hsp70 (Osipiuk *et al.*, 1999) have had their structures determined. The ATPase domain consisted of two lobes with a nucleotide-binding site between them. The two lobes were similar in size, and each lobe could be further subdivided into two subdomains, giving subdomains IA, IB, IIA and IIB (Flaherty *et al.*, 1990). It bore structural similarity to the ATPase domains of hexokinase and actin, despite low sequence similarity between these proteins (Flaherty *et al.*, 1991; Bork *et al.*, 1992). The presence of adenosine triphosphate (ATP) or adenosine diphosphate (ADP) in the nucleotide-binding cleft affected Hsp70s chaperone activity. The structure of the ATPase domain of bovine Hsc70 is depicted below in Figure 1.1A.

1.3.1.2 Substrate binding domain

The structures of substrate binding domains have been determined from *E. coli* DnaK (Zhu *et al.*, 1996; Wang *et al.*, 1998; Pellechia *et al.*, 2000) and rat Hsc70 (Morshauser *et al.*, 1999). The structure of the substrate binding region of DnaK determined in the presence of a model peptide is shown in Figure 1.1B (Zhu *et al.*, 1996). The domain comprised two subdomains. The first was a predominately β-strand rich region, involved in the binding of amino acids in a

substrate. This was followed by a α-helical region, proposed to act as a lid, locking substrate into the domain. The peptide used in the determination of the DnaK substrate binding region (Figure 1.1B) had the sequence NRLLLTG (single letter amino acid code) (Gragerov *et al.*, 1994), and bound to the substrate binding region in an extended conformation.

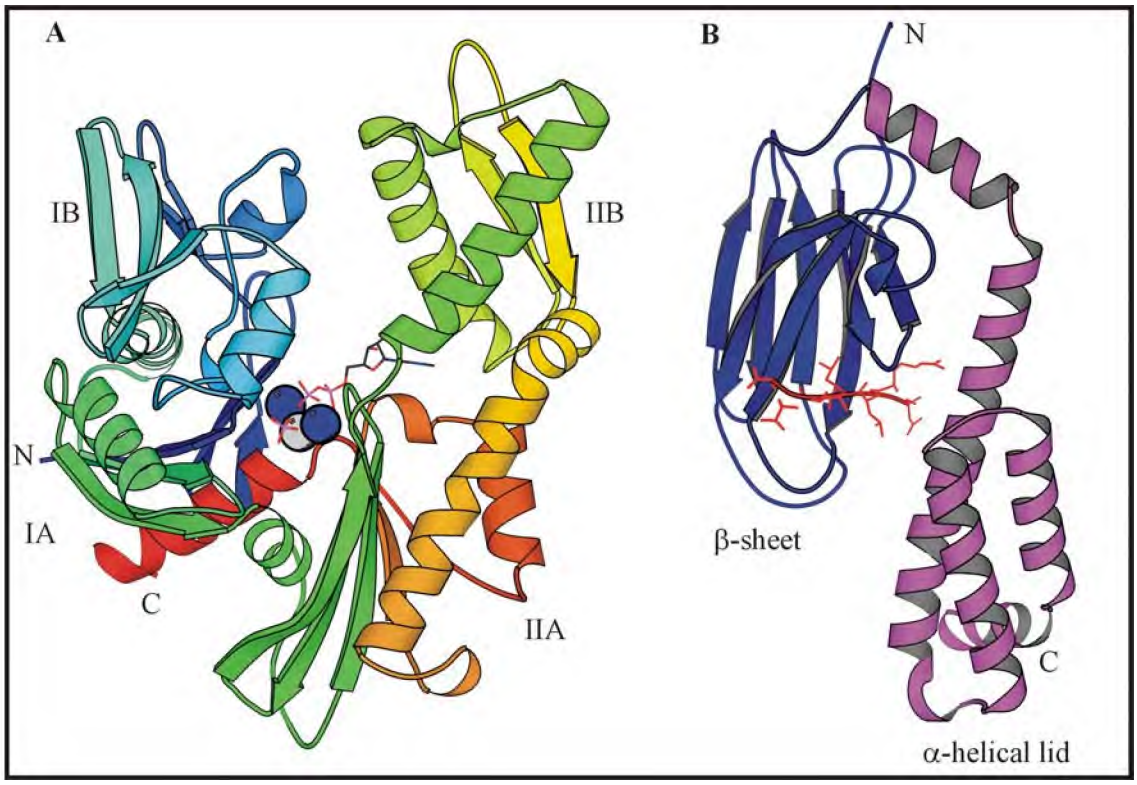

Figure 1.1: Ribbon representation of the ATPase domain and substrate binding region of Hsp70s

Figures were made using Molscript (Kraulis, 1991). (A) ribbon representation of the ATPase domain of the *Bos taurus* Hsc70. The PDB accession code is 3HSC (Flaherty *et al.*, 1990). ADP is represented by sticks and magnesium by spheres. The figure is colour coded with the N terminus being blue, and the C terminus red. (B) ribbon representation of the substrate binding region of *E. coli* DnaK. The PDB accession code is 1DKX (Zhu *et al.*, 1996). The substrate binding β-sheet region is in blue, and the α-helical lid region is in purple. The peptide substrate is coloured red.

1.3.2 Chaperone function of Hsp70

E. coli DnaK has been shown to interact strongly with peptides enriched in phenylalanine, isoleucine, leucine and valine (Richarme and Kohiyama, 1993; Gragerov *et al.*, 1994). Leucine was found in approximately 90 % of all regions specifically recognized by Hsp70, with sites on unfolded polypeptide chains occurring statistically every 36 residues (Rüdiger *et*

al., 1997), and at least one site was likely to be present in all proteins. The optimal length of the substrate has been shown to be at least seven or eight amino acids (Flynn *et al.*, 1991). There are four potential substrate binding sites in the substrate binding region, with an optimum binding sequence on the substrate of X-Hy-(W/X)-Hy-X-Hy-X-Hy, where Hy is a large hydrophobic amino acid (tryptophan, leucine or phenylalanine), X is any amino acid, and W is tryptophan (Blond-Elguindi *et al.*, 1993). Peptides enriched in lysine and arginine were also selected for by DnaK (De Crouy-Chanel *et al.*, 1996). The bias in the recognition of amino acids by the Hsp70s could be important in such functions as mitochondrial import of proteins. Mitochondrial pre-sequences, which target proteins to the mitochondria, generally consist of positively charged and hydrophobic amino acids (Zhang *et al.*, 1999). The substrate specificity between different Hsp70s can vary. *E. coli* DnaK had a higher affinity for peptides enriched in leucine than another *E. coli* Hsp70, HscC; HscC by contrast was more likely to favour proline in its peptide substrates than DnaK (Kluck *et al.*, 2002). Subtle differences such as these may allow for very di stinct pools of substrates in prokaryotes and in eukaryotic organelles where multiple Hsp70s are present. The *E. coli* protein DnaK has been shown to bind exposed hydrophobic residues, regardless of whether they were in unfolded or native proteins (De Crouy-Chanel *et al.*, 1999). This may indicate a conserved mechanism for all Hsp70s, as some Hsp70 functions require an ability to bind folded proteins, e.g. during clathrin uncoating (Chappell *et al.*, 1987; Holstein *et al.*, 1996). The nature of the substrate proteins recognised by Hsp70s may be regulated by the partner DnaJ-like protein (Misselwitz *et al.*, 1998).

DnaK has also been shown to act as a secondary amide peptide bond *cis-trans* isomerase (Schiene-Fischer *et al.*, 2002). DnaK can catalyse the *cis-trans* isomerisation of non-prolyl peptide bonds. This ability was potentially critical in the substrate folding activity of DnaK, as it may allow for the recovery of a polypeptide that has an incorrect backbone fold (Schiene-Fischer *et al.*, 2002). Importantly, the presence of the co-chaperones DnaJ and GrpE enhanced this novel function of DnaK.

1.3.3 ATPase activity of Hsp70s

Hsp70s typically have a low ATPase activity (Russell *et al.*, 1999). This is comprised of two facets; the ATP hydrolysis activity (ATP to ADP and inorganic phosphate [Pi]), mediated by

the DnaJ-like family of proteins (Section 1.5) and the nucleotide exchange activity (ADP replaced by ATP), mediated in some instances by a separate nucleotide exchange factor (Section 1.3.4). The type of nucleotide bound to the ATPase domain determines the efficiency of the Hsp70s chaperone activity. When ATP is present, there is a low affinity for unfolded proteins by the substrate-binding region. By contrast, when ADP is present there is a high affinity for unfolded proteins. This higher affinity allows a longer period of time for stabilisation of the unfolded protein, and hence gives it a greater chance of folding correctly. Hence nucleotide exchange is important for ensuring that Hsp70s are not permanently in the high affinity ADP state. ATP has been shown to bind to DnaK with a high affinity of approximately 1 nM (Russell *et al.*, 1998). There was a far lower affinity for ADP by DnaK than for ATP (Russell *et al.*, 1998). This made the ATP hydrolysis step the rate limiting step in the ATPase cycle in the absence of exogenous Pi for DnaK as well as for the eukaryotic protein BiP (Pierpaoli *et al.*, 1997; Mayer *et al.*, 2003). However when inorganic phosphate was present in concentrations of between 5 and 10 mM (physiological concentrations), then the rate of ADP release was diminished, and this became partially rate limiting with respect to the ATPase cycle (Russell *et al.*, 1998). Therefore a fraction of the DnaK species in the cell would be expected to be bound to ADP.

The chaperone function of Hsp70s is dependent on the ATPase activity. Low ATPase activity implies that the Hsp70 will be predominately in the ATP bound form, and hence will have a relatively poor chaperone activity. Stimulated ATPase activity will therefore enhance the chaperone activity of the Hsp70. Close inter-domain communication between the substrate binding domain and ATPase domain is critical for effective chaperone activity and substitutions in the linker region between these two domains affected this activity (Han and Christen, 2001). The ATPase activity, and its effect on the substrate binding region, required several conformational shifts (Kamath-Loeb *et al.*, 1995), dependent on the presence of ATP, co-chaperone proteins and peptides. Binding of peptide could also stimulate the ATP hydrolysis activity of DnaK (McCarty *et al.*, 1995) and BiP (Flynn *et al.*, 1989; Flynn *et al.*, 1991; Mayer *et al.*, 2003), however binding of protein to BiP decreased the rate of ATP hydrolysis (Mayer *et al.*, 2003).

1.3.4 Nucleotide exchange factors in the DnaK/Hsp70 system

Two predominant groups of co-chaperone proteins have evolved to regulate the ATPase activity of partner Hsp70s, and hence play a role in enhancing their chaperone activity. The first group are members of the DnaJ-like family of proteins, and will be discussed further in Section 1.4.

The second predominant group of Hsp70 co-chaperones are the nucleotide exchange factors. The prokaryotic nucleotide exchange factor was GrpE; this protein was initially purified from *E. coli* due to its ability to form a stable interaction with DnaK which was only disrupted in the presence of ATP (Zylicz *et al.*, 1987). The structure of the DnaK-GrpE complex has been determined, and shows a 2:1 ratio of GrpE:DnaK (Harrison *et al.*, 1997). GrpE bound to monomeric DnaK, causing a conformational shift, and a subsequent decrease in the affinity of DnaK for ADP (Reid and Fink, 1996; Harrison *et al.*, 1997), which allowed for an increase in the rate of nucleotide release (Liberek *et al.*, 1991). The region that GrpE binds to on the ATPase domain appeared to be localised to IIB sub-domain (Sondermann *et al.*, 2001) (Figure 1.1A).

Previous work has proposed a classification of Hsp70 proteins into three classes depending on the presence or absence of critical salt bridges in the ATPase domain (Brehmer *et al.*, 2001). The three subfamilies are the Hsp70 family, the HscA family and the DnaK family. Of these three groups, only members of the DnaK family required GrpE to catalyse nucleotide exchange (Brehmer *et al.*, 2001). GrpE homologues have only been detected in mitochondria or chloroplasts (Naylor *et al.*, 1998; Choglay *et al.*, 2001; Kabani *et al.*, 2003). The Hsp70 family generally required the Bag-1 (Bcl2-associated athanogene 1) nucleotide exchange factor, which stimulates ADP, but not ATP dissociation from Hsc70 (Brehmer *et al.*, 2001). Bag-1 was the first eukaryotic nucleotide exchange protein identified that was not localised to mitochondria. It was shown to interact with the ATPase domain of Hsp70s (Takayama *et al.*, 1997). It appeared to inhibit the protein refolding activity of both Hsp70 and Hsc70, possibly by its stable association with an Hsp70-substrate complex (Takayama *et al.*, 1997). By contrast neither GrpE nor Bag-1 could associate with HscA proteins. Hence there is a clear distinction within the Hsp70 family with respect to the type of nucleotide exchange factor required in the ATPase cycle, if any.

Nucleotide exchange factors have also been recently identified in yeast. Sls1p (*S. cerevisiae* lethal mutations with the 7S RNA mutation) interacted with the yeast endoplasmic reticulum protein Kar2 (Kabani *et al.*, 2000). Homologues of this protein have been described in several higher eukaryotic organisms (Kabani *et al.*, 2003) such as *D. melanogaster*, *M. musculus* and humans. The Sls1p proteins in these organisms will bind to BiP proteins, the homologues of Kar2 in higher eukaryotes. An additional nucleotide exchange factor, termed BAP (BiP associated protein), with similarity to Sls1p was identified which interacts with mammalian BiP (Chung *et al.*, 2002). A cytosolic homologue of Sls1p termed Fes1p (factor exchange for Ssa1p) has also been identified (Kabani *et al.*, 2002). This protein, in contrast to other nucleotide exchange factors, appeared to inhibit the ATPase activity of the yeast cytosolic Hsp70, Ssa1p (Kabani *et al.*, 2002). The wide variety of nucleotide exchange factors, and other co-chaperones, may provide an additional level of regulating Hsp70 function (Kabani *et al.*, 2003). For example recent work has indicated that the nature of the co-chaperone interacting with Hsp70 may affect the nature of the specific functioning of Hsp70 (Höhfeld *et al.*, 2001).

1.4 DnaJ and DnaJ-like proteins

"A protein there was, and that a worthy one,
That from the time that it first began
To fold others, it helped 70
Fold and unfold, bind and degrade."[4]

DnaJ-like proteins are defined by the presence of the J domain, a 70 amino domain with similarity to the initial 73 amino acids of *E. coli* DnaJ. DnaJ-like proteins are involved in facilitating the folding of nascent polypeptides by regulating partner Hsp70 proteins (Cyr *et al.*, 1994), however DnaJ-like proteins have also been implicated in protein translocation (Feldheim *et al.*, 1992), protein degradation (Jubete *et al.*, 1996), clathrin uncoating (Jiang *et al.*, 1997; Ma *et al.*, 2002) and viral infection (Campbell *et al.*, 1997). These proteins have structural features that are conserved from the *E. coli* protein DnaJ (Ohki *et al.*, 1986).

1.4.1 Structural motifs in *E. coli* DnaJ and DnaJ-like proteins

E. coli DnaJ is comprised of four canonical domains (Cheetham and Caplan, 1998). Structural features include a J domain, a glycine-phenylalanine rich region, four cysteine repeats and an uncharacterised C terminal region (Cheetham and Caplan, 1998). These are schematically illustrated in Figure 1.2.

1.4.1.1 The J domain

The J domain is the specific feature that defines a protein as a DnaJ-like protein (Pellechia *et al.*, 1996). The presence of a J domain does not imply that the protein is a true homologue of DnaJ however. Many proteins contain regions that are not found in *E. coli* DnaJ (Cheetham and Caplan, 1998), and the presence of other domains may allow a specific DnaJ-like protein to fulfil a specific function. The J domain had similarity to the initial 73 amino acids of DnaJ, and was normally found at the amino terminus of proteins, although this was not always so (Cyr *et al.*, 1994). This region was believed to be the site of the interaction between DnaJ-

[4] With apologies to Geoffrey Chaucer. The original quote can be found in the Prologue to *The Canterbury Tales*.

like proteins and their partner Hsp70s. It also appeared to be the minimal region required for interaction between DnaJ-like proteins and partner Hsp70s (Corsi and Schekman, 1997).

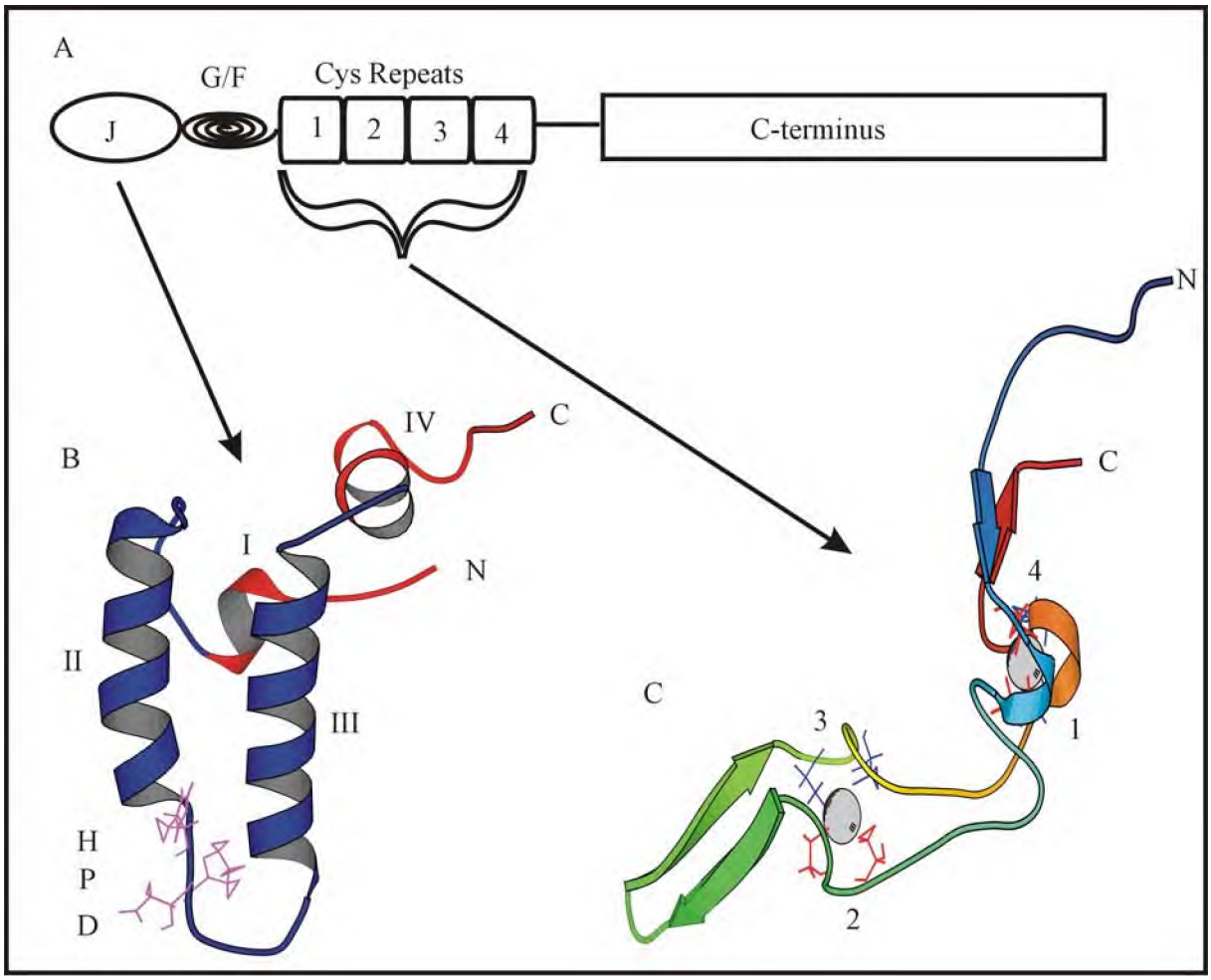

Figure 1.2: Structural motifs found in *E. coli* DnaJ

(A) A diagrammatic representation of the domains present in *E. coli* DnaJ. J represents the J domain, G/F stands for the glycine-phenylalanine rich region and cysteine repeats represents the four zinc finger like motifs. The first three domains correspond to approximately half DnaJ. (B) A ribbon representation of the J domain (1XBL) (Pellechia *et al.*, 1996). The conserved HPD motif is depicted in purple. The four helices are labelled. (C) A ribbon representation of the cysteine repeats (1EXK) (Martinez-Yamout *et al.*, 2000). Cysteine residues are depicted in red (repeats one and two) and blue (repeats three and four). Repeats 1 and 4 form zinc center 1, and repeats 2 and 3 form zinc center 2 (Linke *et al.*, 2003). The co-ordinated zinc atoms are shown in CPK format. This diagram is not to scale. The figures were generated using Molscript (Kraulis, 1991).

The structures of the J domain from six DnaJ-like proteins have been determined using nuclear magnetic resonance (NMR) and X-ray crystallography. These are from DnaJ in *E. coli* (Pellechia *et al.*, 1996; Huang *et al.*, 1998), human HDJ1 (Qian *et al.*, 1996), *E. coli* Hsc20 (Cupp-Vickery and Vickery, 1997; Cupp-Vickery and Vickery, 2000), the large T antigen from murine polyomavirus (Berjanskii *et al.*, 2000), the large T antigen from SV40 in

conjunction with the retinoblastoma tumour suppressor (Kim *et al.*, 2001), and bovine auxilin (Jiang *et al.*, 2003). These structures are depicted in Figure 1.3. The structures contained four α-helices, with a loop region containing a highly conserved tripeptide of histidine, proline and aspartic acid (HPD motif) located between helices II and III (Qian *et al.*, 1996). The HPD motif was present in all known J domains, with the exception of the ring infected erythrocyte surface antigen (RESA) proteins of *Plasmodium falciparum* (Bork *et al.*, 1991), and the recently identified DjlB/DjlC family of proteins described in *E. coli* (Kluck *et al.*, 2002). The minimal interacting region in the J domain was shown to be between amino acids 2 and 35, which included Helices I and II, and the HPD motif (Tsai and Douglas, 1996; Greene *et al.*, 1998). The auxilin J domain was unique in that it contains an extra loop region between Helices I and II, which was proposed to act as an extended interface for interaction with Hsc70 during clathrin uncoating (Jiang *et al.*, 2003). The J domain also appeared to be flexible in structure, with some workers proposing an induced-fit mechanism, with the HPD motif aiding in the alteration of the orientation of the charged residues in Helix II to allow them to interact correctly with the ATPase domain of a partner Hsp70 (Berjanskii *et al.*, 2002).

It appears that the J domain has been sequestered over time to help fulfil a number of functions. An example of such an adaption of function would be the presence of the J domain in the large and small T antigen proteins of viruses (Kelley and Landry, 1994; Brodsky and Pipas, 1998). This J domain was shown to replace functionally the J domain in *E. coli* using an *in vivo* complementation assay with an *E. coli dnaJ* mutant strain (Kelley and Georgopoulos, 1997b). At an eukaryotic level, a mammalian protein, ERdj5/JPD1 (Endoplasmic reticulum DnaJ protein 5 / J protein domain 1) contained a J domain, protein disulphide isomerase like domain, and a thioredoxin domain (Cunnea *et al.*, 2003; Hosoda *et al.*, 2003). As an ER located protein, it presumably required these domains to aid in the correct formation of disulphide bonds in the ER environment, both *via* its ability to promote the formation of disulphide bonds and its ability to sequester BiP to help in protein folding. This was of particular importance during times of cell stress, and this was presumably why ERdj5 has its expression up-regulated during periods of cellular stress (Cunnea *et al.*, 2003).

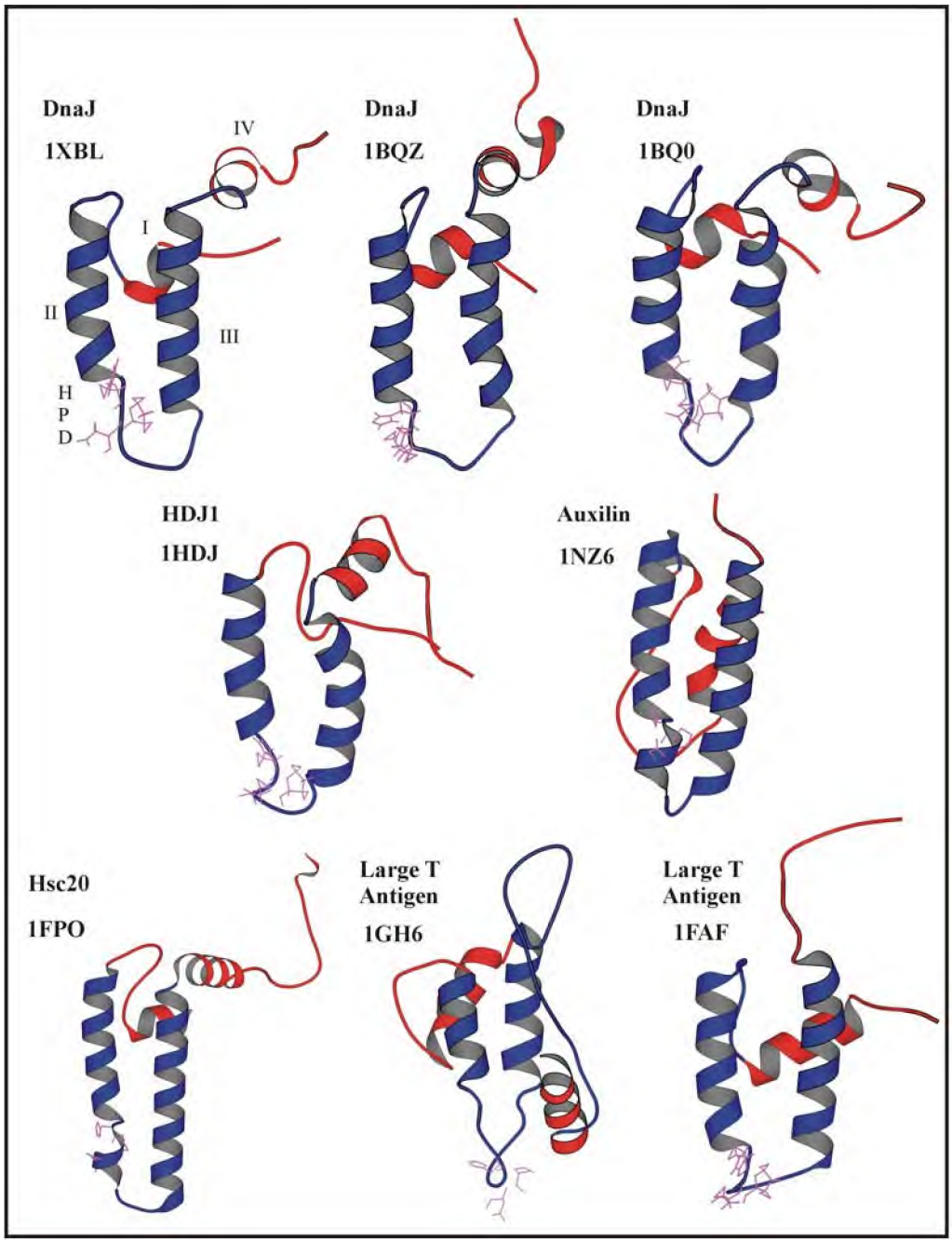

Figure 1.3: Ribbon representation of the structures of J domains from various DnaJ like proteins

The structures for *E. coli* DnaJ (1XBL, 1BQZ, 1BQ0) (Pellechia *et al.*, 1996; Huang *et al.*, 1998), human HDJ1 (1HDJ) (Qian *et al.*, 1996) and the murine polyomavirus T antigen (1FAF) (Berjanskii *et al.*, 2000) J domains were determined using NMR. The structure of *E. coli* Hsc20 (1FPO) (Cupp-Vickery and Vickery, 2000), SV40 T antigen (1GH6) (Kim *et al.*, 2001) and bovine auxilin (1NZ6) (Jiang *et al.*, 2003) were determined using X ray crystallography. The structures were visualised using Molscript (Kraulis, 1991). The HPD motif is depicted in purple. Helices II and III are in blue, with the more mobile helices I and IV in red. The helices and HPD motif are labelled on the DnaJ 1XBL J domain structure. Only the J domain region is shown even if additional structural regions were determined.

1.4.1.2 The Glycine-Phenylalanine rich region

The glycine-phenylalanine rich region (G/F region) was a non-structured region (Huang *et al.*, 1999) believed to be involved in aiding the stimulation of Hsp70s ATPase activity (Cheetham and Caplan, 1998). Some work has shown that the G/F region was essential in concert with the J domain for the stimulation of the ATPase activity. The J domain of DnaJ on its own was incapable of this stimulatory activity, but in the presence of the G/F region the J domain regained its ability to stimulate ATP hydrolysis activity (Wall *et al.*, 1995; Karzai and McMacken, 1996). The J domain of human Hsp40 could not stimulate the ATP hydrolysis activity of Hsp70 in the absence of the G/F region (Minami *et al.*, 1996). In contrast the J domain of the yeast protein Ydj1 without the G/F region was sufficient to stimulate the ATPase activity of Hsp70 (Lu and Cyr, 1998). However a certain element of specificity in the G/F region was shown when the G/F region of Ydj1 in conjunction with the J domain of Sis1 was not able to rescue the lethality of a *Δsis1 S. cerevisiae* strain, whereas the G/F region from Sis1, in conjunction with the J domain from Ydj1 was able to do so (Yan and Craig, 1999). A mutant version of *E. coli* DnaJ with the G/F region deleted was able to stimulate the ATPase activity of DnaK, but was unable to activate DnaK to bind to the substrate protein σ^{32} (Wall *et al.*, 1995). Equally some DnaJ like proteins that do not contain the G/F region have been shown to stimulate the ATPase activity of specific Hsp70s, in particular the cysteine string proteins (csp) (Braun *et al.*, 1996; Chamberlain and Burgoyne, 1997a) and auxilin (Ungewickell *et al.*, 1997). It was likely therefore that the G/F region would mimic a protein substrate, thereby allowing maximal stimulation of DnaK's ATPase activity. However, recent work from the Landry laboratory has indicated that the lack of a detectable interaction when only the J domain was present may be due to the nature of the buffer system used previously by the McMacken laboratory (Wittung-Stafshede *et al.*, 2003). Usage of a high ionic strength buffer may affect the large network of electrostatic interactions that are key in the interaction between Hsp70s and partner DnaJ-like proteins. The region was highly solvent accessible (Greene *et al.*, 2000), and it has been suggested that its highly flexible nature allows for independent folding of the J domain and the zinc finger like region.

1.4.1.3 The zinc finger like repeats

The third domain in DnaJ consists of four repeats of the form Cys-X-X-Cys-X-Gly-X-Gly, where X is any amino acid. The repeats are similar in amino acid sequence to zinc finger DNA binding regions (Berg, 1990; Laity *et al.*, 2001). The structure has been determined using NMR (Martinez-Yamout *et al.*, 2000) (Figure 1.2C), and shows the formation of two zinc centers. The first center was formed from repeats one and four, and the second was formed from repeats two and three (Martinez-Yamout *et al.*, 2000; Linke *et al.*, 2003). DnaJ has been shown to contain two zinc molecules per monomer (Szabo *et al.*, 1996; Banecki *et al.*, 1996). These zinc ions stabilized the molecule (Banecki *et al.*, 1996). Loss of the zinc did not appear to have too marked an effect on the structure, potentially due to the formation of disulphide bonds between the cysteine residues, stabilising this region (Banecki *et al.*, 1996). This region appeared to play a minimal role in DnaK interaction (Banecki *et al.*, 1996), and can not stimulate the ATPase activity of DnaK (Lu and Cyr, 1998). Previous work has suggested that these repeats are involved in peptide binding, possibly aiding DnaJs chaperone function (Szabo *et al.*, 1996). The two zinc binding groups appeared to have differing roles in DnaJ function (Linke *et al.*, 2003). Loss of zinc center 1 appeared to affect negatively the chaperone activity of DnaJ, although loss of this activity did not impact on the interaction of the mutant protein with DnaK (Linke *et al.*, 2003). Zinc center 2, in contrast, was important in binding to DnaK, and loss of this region prevented DnaJ from transforming DnaK into a high substrate affinity state (Linke *et al.*, 2003). Several eukaryotic DnaJ-like proteins contain these repeats, however some of these have a lysine residue substituted for the last glycine. Examples are the yeast proteins Scj1 (Blumberg and Silver, 1991) and Ydj1 (Caplan and Douglas, 1991), and the porcine protein pDJA1 (Depre *et al.*, 2003). Other substitutions in other repeats have also been observed.

1.4.1.4 C terminus

The rest of the DnaJ molecule, approximately half, is relatively unconserved. There are indications that the chaperone activity of *E. coli* DnaJ is localized to this region (Section 1.4.2.1), although other DnaJ-like proteins lacking this domain can exhibit limited chaperone activity. There are conserved amino acids in this section, including a glycine rich region and a conserved gly-asp-leu-tyr-val pentapeptide (Goffin and Georgopoulos, 1998). The C

terminal region is the most variable part of DnaJ-like proteins. The structure of the C terminal part of Sis1, a yeast DnaJ-like protein that does not contain the cysteine repeats found in *E. coli* DnaJ, has been determined (Sha *et al.*, 2000). Similarly to *E. coli* DnaJ (Zylicz *et al.*, 1985; Bardwell *et al.*, 1986) Sis1 formed a dimer, and contained a region proposed to act as a peptide-binding domain. The structure of the C terminus of the yeast protein Ydj1, in conjunction with a peptide substrate, has recently been released (Li *et al.*, 2003). This structure included the cysteine repeats, but not the J domain. The structure of this protein was similar to that of the C terminus of Sis1 (Sha *et al.*, 2000), which possibly explains the overlapping functions of these two proteins in *S. cerevisiae* (Johnson and Craig, 2001).

1.4.2 Characteristics of *E. coli* DnaJ and DnaJ-like proteins

1.4.2.1 Chaperone activity

DnaJ appeared to be able to function as a molecular chaperone in its own right, however its co-chaperone activity was independent of its chaperone activity (Han and Christen, 2003). The chaperone activity generally presents itself as an ability to stabilise unfolded proteins as opposed to actively facilitating the refolding process. The zinc finger domain can bind to denatured proteins (Szabo *et al.*, 1996). Mutations in this region affected DnaJ's ability to refold luciferase, but DnaJ could still stabilize unfolded proteins (Lu and Cyr, 1998). DnaJ has also been shown to prevent aggregation (Langer *et al.*, 1992). A substitution in a conserved glycine prevented DnaJ from suppressing aggregation, and forming complexes with unfolded luciferase (Lu and Cyr, 1998). Deletion of conserved residues in the C-terminus also prevented DnaJ from suppressing aggregation and forming complexes with unfolded luciferase (Goffin and Georgopoulos, 1998). Deletions of either of two conserved regions caused mutant DnaJ proteins that were defective in preventing aggregation of firefly luciferase (Goffin and Georgopoulos, 1998). These mutant proteins were also partially defective in the stimulation of the DnaK ATPase activity (Goffin and Georgopoulos, 1998). A single amino acid substitution of a conserved glycine residue in the yeast protein Ydj1, Gly^{315} to Asp, prevented Ydj1 from assisting in luciferase refolding, and rendered Ydj1 incapable of preventing rhodanese aggregation (Lu and Cyr, 1998). Similarly the deletion of

C terminal residues in another yeast protein, Mdj1, also affected that proteins ability to prevent aggregation of luciferase (Westermann *et al.*, 1996).

DnaJ has also been shown to act as a protein disulphide isomerase (PDI) (De Crouy-Chanel *et al.*, 1995). PDI affected the formation of disulphide bonds, and was regarded as a protein folding catalyst, rather than a chaperone, though it did appear to have some chaperone activity. It may be as a result of this function that DnaJ has been shown to be essential above 43°C (Sell *et al.*, 1990).

DnaJ was capable of binding peptides containing D-amino acids, as well as peptides composed of the biologically significant L-amino acids (Feifel *et al.*, 1998), in contrast to DnaK which will only bind peptides composed of L-amino acids. DnaJ also recognised peptides predominately comprising of large hydrophobic residues and aromatic residues (Rüdiger *et al.*, 2001). This was similar to the residues that DnaK recognised (Rüdiger *et al.*, 1997). However, DnaK favoured the binding of peptides enriched in leucine residues, whereas DnaJ favoured the binding of peptides enriched in aromatic residues (Rüdiger *et al.*, 2001). The chaperone activity of DnaJ, and other DnaJ-like proteins, has been localised to its C terminal region, including the zinc finger region (Szabo *et al.*, 1996; Goffin and Georgopoulos, 1998). DnaJ can bind to the N terminal region of rhodanese in a 1:1 ratio (Kudlicki *et al.*, 1996). Nucleotides do not affect the binding of substrate proteins to *E. coli* DnaJ (Wawrzynów and Zylicz, 1995).

Other DnaJ-like proteins have also been shown to be able to act as molecular chaperones. The human proteins HDJ-1 and HDJ-2 could suppress the aggregation of the polyglutamine protein ataxin-3 (Chai *et al.*, 1999). The presence of the J domain appeared to affect this activity as truncation HDJ-2 mutants lacking the J domain were more effective at the suppression of aggregation than the full length protein (Chai *et al.*, 1999). The yeast mitochondrial protein Mdj1p could prevent the heat inducible aggregation of luciferase *in vitro*, and was also found in a chaperone complex with mitochondrial Hsp70 and nascent polypeptides *in vivo* (Westermann *et al.*, 1996). In general though, DnaJ-like proteins appear to act as holders of unfolded polypeptides; in other words they stabilise unfolded conformations, as opposed to refolding these actively.

Work in yeast with the essential Sis1 and non-essential Ydj1 showed that their disparate C termini shared similar roles in the cell, and that at least one of the C termini must be present for *S. cerevisiae* viability (Johnson and Craig, 2001). The Sis1 C terminus did not contain the cysteine repeats which were present in Ydj1 and DnaJ, and the percentage identity between the Sis1 and Ydj1 C termini was far lower than between their J domains. Hence the substrate binding activity between these two proteins was likely to be different, but it appeared that the substrate specificity would overlap (Johnson and Craig, 2001). The C terminal structures of these two proteins appeared to be similar (Sha *et al.*, 2000; Li *et al.*, 2003), allowing for overlapping functioning.

1.4.2.2 Membrane association

DnaJ has been shown to be membrane associated (Zylicz *et al.*, 1985), however no membrane binding regions have been shown. Many true homologues of DnaJ contained a C terminal motif specifying the addition of a farnesyl group to the protein. Examples of this were the yeast protein Ydj1 (yeast DnaJ) (Caplan *et al.*, 1992a; Caplan *et al.*, 1992b), the porcine protein pDJA1 (pig DnaJ-like protein A1) (Depre *et al.*, 2003) and the plant proteins LDJ1 from *Allium porrum* (Leek DnaJ2) (Bessoule *et al.*, 1994) and ANJ1 from *Atriplex numularia* (*Atriplex numularia* DnaJ1) (Zhu *et al.*, 1993). The farnesyl addition was to a cysteine residue that occured as part of a C-terminal CaaX motif where X represented any amino acid, a represented aliphatic amino acids, and the motif was the last four amino acids of the protein (Clarke, 1992). The farnesylation reaction involved the cleavage of the last three amino acids, and the addition of the 15 carbon farnesyl moiety to form a C-terminal S-farnesyl cysteine methyl ester (Clarke, 1992). This allowed these proteins to be membrane associated. There was some evidence that the prenyl group was important for protein-protein interaction and signal transduction (Marshall, 1993; Zhang and Casey, 1996). Mutating the cysteine residue to a serine thereby preventing farnesylation of Ydj1p caused a temperature sensitive phenotype in yeast (Caplan *et al.*, 1992b). The majority of DnaJ-like proteins with a farnesylation motif received a farnesyl moiety, however the human protein Hsj1b has recently been shown to be the first member of a group of DnaJ-like proteins that receive a geranylgeranyl moiety (Chapple and Cheetham, 2003). Hsj1b does not contain the cysteine repeats (Cheetham *et al.*, 1992), and this in turn may influence the nature of its membrane interaction.

1.4.3 Subdivision and nomenclature of DnaJ-like proteins

DnaJ-like proteins are defined by the presence of the J domain. Due to the nature of this definition of DnaJ-like proteins, there have been several attempts to subdivide them to make analysis easier. The first categorisation proposed the division of the family into two groups, depending on the presence or absence of the Gly/Phe region (Kelley, 1998). The second categorisation involved the division of the family into three groups depending on the presence or absence of the Gly/Phe region and the cysteine repeats (Cheetham and Caplan, 1998). Type I DnaJ-like proteins contain similarity to DnaJ over all the domains present, namely the J domain, the Gly/Phe region and the cysteine repeats. Type II proteins have a J domain and the Gly/Phe rich region. Type III proteins only have a J domain in common with DnaJ (Figure 1.4). The Cheetham and Caplan classification will be employed in this thesis.

There has also been a recent proposal for a revamp of the nomenclature of DnaJ like proteins (Ohtsuka and Hata, 2000), based on the Cheetham and Caplan mechanism of division. This system proposed the naming of mammalian DnaJ-like proteins based on the organism the protein was isolated from and the class they belong to, and allows for chronological order of identification.

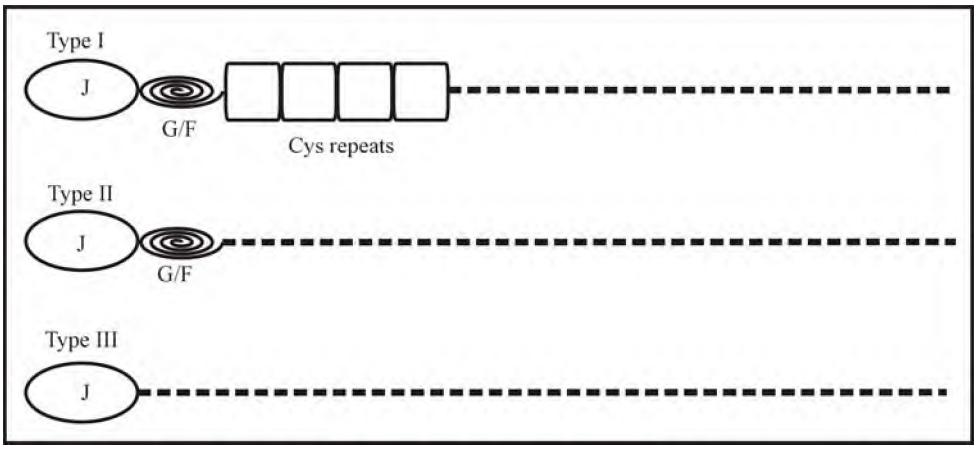

Figure 1.4: Schematic illustrating the canonical domains contained in Type I, II and III DnaJ-like proteins (Cheetham and Caplan, 1998)

J represents the J domain, G/F stands for the glycine-phenylalanine rich region and Cys repeats represents the four zinc finger like motifs. The dotted line represents the remainder of the protein, which may, or may not, bear similarity to the C terminus of *E. coli* DnaJ.

1.5 INTERACTION OF DNAJ-LIKE PROTEINS WITH HSP70S

> *"It is a truth universally acknowledged, that a single co-chaperone in possession of a substrate, must be in want of a chaperone."*[5]

DnaK, the *E. coli* Hsp70, interacted with DnaJ in an ATP dependent fashion in the presence of another co-chaperone, GrpE (Szabo *et al.*, 1994). Hsp70s in general have a low basal ATPase activity, and stimulation of this activity was performed by DnaJ and DnaJ-like proteins (Russell *et al.*, 1999). DnaJ stimulated ATP hydrolysis by DnaK, thereby stimulating the DnaK ATPase activity by approximately 2 fold, and up to 50 fold in the presence of GrpE (Liberek *et al.*, 1991). Maximal enhancement of the ATPase activity of DnaK occurred in the presence of DnaJ and a substrate peptide (Laufen *et al.*, 1999). Stable formation of a DnaK-DnaJ complex was dependent upon ATP hydrolysis (Suh *et al.*, 1999). A similar process occurred in eukaryotes in the presence of Hsp70, DnaJ-like proteins, and other co-chaperones such as Hip (Hsc70 interacting protein) and Hop (Hsp70-Hsp90 organising protein) (Schumacher *et al.*, 1996). GrpE did not appear to be present in the cytoplasm of eukaryotes (Schumacher *et al.*, 1996), although it has been shown to be present in mitochondria (Bolliger *et al.*, 1994). Other nucleotide exchange factors have been identified that appear to have very specific functions in affecting the function of Hsp70s (Kabani *et al.*, 2003) (Section 1.3.4). A theoretical model of the interaction of DnaJ-like proteins and Hsp70s is shown in Figure 1.5.

In prokaryotes, the protein folding cycle starts when DnaJ presents a substrate to DnaK (Szabo *et al.*, 1994; McCarty *et al.*, 1995; Laufen *et al.*, 1999) or binds to an already existing DnaK-substrate complex (Pierpaoli *et al.*, 1997; Gisler *et al.*, 1998; Nagata *et al.*, 1998; Pierpaoli *et al.*, 1998a). It is likely that the nature of the amino acid sequence of the substrate peptide dictated which mechanism was employed for entry of the substrate into the cycle (Pierpaoli *et al.*, 1998a), and that different partner DnaJ-like proteins would deliver different substrates to an Hsp70 (Misselwitz *et al.*, 1998).

[5] With apologies to Jane Austen. The original quote can be found in *Pride and Prejudice.*

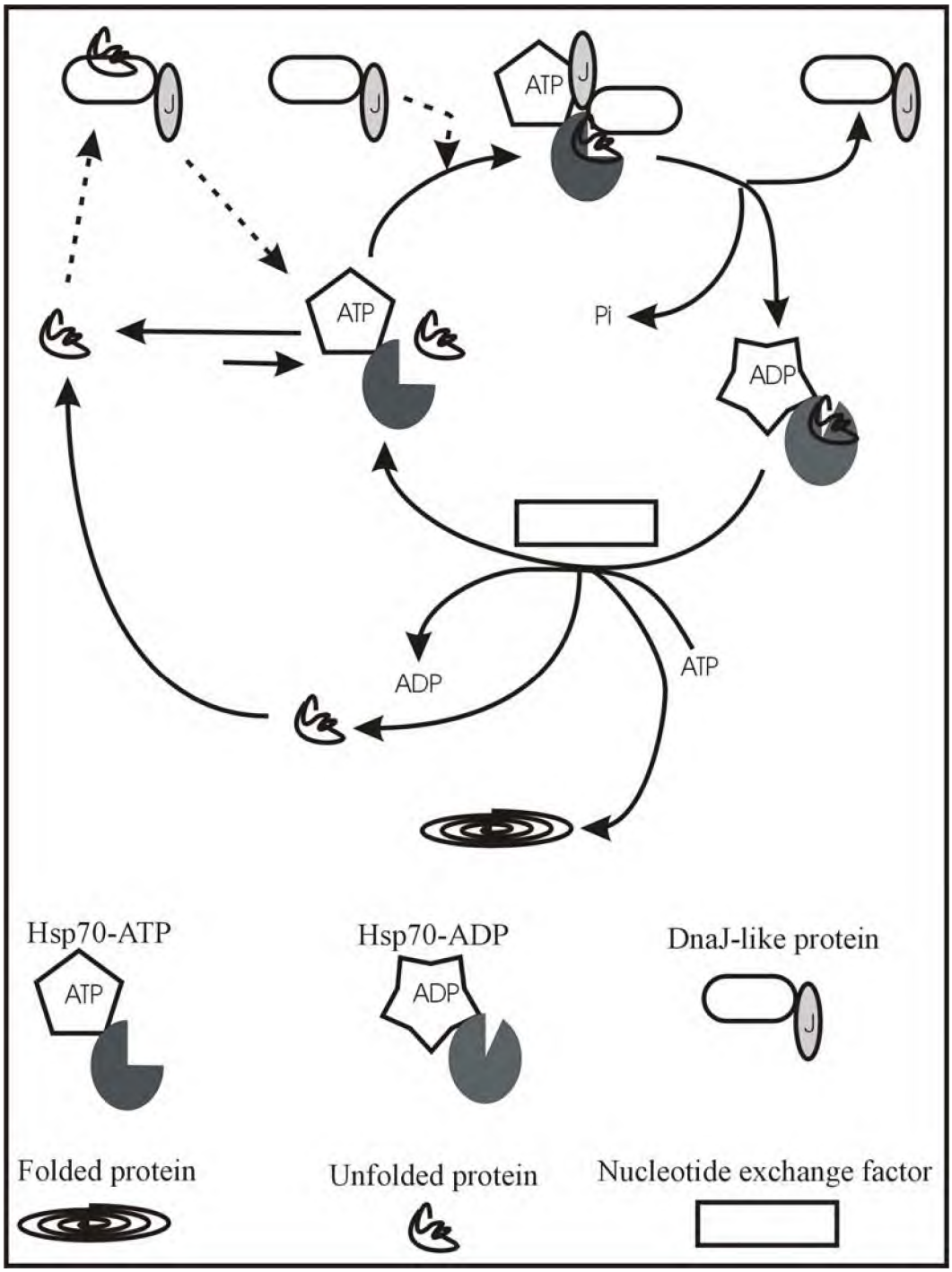

Figure 1.5: Schematic showing the protein folding cycle involving the interaction of DnaJ-like proteins with partner Hsp70s

Pi is inorganic phosphate. Dotted lines indicate the two different paths that substrate proteins and DnaJ-like proteins can enter the cycle. Substrate proteins are either recognised by DnaK, following which DnaJ enters the cycle, or are presented to DnaK by DnaJ. ATP hydrolysis, stimulated by DnaJ, causes a conformational shift in the substrate binding domain, locking the substrate in. Nucleotide exchange then reverses the conformational shift, allowing for the release, and subsequent folding, of the substrate. Alternatively the substrate can re-enter the cycle.

The binding of DnaJ to DnaK has been shown to allow DnaK to attain a high substrate-affinity conformational state after the hydrolysis of ATP (Szabo *et al.*, 1994), and in the process created a [$(DnaK)_n$-substrate$(DnaJ)_m$] ternary complex (Han and Christen, 2003). Upon hydrolysis of ATP, a process stimulated by DnaJ, a conformational shift occurred in DnaK, locking the peptide into the substrate binding cleft (Moro *et al.*, 2003). GrpE then caused the dissociation of ADP from the DnaK ATPase domain, and replacement with ATP. This caused the release of the polypeptide to fold correctly *via* a folding intermediate form or to rebind to DnaK and re-start the cycle.

In eukaryotes the situation was slightly more complicated. The DnaJ-like protein stimulated the ATPase activity of Hsp70, causing the formation of a peptide bound Hsp70 as described above. The Hsp70-substrate conformation could then be further stabilised by Hip. Hip was shown to be involved in the stabilisation of the ADP bound form of Hsp70 in eukaryotic systems (Höhfeld, 1998). This allowed for further time for the unfolded polypeptide to attain a native or native-like conformation.

Once released, the folding intermediate could be passed onto other chaperone systems such as the Hsp60, or fold correctly on its own. There appeared to be a distinct level of interaction between the chaperone families that was tightly co-ordinated to ensure effective protein folding (Frydman *et al.*, 1994). Hsp70 in association with Hsp40 bound to the initial shorter fragments of synthesising polypeptides, whereas the Hsp60 family of proteins bound to larger fragments (at least 100 to 150 residues) (Frydman *et al.*, 1994). An example of a substrate moving between different chaperone systems was shown by using ferredoxin-NADP+ reductase as a model system for investigating the import of proteins into chloroplasts (Tsugeki and Nishimura, 1993). Hsp70 interacted with the ferredoxin-NADP+ reductase before Hsp60. Additional work involved investigating the transport of firefly luciferase and Mas2 into mitochondria (reviewed in Frydman, 2001). The varying types of chaperones in the cell may have developed as a response to different folding mechanisms. Hsp70s in general would prevent protein aggregation by binding to unfolded sections of a polypeptide, whereas chaperones such as Hsp60 functioned by recognising more compact structures, and promote final folding (Frydman, 2001). The differences between these two mechanisms may allow for a tight regulation of the nature of folding of specific newly synthesised proteins.

In general, the ATPase cycle of Hsp70 is stoichiometrically coupled to the cycle of substrate binding and release. The presence of DnaJ-like proteins coupled the two processes, allowing for a more effective chaperone activity (Pierpaoli *et al.*, 1997; Davis *et al.*, 1999). The enhancement of the ATPase activity could be greater than 5 times the initial rate (Liberek *et al.*, 1991), but could be as much as 1000 fold, but only at high DnaJ concentrations (Laufen *et al.*, 1999), with the variation being due to the nature of the ATPase activity being determined (steady state versus single turnover). However high stimulation of the ATPase activity did not lead to enhanced chaperone activity, as maximal enhancement of luciferase refolding by DnaK occurred at substoichiometric levels of DnaJ (Laufen *et al.*, 1999). The complete cycle involving the interaction between DnaK, DnaJ and GrpE in prokaryotes has been predicted to have a half life of approximately 17 seconds at 25°C (Pierpaoli *et al.*, 1997). The affinity of DnaJ for the ATP bound DnaK was low – a KD of approximately 20 μM was determined (Russell *et al.*, 1999). This low affinity was cellularly relevant as it will prevent non-productive hydrolysis of ATP in the absence of polypeptide substrates (Russell *et al.*, 1999).

1.5.1 Binding sites and binding determinants

1.5.1.1 Binding sites and determinants on Hsp70s

The interaction between DnaJ-like proteins and Hsp70s was complex. This was due to the binding of DnaJ-like proteins to multiple sites on Hsp70, *via* multiple sites on the DnaJ, involving at least a two signal mechanism (Mayer *et al.*, 1999). This has been investigated in several recent papers looking at the interaction between DnaK and DnaJ using surface plasmon resonance (SPR) spectroscopy (Suh *et al.*, 1998; Gässler *et al.*, 1998; Mayer *et al.*, 1999). Gässler *et al.* and Suh *et al.* provided evidence that the J domain of DnaJ bound to an underside cleft on the ATPase domain (Figure 1.6). This region contained residues that were structurally equivalent to residues found in actin that were important in actin-actin binding (Gässler *et al.*, 1998). Genetic studies isolated Hsp70s containing mutations that suppress the effect of mutations in the critical HPD motif. For example the single amino acid changes R167H and I169F in the underside cleft suppressed the effect of DnaJ-D35N, a mutant DnaJ that contained a mutation in the aspartic acid of the conserved HPD motif (Suh *et al.*, 1998). The DnaK-R167H and R167A mutants bind more effectively to DnaJ-D35N than to wild type DnaJ (Suh *et al.*, 1998), indicating that Arg^{167}, a highly conserved ATPase domain residue,

was important in the interaction between DnaJ and DnaK. It was further noted that one side of the ATPase channel was more critical than the other, although this could be due to the triple mutation that was performed, in which three residues were all converted to alanines (Gässler *et al.*, 1998). This could cause a large structural change, which might explain the severity of the effect of the substitutions.

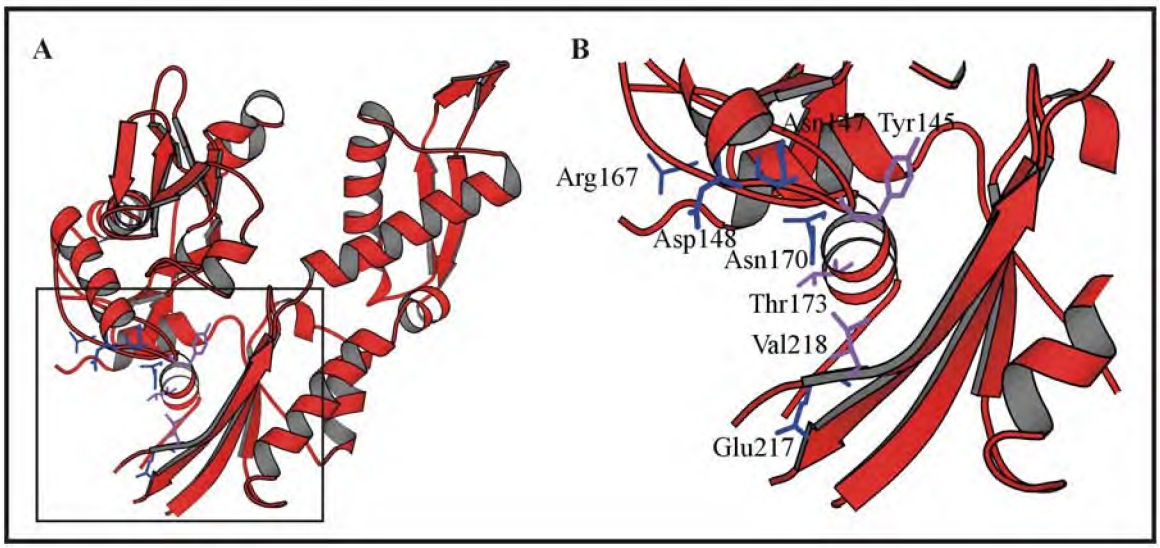

Figure 1.6: Amino acids on the ATPase domain of DnaK proposed to be involved in the interaction with DnaJ

Charged and polar residues are depicted in blue; hydrophobic residues are in purple. (A) The ATPase domain of DnaK, with the conserved underside cleft boxed. (B) shows the region boxed in (A) in detail. The structure used is the 1DKG structure of the DnaK ATPase domain crystallised with GrpE (Harrison *et al.*, 199[7]). GrpE is not shown in this figure. Visualisation of the structures was performed using Molscript (Kraulis, 1991).

In vitro analysis of the carboxy terminal region of rat Hsc70, including the α-helical lid region, and remainder of the C terminus, indicated that it also possessed a binding site for Hsp40 (Demand *et al.*, 1998). This was potentially *via* the C terminal EEVD motif (Freeman *et al.*, 1995). However a truncated DnaK mutant lacking the last 100 residues at the C terminus was still capable of having its ATPase activity stimulated by DnaJ, and of binding to DnaJ albeit at a slightly reduced level than when compared to wild type DnaK (Gässler *et al.*, 1998; Suh *et al.*, 1999). The ATPase core of DnaK (residues 1 to 385) was unable to bind to DnaJ as measured by SPR, nor was DnaJ able to stimulate the ATPase activity of this portion of DnaK (Gässler *et al.*, 1998). Blocking of the substrate binding domain of DnaK also abrogated ATPase stimulation by DnaJ. When Val^{436} in the substrate binding region was mutated to phenylalanine, a mutation which would impede the binding of peptide to the

substrate binding domain, DnaJ was unable to stimulate the ATPase activity as efficiently (Laufen *et al.*, 1999). This work implied that binding of a substrate was important for stimulation, but could also indicate that the V436F substitution was impacting negatively on a DnaJ binding site in that domain.

1.5.1.2 Binding sites on DnaJ-like proteins

The amino acids on DnaJ that are involved in the binding to DnaK are less well defined. The role of the HPD has been well characterized, however other residues are gradually being implicated (Genevaux *et al.*, 2002; Genevaux *et al.*, 2003). DnaJ-like proteins interact with Hsp70s through the J domain and in particular the HPD motif. Mutations of the HPD residues abolished the stimulation of the Hsp70 ATPase activity (Tsai and Douglas, 1996). Mutated residues on the underside of the ATPase domain of DnaK have been shown to suppress the effect of the histidine to glutamine (H34Q) in DnaJ (Suh *et al.*, 1998; Gässler *et al.*, 1998).

The minimal region of interaction has been proposed to be Helix II and the HPD motif. Experiments have been performed examining the ability of peptides corresponding to various portions of the Ydj1 J domain to out-compete full length Ydj1 or the Ydj1 J domain in terms of interaction with Hsp70 (Lu and Cyr, 1998). Helix II and the HPD motif were almost as effective as the full length J domain in blocking rhodanese aggregation by interfering with the Hsp70-full length Ydj1 interaction. However Helix III and the HPD motif were also shown to be almost as capable of affecting the interaction, but a competing peptide of Helix I was not (Lu and Cyr, 1998). Neither Helix II nor Helix III were as effective as the full length J domain. The mutation of His^{34} in the HPD motif also affected Ydj1 functioning. It therefore appeared that the minimal region of interaction was located within residues 2 to 35 of DnaJ, which spans Helix II and the loop region (Greene *et al.*, 1998). Other parts of the J domain were still required, as fragments containing only this region did not bind as effectively to Hsp70 as the full J domain.

The QKRAA motif present in Helix IV of the J domain in DnaJ has also been shown to be recognised by DnaK (Auger and Roudier, 1997), and the lysine and arginine residues present in that motif are well conserved across all J domains (Hennessy *et al.*, 2000) (Section 2.3.1).

Mutations in this region in the DnaJ J domain caused partial abrogation of the interaction of the J domain with DnaK (Suh *et al.*, 1999).

It was likely that the binding of DnaJ to DnaK was bipartite in nature, in other words DnaJ was likely to bind to two distinct binding sites on DnaK with differing affinities (Suh *et al.*, 1998; Mayer *et al.*, 1999). The nature of this interaction was likely to be important in the targeting of substrate proteins to DnaK by DnaJ. The coupling of peptide binding to ATP hydrolysis was mediated by the J domain (Wittung-Stafshede *et al.*, 2003). The presence of a peptide substrate was required for maximal stimulation of the ATP hydrolysis activity (Jordan and McMacken, 1995), although this could have been an artefact due to an electrostatic affect. As stated previously, mutations in the HPD motif abolished the stimulatory effect of DnaJ-like proteins on the ATPase activity. This was likely due to the inability of the mutated J domain to stimulate the conformational change by the Hsp70, hence affecting the locking in of substrate proteins into the substrate binding cleft (Wittung-Stafshede *et al.*, 2003).

1.5.2 Targeting of the substrate to partner Hsp70s

Maximal stimulation of the levels of ATP hydrolysis of DnaK by DnaJ required a full length DnaJ, as well as a substrate protein (Laufen *et al.*, 1999). The protein substrate and DnaJ act in combination to stimulate DnaK effectively (Laufen *et al.*, 1999). Generally there was rapid dissociation and association of unfolded or partially folded polypeptide with the Hsp70 (Höhfeld, 1998). The interaction between the Hsp70 and the polypeptide was stabilised by the presence of ATP. How DnaJ and DnaK interact with each other and with a substrate is only now starting to be elucidated thoroughly. DnaJ has been shown to have the ability to bind unfolded proteins (Langer *et al.*, 1992) and this could provide a paradigm for the entry of substrate proteins into the DnaK/Hsp70 chaperone cycle. There were two potential mechanisms for the entry of unfolded proteins into the Hsp70 cycle (Suh *et al.*, 1999). The first and most likely was that the substrate protein would bind to Hsp70, at which point the DnaJ-like protein would enter the cycle causing stimulation of the ATPase activity, hence enhancing the chaperone activity. The second possibility was that the DnaJ-like partner protein itself would bind the unfolded protein and then present it to Hsp70, and the cycle would continue (Fink, 1999). DnaJ acted initially by scanning the polypeptide sequence and had the ability to recognise hydrophobic patches on a polypeptide sequence (Rüdiger *et al.*,

2001). DnaJ was shown to have low peptide binding affinity, allowing for ease of transfer of peptide substrate from DnaJ to DnaK (Rüdiger *et al.*, 2001). DnaJ and DnaK bind to different parts of the polypeptide chain (Han and Christen, 2003), and the transient nature of the DnaK-DnaJ interaction meant that one DnaJ molecule could stimulate the ATPase activity of several DnaK molecules (Misselwitz *et al.*, 1998). There were also indications that the dual binding of DnaK and DnaJ to the same substrate was important for efficient chaperone activity. A polypeptide made of D-amino acids inhibited the refolding of luciferase by the DnaK/DnaJ/GrpE system (Bischofberger *et al.*, 2003). D-amino acid polypeptides bound to the DnaJ substrate binding region and prevented the binding of other substrate proteins (Bischofberger *et al.*, 2003).

Interestingly DnaK mutants that contain substitutions in the substrate-binding region generally tended to have a far lower level of ATPase stimulation than wild type DnaK. This implied that either DnaJ or GrpE must interact with DnaK in this region (Montgomery *et al.*, 1999). There have been indications that it was DnaJ as opposed to GrpE that binds to this region (Mayer *et al.*, 1999). It was possible that the disordered Gly/Phe region in DnaJ could be involved in the binding to the substrate-binding region, and was displaced during substrate binding. DnaJ and DnaK have different peptide binding specificities, and DnaJ-specific substrates would not necessarily be recognised by DnaK, and DnaJ-specific substrates would not be passed onto DnaK (Han and Christen, 2003). However, DnaK and DnaJ would compete for substrates when there was joint substrate specificity (Han and Christen, 2003).

The variation in types of DnaJ-like proteins could affect the nature of substrates that could be bound by these proteins, and hence in the nature of the substrates that are passed onto specific Hsp70s. Examples were the roles of auxilin in clathrin uncoating (Holstein *et al.*, 1996; Ungewickell *et al.*, 1997; Jiang *et al.*, 1997) and the interaction of the iron/sulphur proteins with their specific chaperones Hsc66 and Hsc20 (Hoff *et al.*, 2000; Landry, 2003).

1.5.3 Specificity of Interaction

It is known that there is not complete inter-changeability between DnaJ-like proteins with respect to their interaction with distinct and separate Hsp70s. For example *in vitro* assays investigating the stimulation of the ATPase activity of Hsp70s by DnaJ-like proteins have

also revealed significant differences in the level of the stimulation attained. MmDjC7, a newly identified murine DnaJ-like protein, stimulated the ATPase activity of DnaK, human Hsc70 and murine BiP to different levels (Kroczynska and Blond, 2001). MmDjC7 stimulated the mammalian Hsc70 and BiP by a factor of three fold, but stimulated the prokaryotic DnaK by a factor of 9 fold. As the J domain in MmDjC7 was approximately half of the entire protein, and there was limited similarity to other DnaJ-like proteins in the remainder of the protein, the regulation of stimulation very likely occurred within the J domain, and must have varied for each of the Hsp70s used in this experiment. Equally *E. coli* DnaJ was capable of stimulating the ATPase activity of Hsp70, whereas Hsp40/Hdj2, the normal partner of Hsp70, was incapable of stimulating the ATPase activity of DnaK (Minami *et al.*, 1996). Hence, there must be a mechanism within the J domain that mediated specificity both of binding between Hsp70s and partner DnaJ-like proteins, and ensured that a productive interaction was the result.

1.5.3.1 J Domain swapping experiments

Substitution of a J domain from one protein by a J domain from another has been a fruitful area in terms of establishing the elements of specificity of interaction. Results from various experiments are shown below in Table 1.1. One of the first experiments conducted showed that the J domain from the yeast mitochondrial protein Mdj1 (Type I) could effectively substitute for the J domain of *E. coli* DnaJ (Deloche *et al.*, 1997). Equally the J domain from the Type III *E. coli* protein DjlA could effectively substitute for DnaJ's J domain (Genevaux *et al.*, 2001b). However, both DnaJ and DjlA interacted with the same Hsp70, namely DnaK. The J domain from another Type III *E. coli* DnaJ-like protein DjlC was not able to replace the J domain of *E. coli* DnaJ (Kluck *et al.*, 2002)

The situation became more complicated when looking at eukaryotic proteins. An analysis of yeast DnaJ-like proteins showed that neither the Sis1 J domain nor the Mdj1 J domain could functionally substitute for the J domain from Sec63, whereas the J domain from Scj1 could (Schlenstedt *et al.*, 1995). Sis1 was a cytosolic Type II DnaJ-like protein and Mdj1 was a mitochondrial Type I DnaJ-like protein, whereas Sec63 and Scj1 are both endoplasmic reticulum located proteins.

Table 1.1: Summary of J domain swapping experiments

J domain source		C terminus source		Function	Method	Ref [#]
Protein [^]	Type	Protein	Type			
Sis1 Scj1 Mdj1	II I I	Sec63	III	No Yes No	Complementation assays in yeast cells not producing Sec63	1
Mdj1	I	DnaJ	I	Yes	Complementation for lack of *E. coli* DnaJ and CbpA	2
DjlA	III	DnaJ	I	Yes	Complementation for lack of *E. coli* DnaJ and CbpA	3
DjlC	III	DnaJ	I	No	Inability to replace *E. coli* DnaJ *in vivo*	4
Large T Ag [*]	III	DnaJ Ydj1	I I	Yes Yes	Complementation for lack of *E. coli* DnaJ and CbpA	5
DnaJ Ydj1	I I	Large T Ag	III	No No	Inability to form infectious virions; ineffective DNA replication; inability to stimulate ATPase activity	6
Hdj1 Hsj1	I II	Large T Ag	III	Yes Yes	Mediating reduction in levels of phosphorylated p130 protein	7
P58 [IPK]	III	DnaJ Ydj1	I I	Yes Yes	Complementation for lack of *E. coli* DnaJ and CbpA; substitute for lack of Ydj1 in yeast	8
Hsj1 DnaJA2	II I	Large T Ag	III	Yes Yes	Allows efficient viral DNA replication	9
Hdj1	I	DnaJ	I	Yes	Complementation for lack of *E. coli* DnaJ and CbpA	10
Hsj1	II	T antigen	III	Yes	Functional in T Ag mediated cellular transformation	11
T Ag	III	Ydj1	I	Yes	Substitute for lack of Ydj1 in yeast	12

[*]: The large T antigen J domains from SV40, JC virus and BK virus were used in this experiment.
[^] - Sis1, Scj1, Mdj1 and Ydj1 are from *S. cerevisiae*; DnaJ, DjlA and DjlC are from *E. coli*; Hdj1, Hsj1, DnaJA2 and P58 [IPK] are from *H. sapiens*.
[#]: 1 - (Schlenstedt *et al.*, 1995); 2 - (Deloche *et al.*, 1997); 3 - (Genevaux *et al.*, 2001b); 4 - (Kluck *et al.*, 2002); 5 - (Kelley and Georgopoulos, 1997b); 6- (Sullivan *et al.*, 2000b); 7 - (Stubdal *et al.*, 1997); 8 - (Yan *et al.*, 2002); 9 - (Campbell *et al.*, 1997); 10 - (Genevaux *et al.*, 2002); 11 - (Zalvide *et al.*, 1998); 12 - (Fewell *et al.*, 2002)

Additional substitutions in the J domain of the Sis1J-Sec63 chimera did result in a functional chimeric protein (K42V, Q13R, K17S) (Schlenstedt *et al.*, 1995), implying differences in the amino acid specificity determinants that would affect the interaction of these proteins with

their partner Hsp70s. Scj1 and Sec63 were likely to both be interacting with the endoplasmic reticulum Hsp70, Kar2p, whereas Sis1 would be interacting with a cytosolic Hsp70, and Mdj1 with a mitochondrial Hsp70.

Domain swapping experiments have also been performed using the J domains from the viral large T antigen proteins. J domains from these proteins used to substitute for the J domains from DnaJ and the yeast protein Ydj1 gave fully functional chimeric DnaJ and Ydj1 proteins (Kelley and Georgopoulos, 1997b). The converse however did not apply. Substitutions of either the DnaJ or the Ydj1 J domains in the large T antigen only gave partially functional chimeras, with the inability to perform some of their cellular roles (Sullivan *et al.*, 2000b). SV40 T antigen chimeras containing the J domains from Hdj1 (human DnaJ protein 1) or Hsj1 (*Homo sapiens* DnaJ 1) were able to function in the reduction of p130, and phosphorylation of p130 (Stubdal *et al.*, 1997) which was important for T Ag mediated cellular transformation. Large T antigen proteins containing the J domains from Hsj1 or DnaJ2 (Hsj2/Hdj2; human DnaJ protein 2) could promote viral replication (Campbell *et al.*, 1997). Recently a mammalian J domain was shown to be able to substitute for the J domains from DnaJ and Ydj1 (Yan *et al.*, 2002). The J domain from $P58^{IPK}$ a mammalian protein that also contained the chaperone interaction tetratricopeptide repeat motifs could functionally replace the DnaJ and Ydj1 J domains *in vivo* using complementation assays. Mutations in the HPD motif of the chimeric proteins prevented successful complementation for the lack of DnaJ in *E. coli*, or Ydj1 in yeast knockout strains (Yan *et al.*, 2002).

1.5.3.2 Specificity of Interaction in *E. coli*

Whole protein experiments using the defined systems in *E. coli* also imply some level of specificity of interaction. As *E. coli* has no intracellular compartments, proteins containing J domains can potentially interact with all Hsp70s in the cell. This could allow for a situation whereby the levels of non-specific DnaJ-like proteins interaction with Hsp70s swamp the levels of productive interactions. *E. coli* has three identified Hsp70s, DnaK, Hsc66 (Heat shock cognate protein 66 kDa) (Lelivelt and Kawula, 1995; Vickery *et al.*, 1997; Silberg and Vickery, 2000) and HscC/Hsc62 (Heat shock cognate protein C) (Yoshimune *et al.*, 2002; Kluck *et al.*, 2002). It also has several identified DnaJ-like proteins, DnaJ, CbpA (curved DNA binding protein) (Ueguchi *et al.*, 1994), DjlA (DnaJ-like protein A) (Clarke *et al.*,

1996), Hsc20 (Heat shock cognate protein 20 kDa) (Silberg *et al.*, 1998), DjlB (DnaJ-like protein B) and DjlC/Hsc56 (DnaJ-like protein C / Heat shock cognate protein 56 kDa) (Yoshimune *et al.*, 2002; Kluck *et al.*, 2002). All these proteins have defined partnerships (Figure 1.7).

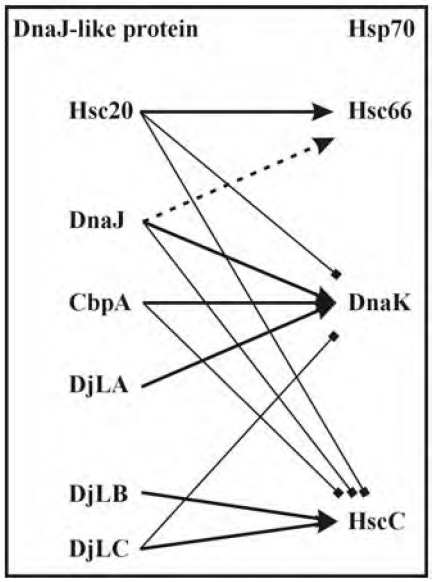

Figure 1.7: Specificity of Interaction between Hsp70 – DnaJ-like proteins in *E. coli*

Thick arrows indicate known partnerships; dotted lines and arrows indicate lower levels of ATP hydrolysis stimulation *in vitro*. Thin lines with squares replacing arrowheads indicate no detectable levels of ATPase stimulation, and no known partnership.

DnaJ, CbpA (Ueguchi *et al.*, 1995; Wegrzyn *et al.*, 1996) and DjlA (Genevaux *et al.*, 2001b) interacted with DnaK, Hsc20 (HscB) interacted with Hsc66 (HscA) (Vickery *et al.*, 1997; Silberg *et al.*, 1998), and DjlB and DjlC interacted with HscC (Kluck *et al.*, 2002). DnaJ could stimulate the ATPase activity of Hsc66 to a certain extent, whereas Hsc20 could not stimulate the ATPase activity of DnaK (Silberg *et al.*, 1998). DnaJ, CbpA and Hsc20 could not stimulate the ATPase activity of HscC (Kluck *et al.*, 2002), and DjlC could not stimulate the ATPase activity of DnaK. This was most likely due to the glutamic acid substitution in the HPD motif in DjlB and DjlC, which contained a HPE motif instead. GrpE could however stimulate the ATPase activity of HscC in the presence of DjlC (Yoshimune *et al.*, 2002). DnaJ also could not stimulate the ATPase activity of the mammalian endoplasmic reticulum (ER) Hsp70, BiP (Chevalier *et al.*, 2000). It could stimulate activity of the mammalian Hsc70 (Minami *et al.*, 1996), although mammalian Hsp40 could not stimulate the activity of DnaK.

The primary site of interaction between DnaJ-like proteins and their partner Hsp70s is the J domain; hence the inability of all DnaJ-like proteins to interact with all Hsp70s implies a level of mechanistic discrimination. These determinants could be in the poorly conserved remainder of the DnaJ-like protein. However for several reasons the major site of interaction must be in the J domain. Firstly, substitutions in the HPD motif of the J domain abolish functional interactions with partner Hsp70s (Tsai and Douglas, 1996; Mayer *et al.*, 1999). Secondly, there are DnaJ-like proteins that consist almost solely of a J domain such as the tiny T antigen (Riley *et al.*, 1997), implying that any ability for these DnaJ-like proteins to interact with a partner Hsp70 must reside solely in the J domain. Thirdly, if a conserved mechanism of interaction between DnaJ-like proteins and partner Hsp70s has been retained, conserved residues in both the DnaJ-like protein and the partner Hsp70 must also be present, and only residues in the J domain are conserved across all DnaJ-like proteins. However, these reasons will only apply to a general J domain – Hsp70 interaction. Other potential ways for a cellular regulation of specific interactions to occur include localisation of partnerships to specific organelles or tissues at a eukaryotic level, and by production of specific partnerships at different times in the cellular cycle. While the subtle level of distinguishing between partnerships could also be influenced by the C terminal portion of DnaJ-like proteins most of the evidence to date indicates that this level of specificity was likely to reside primarily in the J domain. This can be seen in domain swapping experiments (Schlenstedt *et al.*, 1995) where the C terminus was the same, but the J domain changed, forming a non-functional chimera, and in the existence of proteins such as the tiny T antigen, where specificity determinants must reside in the J domain (Riley *et al.*, 1997). If the sole function of the J domain is to stimulate the ATPase activity of a partner Hsp70, particularly through the HPD motif, it is to be expected that it would be promiscuous in its action. It is evidently not promiscuous, hence there must be a level of specificity inherent in its sequence or structure. Alternatively, there is the possibility that there exist proteins whose functions are to keep DnaK and DnaJ in a complex together, and this clamp provides specificity. An example is found in *Thermus thermophilus* (Motohashi *et al.*, 1994; Motohashi *et al.*, 1996). It is important to distinguish between general binding determinants that will be important in the majority of DnaJ-like – Hsp70 partnerships, and specific binding determinants, that will be important in specific DnaJ-like – Hsp70 partnerships. It is an attempt to elucidate some of these binding determinants, that is one of the aims of this thesis.

1.6 HYPOTHESIS AND BROAD OBJECTIVES

"All J domains are created equal, but some are more equal than others."[6]

Despite the high level of sequence and structural similarity between J domains, J domains cannot be replaced generically. This implies a level of control mediating DnaJ-Hsp70 interactions, with the likelihood being that this level of control occurs at the amino acid level in the J domain. However, determining which residues are involved in this interaction, outside the HPD motif, has not been elucidated thoroughly. Based on sequence alignments of J domains, several charged amino acids outside the HPD motif have been shown to be conserved (Hennessy *et al.*, 2000). The hypothesis for this thesis is that certain conserved, charged residues, separate from the HPD motif, are important in mediating the interaction of DnaJ-like proteins with partner Hsp70s at a general level (i.e. the majority of interactions between DnaJ-like proteins and Hsp70s will require these residues), and that some of these residues may contribute towards specificity of an interaction (i.e. these residues will determine if a particular interaction can actually occur).

The broad objective of this thesis was to determine which amino acids in the J domain, apart from the HPD motif, were important for the interaction with partner Hsp70s. The predominant focus was on those residues important in the interaction between DnaJ and DnaK proteins. Distinguishing between those residues important for general binding, and those required for a specific DnaJ-DnaK interaction was a secondary objective.

[6] With apologies to George Orwell. The original quote can be found in *Animal Farm*.

1.7 APPROACH

1.7.1 Bioinformatic analysis of the J domain

Target residues of interest were identified through the determination of levels of conservation of amino acids in the J domain. This required expansion of a previously performed multiple sequence alignment (Hennessy *et al.*, 2000), and further analysis of the data obtained. Other techniques employed included structural alignments of known J domain structures, in an attempt to determine regions of conservation of charge and hydrophobicity. Contact analysis of known J domain structures was also performed to determine regions critical for maintaining the structural integrity of the J domain.

1.7.2 Development of a model system

A model system was developed to test the effect of the substitutions. The model system chosen was the *Agrobacterium tumefaciens* DnaJ and DnaK. This system was chosen as an alternative to the *E. coli* system. Using an alternative system would provide a mechanism for determining if residues found to be important in the *E. coli* system were important in other prokaryotic systems, or only in the *E. coli* system.

1.7.3 Replacement of the J domain of *A. tumefaciens* DnaJ with J domains from alternative *A. tumefaciens* DnaJ-like proteins

Additional DnaJ-like proteins were identified from the *A. tumefaciens* C58 genome. The open reading frames encoding these proteins were isolated. The regions encoding the J domains were used as replacements for the J domain from the *A. tumefaciens* DnaJ protein. This would provide a framework for investigating which amino acids were potentially important at both a general and a specific level for the interaction of *A. tumefaciens* DnaJ with a partner DnaK.

1.7.4 Rational site-directed mutagenesis of the J domain

Conserved amino acids in the J domain were substituted using site-directed mutagenesis, and the effect on the functioning of the J domain was investigated using *in vivo* and *in vitro* techniques. Suitable amino acids were identified using the bioinformatics-based approach and the results from the J domain replacement experiments, as well as choosing amino acids previously described in the literature as being important. Two distinct groups of amino acids were substituted. The first was the conserved, hydrophobic amino acids. These tended to be the most conserved amino acids, and were thought to play a key role in maintaining the structural integrity of the J domain. The second group was the conserved, charged amino acids. These amino acids were thought to be important in specifying interactions with partner Hsp70s.

Chapter 2: Bioinformatic Analysis of the J domain

2 BIOINFORMATIC ANALYSIS OF THE J DOMAIN

2.1 INTRODUCTION

2.1.1 Current perspective on bioinformatics, genomics and proteomics

Bioinformatics is a relatively new area of research that entails the usage of computers and specialised software programmes in order to perform *in silico* analyses of protein and nucleic acid sequence and structural data. Bioinformatics can be used in the determination of possible functions for novel proteins, additional characterisation of known genes and proteins, and the generation of three dimensional homology models of novel proteins. There are many programmes currently available, both in the public domain and commercially available, that can be used for the generation of bioinformatic data. Equally, many databases have been generated that collate information on specific families of proteins, or on specific organisms. Many of these are referenced in the Molecular Biology Database (Baxevanis, 2002; Baxevanis, 2003).

The reason for the steady growth in the field of bioinformatics can be ascribed to the exponential increase in the number of uncharacterised nucleotide and derived protein sequences in the databases, due to the vast number of genome sequence projects currently underway. There are over 18 million nucleic acid sequences (August 2002) that are accessible in the GenBank database (Boguski, 1998; Benson *et al.*, 1999; Benson *et al.*, 2003). Approximately half of these are automatically annotated data produced by the various genome sequencing projects. Currently the genome repository at the National Centre for Biotechnology Information (NCBI) has over 700 partial and complete genomes represented in its database (181 bacteria, 36 archaea, and 559 eukaryotes), although there are still many organisms whose genetic makeup still needs to be deciphered. Many genomes have been sequenced completely, or are in the process of being sequenced. Important genomes that have been sequenced include the plant *Arabidopsis thaliana* (The Arabidopsis Genome Initiative., 2000), the bacterial *Mycobacterium tuberculosis* (Cole *et al.*, 1998), *Borrelia burgdorferi* (Fraser *et al.*, 1997), *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Mycoplasma genitalium* (Fraser *et al.*, 1995) the archaeal *Methabobacterium thermoautotrophicum* (Smith

et al., 1997), the plant pathogen *Agrobacterium tumefaciens* (Goodner *et al.*, 2001; Wood *et al.*, 2001), the nematode *Caenorhabditis elegans* (Hodgkin *et al.*, 1995) and the human genome (International Human Genome Sequencing Consortium, 2001). The number of sequences deposited in GenBank has increased exponentially since 1982. The amount of sequences and number of nucleotides in GenBank is shown in Table 2.1. Swiss-Prot, the database for protein sequences, currently contains 138 347 sequence entries, comprising no less than 47 026 705 amino acids (Table 2.2).

Table 2.1: Exponential increase in data deposited in GenBank (1982 – 2002)

Year	Base Pairs	Sequences	Year	Base Pairs	Sequences
1982	680 338	606	1993	157 152 442	143 492
1983	2 274 029	2 427	1994	217 102 462	215 273
1984	3 368 765	4 175	1995	384 939 485	555 694
1985	5 204 420	5 700	1996	651 972 984	1 021 211
1986	9 615 371	9 978	1997	1 160 300 687	1 765 847
1987	15 514 776	14 584	1998	2 008 761 784	2 837 897
1988	23 800 000	20 579	1999	3 841 163 011	4 864 570
1989	34 762 585	28 791	2000	11 101 066 288	10 106 023
1990	49 179 285	39 533	2001	15 849 921 438	14 976 310
1991	71 947 426	55 627	2002	28 507 990 166	22 318 883
1992	101 008 486	78 608	**Dec 15 2003**	**36 553 368 485**	**30 968 418**

http://www.ncbi.nih.gov/Genbank/genbankstats.html

By contrast the Protein Data Bank (Berman *et al.*, 2000) has slightly over 20 000 structures available (Table 2.2; Table 2.3) (Westbrook *et al.*, 2003), which is approximately a tenth of the number of protein sequences, and the majority of these known structures are of the same protein containing different mutations. There are, therefore, only relatively few proteins with a known sequence, that have a known function and known three dimensional structure. Consequently a vast amount of work remains to be done to extrapolate available data to unknown proteins (Andrade and Sander, 1997), particularly given the huge magnitude of un-annotated data being generated by the genome sequencing projects.

Table 2.2: Comparison of numbers of protein sequences in various databases

Database	Release	Date	No. of Entries
PIR	78.03	24-Nov-2003	283 366
Swiss-Prot	42.7	15-Dec-2003	141 681
TrEMBL	25.7	15-Dec-2003	1 078 339
GenPept	139.0	15-Dec-2003	1 650 733
RefSeq		30-Dec-2003	622 048
PDB		29-Dec-2003	23 750

http://pir.georgetown.edu/pirwww/search/pirnref.shtml

Table 2.3: PDB Holdings List: 6 January 2004

		Molecule Type			
Technique	**Proteins, Peptides; Viruses**	**Protein/Nucleic Acid Complexes**	**Nucleic Acids**	**Carbohydrates**	**Total**
X-ray Diffraction and other	18 632	887	712	14	20 245
NMR	2 905	95	564	4	3 568
Total	21 537	982	1276	18	**23 813**

http://www.rcsb.org/pdb/holdings.html

The premise of bioinformatics is that biological questions can be answered *in silico*. The types of questions that can be answered can range from prediction of function for a polypeptide sequence, to prediction of a tertiary fold. The ultimate "Holy Grail" for structural biology and bioinformatics is a method of determining the structure of a protein from its primary amino acid sequence. The majority of the results that will be presented in this section were generated using public domain programmes. Uniform resource locations (URL's) for all sites used are given in the Appendix (Table 6.1 in Section 6.1). Further URL's can be found in the Molecular Biology Database Collection *via* Nucleic Acids Research (http://www.nar.oupjournals.org/cgi/content/full/31/1/1/DC1) (Baxevanis, 2001; Baxevanis, 2002; Baxevanis, 2003).

2.1.2 Bioinformatic analysis of DnaJ-like proteins and the J domain

Several papers have used a bioinformatic approach to look at the importance of specific amino acids in the J domain. The first (Hennessy *et al.*, 2000) looked at the J domains found in all DnaJ-like proteins then currently known. From there, a multiple sequence alignment (MSA) of these was generated, and the percentage occurrence of all amino acids at all positions in the alignment was calculated. This method allowed for the determination of highly conserved residues, and showed the presence of previously unidentified motifs, such as the KFK motif, located in Helix III. The conserved phenylalanine in this motif was shown to have the potential to interact with His^{33} in the HPD motif. This was proposed to be a critical interaction stabilising the lower portion of the J domain. The analysis also identified hydrophobic residues likely to be important for maintaining the structure of the J domain, as well as conserved charged residues that were potentially important from a functional perspective. However some questions still remained. Were non-conserved residues structurally or functionally important, and was their lack of conservation a mechanism of ensuring specificity? Equally, was the conservation of other residues significant, and what was their relevance in J domains where they were not conserved?

An organism specific method was used to examine in particular the DnaJ-like proteins occurring in *Arabidopisis thaliana* (Miernyk, 2001). Miernyk found 89 J domain containing proteins in *A. thaliana*, and 17 Hsp70s. This was unexpectedly high, as comparisons of DnaJ-like proteins and Hsp70s from other organisms indicated a ratio of approximately 2:1 DnaJ-like proteins : Hsp70s (Miernyk, 2001). The *A. thaliana* DnaJ family was divided into 51 families on the basis of the occurrence of structural motifs and domains. Many of these families bore similarity to previously described proteins from other organisms, however the majority appear to have novel functions (Miernyk, 2001). The likelihood was that the majority of these families represent ancestral proteins that sequestered the J domain, in order to fulfil highly specific plant related functions. The high number of DnaJ-like proteins implied that the Hsp70s present must be interacting with multiple DnaJ-like proteins, and that there must be a mechanism in place to ensure that only productive interactions occur.

Both of the papers described above represent differing methods of analysing large quantities of J domains either by focusing specifically on one organism, or by focusing on the differences between the types of DnaJ-like proteins. The aim for this section of work was to update the multiple sequence alignment generated previously (Hennessy *et al.*, 2000), and to look more closely at the differences between J domains. This would allow the identification of residues that could be postulated to be critical in the structure and function of the J domain.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 Updating of the multiple sequence alignment

As stated previously, work had been done initially on generating a multiple sequence alignment that contains all identified J domains (Hennessy *et al.*, 2000). The aim of this section of the work was to update this alignment, and from there to determine conserved functional and structural residues in addition to those described previously. The generation of the alignment needed to be as simple as possible to deal with the volume of sequences that could be aligned. Further data mining from the database could then be attempted. Secondly, the presence of non-HPD containing J domains would also be investigated in order to investigate whether these could be considered true J domains.

Initially J domain containing proteins were identified by running a BLAST search (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990) of the *E. coli* DnaJ J domain, and the derived consensus sequence (Hennessy *et al.*, 2000) against the non redundant database present at NCBI. BLAST searches were also performed against the additional databases present at the NCBI in order to identify as many J domain containing proteins as possible.

A database of J domain containing proteins was established (see Appendix, Section 6.1), based on data obtained from the various protein sequence databases. All proteins that appeared to contain a J domain were inserted into the database, and redundant sequences were removed as far as was possible. The J domains from these proteins were inserted into a Microsoft ExcelTM spreadsheet and aligned manually, using the HPD motif as an anchor for Helices I and II, and the KFK motif (Hennessy *et al.*, 2000) and a conserved Ala-Tyr dipeptide as the anchors for Helices III and IV. The enclosed CD contains the complete alignment, and databases from which the sequences were derived.

Similar searches were performed using the J domains from cysteine string proteins (Chamberlain and Burgoyne, 1997b), the T antigen proteins (Kelley and Georgopoulos, 1997b), the Hsc20 family of proteins (Silberg *et al.*, 1998) and the RESA proteins (Bork *et al.*, 1991). The J domains from these proteins were also added to the overall alignment.

The levels of conservation of each amino acid at each point were also determined using function parameters available in Excel. The command used was the "Countif" command, which allowed for the determination of the amount of each amino acid at each position in the alignment. Using this data, a consensus sequence could be determined, and other data could be obtained relatively easily from the alignment.

A comparative alignment was performed using the ClustalX package (Thompson *et al.*, 1994; Jeanmougin *et al.*, 1998). The overall alignment was subdivided into three on the basis of the classification previously described (Cheetham and Caplan, 1998), and internal comparisons were performed.

2.2.2 Determination of structurally and functionally conserved residues

Residues were divided into several categories in the alignment. The subdivision criteria are defined below in Table 2.4.

Table 2.4: Definition of groupings used for analysis of multiple sequence alignment

Definition	Criteria
Positively charged	lysine, arginine or histidine
Negatively charged	aspartic acid or glutamic acid
Polar	serine, threonine, asparagine, glutamine, tyrosine
Structural	leucine, isoleucine, alanine, valine, phenylalanine, tryptophan
Highly conserved	Greater than 75 % conservation
Conserved	Greater than 60 % conservation
Non conserved / variable	Less than 40 % conservation

2.2.3 Contact analysis

Analysis of residues involved in contacts with conserved residues such as the HPD motif was performed using the molecular analysis package Whatif (Vriend, 1990).

2.2.4 Modelling of the consensus sequence

Two methods of modelling the consensus sequence were used. The first involved SWISS-MODEL (Peitsch, 1996), based at the Expasy server (http://www.expasy.ch). The second involved the modelling software included in the analysis programme Whatif (Vriend, 1990).

Visualisation of structures was performed using the Swiss PDB viewer, Rasmol (Sayle and Milner-White, 1995) and Molscript (Kraulis, 1991).

2.2.5 Identification of pseudo J domain structures

The derived overall consensus sequence was used as a probe to conduct a BLAST analysis of the amino acid sequences of proteins in the PDB. Structures that were detected but were not of known J domain structures, were examined further to investigate the presence of conserved amino acids.

2.3 RESULTS

2.3.1 Updating of the multiple sequence alignment

The number of sequences aligned in this work was 907, compared to the little over 200 used in the previous alignment. This was a reflection of the marked increase in the number of genomes that have been sequenced over the intervening two years. The complete database is included as an appendix on the attached CD.

As expected there was minimal variation in the overall consensus sequences between the work previously described, and the work presented herein. The most significant difference occurred in the Type III alignment where the levels of sequence conservation decreased markedly, however overall patterns of conservation remained the same. The results are summarised below in Figure 2.1.

Highly conserved residues (>75 %) were located at positions 3, 6, 12, 21, 29, 30, 31, 64, 70, 71, 74, 83 and 84 (consensus numbering). These are predominately hydrophobic residues, with the exception of the aspartic acid at position 84, located in Helix IV. In general, Type I J domains had a higher level of conservation than Type II, and both of these were more conserved than Type III. The most notable exception occurs at position 32, where Type II and Type III J domains are significantly more conserved than Type I.

Several residues were not conserved, and varied between the four consensus sequences obtained. These residues are indicated by asterisks in Figure 2.1, and tabulated in Table 2.5. These residues occurred in all helices, with the exception of Helix III.

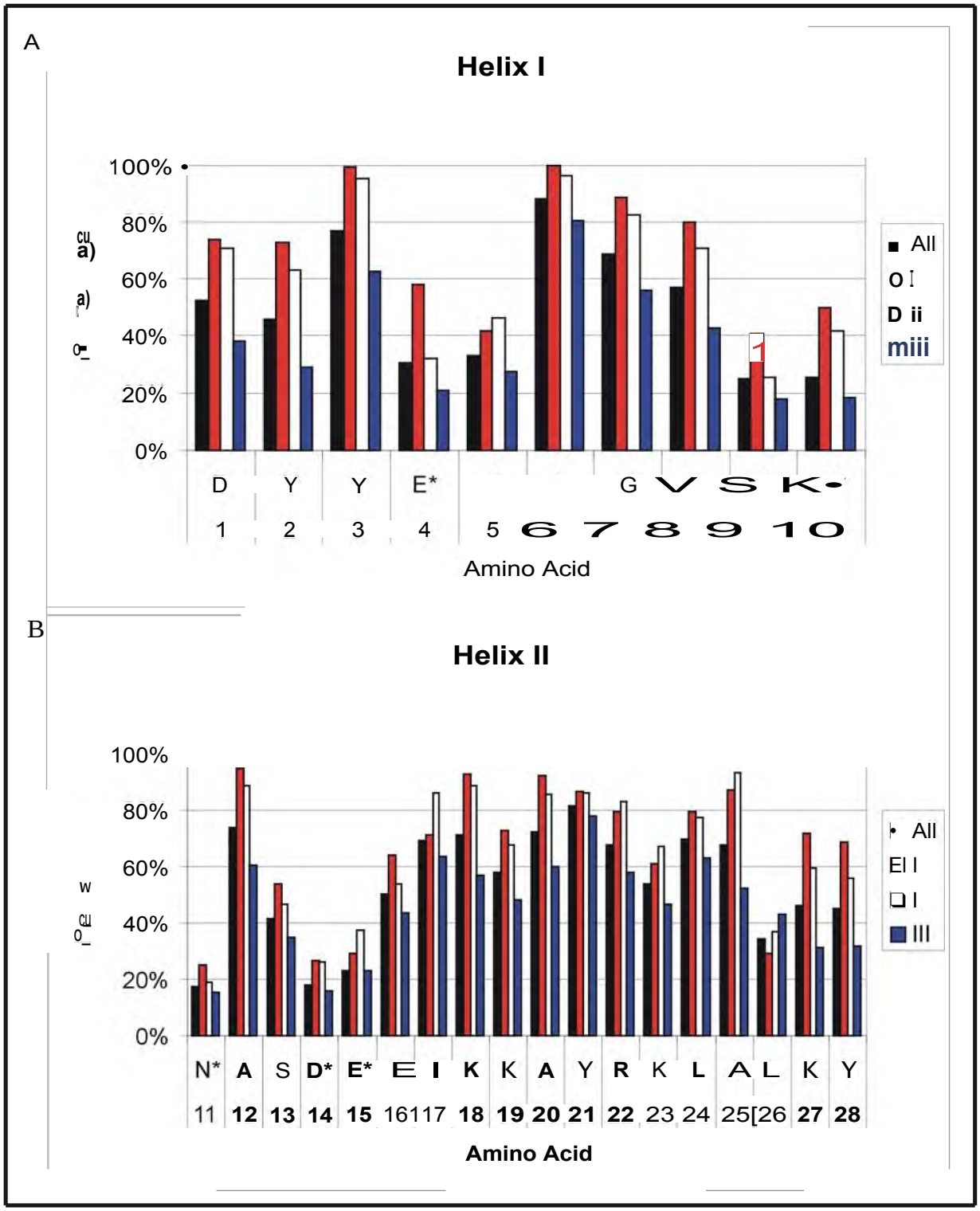

Figure 2.1 A and B

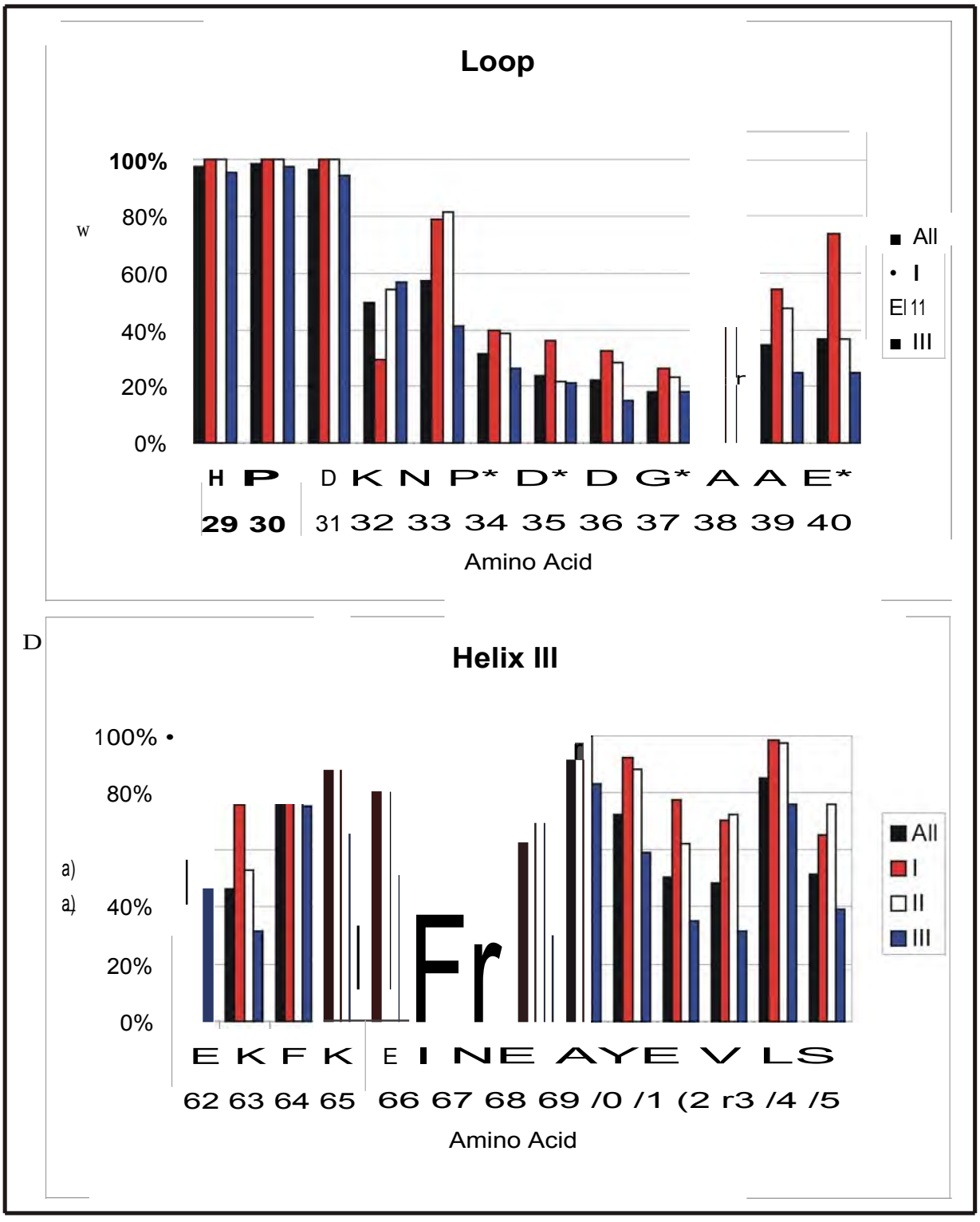

Figure 2.1: C **and D**

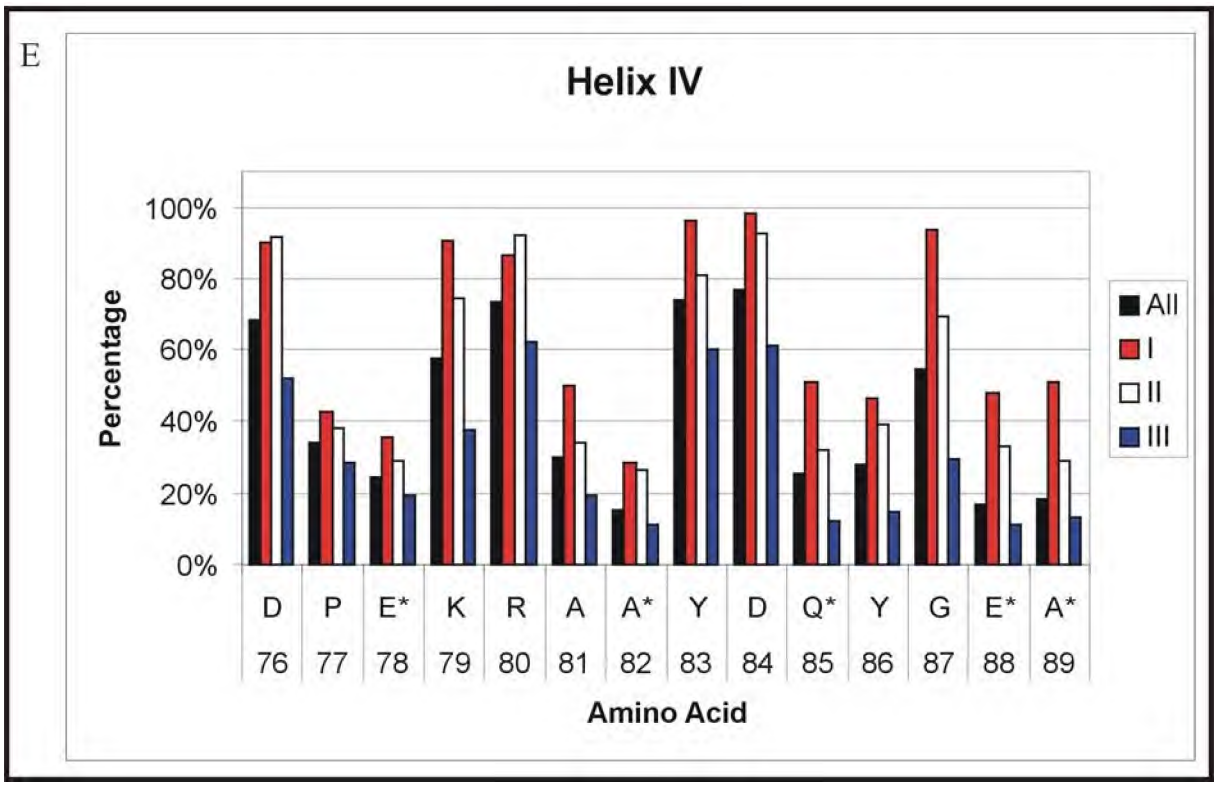

Figure 2.1 - E

Figure 2.1: Summary of the residues most prevalent at each position of the multiple sequence alignment of all J domains

The sequences of 907 J domains were aligned. The percentage conservation of the most highly conserved residue at each position for the overall alignment is depicted in black, for Type I residues in red, for Type II in white, and for Type III in blue. The results are divided with (A) representing Helix I and flanking loop region, (B) represents Helix II, (C) represents the loop region, (D) represents Helix III and (E) represents Helix IV and flanking regions. Numbering is for the overall consensus sequence. Asterisks next to an amino acid indicate variation in the type of most conserved amino acid at that position between the four consensus sequences; the percentage given is for the most conserved residue at that point and further details are given in Table 2.5. A consensus sequence is not given for positions in the loop region after the number of sequences used for the alignment dropped below 50 % of the total number of sequences at that point.

Table 2.5: Variation in amino acids in the multiple sequence alignment

Position	Helix	All		Type I		Type II		Type III	
		aa	%	aa	%	aa	%	aa	%
4	I	E	30.07	E	58.00	K	31.91	E	20.70
5	I	I	33.22	S	41.99	V	46.10	I	27.59
10	I	K	25.58	K	49.79	R	41.84	P	18.48
11	II	N	17.60	N	24.90	S	19.15	T	15.20
14	II	D	17.89	E	26.50	D	26.24	D	15.62
15	II	E	22.95	D	29.06	D	37.59	E	23.15
26	II	L	34.07	M	26.07	L	36.62	L	43.13
34	Loop	P	31.30	P	39.70	K	38.73	P	26.30
35	Loop	D	23.60	G	35.90	D	21.80	D	20.90
37	Loop	G	18.29	K	26.51	K	23.24	G	17.94
39	Loop	A	34.57	A	54.36	E	47.41	A	24.89
40	Loop	E	36.46	E	73.77	A	36.59	A	24.86
78	IV	E	24.60	Q	35.47	E	29.10	E	19.30
82	IV	A	15.38	A	28.63	I	26.24	I	11.06
85	IV	Q	25.33	Q	50.85	Q	31.91	K	12.45
88	IV	E	16.93	H	47.86	E	33.33	S	11.28
89	IV	A	14.55	A	50.85	E	29.08	E	13.11

Numbering is according to the consensus sequence numbering. Amino acid is abbreviated to aa, and % depicts the level of percentage conservation of that specific amino acid at that position. The structural element that the residues are located on is also indicated.

The level of conservation was variable between types, and between different positions. Analysis of the conservation of the residues allowed sub-division into several categories. Firstly, those charged residues that were conserved, and were hence likely to be important from a functional perspective. Secondly, those hydrophobic residues that were conserved, and were therefore likely to be important from a structural perspective. Thirdly, those residues that had low levels of conservation, and could therefore potentially play a role in the determination of specificity of interaction between specific DnaJ-like proteins and partner Hsp70s. These categories are summarised below in Figure 2.2 and Figure 2.3.

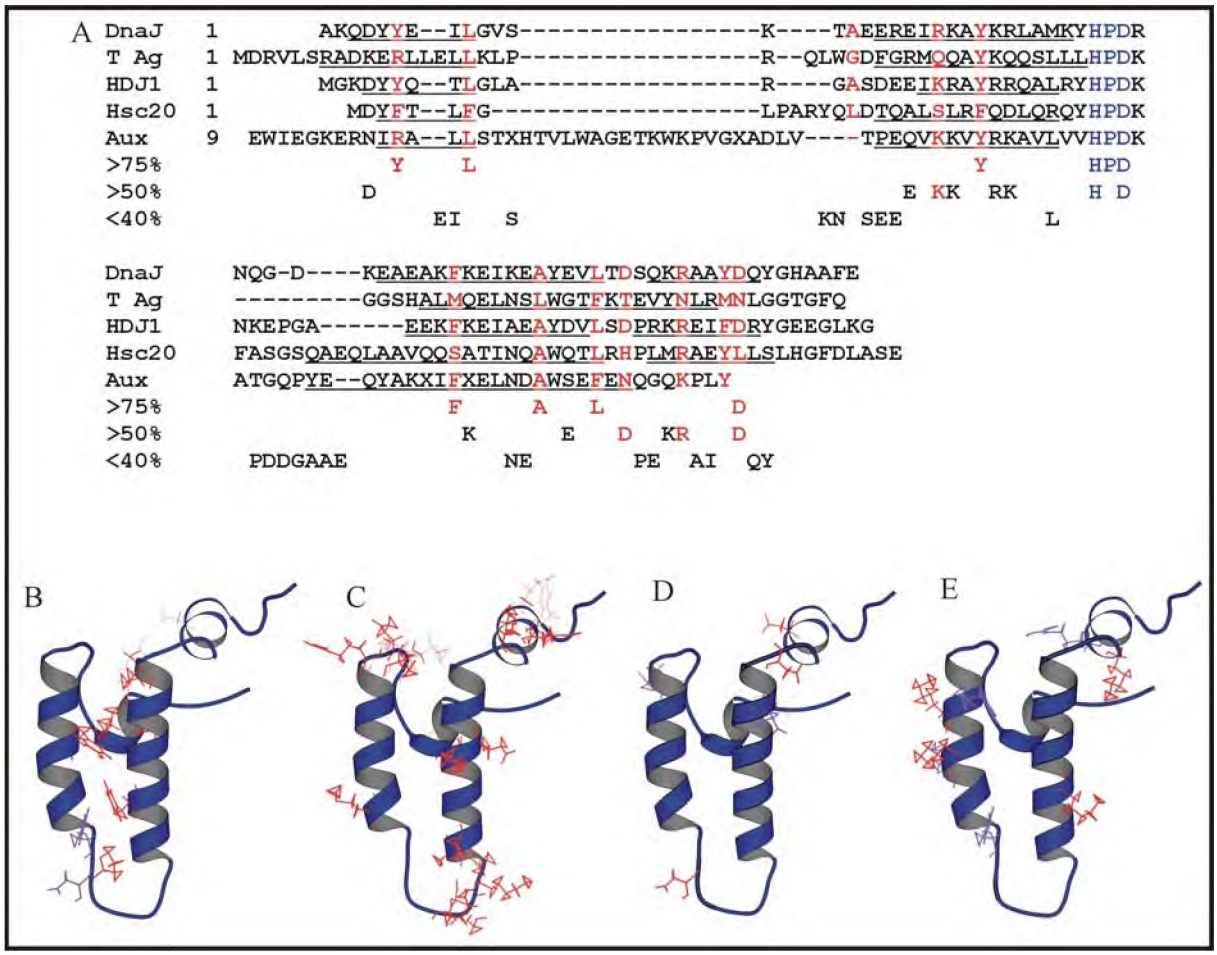

Figure 2.2: Orientation of conserved and poorly conserved residues in the overall consensus sequence

(A) depicts a structure based alignment of the known J domain structures performed using the SPDBV programme. Auxilin is abbreviated as Aux. The T antigen sequence is from Polyomavirus. The position of residues that occur greater than 75 % of the time in the overall consensus are highlighted in red. The HPD motif is depicted in blue. Conserved charged residues that occur greater than 50 % of the time are shown using > 50%. Poorly conserved residues that occur less than 40 % of the time are shown using < 40%. (B) – (E) show amino acids that occur at various levels of conservation. Structures were visualised using the DnaJ J domain structure (1XBL) (Pellechia *et al.*, 1996) in Molscript (Kraulis, 1991). (B) shows all residues that occurred at levels of greater than 75 % conservation. Charged residues are in purple and hydrophobic residues are in red. (C) shows all unconserved residues that occurred at levels of less than 40 % conservation. (D) shows all negatively charged residues that occurred at levels of greater than 50 % conservation. Aspartic acids are depicted in red, and glutamic acids in purple. (E) shows all positively charged residues that occurred at levels of greater than 50 % conservation. Histidines and arginines are depicted in purple, and lysines in red.

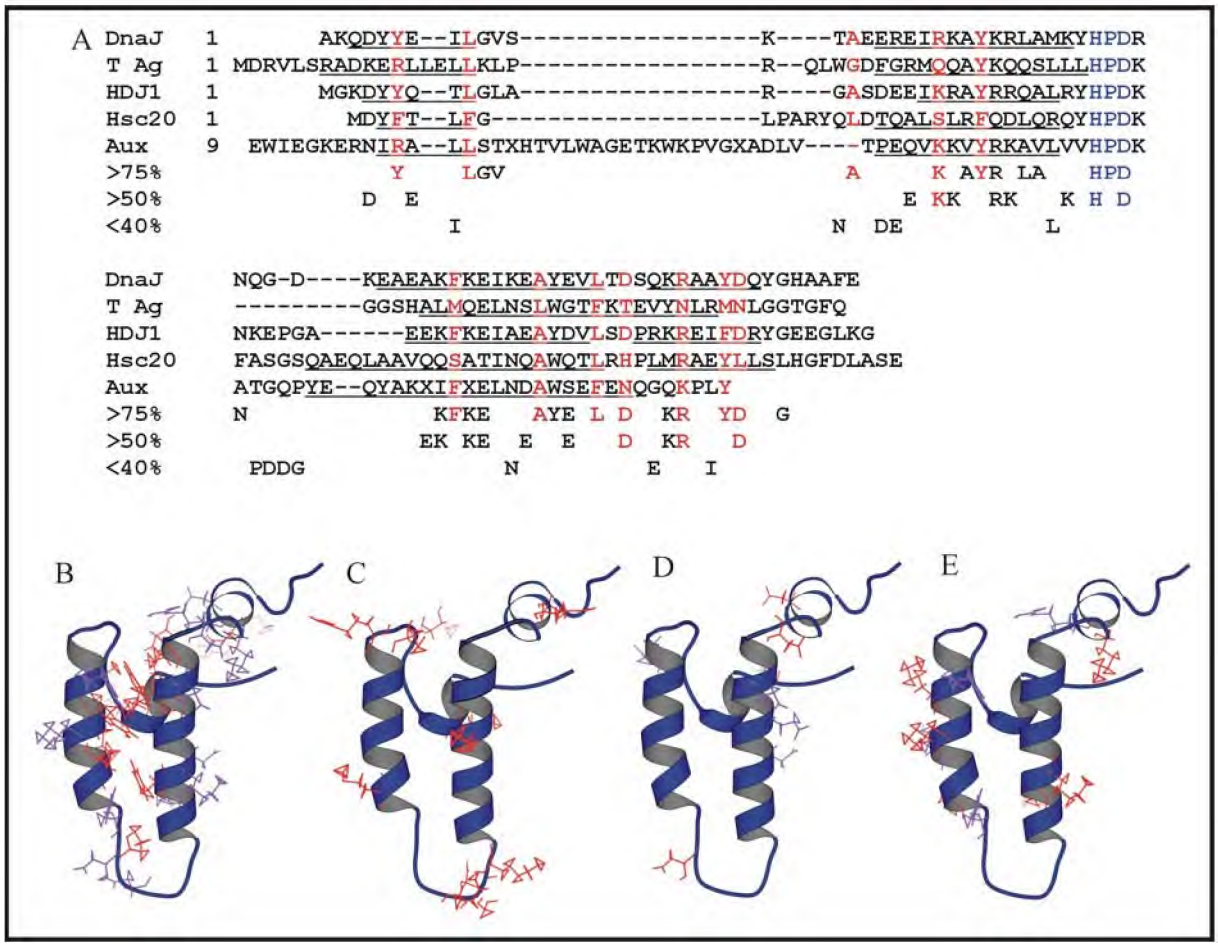

Figure 2.3: Orientation of conserved and poorly conserved residues in Type I J domains

(A) A structure based alignment of the known J domain structures performed using the SPDBV programme. Auxilin is abbreviated as Aux. The T antigen sequence is from Polyomavirus. The position of residues that occur greater than 75 % of the time in the overall consensus are highlighted in red. The HPD motif is depicted in blue. Conserved charged residues that occur greater than 50 % of the time are shown using > 50%. Poorly conserved residues that occur less than 40 % of the time are shown using < 40%. (B) – (E) show amino acids that occur at various levels of conservation. Structures were visualised using the DnaJ J domain structure (1XBL) (Pellechia *et al.*, 1996) in Molscript (Kraulis, 1991). (B) shows all residues that occurred at levels of greater than 75 % conservation. Charged residues are in purple and hydrophobic residues are in red. (C) shows all unconserved residues that occurred at levels of less than 40 % conservation. (D) shows all negatively charged residues that occurred at levels of greater than 50 % conservation. Aspartic acids are depicted in red, and glutamic acids in purple. (E) shows all positively charged residues that occurred at levels of greater than 50 % conservation. Histidines and arginines are depicted in purple, and lysines in red.

Type I J domains generally had more highly conserved residues, as well as more conserved negatively and positively charged residues than the overall consensus sequence; conversely it had fewer poorly conserved residues. This was a reflection of the impact that the wide variety of Type III J domains has on the overall consensus sequence. Equally it was apparent that the most conserved residues in the overall consensus sequence are hydrophobic in nature, and therefore likely to play a structural role in the J domain. These hydrophobic residues included Tyr^{3}, Leu^{6}, Tyr^{21}, Phe^{64}, Ala^{70} and Leu^{74} (consensus numbering). The only charged residue that was highly conserved in the overall consensus sequence was Asp^{84}, which was located on Helix IV or its equivalent. Additional highly conserved residues in Type I J domains included Gly^{7}, Val^{8}, Ala^{12}, Lys^{18}, Ala^{20}, Arg^{22}, Leu^{24}, Ala^{25}, Asn^{33}, Lys^{63}, Lys^{65}, Tyr^{71}, Glu^{72}, Asp^{76}, Lys^{79}, Arg^{80}, Tyr^{83} and Gly^{87}. These were mostly hydrophobic residues, but there were several charged residues that were highly conserved as well. This may be indicative of a highly conserved mechanism of interaction with a partner Hsp70, which was less conserved among the other Types of DnaJ-like proteins. There were a significant number of positions that have charged residues occurring greater than 50 % of the time in all J domains. This indicated the role that charge plays in the J domain structure and function. The majority of variable residues occurred on Helices I and IV, possibly due to residues in these helices not playing as key a role in the structure of the J domain, and interactions between these helices and the remainder of the DnaJ-like protein.

2.3.2 Backbone conservation of the J domain

Structures of the available J domains were aligned on the DnaJ 1XBL structure (Pellechia *et al.*, 1996)to determine the area of greatest structural conservation, and the results are shown in Figure 2.4. Structures were coloured according to their root mean square deviation (RMSD) from the reference 1XBL structure. Blue indicated structural conservation, and red indicated high variation in structures at that point. The main region of backbone conservation occurred on Helix II of the J domain (Figure 2.4). This region had been specifically implicated in the interaction with partner Hsp70s. The major loss of structural conservation occurred in Helices I and IV. These were generally more mobile than Helices II and III, and may not align at a structural level, even if there was amino acid similarity.

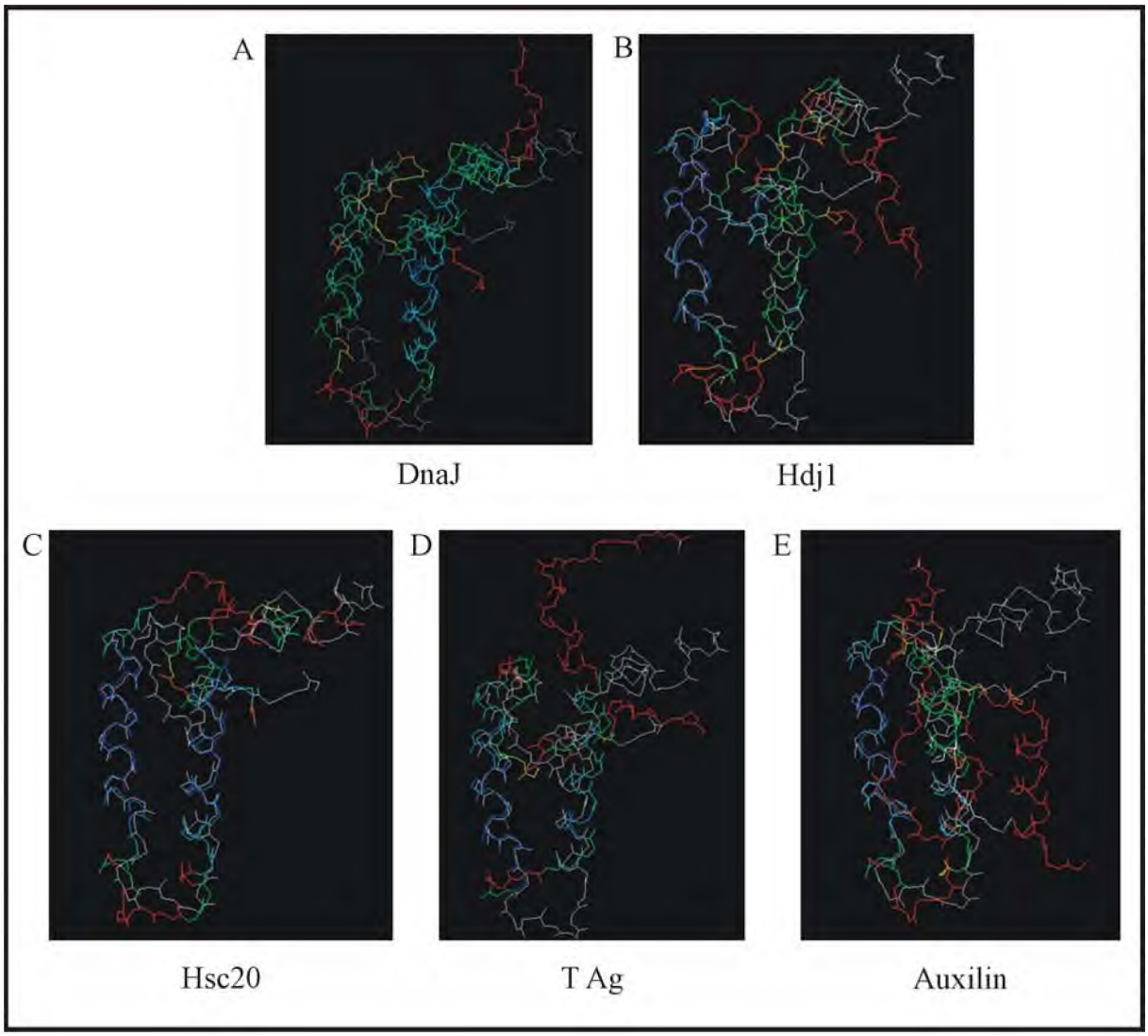

Figure 2.4: Backbone conservation of J domain structures

This figure was generated using Swiss Protein Data Bank viewer (SPDBV) (Guex and Peitsch, 1997). The DnaJ J domain structure was used as a reference and is coloured in grey. (A) Alignment of the *E. coli* DnaJ 1BQZ structure (Huang *et al.*, 1998), (B) represents the HDJ1 structure (1HDJ) (Qian *et al.*, 1996), (C) the Hsc20 structure (1FAF) (Berjanskii *et al.*, 2000), (D) the Large T antigen structure (1FPO) (Cupp-Vickery and Vickery, 2000) and (E) the auxilin structure (1NZ6) (Jiang *et al.*, 2003), structurally aligned on 1XBL. The backbone structures are coloured according to their RMSD with respect to the *E. coli* DnaJ 1XBL structure.

Despite the long loop between Helices I and II in the auxilin J domain structure (Jiang *et al.*, 2003), the helices still overlapped with the helices from the DnaJ J domain structure. This extra loop was also positively charged and was proposed to be involved in binding to Hsc70 (Jiang *et al.*, 2003). The slight change in bulk of the auxilin J domain as compared to other J domain structures may affect the level of specificity of interaction with Hsc70. Interestingly, the second *E. coli* DnaJ J domain structure (1BQZ) did not align as well with 1XBL as did some of the other structures (eg the human 1HDJ structure). The J domain was a highly

mobile structure, and slight changes in the protocol during structural determination may have had a marked effect on the structure that was determined. The homology modelled structure of the consensus J domain also showed structural similarity to *E. coli* DnaJ (Figure 2.5). However the HPD motif I was in a vastly different conformation , concurrent with the apparent shortening of Helix II.

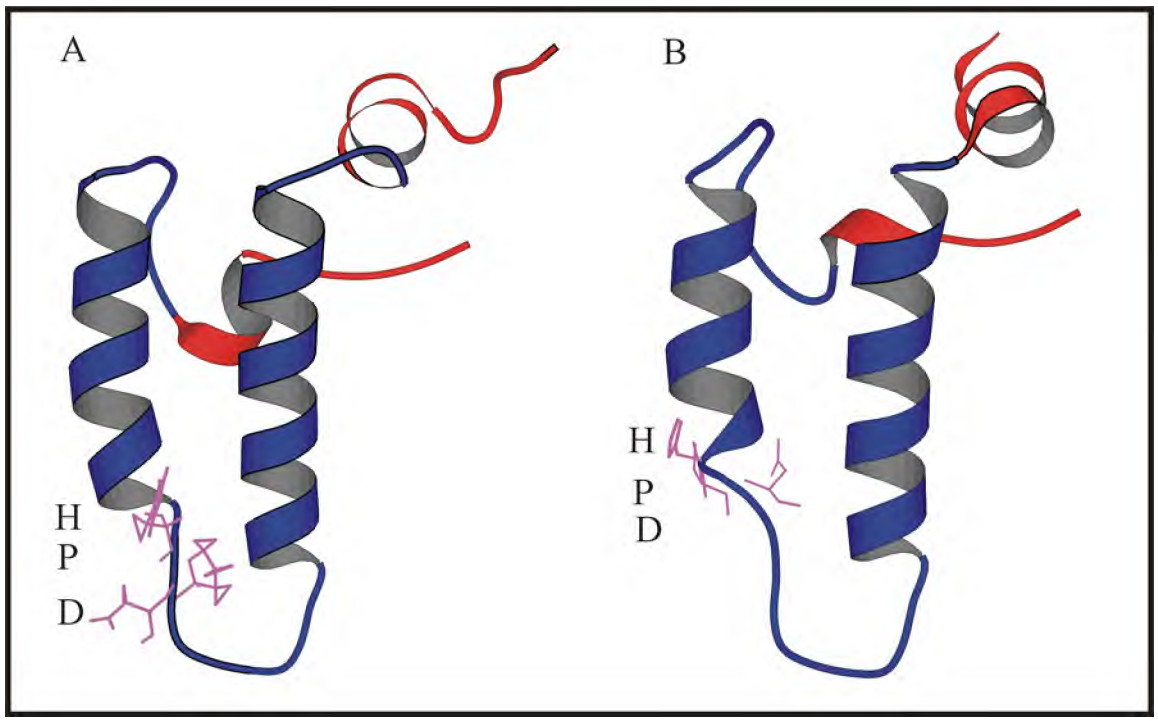

Figure 2.5: Homology model of the derived consensus sequence

The derived consensus sequence was homology modelled using Whatif (Vriend, 1990) with the *E. coli* DnaJ 1XBL (Pellechia *et al.*, 1996) (A) structure as a template. The resulting structure (B) is shown in comparison with the J domain from *E. coli* DnaJ. The figures were generated using Molscript.

2.3.3 Contact Analysis

The J domain structure of *E. coli* DnaJ (1XBL) (Pellechia *et al.*, 1996) was analysed using the analysis package Whatif (Vriend, 1990). The results are demonstrated graphically in Figure 2.6, and are tabulated in the Appendix, (Section 6.1, Table 6.2). Almost 30 % of the potential contacts seen involved highly conserved residues (28.9 %) as found in the overall consensus sequence. The additional highly conserved residues located in Type I J domains also accounted for a high proportion of potential contacts. Potential interactions involving highly conserved residues found in the overall consensus sequence are given in Table 2.6.

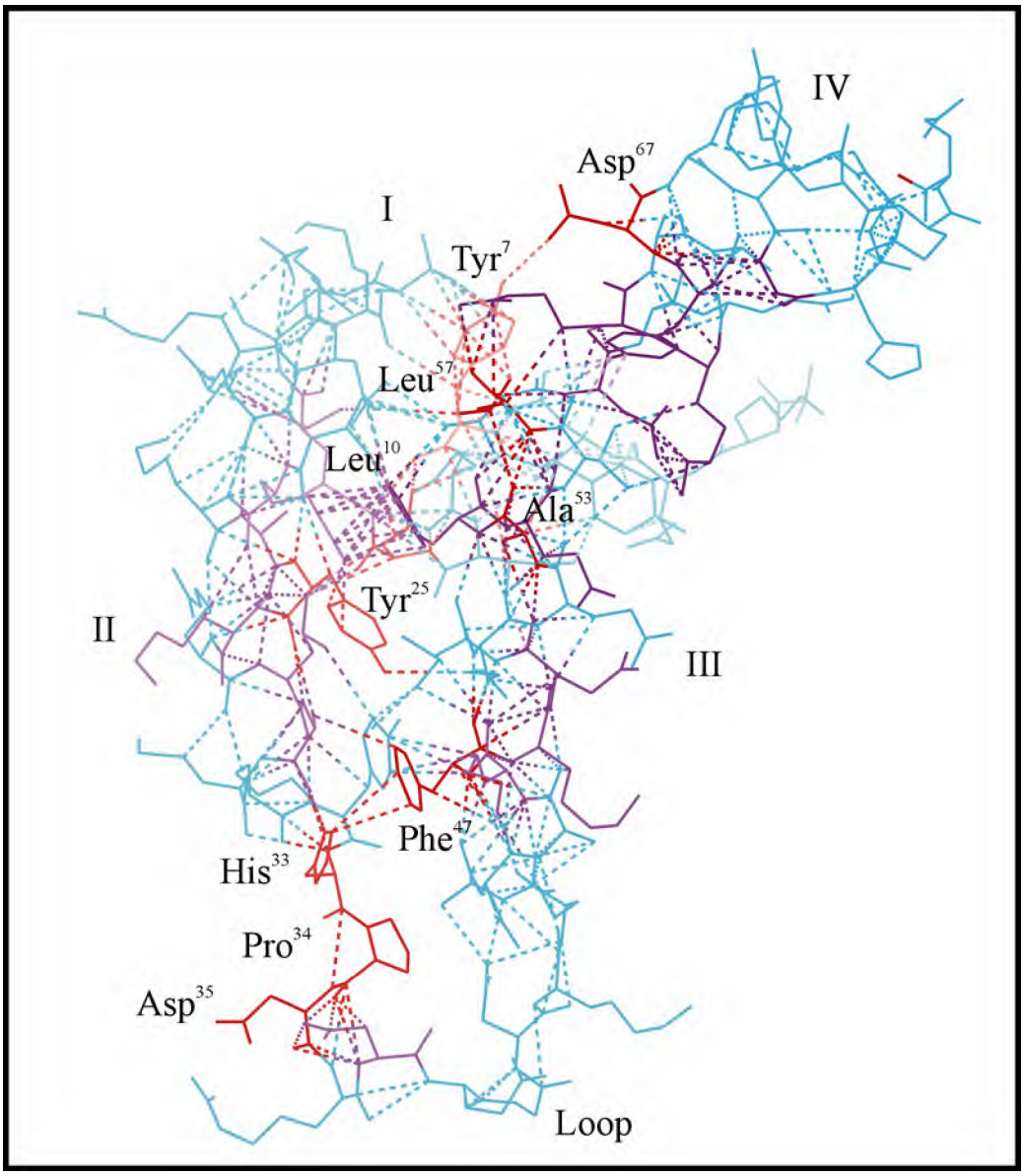

Figure 2.6: Contact analysis of the *E. coli* DnaJ J domain

Contact analysis was performed on the DnaJ J domain using the programme Whatif (Vriend, 1990). All potential contacts of a distance less than 4 Å are highlighted. The J domain is in the standard orientation (Figure 1.2 and Figure 1.3), and helices (I to IV), the loop region, and the BPD motif are labelled. Residues that occur greater than 75 % of the time in all J domains are in red and are labelled, additional residues that occur greater than 75 % of the time in Type I J domains are in purple, and the remaining residues are in blue. Numbering was done according to the *E. coli* DnaJ numbering. The data used in the generation of this figure is given in the Appendix (Section 6.1, Table 6.2).

Table 2.6: Potential interactions involving highly conserved residues in the *E. coli* DnaJ 1XBL structure

Residue	Helix	Interactions	
		Intra-Helix	**Inter-Helix**
Tyr[7]	I	Asp[5], Ile[9], Leu[10], Gly[11]	Lys[14], Thr[15], Leu[57], Asp[67]
Leu[10]	I	Tyr[6], Tyr[7], Glu[8]	Ile[21], Leu[28]
Tyr[25]	II	Ile[21], Arg[22], Tyr[23], Arg[27], Ala[29]	Lys[51]
His[33]	Loop	Ala[29], Met[30], Lys[31], Asp[35]	Phe[47]
Pro[34]	Loop	Arg[36], Asn[37]	
Asp[35]	Loop	Asn[37]	
Phe[47]	III	Ala[43], Glu[44], Ala[45], Glu[49], Ile[50], Lys[51]	Ala[29], His[33]
Ala[53]	III	Glu[49], Lys[51], Glu[55], Val[56], Leu[57]	Tyr[6], Ile[9]
Leu[57]	III	Ala[53], Tyr[54], Glu[55], Asp[59]	Tyr[6], Tyr[7], Ile[21], Arg[63]
Asp[67]	IV	Ala[64], Ala[65]	Tyr[7]

Numbering is given as for the *E. coli* DnaJ numbering; this differs by one from structural numbering given in Table 6.2, which lacked the initial methionine, and started numbering from Ala[2]. Intra-helix refers to potential interactions occurring on the same structural element; Inter-helix refers to potential interactions occurring between different structural elements.

The highly conserved residues involved in making the most potential interactions were the Tyr[7], Phe[47] and the Leu[57]. There appeared to be a network of interactions that co-ordinated the internal structure of the J domain, presumably providing a conserved structural lattice that can position conserved charged residues correctly. Pro[34] and Asp[35] made relatively few internal interactions, possibly as a result of their key function being an interaction with a partner Hsp70. The majority of the interactions are intra-helical; that is they occur between residues on the same structural element. This may be more important for maintaining the structural integrity of the J domain.

2.3.4 Relative Occurrence of Amino Acids

Analysis of the conservation of the relative occurrence of each amino acid in the database was performed (Figure 2.7). The total amount of each amino acid in all J domains was calculated, and was compared to previously defined standards (Klapper, 1977), as well as to the proportions of amino acids currently found in the Swiss-Prot database (http://us.expasy.org/sprot/relnotes/relstat.html).

Analysis of the classes of amino acids showed a distinct trend in enhancing the presence of charged amino acids, and the disfavouring of hydrophobic amino acids. There were approximately 5 % fewer hydrophobic amino acids (glycine, alanine, valine, isoleucine and leucine) in J domains than when compared to standards. Type III J domains were slightly more hydrophobic than Types I and II.

The proportions of aromatic amino acids were very similar to those seen in the published standards (Klapper, 1977), but were approximately double those seen in the Swiss-Prot database (7.2 % for the total consensus, against 4.3 % in the Swiss-Prot database). Again, there were a lower proportion of aromatic residues in Type III J domains than in Types I and II. The proportion of positively charged residues in the J domain was almost double that found in the standards used (20.6% in the overall consensus against 13.5 % in the Swiss-Prot database and 14.3 % in the Klapper standards). This also applied to the negatively charged residues, although this was more pronounced in Type I and II J domains (Type I – 20.6 %, Type II – 19.8 %, Type III – 15.5 %, Swiss-Prot – 11.8 % and Klapper standards – 11.7 %). In other words, Type III J domains were more hydrophobic, and had an overall lower negative charge than the Type I and II J domains.

The frequency of the occurrence of amino acids in the J domain was ranked and is given below in Table 2.7.

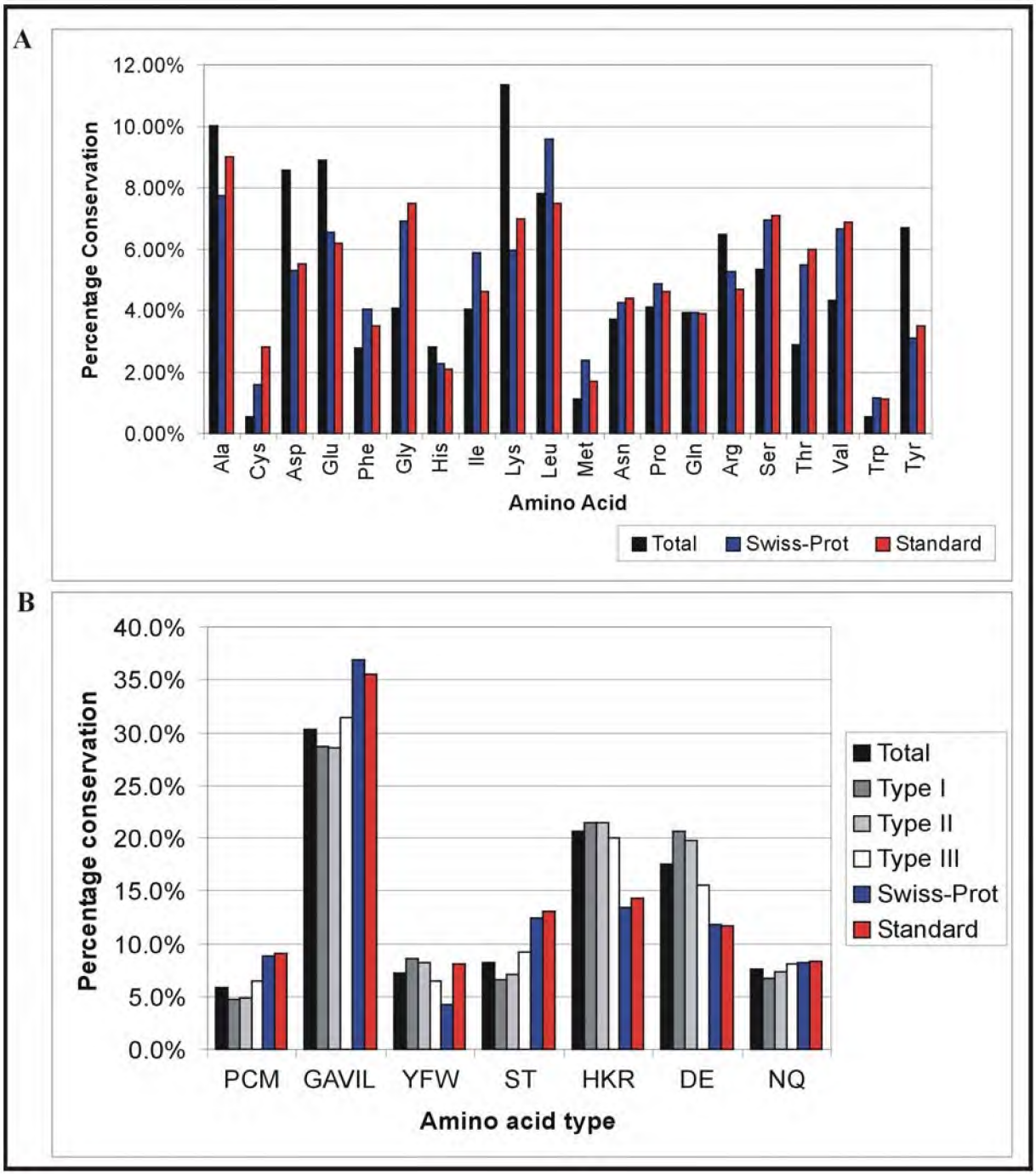

Figure 2.7: Relative occurrence of amino acids in the J domain

(A) shows the proportions of each amino acid in the overall J domain consensus compared to known standards (Standard) (Klapper, 1977), and to the percentage of each amino acid occurrence in Swiss-Prot (Swiss-Prot). Three letter amino acid code was used. (B) shows the proportion of each amino acid grouping in the overall consensus, Type I, Type II and Type III J domains in comparison to the proportions found in the standards used. Single letter amino acid code was used.

Table 2.7: Classification of the prevalence of amino acids in the J domain by their frequency

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Total	Lys	Ala	Glu	Asp	Leu	Tyr	Arg	Ser	Val	Pro	Gly	Ile	Gln	Asn	Thr	His	Phe	Met	Cys	Trp
Type I	Lys	Ala	Glu	Asp	Tyr	Arg	Leu	Gly	Ser	Val	Pro	Gln	Ile	Asn	His	Phe	Thr	Met	Cys	Trp
Type II	Lys	Ala	Asp	Glu	Tyr	Arg	Leu	Ser	Ile	Pro	Val	Gln	Gly	Asn	Phe	His	Thr	Met	Trp	Cys
Type III	Lys	Ala	Leu	Glu	Asp	Arg	Ser	Tyr	Val	Ile	Pro	Gln	Asn	Gly	Thr	His	Phe	Met	Cys	Trp
Swiss-Prot	Leu	Ala	Ser	Gly	Val	Glu	Lys	Ile	Thr	Asp	Arg	Pro	Asn	Phe	Gln	Tyr	Met	His	Cys	Trp

Amino acids were sorted on the basis of their prevalence in the J domain, where 1 is the most common amino acids, and 20 is the least common.

Lysine was the most common amino acid, followed by alanine. Types I and II then had negatively charged residues at positions 3 and 4, whereas Type III had a glutamic acid at position 4, but the hydrophobic residue leucine at position 3. This may explain the relative decrease in the proportion of negatively charged amino acids, and corresponding increase in hydrophobic amino acids in Type III J domains relative to Types I and II.

It was also possible to focus on the conservation of amino acids in each helix and loop of the J domain. This is shown in Figure 2.8. As was expected, the aliphatic and hydrophobic amino acids (glycine, alanine, valine, isoleucine and leucine) and the aromatic amino acids (tyrosine, phenylalanine and tryptophan) were selected against in the loop region, and proline was selected for (HPD motif). However the significant differences were in the positively and negatively charged residues. The proportion of negatively charged to positively charged residues in Helices III and IV and the loop region was approximately equal for all J domains. However, there were markedly more positively charged residues in Helix II than negative, and more negatively charged residues in Helix I than positive. In Type I J domains, the situation differed slightly, as the proportion of negatively charged residues in Helix III increased markedly in relation to positively charged residues. There was also a marked diminishment of hydrophobic amino acids in Helix IV.

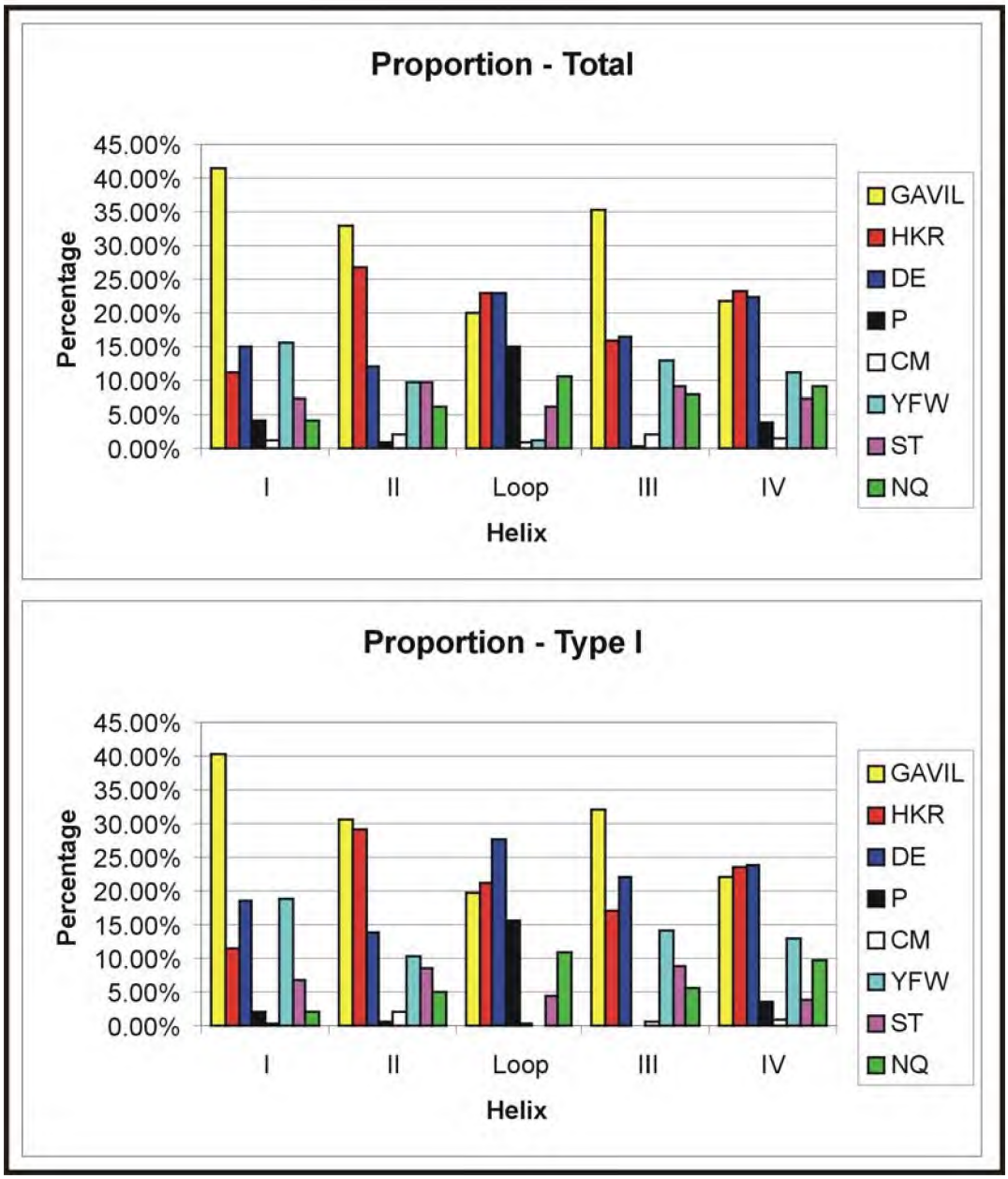

Figure 2.8: Comparison of amino acid composition in each helix of the J domain

The percentage composition of amino acids in the J domain seen in Figure 2.7b was subdivided to give the composition of each amino acid type in each helix. The total proportions are contrasted to those in Type I J domains. Single letter amino acid code was used.

There were also slight changes in the proportions of amino acid types in each helix between all J domains and Type I J domains (Table 2.8). The most significant difference was in the proportion of negatively charged amino acids in Helix III and the loop region, where there was an approximately 5 % change between total J domains and Type I J domains.

Table 2.8: Difference in percentage composition of amino acid type between total J domains and Type I J domains

	P	CM	GAVIL	YFW	ST	HKR	DE	NQ
I	2.16%	1.08%	1.17%	-3.17%	0.45%	-0.40%	-3.52%	2.22%
II	0.27%	0.08%	2.33%	-0.50%	1.10%	-2.54%	-1.66%	0.91%
Loop	-0.60%	0.46%	0.28%	1.02%	1.83%	1.85%	-4.51%	-0.33%
III	0.15%	1.35%	3.15%	-0.99%	0.44%	-1.05%	-5.40%	2.34%
IV	0.19%	0.63%	-0.18%	-1.64%	3.54%	-0.43%	-1.71%	-0.41%

A positive change indicates a higher proportion of that group of amino acids in total J domains; a negative change indicates a higher proportion of that group of amino acids in Type I J domains. Changes of greater than 1.5 % are shaded.

Negatively charged residues appeared to be more prevalent in Type I J domains than in total J domains, consistent with the apparent increased hydrophobicity of Type III J domains.

2.3.5 Length and conservation of the loop region

The length of the loop region was calculated. The length was calculated from the amino acid after the HPDK region, to one residue prior to the previously defined KFK motif (Hennessy *et al.*, 2000). Figure 2.9 shows the variable nature of the length of the loop region. The vast majority of J domains had a loop region of between 9 to 11 amino acids in length; however, this was highly variable and could range from two amino acids to 30. The mean lengths of the loop region were 9.25 amino acids for Type I J domains, 9.81 amino acids for Type II J domains, and 11.37 amino acids for Type III J domains. As expected the majority of the variance occurred in Type III J domains. The reason for the variability is as yet unclear.

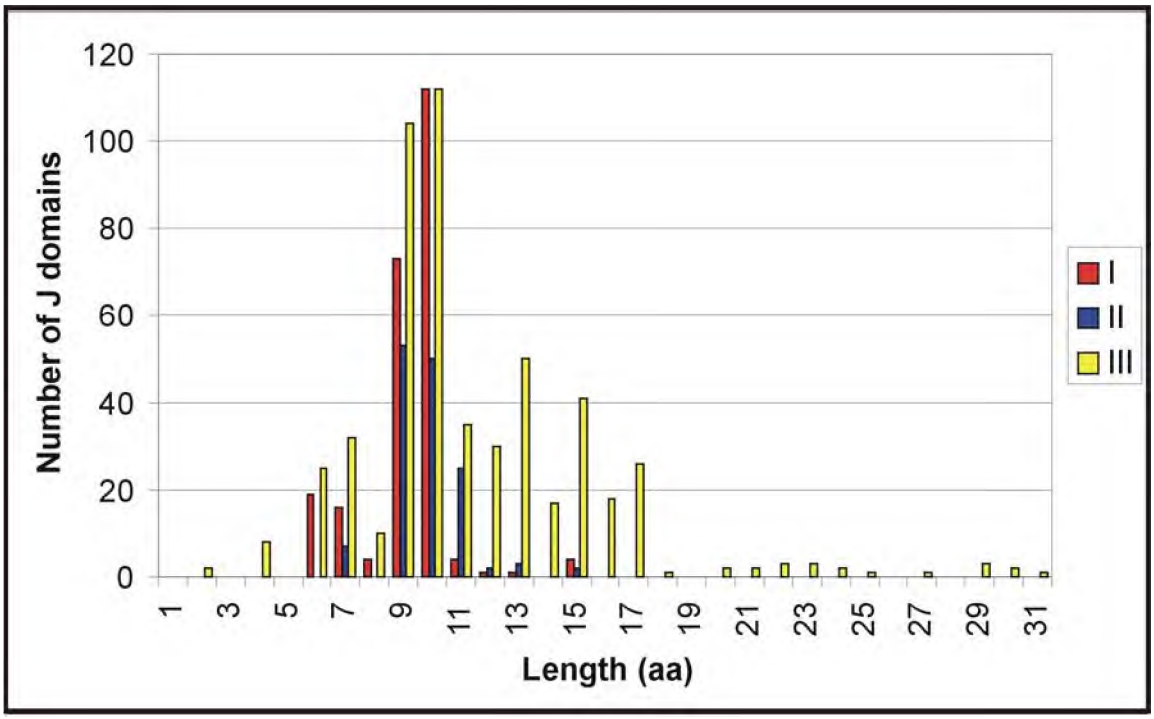

Figure 2.9: Analysis of the length of the loop region in DnaJ-like proteins

The length of the loop region was determined for each type of J domain in order to determine the average length for this region. The loop region was calculated as being from immediately after the HPDK region to one residue before the previously defined KFK motif.

2.3.6 Non-HPD containing J-like domains

J-like domains that do not contain a HPD motif were also identified during the course of this work. These included the RESA proteins from *P. falciparum* (Bork *et al.*, 1991; Cappai *et al.*, 1992) and the DjlB/DjlC proteins from *E. coli* (Kluck *et al.*, 2002). Thirty six proteins in total were identified (4.74 % of total alignment). A comparison between this alignment and the total consensus alignment is shown below (
Figure 2.10).

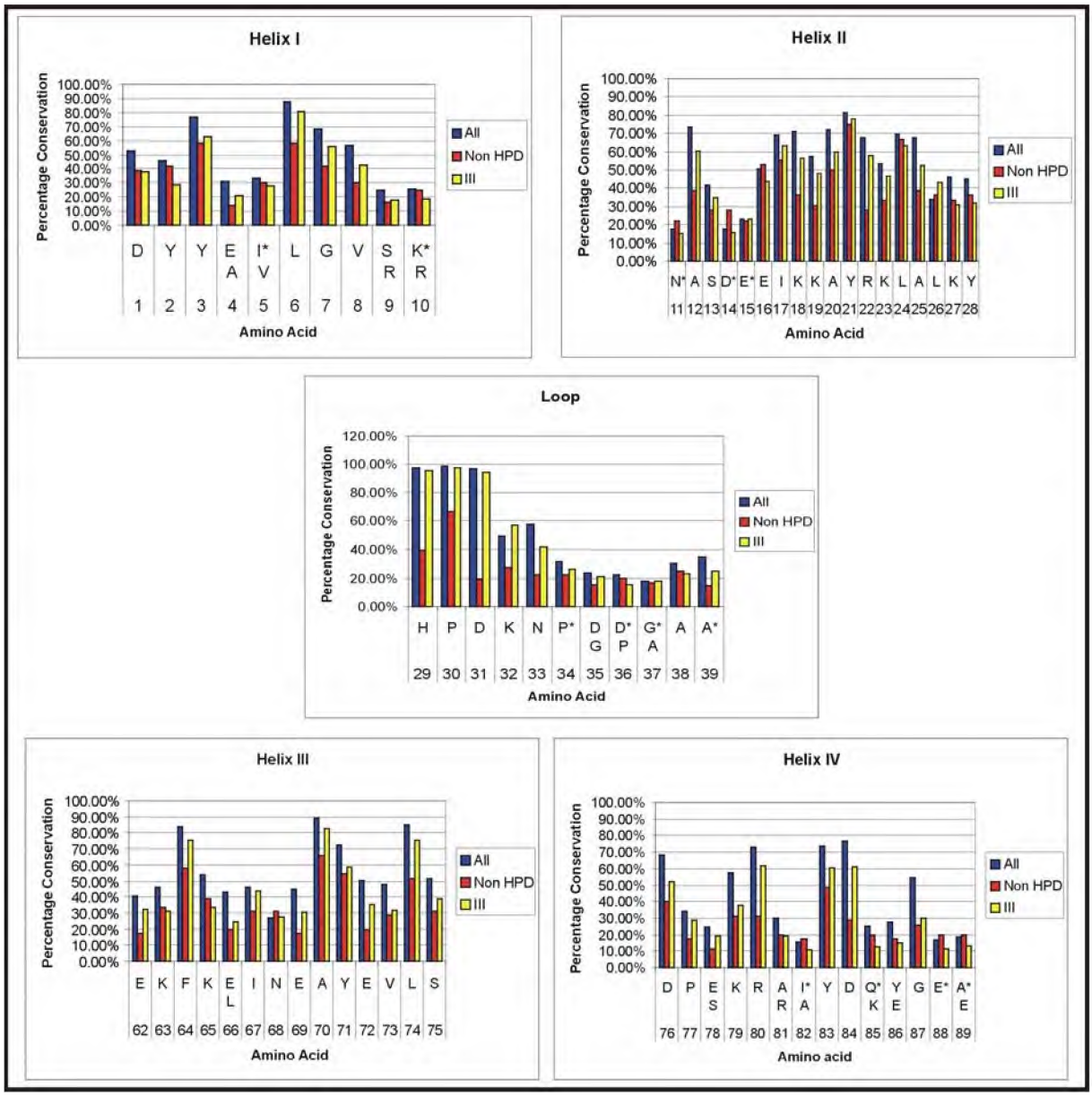

Figure 2.10: Comparison of the overall consensus sequence against a consensus sequence for non-HPD containing J-like domains

The overall consensus sequence (all) is shown in blue, the Type III sequence (III) in yellow and the non-BPD sequence (Non-BPD) in red. Where two amino acids are shown, the first is the amino acid at that position in the overall consensus and the second is the amino acid at that position in the non-BPD J-like domains. An asterisk indicates variability at that position, in terms of the amino acid present, in the other consensus sequences (Figure 2.1 and Table 2.5).

Generally, most of the conserved residues present in the total alignment were still conserved in these domains, although at lower conservation levels. The highest level of variation occurred in Helix IV, where even amino acids conserved in the overall alignment were shown to be poorly conserved. In general, conserved hydrophobic residues (Tyr^{21}, Leu^{24}, Phe^{64}, Ala^{70}, Tyr^{71}, Leu^{74}, consensus numbering) retained conservation to a higher degree than charged residues (Arg^{22}, Lys^{23}, Lys^{63}, Lys^{65}).

2.3.7 Degenerate J domains

The discovery that there are J domains that do not contain an HPD motif implied the possibility that there are J-like domains, that will retain conserved hydrophobic residues, but may not be recognised as a true J domain. In order to determine if the J domain could have been sequestered over time to fulfil other functions, the consensus sequence was used as a probe to look for similar sequences in the PDB database. Structures that bore some similarity at the amino acid level were also compared at the tertiary level. One structure was identified that merited further analysis. This was that of the DNA binding domain of the *D. melanogaster* Dead Ringer protein (Iwahara and Clubb, 1999; Iwahara *et al.*, 2002). The structure of this protein is shown in Figure 2.11.

As expected, many of the conserved hydrophobic residues are retained in this structure (Tyr^{3}, Leu^{6}, Tyr^{71}, Leu^{74}; consensus numbering), and some charged residues, in particular the first pair of lysines in Helix II (Lys^{18}, Lys^{19}) and the conserved aspartic acid in Helix IV (Asp^{84}). Despite the retention of these amino acids, it was not possible to superimpose this structure on the DnaJ J domain structure. Other amino acids have altered the structure, such that the basic finger-like structure has been retained, but with a twist in Helix III. This will presumably abrogate any potential interaction with a partner Hsp70, preventing this protein acting as a DnaJ-like co-chaperone.

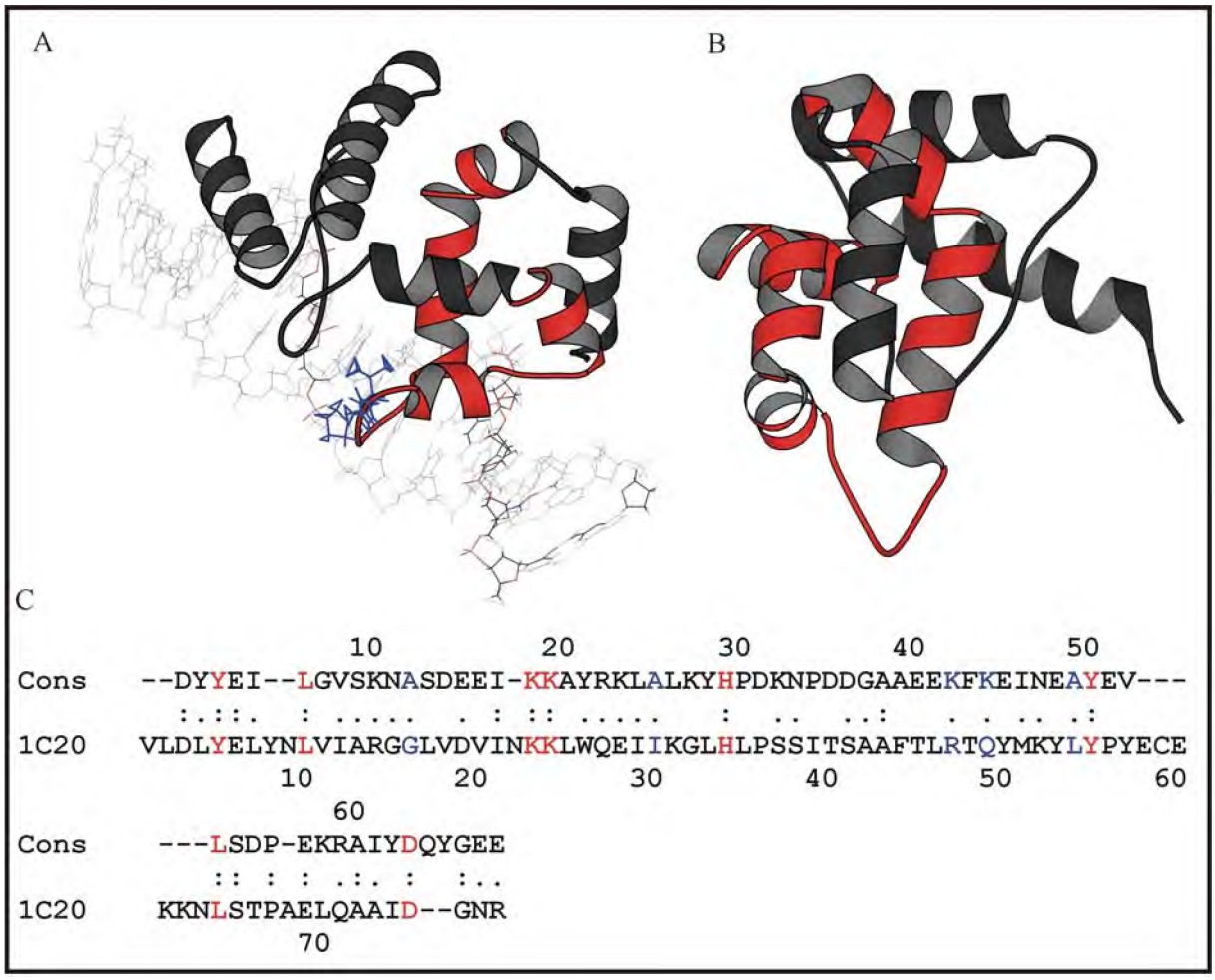

Figure 2.11: Analysis of the Dead Ringer protein from *D. melanogaster*

(A) Ribbon representation of the 1KQQ structure (Iwahara *et al.*, 2002); (B) Ribbon representation of the 1C20 structure (Iwahara and Clubb, 1999). The J domain-like region is in red. The HPD-like motif is depicted in (A) in purple. (C) Pairwise alignment of the amino acid sequence of the dead ringer protein (1C20) and the J domain consensus sequence (cons). Highly conserved residues from the consensus that are identical in this alignment are depicted in red, and highly conserved residues from the consensus that are similar in this alignment are depicted in blue.

2.3.8 Analysis of the loop region between Helices I and II

When focusing on the primary amino acid sequence, it became apparent that there was a potential Hsp70 peptide recognition motif between Helix I and Helix II. Analysis of amino acids in this region is shown in Figure 2.12.

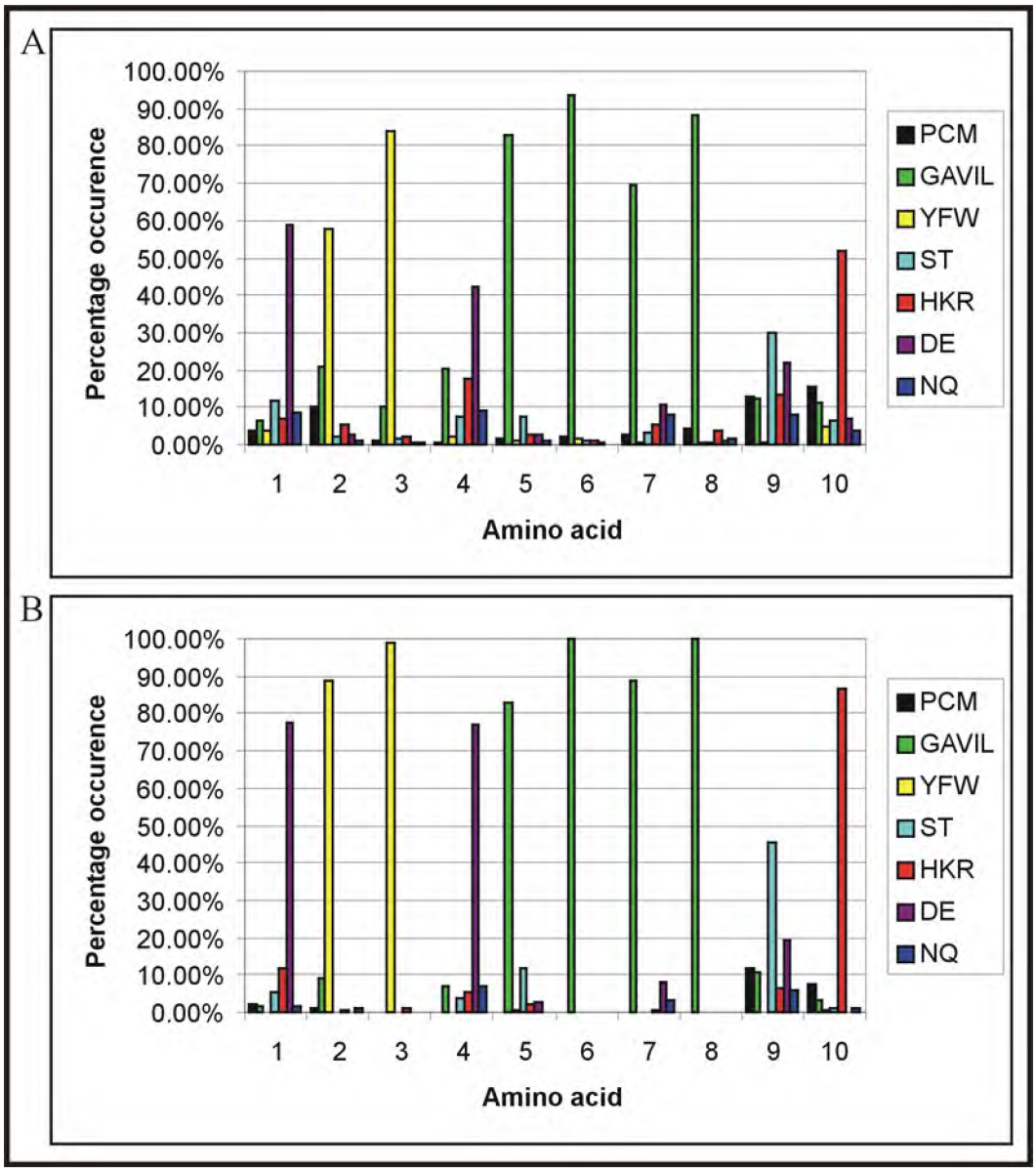

Figure 2.12: Proportions of amino acids in Helix I and the subsequent loop region

Proportions of amino acids in each of the first ten consensus residues were analysed. (A) shows amino acids in this region in the overall consensus sequence. (B) shows amino acids in this region in Type I J domains.

There was a strongly hydrophobic patch at the end of Helix I leading into Helix II. This contained hydrophobic residues including an isoleucine, leucine and valine (amino acids 5, 7 and 8) and a pair of aromatic residues at positions 2 and 3. A flanking positive charge was found at position 10. There was however a negative charge located at position 4. This bore some similarity to the Hsp70 peptide recognition motif - X-Hy-(W/X)-Hy-X-Hy-X-Hy, which was often flanked by positive charged residues, with negative residues being disfavoured.

2.4 DISCUSSION

The purpose of this alignment was to provide a mechanism by which conserved amino acids could easily be identified, and provide a rational basis for mutagenesis of the J domain.

2.4.1 Generation of a multiple sequence alignment of J domains

Despite the increase in size of the alignment, the overall level of conservation of Type I and Type II J domains remained consistent, but decreased in the Type III J domains. Type III J domains generally had far lower levels of conservation than Type I and II J domains (Hennessy *et al.*, 2000). These J domains are likely to be involved in highly specialised functions, and are there to sequester specific Hsp70s to play specific cellular functions. Type I J domains can interact with more Hsp70s leading to the conservation of those general residues involved in this interaction. Type III J domains are more likely to be able to interact with one specific Hsp70, hence they are likely to have evolved residues that will enhance the specificity of that interaction.

Some residues were shown to be important in J domain function in *E. coli* DnaJ; these include Tyr^{25}, Lys^{26} and Phe^{46} (Genevaux *et al.*, 2002). All of these residues were shown here to be conserved across J domains from the various types of DnaJ-like proteins, which implied a conserved mechanism of action. This mechanism of action was potentially due to the highly charged nature of Helix II of the J domain, as well as the highly conserved nature of the backbone structure of the J domain (Figure 2.4). Other residues implicated in DnaJ function included the conserved lysine-arginine dipeptide in Helix IV which was initially identified as being part of a QKRAA motif (Auger and Roudier, 1997). Expression of certain motifs on human leukocyte antigens have been shown to confer susceptibility to specific diseases, and this motif was shown to promote susceptibility to rheumatoid arthritis, and has been shown to be a motif recognised by DnaK (Auger and Roudier, 1997).

The majority of poorly conserved residues occurred on Helices I and IV, which implied interaction between these residues and the remainder of the DnaJ-like protein. Interestingly there were highly conserved residues located on Helix IV. This could imply that Helix IV was making conserved contacts within the J domain, that Helix IV was making interactions

with a partner Hsp70, or that conserved residues on the remainder of the DnaJ-like protein, regardless of type, must be retained in order to allow for correct functioning of the protein. Due to the wide variability of the remaining portions of DnaJ-like proteins, the first two theories are more likely than the third.

2.4.2 A network of interactions allows for correct J domain structure

There appeared to be a complex network of interactions at the top of the J domain, involving residues in all four helices, particularly conserved residues. There appeared to be far fewer contacts towards the tip of the J domain in the region of the HPD motif. This is likely reflecting the presence of more hydrophobic residues in the core of the J domain at the top of Helices II and III, and the more hydrophilic residues present in the loop region. The tip of the J domain appeared to be mobile, and not structurally restrained. The close network of interactions between Leu^{57}, Ala^{53} and Tyr^{7} (*E. coli* numbering) are likely to be the key residues in maintaining the structural integrity of the J domain. The proline and the aspartic acid in the DnaJ J domain structure appeared to make very few internal interactions. These two residues were important for interacting with Hsp70s, and were unlikely therefore, to be playing a structural in the J domain. The conserved charged residues made fewer interactions within the J domain, consistent with these residues forming potential interactions with partner Hsp70s. The conserved nature of the backbone structure of the J domain led to the supposition that several highly conserved residues are critical in conserving the finger-like structure of the domain, thereby allowing the formation of a correct scaffold for the positioning of the charged residues which interact with partner Hsp70s. The majority of the interactions are intra-helical. This may imply that the helices of the J domain can attain structural integrity in the absence of additional J domain structural elements. These intra-helical interactions will allow for the correct formation of the α-helical elements in the J domain, before the J domain itself is correctly formed. Subsequent to the formation of the α-helices, the inter-helical interactions can happen, allowing for a correct J domain structure.

2.4.3 Charge plays a key role in specifying DnaJ-like – Hsp70 interactions

The role that charge plays in determining the structure and function of the J domain was apparent from investigation the proportions of charged residues in the J domain. A comparison of the proportions of amino acids in the J domain was performed against known standards. Positively and negatively charged residues, both individually and as a group, showed the greatest increase with respect to the standards (Klapper, 1977), whereas hydrophobic and polar residues such as leucine, threonine, serine and valine showed a marked decrease. It is possible that the level of charge seen in the J domain will play a vital role in determining whether a specific DnaJ-like protein will interact with a specifi c Hsp70.

Helix II was predominately positively charged, presumably allowing for interactions with the ATPase domain of Hsp70s, whereas Helix III had equal proportions of negative and positive charged residues. Helix IV was likely to be solvent exposed as greater than 50 % of its residues were charged; this was in comparison to Helix I where almost 60 % of the residues were hydrophobic or aromatic. This could imply a more functional role for Helix IV, consistent with structural data (Pellechia *et al.*, 1996), which showed the highly mobile nature of this region, whereas Helix I is forming the remaining portion of the structural lattice. There was also a distinct drop in the proportion of hydrophobic amino acids in the J domain, as compared to known standards, and a corresponding increase in charged residues. Affecting the charge ratio of the J domain may be the defining method that specifies particular DnaJ-Hsp70 interactions.

2.4.4 Non-HPD J domains

The existence of J domains that do not contain HPD motifs could imply that these J-like domains retain the overall J domain fold, but have developed altered functions, which do not necessarily include interaction with a partner Hsp70. Alternatively, a partner Hsp70 could have evolved in tandem with a DnaJ-like protein, causing shifts in amino acid sequence. An example of this could be the HscC and DjlB/DjlC family in *E. coli* (Kluck *et al.*, 2002). The possibility also arises therefore that a J domain containing protein has evolved its J domain to such an extent that it will no longer be recognised as a J domain, but re tains the overall

tertiary fold, as well as conserved amino acids. Such an event would not be without precedent in the chaperone familie s; the ATPase domains of Hsp70s bear marked structural resemblance to the hexokinase structure, despite low levels of sequence similarity, and dissimilar functions (Bork *et al.*, 1992). The conservation of hydrophobic amino acids in preference to charged amino acids as seen in the Type III J domains may also imply a variation in the mechanism of interaction with partner Hsp70s.

A potential J-like domain that has a similar finger-like structure to the J domain, while retaining significant structural residues was identified. If this domain is a relic of an ancestral J domain, it could provide credence for the proposal that J domains are sequestered for additional functions not always related to general chaperone functioning.

2.4.5 Analysis of the loop region between Helices I and II

The presence of a strongly hydrophobic region in Helix I and the subsequent loop region was reminiscent of an Hsp70 binding site. The major difference was the presence of a negative charged residue in this region. This region may be recognised by a partner Hsp70, *via* its substrate binding region, but not as strongly as a peptide substrate that does not contain negatively charged residues. This may imply that the multiple determinants mediating DnaJ-DnaK interaction are all located on the J domain. It may also explain why DnaJ was unable to stimulate the ATPase activity of the ATPase domain of DnaK, and why the DnaK ATPase domain on its own was incapable of binding to DnaJ (Gässler *et al.*, 1998).

Hence, a slightly revised mechanism of interaction of DnaJ-like proteins with Hsp70s may be proposed: (i) the HPD motif and Helix II interact with the underside cleft on the ATPase domain (Figure 1.6); (ii) the hydrophobic loop separating Helices I and II interacts with the substrate binding region in the absence of a peptide substrate; (iii) The α-helical lid region is unable to lock onto the hydrophobic loop due to the presence of the rest of the J domain; (iv) the presence of the tyrosine residues (positions 2 and 3 in Figure 2.12) and the negatively charged residues (positions 1 and 4) will disfavour binding, and should allow easy displacement of the loop in the presence of substrate or during presentation of correct substrate protein by the DnaJ. This suggestion is hypothetical at present, particularly in the absence of structural data, which may confirm or falsify such a hypothesis.

2.4.6 Conclusion

This section has provided a mechanism for the rational choice of amino acids for future mutagenic analysis of the J domain. The work in the rest of this thesis will focus on two different methods of determining residues important for the interaction of DnaJ-ike proteins with partner Hsp70s. A model system was established using the DnaJ-like proteins present in *Agrobacterium tumefaciens*, and the ability of these proteins to interact with the *E. coli* protein DnaK was assessed. The technique focused on conserved amino acids present in the J domain, and rational site-directed mutagenesis was performed on these residues to investigate their effect on J domain function.

Chapter 3: Isolation and characterisation of the DnaJ-like protein s from *A. tumefaciens*

3 ISOLATION AND CHARACTERISATION OF DNAJ-LIKE PROTEINS FROM *AGROBACTERIUM TUMEFACIENS*

3.1 INTRODUCTION

3.1.1 Importance of *A. tumefaciens*

Agrobacterium tumefaciens is an economically important plant pathogen, whose genome sequence has recently been released (Goodner *et al.*, 2001; Wood *et al.*, 2001). *A. tumefaciens* is instrumental in the formation of Crown Gall disease in plants which causes the formation of invasive tumours in the plant. Certain strains of *A. tumefaciens,* such as *A. tumefaciens* RUOR have been found to be of industrial importance for the over-production of the commercially important hydantoinase enzymes that are involved in the synthesis of optically pure amino acids (Hartley *et al.*, 1998; Burton *et al.*, 1998; Hartley *et al.*, 2001). This thesis makes use of the *A. tumefaciens* RUOR strain (Hartley *et al.*, 1998; Burton *et al.*, 1998).

3.1.2 Usage of *E. coli* DnaK/DnaJ as a model system

The majority of work that has been done on defined DnaK-DnaJ systems at a prokaryotic level has involved analysis of the *E. coli* system (Liberek *et al.*, 1991; Langer *et al.*, 1992; Gamer *et al.*, 1996; Banecki and Zylicz, 1996; Gässler *et al.*, 1998; Rüdiger *et al.*, 2001; Han and Christen, 2003). Minimal work has been performed on other prokaryotic systems (Tilly *et al.*, 1993; Motohashi *et al.*, 1994; Zuber *et al.*, 1995; Motohashi *et al.*, 1996). However, it is still not clear as to how far conclusions made using this system can be extrapolated to other systems. For example, the DnaK and DnaJ proteins from *Thermus thermophilus* appears to form a stable complex containing three molecules each of DnaK and DnaJ, as well as three molecules of an assembly factor (Motohashi *et al.*, 1994; Motohashi *et al.*, 1996). This appears to be different to the *E. coli* system. Equally the DnaK homologue from *Borrelia burgdorferi* was unable to complement for lack of *E. coli* DnaK, whereas the DnaJ was able to complement for lack of *E. coli* DnaJ (Tilly *et al.*, 1993). This might imply that the *B.*

burgdorferi DnaK was incapable of interacting with the *E. coli* DnaJ, or that it was incapable of interacting correctly with *E. coli* protein substrates.

The full sequence of the *dnaK* gene and a partial sequence of the *dnaJ* gene from *A. tumefaciens* C58 were released previously (Segal and Ron, 1995), followed by the genome sequence (Goodner *et al.*, 2001; Wood *et al.*, 2001) and consequently it was decided to use this system as a model system in this thesis, as a n alternative to the *E. coli* system. An attempt was made to isolate the coding regions for all DnaJ-like proteins from *A. tumefaciens* and the ability of the resultant proteins to interact with partner Hsp70s was investigated. *In vivo* and *in vitro* techniques were employed to investigate the functional ability of the *Agt* DnaJ-like proteins, and *Agt* DnaJ chimeric proteins. These proteins were assayed for their ability to complement for lack o f *E. coli* DnaJ and CbpA, and for their ability to stimulate the ATPase activity of *Agt* DnaK.

3.1.3 Aims and objectives

The aim of this section of work was to institute a suitable model system for investigating conserved amino acids, and their role in J domain functioning. As explained above, it was decided to focus on the system from *A. tumefaciens* as an alternative to the *E. coli* DnaK/DnaJ system . In *E. coli* DnaK, DnaJ and GrpE form a close, functional association, and are encoded in a single operon. In *A. tumefaciens* the genes encoding DnaK and DnaJ are located on a single operon implying the close interaction between these two proteins, while there is perhaps a different association with GrpE which is encoded elsewehere on the genome. Secondly, an analysis of additional DnaJ-like proteins present in *A. tumefaciens* was undertaken. Isolating the coding regions for additional DnaJ-like proteins would allow experimental analysis of the interaction of these proteins with Hsp70s. Hence coding regions for DnaJ-like proteins from *A. tumefaciens* were amplified and inserted into suitable expression vectors, allowing for future experimental analysis. The DnaJ homologue was termed *Agt* DnaJ, coded for by *Agt dnaJ.* Naming of open reading frames amplified for Type III DnaJ-like proteins present in *A. tumefaciens*, and the proteins coded for by these open reading frames was done by abbreviating *Agt* for the organism name, Dj for DnaJ-like protein and C for Type III protein, and proteins were numbered consecutively. Hence *Agt* DjC1 is the first Type III DnaJ-like protein described in *A. tumefaciens*, and is coded for by *Agt djC1.*

The nomenclature is based on the method described by Ohtsuka and Hata, (2000). The nomenclature for the open reading frames amplified, and the proteins coded for by these open reading frames is given in Table 3.1.

Table 3.1: Nomenclature for the open reading frames encoding DnaJ-like proteins present in *A. tumefaciens* RUOR, and the resultant protein products

Gene	Protein	Type
Agt dnaJ	*Agt* DnaJ	I
Agt djC1	*Agt* DjC1	III
Agt djC2	*Agt* DjC2	III
Agt djC3	*Agt* DjC3	III
Agt djC4	*Agt* DjC4	III
Agt djC5	*Agt* DjC5	III

Subsequently, the ability of the *Agt* DnaJ-like proteins, and *Agt* DnaJ chimeric proteins to complement for lack of *E. coli* DnaJ and CbpA was investigated. The basal enzyme kinetics of *Agt* DnaK, and the effect of *Agt* DnaJ and *Agt* DnaJ-H33Q on the basal enzyme activity of *Agt* DnaK were also investigated.

3.2 EXPERIMENTAL PROCEDURES

Standard molecular biology techniques were used throughout this thesis (Sambrook and Russell, 2001). Common techniques are described in the Appendix (Sections 6.4.1 and 6.4.2).

3.2.1 Polymerase chain reaction (PCR) amplification of the *A. tumefaciens dnaJ* (*Agt dnaJ*) coding region

Primers were designed to amplify the coding region for the DnaJ protein from the *A. tumefaciens* strain RUOR. The primers were based on the *A. tumefaciens* C58 genomic sequence and the previously described 5' end of the coding region (Segal and Ron, 1995; Goodner *et al.*, 2001; Wood *et al.*, 2001). A degenerate primer was designed for the 3' end of the sequence based on other members of the Rhizobium family (Figure 3.1).

```
                    :  *  .  *  *  :
B.ovis               R  M  K  E  F  F  E  G  I  G  E  &
B.ovis              CGGATGAAGGAGTTTTTTGAAGGAATCGGCGAATAG---

M.loti               R  M  K  D  F  F  E  S  F  G  E  R  &
M.loti              CGCATGAAGGATTTTTTCGAGTCTTTCGGCGAGCGTTGA

R.fredii             R  M  K  D  F  F  D  T  L  S  E  &
R.fredii            CGCATGAAAGATTTCTTCGATACACTGAGCGAATAG---

A.tumefaciens        R  M  K  E  F  F  D  G  &
A.tumefaciens       CGGATGAAGGAATTCTTCGACGGCTGA------------

E.coli               G  V  K  K  F  F  D  D  L  T  R  &
E.coli              GGTGTGAAGAAGTTTTTTGACGACCTGACCCGCTAA---
                     *  ****  * ** ** **

          5'        CGGATGYAYYAATTCTTTGASGGCTGA  3'

Primer    5'        GTC GAC TCA GCC STC AAA GAA TTY YTY CAT CCG  3'
```

Figure 3.1: Design of the AtDnaJR primer

The AtDnaJR primer was designed using a multiple sequence alignment of the DnaJ proteins from *Brucella ovis* (DNAJ_BRUOV), *E. coli* (DNAJ_ECOLI), *Mesorhizobium loti* (NP 105553) and *Rhizobium fredii* (NolC; NOLC _RHIFR) and the corresponding nucleotide sequences. There appears to be a frame shift error in the published protein sequence for NolC, hence the nucleotide sequence at the 3' end was translated separately to give an alternative translation. Asterisks indicate conservation of the nucleotide (below the alignment) and protein (above the alignment) sequences. The primer sequence at the bottom is complementary to the other sequences, and includes a *Sal* I site (in italics). Conserved amino acids are underlined. Y represents A or G; S represents C or G.

The degeneracy was incorporated here because of the higher levels of difference in the 3' region of DnaJ – like proteins as opposed to the 5' region. The primer was also designed prior to the release of the *A. tumefaciens* C58 genome sequence, hence no 3' sequence data was available. The primers used were AtDnaJF and AtDnaJR (see Table 6.5, Section 6.2 in the Appendix for sequences). AtDnaJF included a *Bam* HI site, and AtDnaJR included a *Sal* I site to allow for insertion into the plasmid pQE30 (Qiagen, USA). Primers were synthesised by Integrated DNA Technologies (IDT; Iowa, USA). Amplification was done on genomic DNA isolated from *A. tumefaciens* RUOR kindly provided by Dr Carol Hartley and Ms Mez Jiwaji, (Department of Biochemistry, Microbiology and Biotechnology, Rhodes University). The Expand High Fidelity PCR kit (Roche, Germany) was used for amplification of the coding region. Standard PCR techniques were used. Final primer concentrations used were 1 μM. The parameters used were the following: an initial step of 1 minute at 95°C, followed by 25 cycles of 1 minute at 95°C, 1 minute at 55°C, 1 minute at 72°C, followed by a final hold at 72°C for 5 minutes.

3.2.2 Insertion of the *Agt dnaJ* coding region into pGEM-T Easy

Following successful amplification of the coding region, the product was ligated into the pGEM-T Easy plasmid (Promega, USA) using the manufacturers protocols. The standard protocol for insertion into pGEM was followed as described in the Appendix (Section 6.4.1.7). Plasmid DNA was recovered from colonies generated after transformation into *E. coli* XL1-Blue, and confirmation of the presence of an insert was performed with by restricting plasmid DNA with *Eco* RI, which released inserts corresponding in size to the expected *Agt dnaJ* coding region. Constructs were termed pGEM-RJ for *A. tumefaciens* RUOR *dnaJ* in pGEM. A plasmid map of pGEM-T Easy is given in the Appendix (Section 6.2, Figure 6.1)

3.2.3 Insertion of the *Agt dnaJ* coding region into pQE30

A single colony that contained plasmid DNA with an insert of the expected size was grown overnight in YT (yeast-tryptone) broth containing 100 μg/ml ampicillin, and the plasmid DNA was isolated using the Qiagen miniprep kit (Qiagen, USA) (Appendix, Section 6.4.1.1). This DNA was restricted using *Bam* HI and *Sal* I, and the DNA fragment released was ligated

into pQE30 that had also been restricted with *Bam* HI and *Sal* I (Appendix, Section 6.4.1.6). The construct was termed pRJ30 for *dnaJ* coding region from *A. tumefaciens* RUOR in pQE30. The plasmid map of the resultant plasmid is shown below in Figure 3.2. A plasmid map of pQE30 is given in the Appendix (Section 6.2, Figure 6.2). The nucleotide sequence of *Agt dnaJ* has been deposited into GenBank with accession number AY494599.

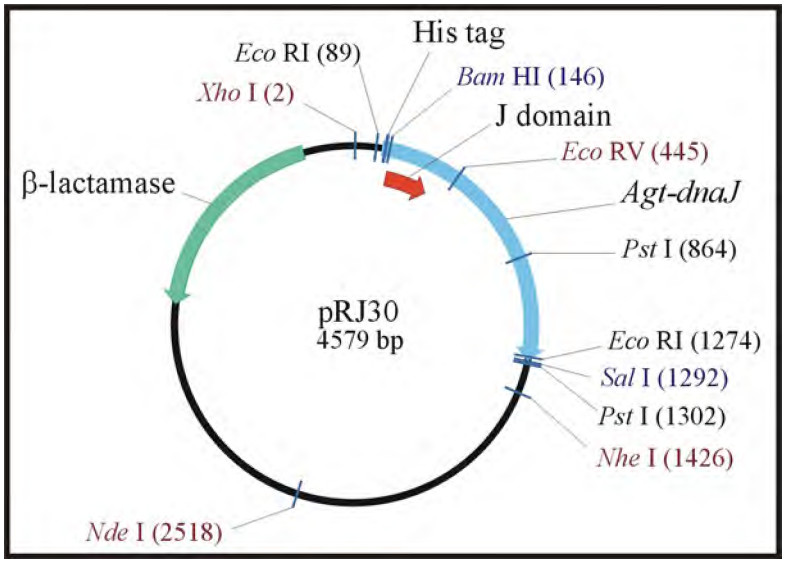

Figure 3.2: Plasmid map of pRJ30

The *Agt dnaJ* coding region is shown in cyan, and its J domain region is shown in red. The *Bam* HI and *Sal* I sites are given in blue, unique restriction sites are given in red, and multiple restriction sites are given in black. The gene encoding β-lactamase, which allows ampicillin resistance, is shown in green. The map was generated using Vector NTI (Informax, USA).

3.2.4 Alteration of the codon for His^{33} to a codon for glutamine

His^{33} in the HPD motif of DnaJ was replaced by a glutamine in order to give a negative control for future work (Tsai and Douglas, 1996). Site directed mutagenesis was performed using the Stratagene QuikChange Site Directed Mutagenesis Kit (Stratagene, USA). Standard protocols were followed, and are given in the Appendix (Section 6.4.1.9). The primers used were DH33QF and DH33QR, and the sequences of these are given in the Appendix (Section 6.2, Table 6.5).

Plasmid DNA from the resultant transformant cells was isolated and restricted with the appropriate restriction endonuclease to confirm the presence of the substitution. All derived constructs were sequenced to ensure the absence of second site mutations.

3.2.5 Amplification of coding regions for DnaJ-like proteins in *A. tumefaciens*

Additional J domain containing proteins in *A. tumefaciens* were identified by using BLAST to search the *A. tumefaciens* genome, using the *Agt* DnaJ J domain sequence as a probe. Five Type III proteins were identified. Further BLAST searches were performed to establish the presence of coding regions, if any, for homologues of CbpA, Hsc20 and DjlB.

Primers for the amplification of coding regions for DnaJ-like proteins from *A. tumefaciens* were then designed, based on the *A. tumefaciens* C58 genome sequence. Primers used are given in Table 6.5. Forward primers contained a *Bam* HI or *Bgl* II site, and the reverse primers contained a *Sal* I or a *Xho* I site. These sites were used to insert the coding regions into pQE30. Primers used for the amplification of *Agt djC1* were C1F and C1R, for *Agt djC2* were C2F and C2R, for *Agt djC3* were C3F and C3R, for *Agt djC4* were C4F and C4R, and for *Agt djC5* were C5F and C5R. Primer sequences are given in the Appendix (Section 6.2, Table 6.5).

3.2.5.1 Isolation of *A. tumefaciens* RUOR genomic DNA

A. tumefaciens strain RUOR was a kind gift of Ms. Mez Jiwaji (Department of Biochemistry and Microbiology, Rhodes University). The strain was grown at 20°C in 5 ml YT broth until stationary phase was achieved. The cells were pelleted at top speed (16 000 x g) in a microcentrifuge for two minutes. The cell pellet was re-suspended in TE buffer (567 µl), and SDS (0.5 % final concentration) and Proteinase K (0.1 mg/ml final concentration) were added. The mixture was incubated at 37°C for one hour. This was followed by the addition of NaCl (0.07 M final concentration), followed by thorough mixing. CTAB (Hexadecyltrimethyl-ammonium bromide) made in NaCl was then added (1 % final concentration) and the mixture was incubated at 65°C for 10 minutes. An equal volume of chloroform/isoamyl alcohol (24:1) was added. The mixture was centrifuged for 5 minutes at

maximum speed. The aqueous supernatant was removed, and an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added to this. The mixture was centrifuged for 5 minutes at maximum speed (16 000 x g) in a microcentrifuge. The supernatant was removed, and 0.6 volume isopropanol was added to this. The genomic DNA pellet was recovered by centrifugation at maximum speed (16 000 x g) for 5 minutes in a microcentrifuge. This pellet was washed with 70 % ethanol to remove any residual CTAB, and the pellet was recovered by centrifugation at maximum speed (16 000 x g) for 5 minutes in a microcentrifuge. The pellet was re-suspended in TE buffer (100 µl).

3.2.6 PCR amplification of the coding regions for Type III DnaJ like proteins from *A. tumefaciens*

The PCR reaction was set up as described in Section 3.2.1. The PCR protocol used was as follows: an initial step of 1 minute at 95°C, followed by 25 cycles of 1 minute at 95°C, 1 minute at 50°C, 1 minute at 72°C, followed by a final hold at 72°C for 5 minutes. PCR products were ligated into pGEM-T Easy as described previously (Section 3.2.2). The resulting plasmids were restricted with *Bam* HI and *Sal* I, and fragments were ligated into pQE30 to give plasmids pC1-30, pC2-30 and pC5-30 (Figure 3.3). These plasmids contain the coding regions for *Agt djC1*, *Agt djC2* and *Agt djC5* respe ctively.

3.2.7 Replacement of the *Agt* DnaJ J domain with the J domains from *A. tumefaciens* Type III DnaJ-like proteins

A *Sfu* I site was introduced into pRJ30 shortly after the end of the J domain at codon 74 using site directed mutagenesis as described in Section 3.2.4 to give plasmid pRJ30-F74. This was a silent mutation that does not affect the coding sequence of *Agt dnaJ.* The region encoding the *Agt* DnaJ J domain was removed using the restriction enzymes *Bam* HI and *Sfu* I. The position of the *Sfu* I site is shown in Figure 3.4.

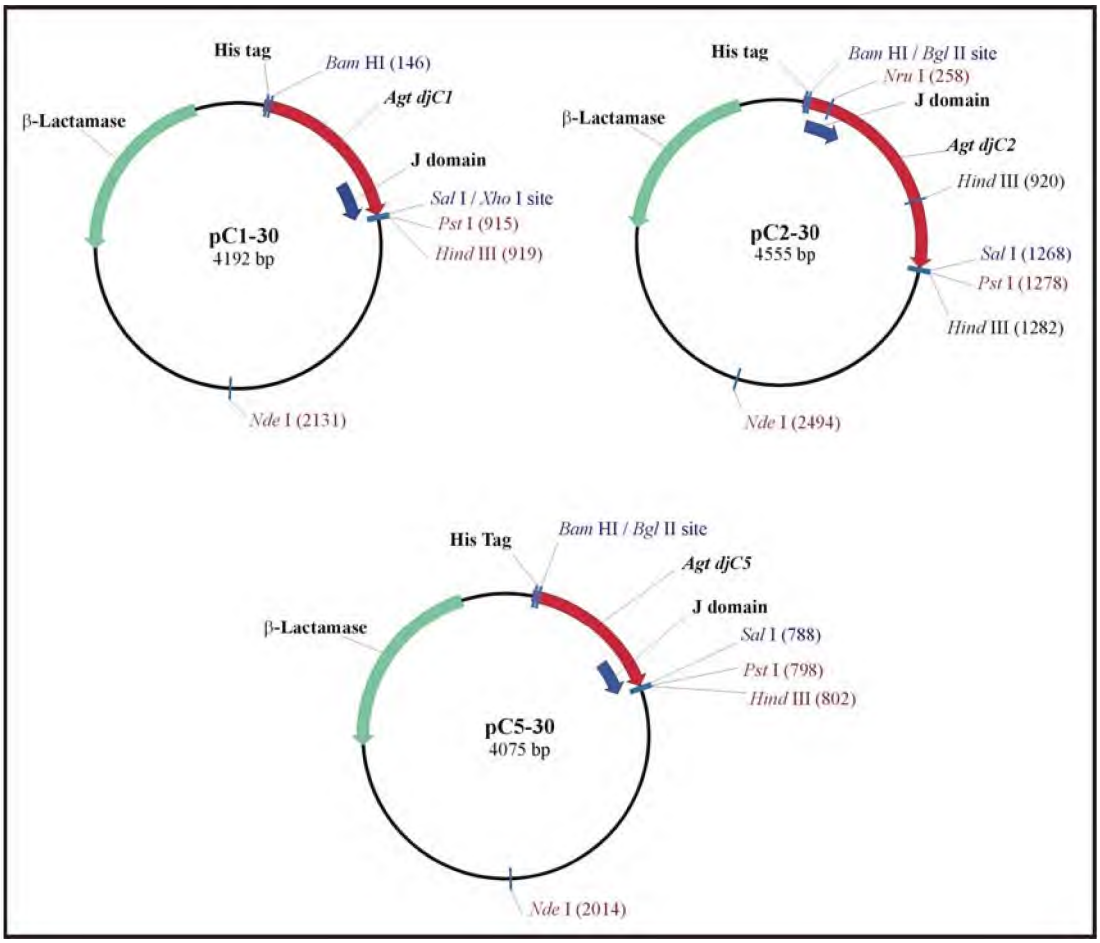

Figure 3.3: Plasmid maps of pC1-30, pC2-30 and pC5-30

The coding regions for the Type III proteins are in red, and for ampicillin resistance in green. The position of the J domain is indicated in blue. Unique restriction sites are in red, and multiple sites are in black. The *Bam* HI and *Sal* I sites are in blue.

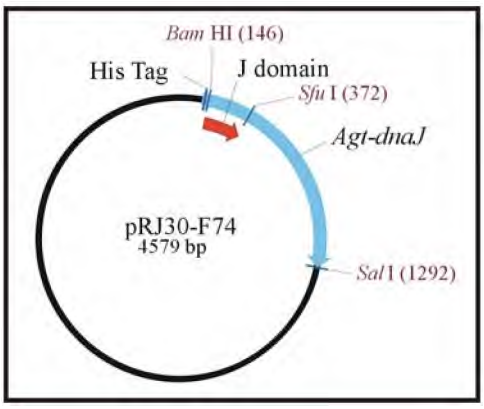

Figure 3.4: Plasmid map of pRJ30-F74 showing the inserted *Sfu* I site

The blue ORF represents the full length *Agt dnaJ* coding region; the red arrow represents the J domain bordered by *Bam* HI and *Sfu* I sites.

The regions encoding the J domains from the *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5 proteins were amplified using PCR off the plasmids containing the coding regions for these proteins (pC1-30, pC2-30 and pC5-30). Primers used were CJ1F and CJ1R for the amplification of the coding region for the J domain from *Agt* DjC1, C2F and CJ2R for the amplification of the coding region for the J domain from *Agt* DjC2, and CJ5F and CJ5R for the amplification of the coding region for the J domain from *Agt* DjC5. Primer sequences are given in the Appendix (Section 6.2, Table 6.5). Standard PCR techniques were used. Final primer concentrations used were 1 μM. The parameters used were the following: an initial step of 1 minute at 95°C, followed by 25 cycles of 1 minute at 95°C, 1 minute at 50°C, 1 minute at 72°C, followed by a final hold at 72°C for 5 minutes. The forward primer included a *Bam* HI site, and the reverse primer included a *Sfu* I site, except in the instance of the DjC2 forward primer. C2F was used (the forward primer used for the amplification of full length *Agt djC2*), and this contained a *Bgl* II site, which possesses an identical overhang to a *Bam* HI restricted fragment. Primer sequences are given in Table 6.5 in the Appendix (Section 6.2). PCR fragments were ligated into pGEM-T Easy as described previously (Section 3.2.2) and the regions encoding the J domains were excised using *Bam* HI or *Bgl* II, and *Sfu* I. The regions encoding the J domain fragments were then ligated into pRJ30-F74 lacking the region encoding the *Agt* DnaJ J domain to give plasmids pRJ-CJ1, pRJ-CJ2 and pRJ-CJ5 (Figure 3.5) encoding *Agt* DnaJ-CJ1, *Agt* DnaJ -CJ2 and *Agt* DnaJ-CJ5 respectively.

3.2.8 DNA sequencing

All constructs that were sequenced were prepared using the Qiagen Miniprep kit using the manufacturer's protocol (Qiagen, USA) (Appendix, Section 6.4.1.1). The plasmid DNA was eluted in 50 μl sterile distilled water. The DNA was sequenced using the BigDye V3.1 Terminator cycle sequencing kit (Applied Bioystems, USA). Primers used for sequencing were the generic pUCR and pUCF primers when inserts were in pGEM-T based constructs, and the generic pQEforward and pQEreverse primers when inserts were in pQE30 based constructs. The internal, specific DnaJ primers DnaJF1, DnaJF2, DnaJR1 and DnaJR2 were used for sequencing *Agt dnaJ*. The sequences for all primers are given in the Appendix (Section 6.2, Table 6.5). Sequencing of coding regions was performed on PCR-amplified

coding regions from three independent PCR reactions, and sequence information was assembled using Vector NTI (Informax, USA).

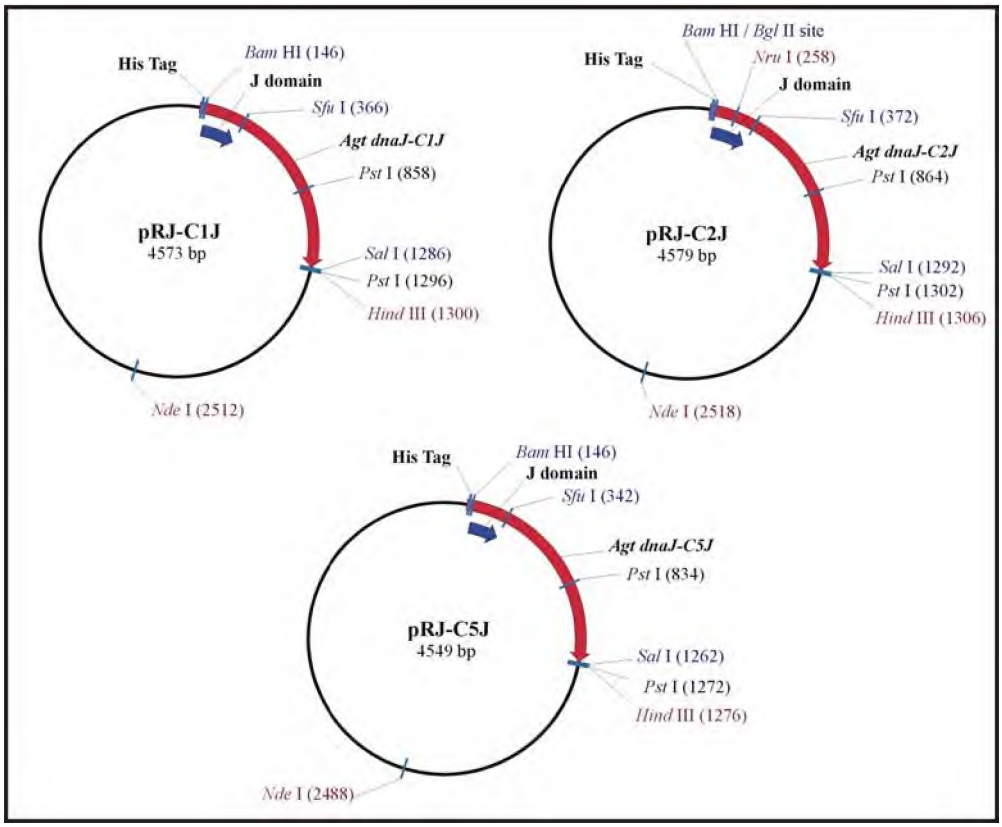

Figure 3.5: Plasmid maps of pRJ-CJ1, pRJ-CJ2 and pRJ-CJ5 showing inserted Type III J domains

The coding regions for the chimeric *Agt* DnaJ proteins are indicated in red. The regions encoding the J domains are in blue. *Bam* HI, *Sal* I and *Sfu* I restriction sites are indicated in blue; multiple restriction sites are indicated in black and unique restriction sites in red.

3.2.9 Complementation assays

E. coli OD 259 (MC4100 *araD139* Δ*ara714* Δ*cbpA*::*ka*n *dnaJ*::Tn*10*-42) was donated by Dr Olivier Deloche (Département de Biochimie Médicale, Centre Médicale Universitaire, Faculté de Médecine, Université de Genève, Geneva, Switzerland). It was initially described as *E. coli* WKG190 (Kelley and Georgopoulos, 1997b); (Deloche *et al.*, 1997). It was temperature sensitive at temperatures below 16°C and above 37°C. Control plasmids pBAD22A (ampicillin resistance marker (Amp^R), *colE1 ori*) and pWKG90 (pBAD22A*dnaJ*+) were also donated for use as negative and positive controls respectively, with the *E. coli dnaJ* coding

region under the control of the arabinose promoter. The control plasmids pBAD22A and pWKG90, and the experimental plasmids, pQE30, pRJ30, pC1-30, pC2-30, pC5-30, pRJ-C1J, pRJ-C2J and pRJ-C5J were transformed into *E. coli* OD259. Overnight cultures (5 ml) were set up containing 100 µg/ml ampicillin using fresh transformants, and grown at 30°C. The overnight cultures were diluted one in 10 into fresh YT broth, and these diluted cultures were grown at 30°C for four hours. Cultures were diluted to an $_{A600}$ of 0.3 to ensure that all cultures were at the same cell density for plating purposes. The diluted cultures were then further diluted 1 in 100, 1 in 10^4, 1 in 10^6, 1 in 10^8, and 1 in 10^{10}. An aliquot of each of these dilutions (3 µl) was spotted onto four plates, two with and two without the appropriate inducer, 6.6 mM arabinose for pWKG and pBAD, and 50 µM isopropyl β D-thiogalactoside (IPTG) for the pQE30 based vectors. Plates were grown at 30°C and at 40°C to ascertain whether *Agt* DnaJ could complement for the lack of *E. coli* DnaJ and CbpA.

3.2.10 Purification of *Agt* DnaJ

Agt DnaJ was purified using Nickel affinity chromatography as described in the Appendix (Section 6.4.2.5). Eight elutions of one bed volume were generally obtained, and these were pooled for buffer exchange. Buffer exchange was performed on a 10 ml Sephadex G-25 gel filtration column equilibrated in 100 mM Tris, pH 8.0, 300 mM NaCl. S amples (1.5 ml) were eluted from the column, and fractions were analysed using the Bradford method of protein determination (Appendix, Section 6.4.2 .7) (Bradford, 1976) and readings at 280 nm to confirm the removal of imidazole from the protein fractions. Proteins were stored at 4°C for immediate use, and at - 80°C for long term storage in 100 mM Tris, pH 8.0, 300 mM NaCl, 1mM PMSF.

3.2.11 Purification of *Agt* DnaK

A plasmid encoding *Agt* DnaK (pRK30, for *A. tumefaciens* RUOR *dnaK* in pQE30) was a kind gift of Dr Aileen Boshoff. A 100 ml starter culture of *E. coli* XL1 Blue [pRK30] was grown overnight at 37°C. The starter culture was diluted 1 in10 into fresh yeast-tryptone broth (YT broth) and grown for a further two hours. Induction was performed using 1 mM IPTG and growth was continued for a further four hours. The cells were pelleted and

resuspended in 1/50th culture volume in lysis buffer (100 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF) and stored overnight at – 20°C. The cells were thereafter thawed and lysed using lysozyme (1 mg/ml final concentration). The cell debris was removed by centrifugation in a microcentrifuge at 16 000 x g for 30 minutes at 4°C. The cell lysate was added to 1.0 ml suspended 50 % slurry of Ni-charged Sepharose beads, and left to bind overnight at 4°C with shaking. The beads were washed twice with 10 column volumes wash buffer I (100 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM PMSF) and twice with 10 column volumes wash buffer II (100 mM Tris, pH 8.0, 300 mM NaCl, 80 mM imidazole, 1 mM PMSF). Elution of protein was performed using 0.5 column volumes elution buffer (100 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole, 1 mM PMSF) seven times, and three times using 1.5 column volumes elution buffer to ensure that all the protein was removed from the beads.

Elution samples 1 to 7 were applied to a 10 ml Sephadex G-25 column equilibrated in 100 mM Tris, pH 8.0, 300 mM NaCl, 1 mM PMSF and eluted in fractions (1.5 ml) into this buffer in order to remove the imidazole. Proteins were stored at 4°C for immediate use, and at -80°C for long term storage.

3.2.12 Purification of *Agt* DnaJ-H33Q

The plasmid encoding *Agt* DnaJ-H33Q was transformed into *E. coli* XL1 Blue. Purifications were performed as described previously for *Agt* DnaJ (Section 3.2.10).

3.2.13 ATPase assays

A colourimetric assay for following the release of inorganic phosphate (Lanzetta *et al.*, 1979; Chamberlain and Burgoyne, 1997a; Chamberlain and Burgoyne, 1997b) was used for following the kinetics of the ATPase activity of *Agt* DnaK as well as the stimulation of the basal ATPase activity of *Agt* DnaK by *Agt* DnaJ.

DnaK (0.4 μM) was incubated in phosphate assay buffer (10 mM Hepes, pH 8.0; 0.1 M KCl, 2 mM $MgCl_2$; 0.5 mM dithiothreitol) for five minutes at 37°C. The reaction was initiated by

the addition of ATP. Samples (50 µl) were removed at specific time intervals, and placed in a microtitre plate and the reaction was stopped by the addition of an equal volume of 10 % sodium dodecyl sulphate (SDS). At the end of the reaction an ammonium molybdate solution (1 .25 % ammonium molybdate in 6.5 % H_2SO_4) (50 µl) and a 9 % ascorbic acid solution (50 µl) was added. After one hour the samples were read at $_{A660,}$ in a microtitre plate reader (PowerWave, Biotek, USA), and data was captured using KCjunior software (Biotek). A standard curve was set up using sodium phosphate in a range from 0 to 300 µM (100 µl volumes). The standards had the ammonium molybdate solution (50µl) and the citric acid solution (50 µl) added to them. The colour was left to develop for one hour. The samples were read at $_{A660,}$ in a microtitre plate reader. The units of specific activity used were nmol phosphate released/min/mg *Agt* DnaK. The kinetics of the ATPase activity was determined using Lineweaver-Burke, Michaelis-Menten, Hanes-Wolfe and Eadie -Hofstee plots. The assay was performed under various conditions including change in ATP concentration (0 to 500 µM), and varying the concentration of *Agt* DnaJ (0 to 2 µM).

The ability of *Agt* DnaJ and *Agt* DnaJ-H33Q to stimulate the ATPAse activity *Agt*-DnaK was investigated by using ATPase assays. Equimolar concentrations (0.4 µM) of *Agt* DnaJ or *Agt* DnaJ-H33Q were incubated with *Agt* DnaK. The concentration of ATP used was 400 µM. Activity was calculated with units of nmol phosphate released/min. All ATPase assays were corrected for spontaneous ATP hydrolysis, and the presence of contaminating phosphate.

3.3 RESULTS

3.3.1 Insertion of *Agt dnaJ* into pGEM-T Easy

PCR amplification of the *Agt dnaJ* coding region was successful. The most successful amplification occurred at a dilution of genomic DNA of 1 in 100 (Figure 3.6). This presumably reflected a dilution in which sufficient DNA was present to allow for amplification, but any inhibitors present had been diluted out.

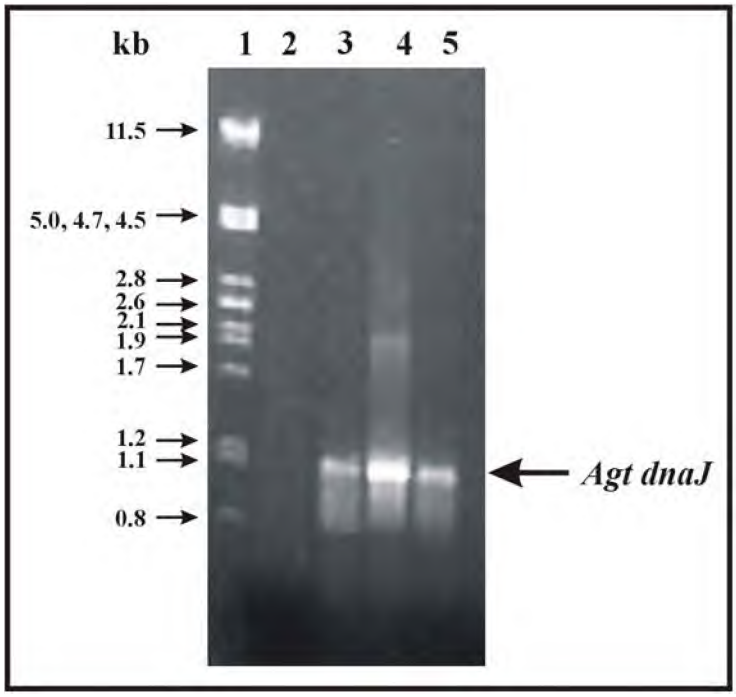

Figure 3.6: PCR-based amplification of the *Agt dnaJ* coding region

A. tumefaciens RUOR genomic DNA was diluted to various concentrations (neat, 1 in 10, 1 in 100 and 1 in 1000). Amplification of the coding region was performed using the various dilutions of genomic DNA, and PCR amplification products were resolved on a 0.9 % agarose gel. Lane 1 - *λ-Pst* I marker; Lane 2 – PCR using neat genomic DNA; Lane 3 – PCR using genomic DNA diluted 1 in 10; Lane 4– PCR using genomic DNA diluted 1 in 100; Lane 5 – PCR using genomic DNA diluted 1 in 1000.

The *Agt dnaJ* coding region was sequenced on both strands using plasmid sequencing primers, and internal *Agt dnaJ* specific primers. The *Agt dnaJ* coding region was successfully inserted into pQE30, as shown in Figure 3.7. The ends of the construct were confirmed by sequencing using the pQE forward and pQE reverse primers (sequences in the Appendix, Section 6.2, Table 6.5). The resultant plasmid was termed pRJ30.

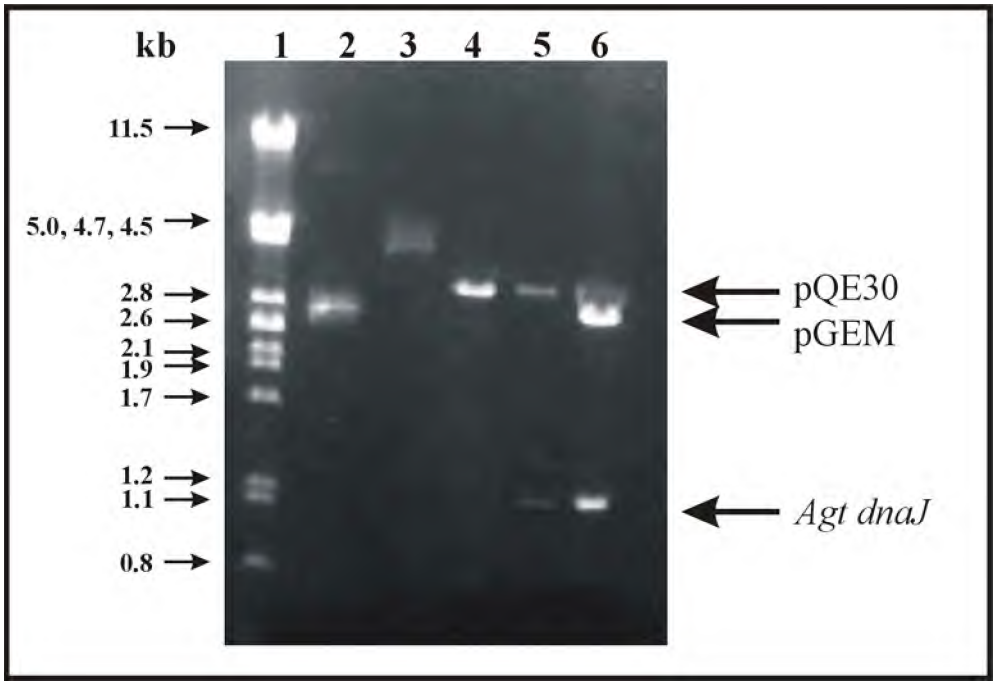

Figure 3.7: Insertion of the *Agt dnaJ* coding region into pGEM-T Easy and pQE30

Plasmids pQE30, pGEM-T Easy, pGEM-RJ and pRJ30 were restricted with *Bam* HI, *Sal* I and a combination thereof. Restricted fragments were resolved in a 0.8% agarose gel. Lane 1 - *λ-Pst* I marker; Lane 2 – Uncut pQE30; Lane 3 – Uncut pRJ30; Lane 4 – pQE30 restricted with *Bam* HI and *Sal* I; Lane 5 – pRJ30 restricted with *Bam* HI and *Sal* I; Lane 6 – pGEM-RJ restricted with *Bam* HI and *Sal* I.

The nucleotide sequence of *Agt dnaJ* and translated protein sequence, along with the position of the internal sequencing primers, is shown below in Figure 3.8. The length of the open reading frame was 1140 bp, encoding a protein with a predicted molecular mass of 41.1 kDa, and a predicted isoelectric point of 7.56. As can be seen the *Agt* DnaJ sequence contained the J domain, a glycine-phenylalanine rich region, and the cysteine repeats. This classified it as a Type I DnaJ-like protein (Cheetham and Caplan, 1998). It also contained previously described conserved regions in its C terminus (Goffin and Georgopoulos, 1998). It was 98.4 % identical to the DnaJ protein from *A. tumefaciens* C58 at the amino acid level (Figure 3.9), and 91.8 % identical at the nucleotide level, where most of the changes occurred at the third codon position. *Agt* DnaJ was 56.7 % identical to *E. coli* DnaJ at the amino acid level.

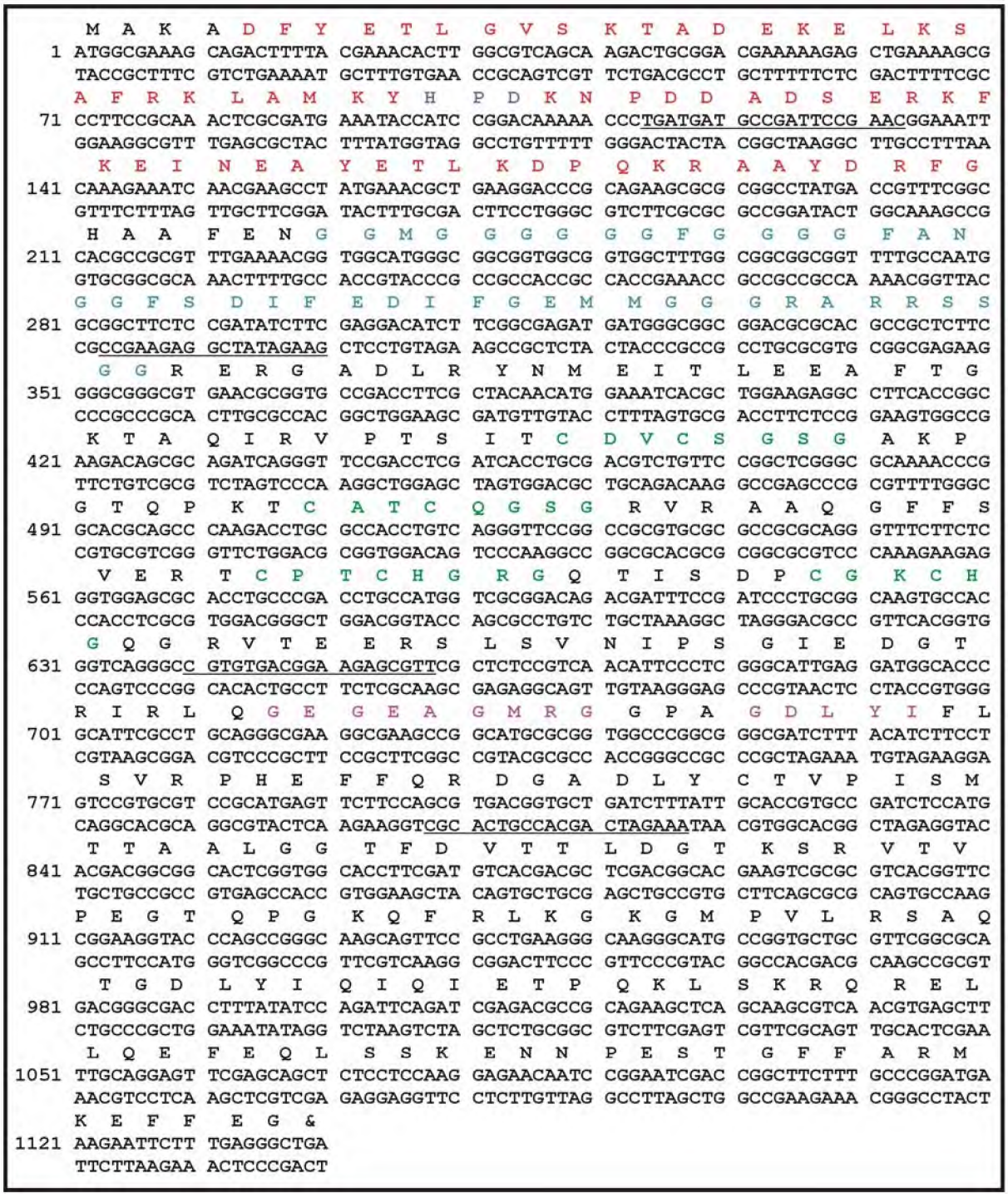

Figure 3.8: The nucleotide sequence of the *Agt dnaJ* coding region and the predicted *Agt* DnaJ protein sequence

The J domain is in red, with the conserved HPD in purple. The Gly/Phe region is in blue, and the cysteine repeats in green. Conserved C terminal regions are in pink (Goffin and Georgopoulos, 1998). Regions used for internal primers for sequencing are underlined, with forward primers indicated on the top strand, and reverse primers indicated on the complementary strand. The generation of the predicted protein sequence was performed in GeneRunner (Hastings Software, USA).

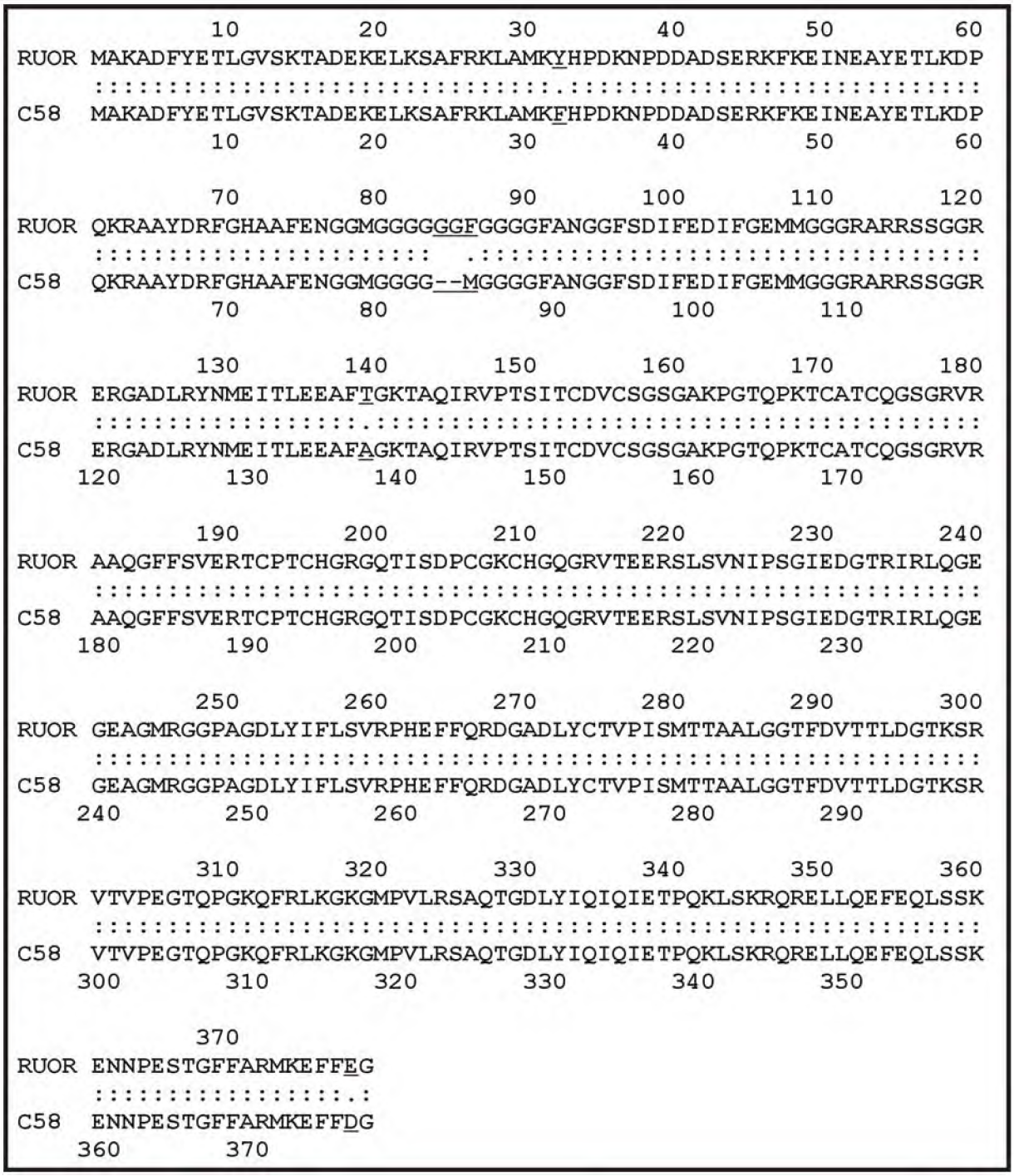

Figure 3.9: Pairwise alignment of the amino acid sequence of *Agt* DnaJ from *A. tumefaciens* C58 and *A. tumefaciens* RUOR

The pairwise alignment was generated using the Align package at Genestream (http://xylian.igh.cnrs.fr/bin/align-guess.cgi). Underlined amino acids indicate lack of conservation between the two proteins.

3.3.2 Homology modelling of the structure of *Agt* DnaJ

Homology models were generated of the J domain and the cysteine repeats of *Agt* DnaJ, using structures from *E. coli* DnaJ as a template. The models are depicted in Figure 3.10.

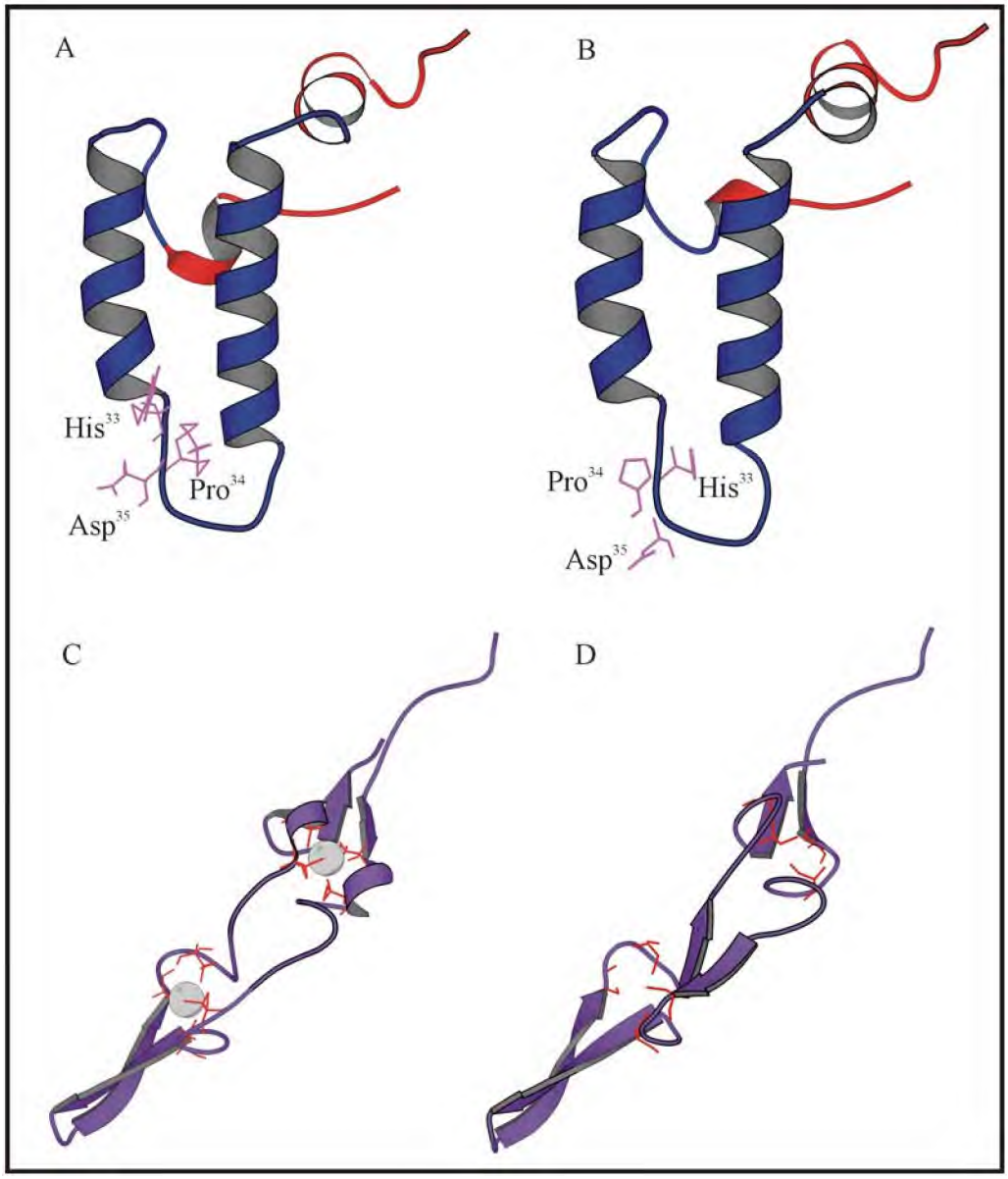

Figure 3.10: Homology models of the J domain and the cysteine repeats of *Agt* DnaJ

The J domain and the cysteine repeats of *Agt* DnaJ were homology modelled using Whatif (Vriend, 1990), using the *E. coli* DnaJ J domain (1XBL) (Pellechia *et al.*, 1996) and cysteine repeats (1EXK) (Martinez-Yamout *et al.*, 2000) structures as templates. (A) and (C) show the *E. coli* DnaJ structures and (B) and (D) show the respective homology models for *Agt* DnaJ. (A) and (B) show the J domain, with the HPD motif depicted in sticks and labelled. (C) and (D) show the cysteine repeats, with the cysteine residues indicated in red. The zinc molecules in (C) are depicted in grey and are space filled. The figures were generated using Molscript (Kraulis, 1991).

The structures of the *Agt* DnaJ domains were similar to those from *E. coli* DnaJ. This would be expected, as the homology models were generated using the *E. coli* domains as templates. The BPD motif of the *Agt* DnaJ J domains was in a slightly different orientation, relative to the *E. coli* DnaJ J domains BPD motif. This was presumably a result of the method of generating the model, but possibly also reflects the mobile nature of the loop region.

Analysis of the C terminal region of *Agt* DnaJ led to a possible further subdivision of Type I DnaJ-like proteins.

3.3.3 Further subdivision of Type I DnaJ-like proteins

Type I DnaJ-like proteins are distinguished from Type II DnaJ-like proteins by the presence of the cysteine repeats (Cheetham and Caplan, 1998). A significant proportion of Type I proteins have also been shown to contain an isoprenylation motif at their C terminus (1.4.2.2). Analysis of the different type I C-terminal domains was performed to investigate conservation of residues in this regi on (Figure 3.11 and Figure 3.12).

An alignment of Type I DnaJ-like proteins that did not contain a CaaX box is shown in Figure 3.11. These DnaJ-like proteins all contained true cysteine repeats, i.e. there were no substitutions in the last amino acid of the last repeat. The C terminus of these proteins was generally ric h in aromatic residues such as phenylalanine, and in negatively charged residues such as aspartic acid.

An alignment of Type I DnaJ-like proteins that contain a CaaX box is shown in Figure 3.12. These are all eukaryotic proteins. Interestingly all proteins that contained a glycine to lysine substitution in the last cysteine repeat (Section 1.4.1.3) contained a CaaX box (Figure 3.12). As these proteins are only found in eukaryotes the lysine substitution may play a role in membrane localisation.

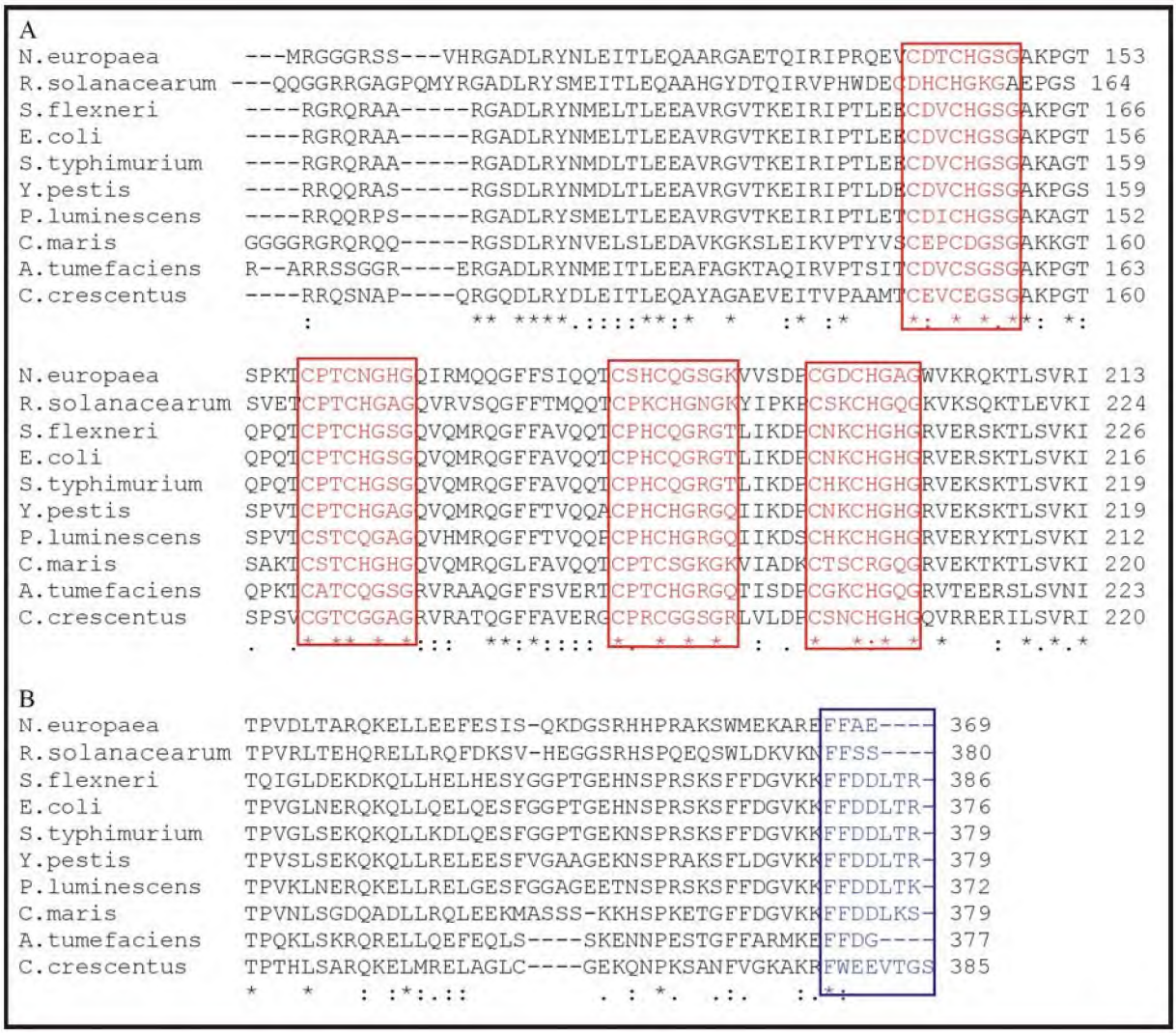

Figure 3.11: Alignment of the C terminal portion of Type I DnaJ-like proteins that lack a CaaX box

Ten representative Type I DnaJ-like proteins lacking a CaaX box were aligned using ClustalW (Thompson *et al.*, 1994). (A) depicts the cysteine rich region, with the cysteine repeats blocked in red. (B) depicts the C terminus, with the last eight residues blocked in blue. DnaJ proteins from the following organisms were used: *Nitrosomonas europaea* (O06431), *Ralstonia solanacearum* (CAD16341), *Shigella flexneri* (AAN41681) *E. coli* (P08622), *Salmonella typhimurium* (Q60004), *Yersinia pestis* (NP_404111), *Photorhabdus luminescens* subsp. *Laumondii* (NP_927928), *Colwellia maris* (BAB91324), *Agrobacterium tumefaciens* C58 (P50018) and *Caulobacter crescentus* (P22305).

Hence there are three distinct classes of Type I DnaJ-like proteins. Type Ia proteins contain normal cysteine repeats, are not membrane associated *via* a farnesyl group, and are bacterial. Type Ib proteins contain normal cysteine repeats, are not membrane associated *via* a farnesyl group, and are eukaryotic. Type Ic proteins contain the variant cysteine repeats, contain the CaaX box motif, and are eukaryotic. Why there is this distinction at the eukaryotic level in Type I DnaJ-like proteins is unclear at present.

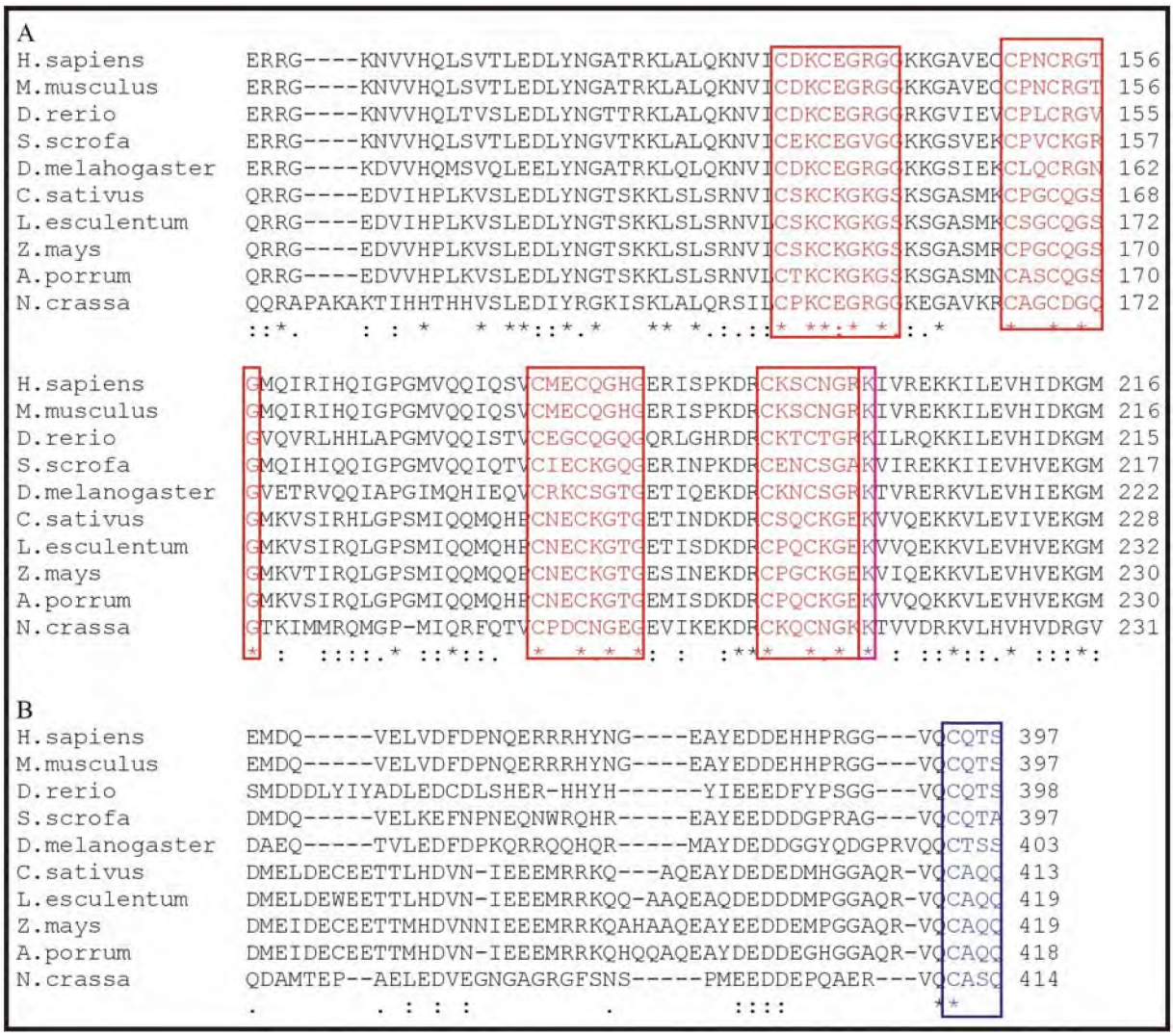

Figure 3.12: Alignment of the C terminal portion of Type I DnaJ-like proteins that contain a CaaX box

Ten representative Type I DnaJ-like proteins containing a CaaX box were aligned using ClustalW (Thompson *et al.*, 1994). (A) depicts the cysteine rich region, with the cysteine repeats blocked in red. (B) depicts the C terminus, with the last four residues forming the CaaX box blocked in blue. The lysine residue that replaces a glycine in the last cysteine repeat is highlighted in pink. DnaJ proteins from the following organi sms were used: *Homo sapiens* DnaJA1 (P31689), *Mus musculus* DnaJA1 (P54102), *Danio rerio* (AAH44445), *Sus scrofa* pDJA1 (AAP22730), *Drosophila melanogaster* (AAL28530), *Cucumis sativus* DnaJ-1 (Q04960), *Lycopersicon esculentum* (AAF28382), *Zea mays* ZMDJ1 (T01643), *Allium porrum* LDJ2 (S42031) and *Neurospora crassa* (XP _327700)

3.3.4 DnaJ-like proteins on the *A. tumefaciens* C58 genome

Five Type III DnaJ-like proteins were identified based on a BLAST search on the *A. tumefaciens* C58 genome. The accession numbers and other information are given in Table 3.2. No Type II and no additional Type I DnaJ-like proteins were identified. There did not appear to be any DnaJ-like proteins that were potentially homologous to the *E. coli* proteins CbpA, Hsc20, DjlB and DjlC.

Table 3.2: Type III DnaJ-like proteins present on the *A. tumefaciens* C58 genome

Protein [1]	Accession Numbers	J domain location	Chromosome
Agt DjC1/DjlA	AGR_C_3834/NP_532788	C terminal	Circular
Agt DjC2	NP_531458/AGR_C_1376	N terminal	Circular
Agt DjC3 [2]	AGR_C_3642p	N terminal	Circular
Agt DjC4	NP_535884/AGR_pAT_763	C terminal	Plasmid
Agt DjC5	NP_534119/AGR_L_2405	C terminal	Linear

[1] – Naming was performed based on the proposed nomenclature by Ohtsuka and Hata, (2000)
[2] – Potential pseudogene; no obvious Methionine start codon

Amplification of *Agt djC1*, *Agt djC2* and *Agt djC5* was successful, however amplification of *Agt djC3* and *Agt djC4* was not successful. The coding regions for *Agt djC1*, *Agt djC2* and *Agt djC5* were then inserted in pQE30 (Figure 3.13). Plasmids were digested with *Xho* I (which restricts prior to the coding regions) and *Pst* I (which restricts after the coding regions) to release inserts.

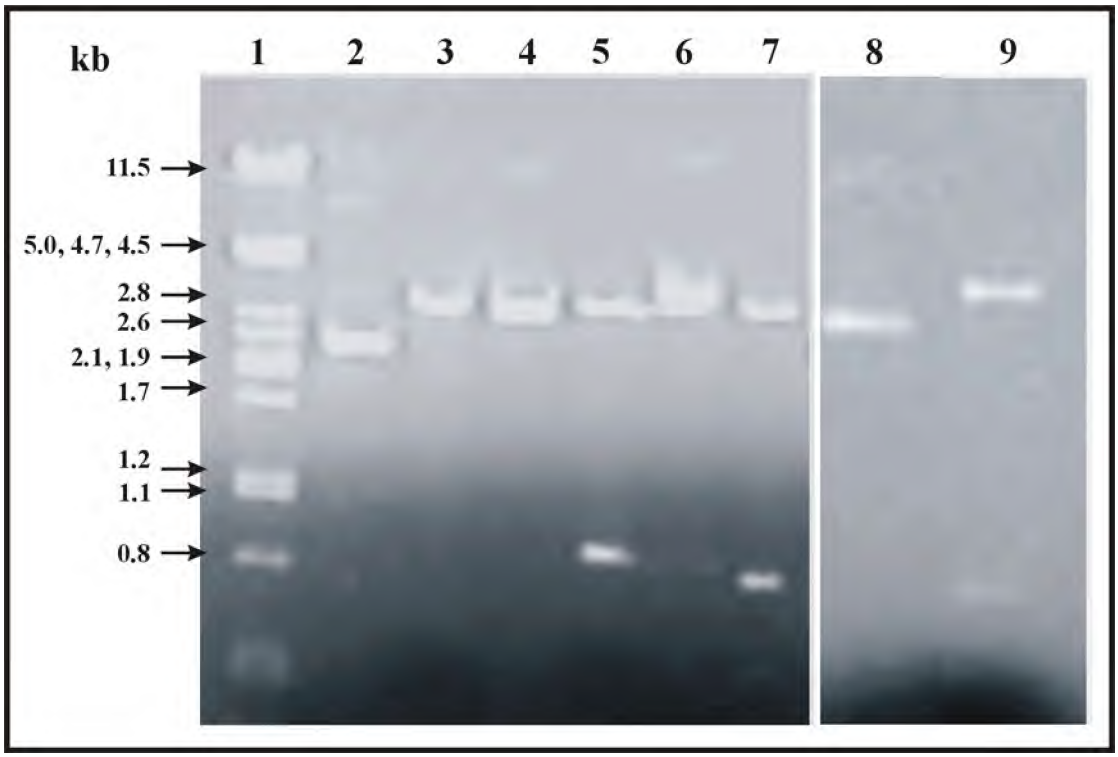

Figure 3.13: Confirmation of the insertion of the *Agt djC1, Agt djC2* and *Agt djC5* coding regions into pQE30

Lane 1: λ DNA restricted with *Pst* I marker; Lane 2: pQE30, uncut; Lane 3: pQE30 restricted with *Xho* I and *Pst* I; Lane 4: pC2-30, uncut; Lane 5: pC2-30 restricted with *Xho* I and *Pst* I; Lane 6: pC1-30, uncut; Lane 7: pC1-30 restricted with *Xho* I and *Pst* I; Lane 8: pC5-30, uncut; Lane 9: pC5-30 restricted with *Xho* I and *Pst* I.

3.3.4.1 Analysis of *Agt* DjC1/DjlA

The coding region for *Agt* DjC1, a DjlA homologue, was amplified from *A. tumefaciens* RUOR genomic DNA. The length of the coding region was 753 bp, encoding a protein of 250 amino acids, with a theoretical molecular mass of 27.8 kDa, and an isoelectric point of 5.74 (Figure 3.14). There were two amino acid changes between the proteins from *A. tumefaciens* RUOR and *A. tumefaciens* C58 (99.2 % identical; Figure 3.15).

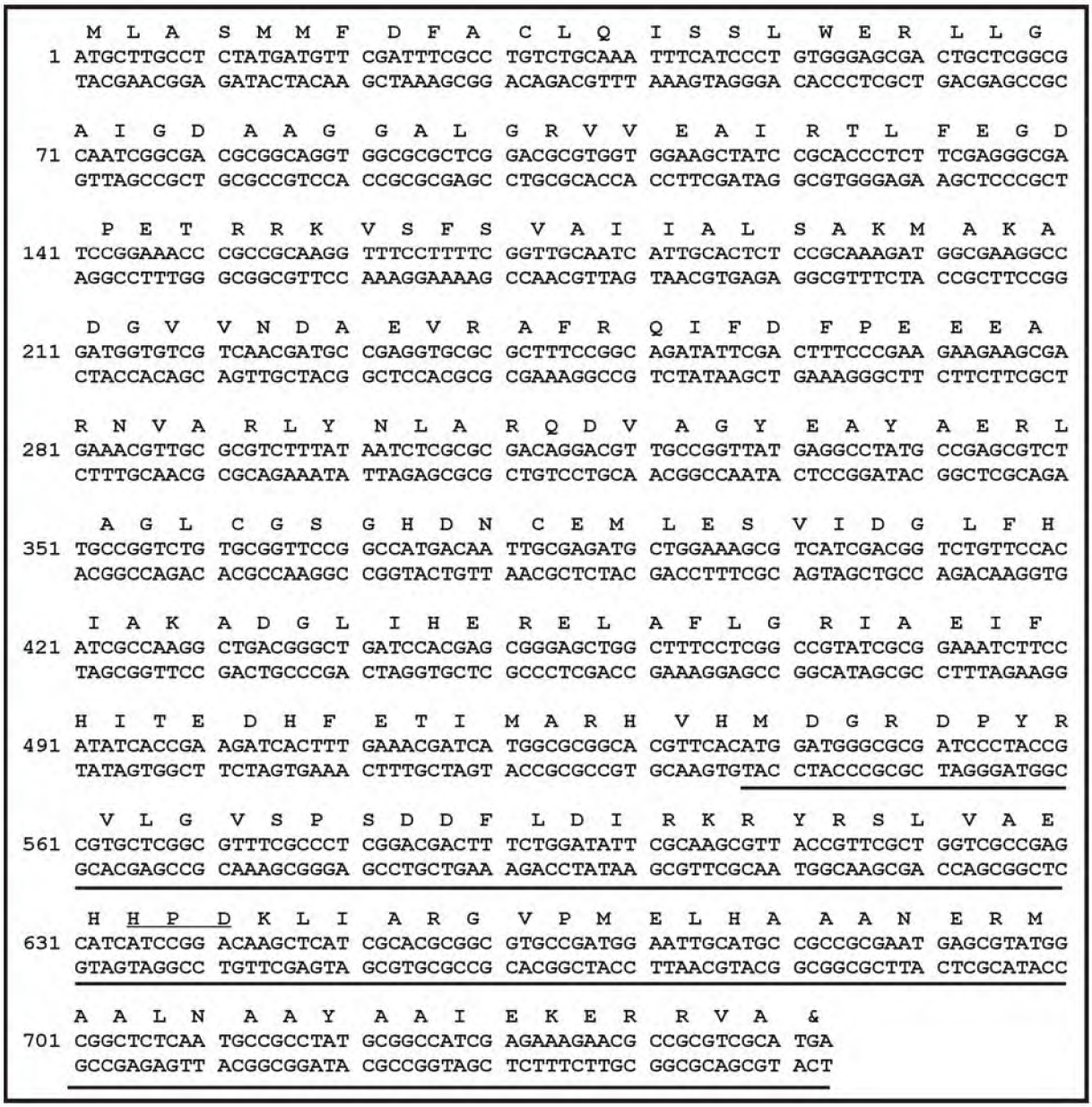

Figure 3.14: The nucleotide sequence of the *Agt djC1* coding region and the predicted *Agt* DjC1 protein sequence

The J domain is underlined on the nucleotide sequence, and the HPD motif is also indicated. The generation of the predicted protein sequence was performed in GeneRunner (Hastings Software, USA).

An alignment of DjlA protein sequences from various organisms is shown in Figure 3.15. The major difference between the DjlA proteins used was the loss of a twenty amino acid region immediately prior to the J domain in the *A. tumefaciens* and *Pseudomonas aeruginosa* DjlA proteins. This was a region rich in glutamine residues, and the significance of its loss is

not apparent. *Agt* DjC1 does not appear to contain the conserved glycine residues that are found in the transmembrane region of *E. coli* DjlA (Clarke *et al.*, 1996).

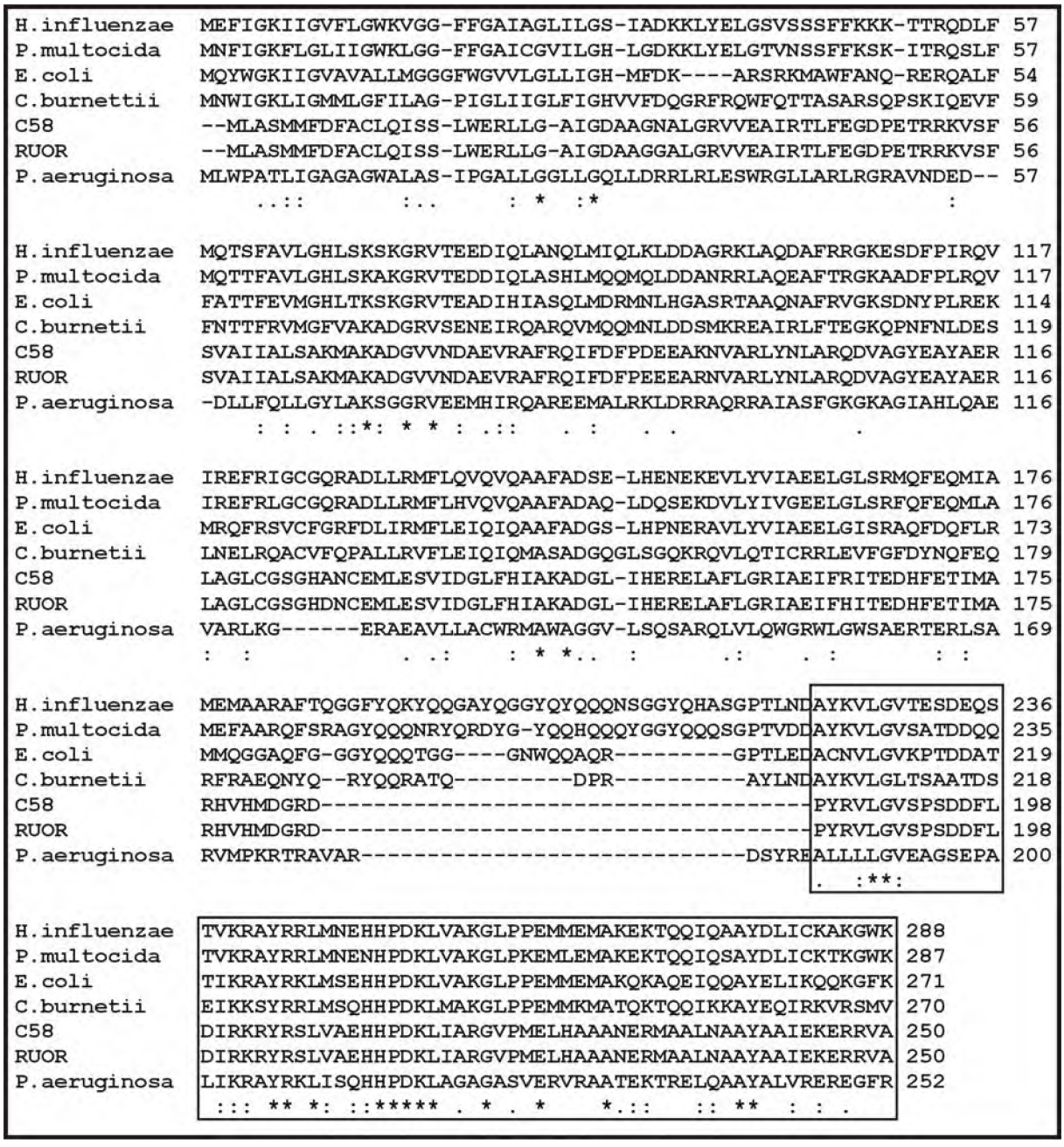

Figure 3.15: Comparison of Agt DjC1 with the DjlA proteins from other organisms

A multiple sequence alignment of the protein sequences of DjC1/DjlA proteins from *A. tumefaciens* C58 (AB2836) and *A. tumefaciens* RUOR, *Haemophilus influenzae* (P44607), *Pasteurella multocida* (NP_246817), *Coxiella burnetii* (Q45885), *Pseudomonas aeruginosa* (AAD22457) and *E. coli* DjlA (P31680) was performed using ClustalW (Thompson *et al.*, 1994). The J domain is boxed. Asterisks indicate an identical match and colons and full stops indicate partial matches.

3.3.4.2 Analysis of *Agt* DjC2

The coding region for *Agt* DjC2 was amplified from *A. tumefaciens* RUOR genomic DNA. The nucleotide sequence and its translated protein sequence are shown in Figure 3.16. The length of the coding region was 1116 bp, encoding a protein of 371 amino acids, with a theoretical molecular mass of 40.3 kDa, and an isoelectric point of 6.18.

Agt DjC2 was 86.7 % identical to the homologous protein in *A. tumefaciens* C58. The closest relation was found in *Sinorhizobium meliloti*, with other potential homologues being found in *Mesorhizobium loti* and *Nostoc* sp. (Figure 3.17). The majority of the conservation occurs in the N and C terminal regions. However there appears to be an insertion of amino acid sequence in the central portions of *Agt* DnaJ relative to the comparative proteins from *S. melitoti, M. loti* and *Nostoc* sp. This potentially implies a loss of sequence information from the *M. loti* and *Nostoc* sp. proteins, or a gain of sequence information in *Agt* DjC2. This may imply that *Agt* DjC2 has slightly differing functions as compared to the homologous proteins in *M. loti* and *Nostoc* sp. Equally of interest was the observation that the *M. loti* homologue has a Gly/Phe rich region, which was not present in any of the other proteins examined. Hence this was an instance where a Type II DnaJ-like protein was potentially homologous to Type III DnaJ-like proteins.

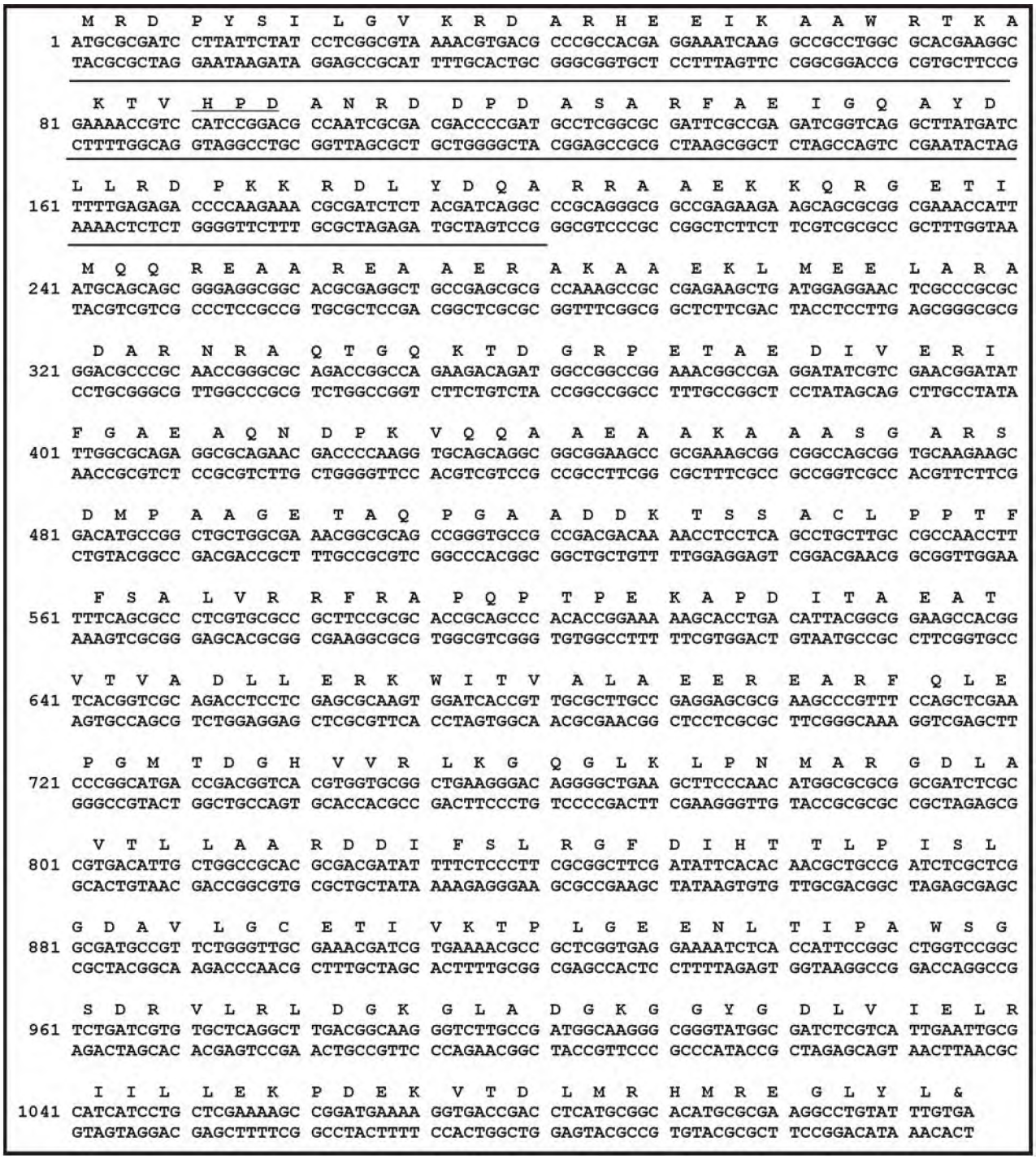

Figure 3.16: The nucleotide sequence of the *Agt djC2* coding region and the predicted *Agt* DjC2 protein sequence

The J domain is underlined on the nucleotide sequence, and the HPD motif is also indicated. The generation of the predicted protein sequence was performed in GeneRunner (Hastings Software, USA).

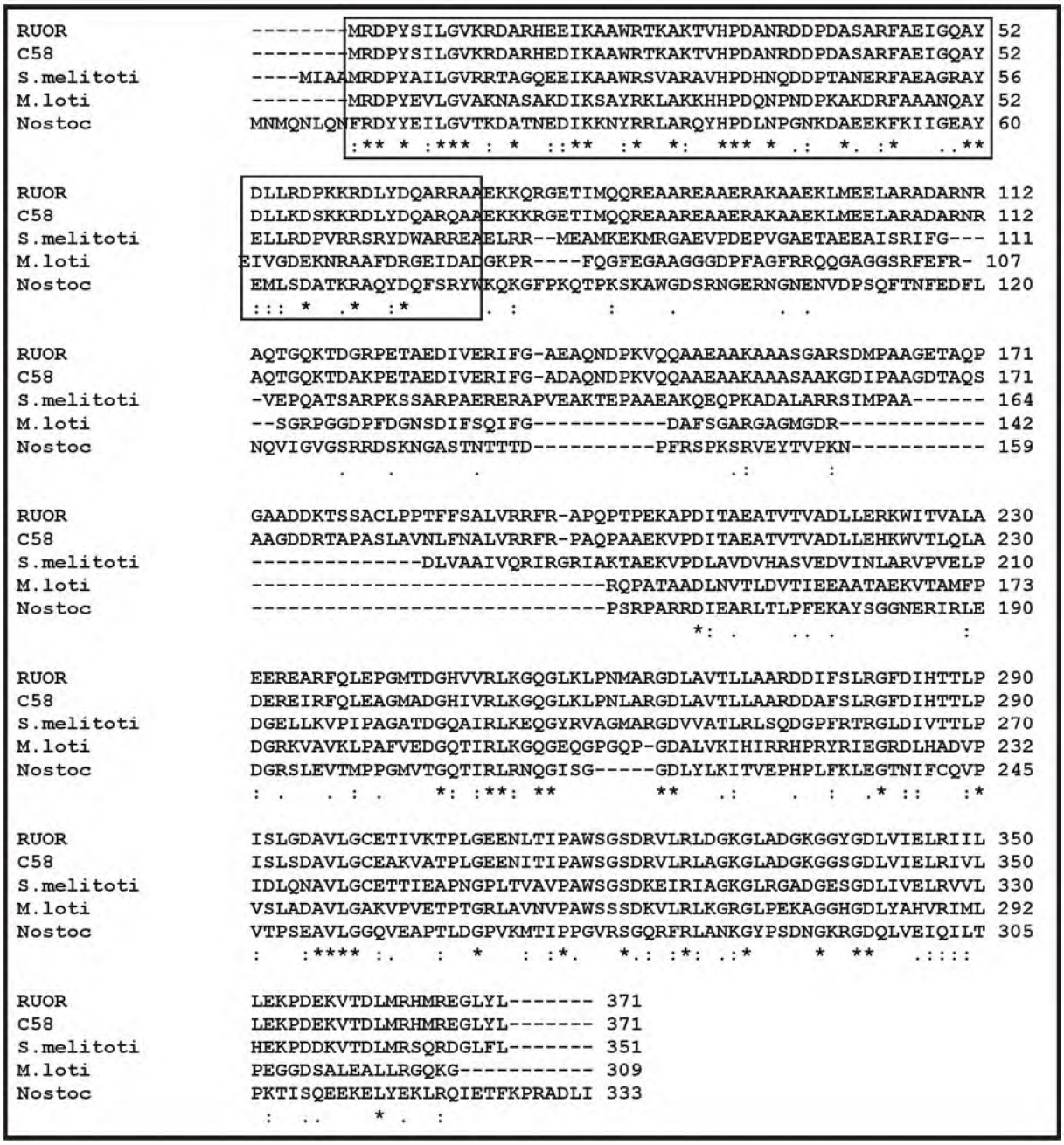

Figure 3.17: Multiple sequence alignment of *Agt* DjC2 potentially homologous proteins

A multiple sequence alignment of the amino acid sequences of *Agt* DjC2 from *A. tumefaciens* RUOR and *A. tumefaciens* C58, as well as homologous proteins from *S. melitoti* (CAC45469), *M. loti* (NP 107769) and *Nostoc* sp (NP 48703). Asterisks indicate an identical match and colons and full stops indicate partial matches. The J domain is boxed, and the BPD motif is underlined. No homologue for this protein was detected *in silico* in *E. coli*.

Despite its lack of a Gly/Phe region, this protein could be the nearest equivalent of CbpA in *A. tumefaciens.* There was some similarity at the C terminus of *Agt* DjC2 to both *E. coli* DnaJ

and CbpA, despite having low levels of overall identity with these two proteins (24.7% and 25.1% identity respectively) (Figure 3.18).

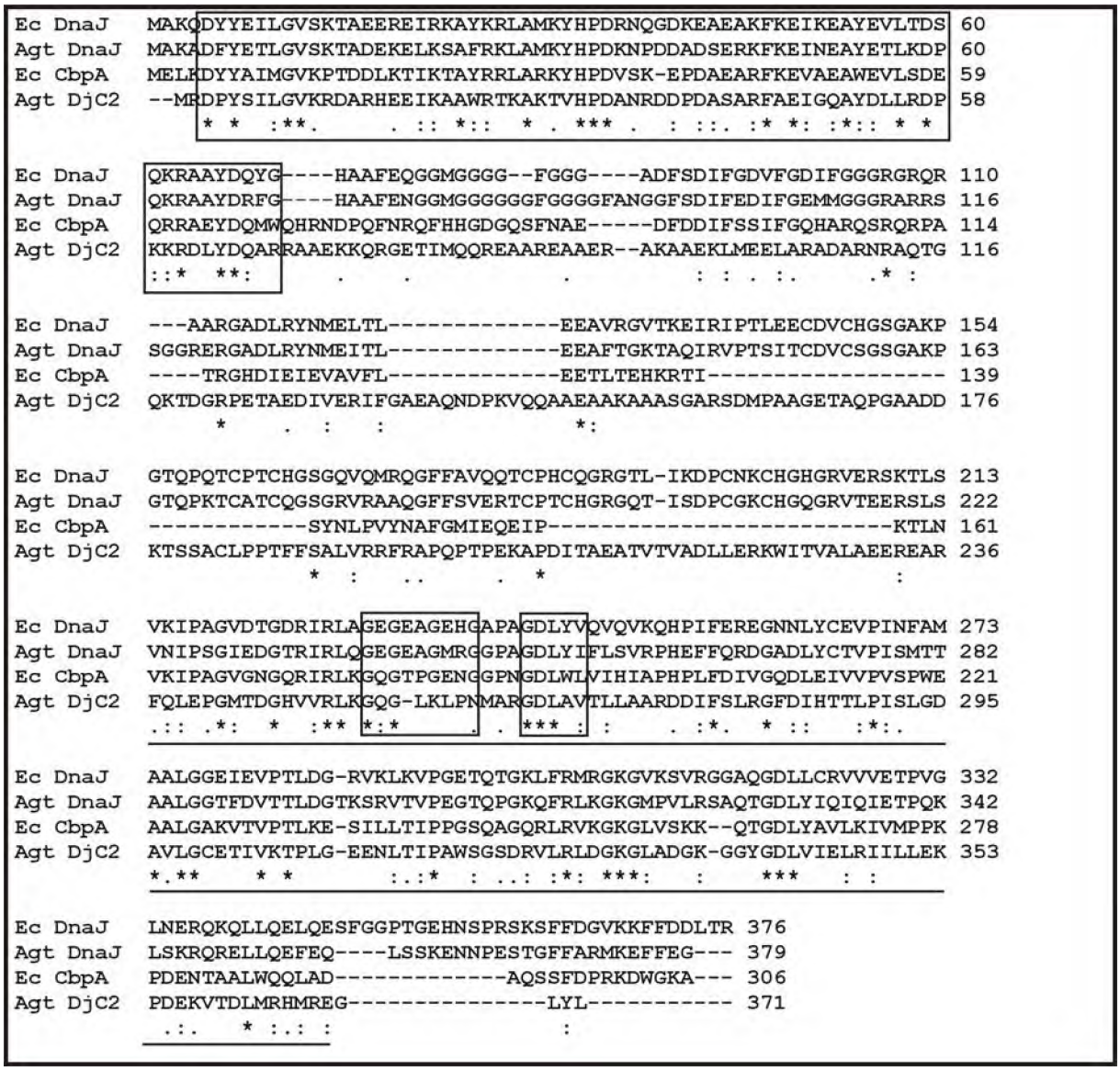

Figure 3.18: Multiple sequence alignment of the *E. coli* proteins DnaJ and CbpA, and *Agt* DnaJ and *Agt* DjC2

ClustalW was used to align *E. coli* DnaJ (Ec DnaJ), *E. coli* CbpA (Ec CbpA), *Agt* DnaJ and *Agt* DjC2. The J domain is boxed, as are conserved regions in the C terminus, and additional regions of similarity in the C terminus are underlined. Asterisks indicate an identical match and colons and full stops indicate partial matches.

Agt DjC2 also contains motifs shown to be important in *E. coli* DnaJ chaperone activity (Goffin and Georgopoulos, 1998) (boxed regions in the C terminus of Figure 3.18), with the latter repeat being well conserved.

3.3.4.3 Analysis of *Agt* DjC5

The coding region for *Agt* DjC5 was amplified from *A. tumefaciens* RUOR genomic DNA. The length of the coding region was 636 bp, encoding a protein of 211 amino acids, with a theoretical molecular mass of 23.7 kDa, and an isoelectric point of 9.70. The nucleotide sequence of the *Agt djC5*, and its translated protein sequence, is shown in Figure 3.19. It contains a C terminal J domain.

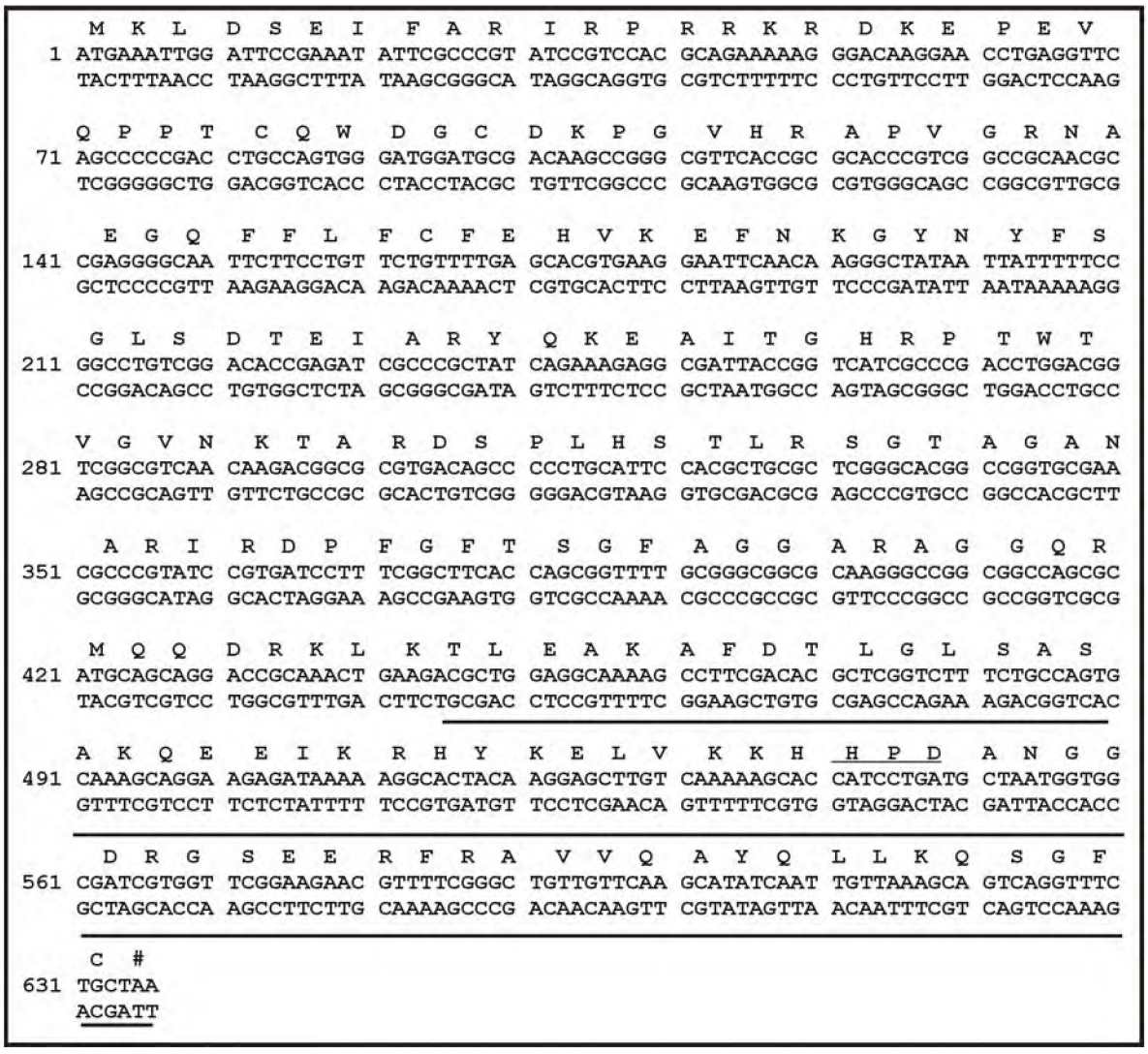

Figure 3.19: The nucleotide sequence of the *Agt djC5* coding region and the predicted *Agt* DjC5 protein sequence

The J domain is underlined on the nucleotide sequence, and the HPD motif is also indicated. The generation of the predicted protein sequence was performed in GeneRunner (Hastings Software, USA).

Agt DjC5 was an unique protein, not previously described at the protein level. *Agt* DjC5 was 91.2 % identical to the homologous protein from *A. tumefaciens* C58. Potential homologues were found in the related organism *S. melitoti*, as well as in *Pseudomonas dentrificans*, and *Brucella melitensis*. An alignment of these proteins is given in Figure 3.20. No homologue to this protein appeared to be present in *E. coli* as determined by performing a BLAST search against the *E. coli* genome using the *Agt* DjC5 sequence.

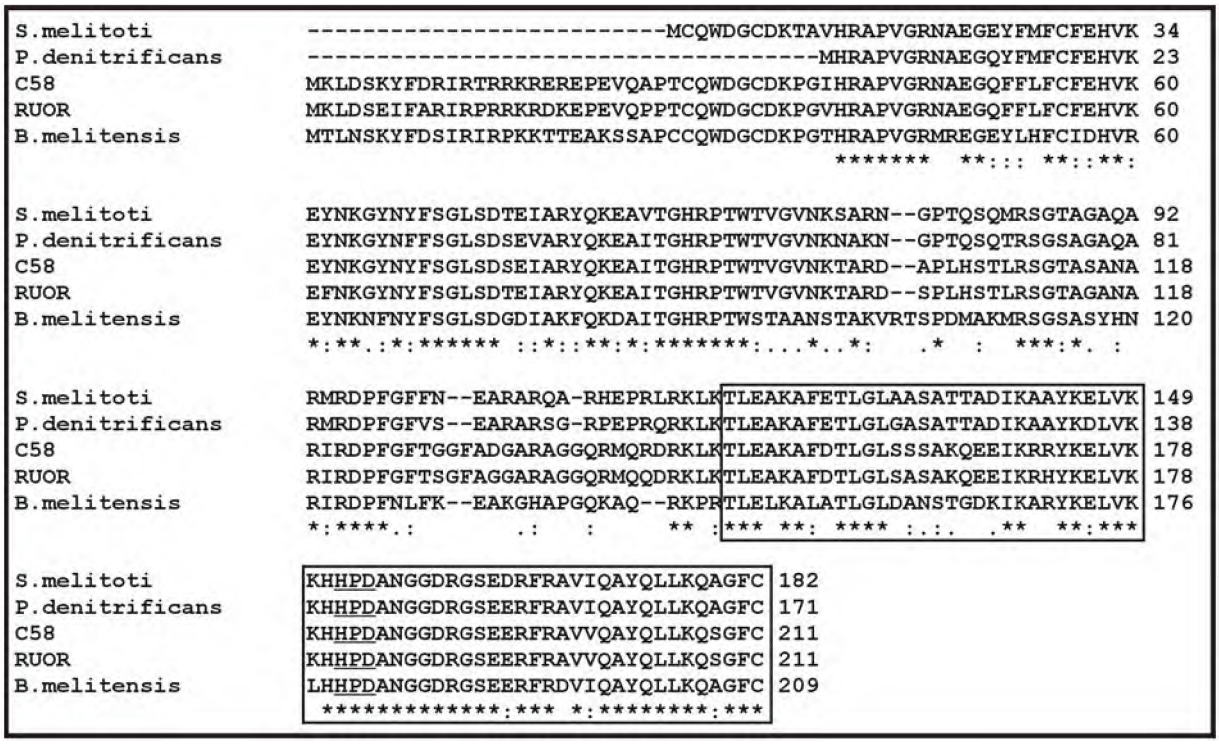

Figure 3.20: Multiple sequence alignment of *Agt* DjC5 potentially homologous proteins

A multiple sequence alignment of the amino acid sequences of *Agt* DjC5 from *A. tumefaciens* RUOR and *A. tumefaciens* C58, as well as potential homologous proteins from *S. melitoti* (CAC47252), *P. dentrificans* (YCB2_PSEDE) and *B. melitensis* (NP 538965) was performed using ClustalW (Thompson *et al.*, 1994). Asterisks indicate an identical match and colons and full stops indicate partial matches. The J domain is boxed, and the HPD motif is underlined. No homologue for this protein was detected *in silico* in *E. coli.*

3.3.5 Construction of chimeric constructs for the replacement of the J domain of *Agt* DnaJ with J domains from Type III DnaJ-like proteins

Primers were designed to amplify the coding regions for the J domains from *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5. The products obtained are shown in Figure 3.21.

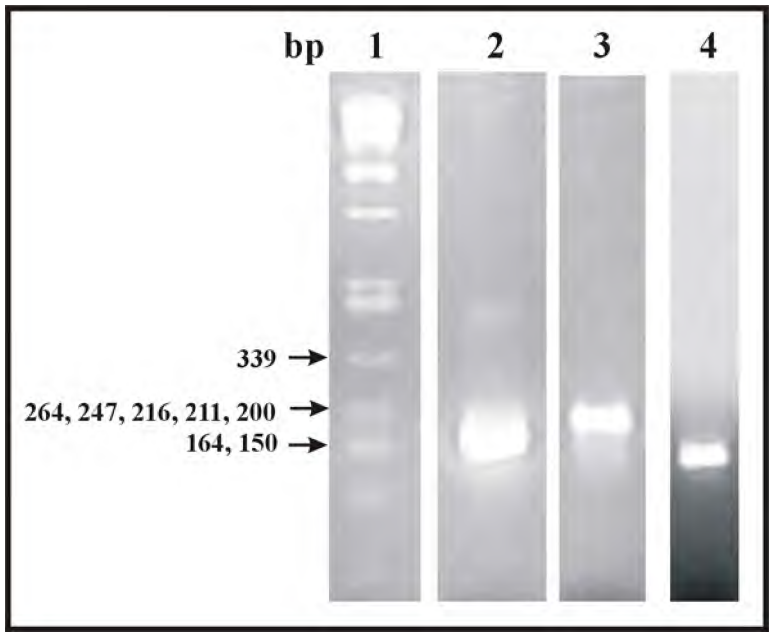

Figure 3.21: Amplification of the coding regions for the J domains from *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5

Amplification products were resolved on a 2 % agarose gel. Markers are given in bp between 150 and 400 bp. Lane 1: λ DNA restricted with *Pst* I marker; Lane 2: Amplification of the coding region for the J domain from *Agt* DjC1; Lane 3: Amplification of the coding region for the J domain from *Agt* DjC2; Lane 4: Amplification of the coding region for the J domain from *Agt* DjC5.

Successful amplification of the coding regions for the J domains from *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5 occurred. These coding regions were inserted into pRJ30-F74. A *Sfu* I was inserted at codon 74, which corresponds to the intervening region between the J domain and the Gly/Phe rich region. A comparison of the 5' and 3' ends of the inserted sequences is shown in Figure 3.22.

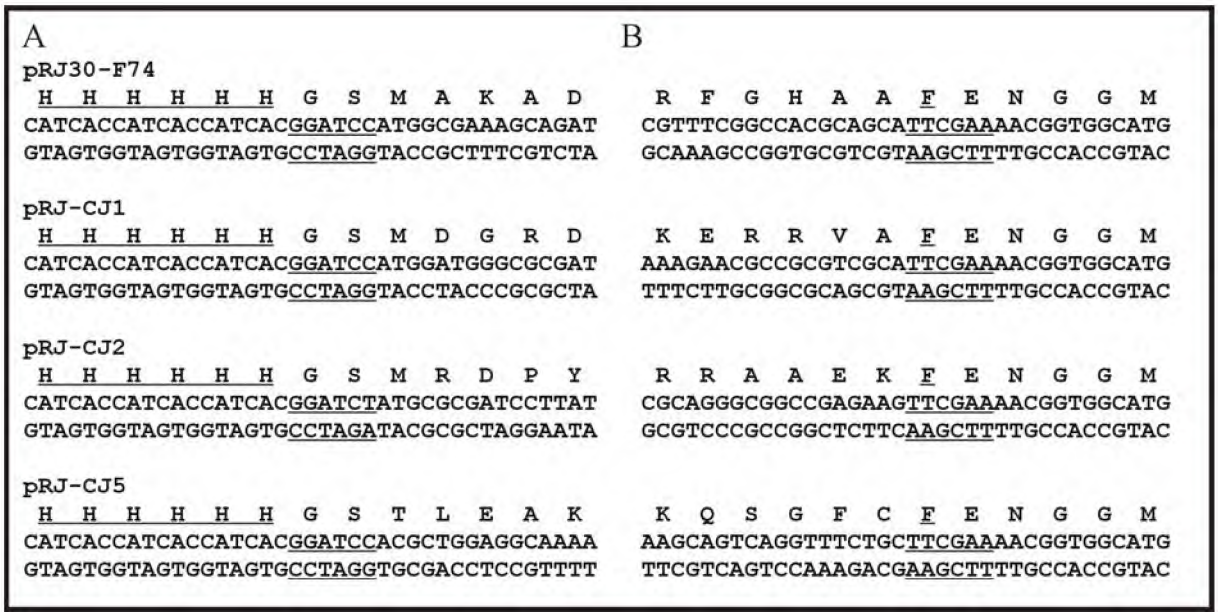

Figure 3.22: Comparison of the 5' and 3' ends of the coding regions for the J domain sequences in pRJ30-F74, pRJ-CJ1, pRJ-CJ2 and pRJ-CJ5

(A) shows the nucleic acid sequence, and translated protein sequence for the 5' position of the J domain coding region. The *Bam* HI site is underlined on the nucleotide sequence, and the translated histidine region is underlined on the translated protein sequence. Plasmid pRJ-CJ2 has a partial *Bam* HI site due to the usage of *Bgl* II for releasing the insert from the pGEM-T Easy construct. (B) shows the nucleic acid sequence, and translated protein sequence for the 3' position of the J domain coding region. The *Sfu* I site is underlined on the nucleotide sequence, and Phe residue which denotes the start of remaining *Agt* DnaJ protein sequence is underlined on the translated protein sequence.

3.3.6 Complementation by *E. coli* DnaJ in *E. coli* OD259

E. coli OD259 lacks genes for DnaJ and CbpA, and hence was sensitive to growth at temperatures above 37°C and below 16°C (Deloche *et al.*, 1997). It was therefore a useful strain to employ to determine whether *Agt* DnaJ was a functional DnaJ protein.

As can be seen in Figure 3.23, plasmid encoded *E. coli* DnaJ could complement for the lack of chromosomal *E. coli* DnaJ and CbpA. At 30°C *E. coli* OD259 containing either pBAD22A (plasmid that does not encode DnaJ) or pWKG90 (pBAD22A encoding DnaJ) could grow successfully, suggesting that DnaJ and CbpA were not essential for growth at normal cellular conditions (data not shown). However at 40°C, and in the absence of the inducer arabinose, *E. coli* OD259 containing pBAD22A or pWKG90 could not grow. The presence of arabinose was required to induce production of the *E. coli* DnaJ, and in its absence, there was no production of *E. coli* DnaJ.

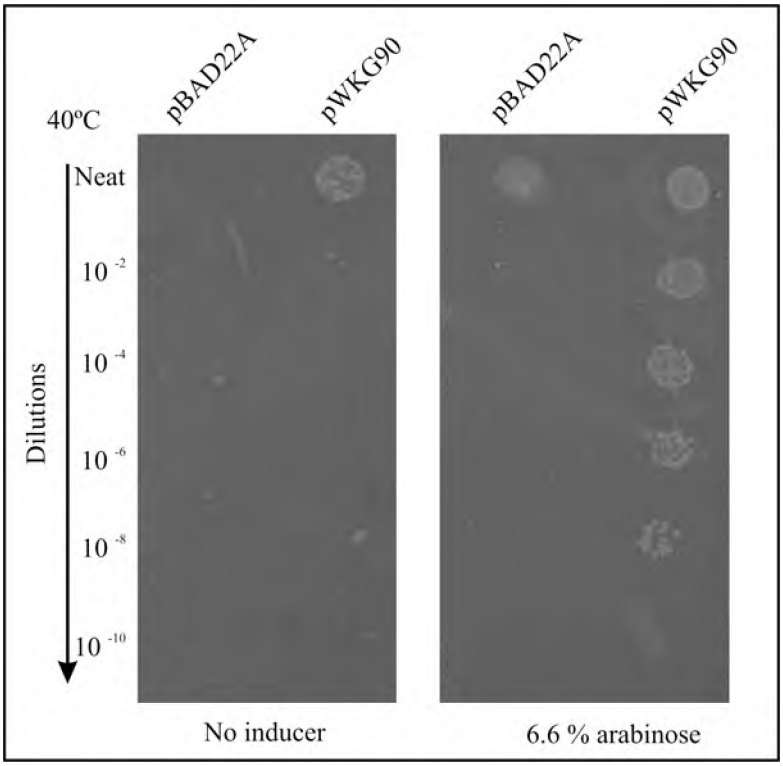

Figure 3.23: Plasmid encoded *E. coli* DnaJ can successfully complement for lack of chromosomally encoded *E. coli* DnaJ and CbpA

Plasmids pBAD22A (negative control) and pWKG (pBAD22A containing the *dnaJ* gene) were transformed into *E. coli* OD259 (Deloche *et al.*, 1997), and their ability to complement for lack of DnaJ and CbpA was investigated under stress conditions 40°C (with and without the inducer arabinose). The dilution series is indicated on the left.

At 40°C and in the presence of arabinose, in order to induce the production of plasmid protein, only *E. coli* OD259 containing pWKG90 could successfully grow, whereas *E. coli* OD259 containing pBAD22A could not. DnaJ was therefore a requirement for growth of *E. coli* at temperatures above 37°C. The ability of the controls to show that a DnaJ protein was required above 40°C made this a suitable system for investigating the function of *Agt* DnaJ *in vivo*.

3.3.7 Complementation by *Agt* DnaJ in *E. coli* OD259

Figure 3.24 shows that *Agt* DnaJ could complement for the lack of *E. coli* DnaJ and CbpA. At 40°C, *E. coli* OD259 containing pRJ30 could grow successfully, whereas *E. coli* OD259

containing pQE30 or pRJ30-H33Q could not. However, presence or absence of the inducer IPTG made no difference to the results. This was possibly due to the low basal levels of *Agt* DnaJ being produced in the absence of inducer being sufficient to allow for complementation.

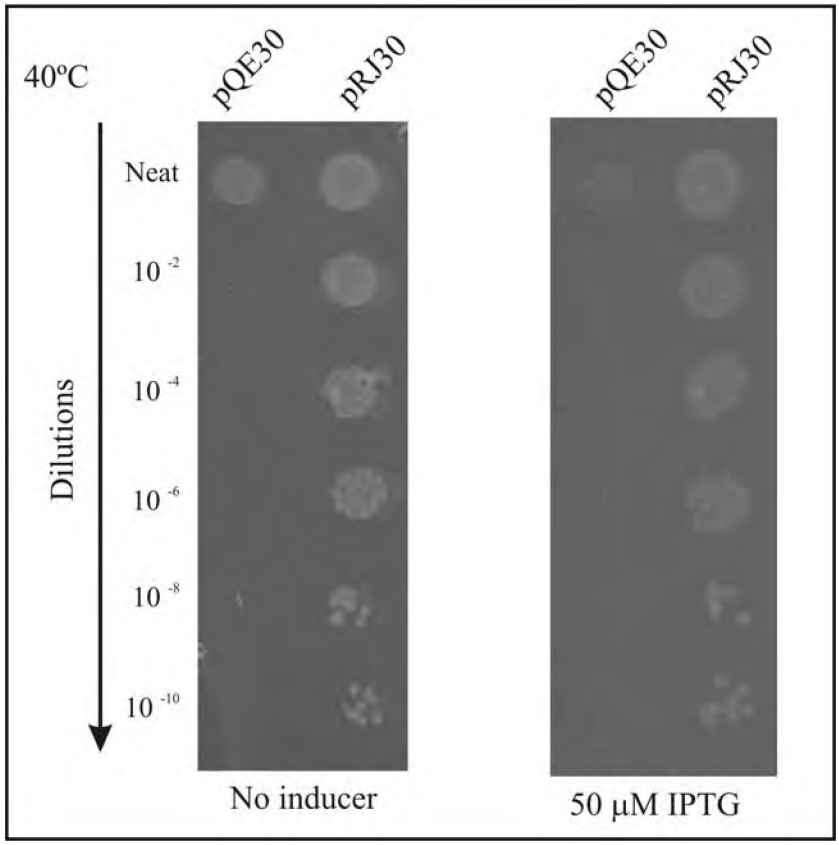

Figure 3.24: Plasmid encoded A*gt* DnaJ could successfully complement for the lack of chromosomally encoded *E. coli* DnaJ and CbpA

Plasmids pQE30 (negative control), pRJ30 (pQE30 containing the *Agt dnaJ* coding region) and pRJ30-H33Q (expressing *Agt* DnaJ-H33Q) were transformed into *E. coli* OD259, and their ability to complement for lack of *E. coli* DnaJ and CbpA was investigated under stress conditions at 40°C (with and without the inducer 50 mM IPTG). The dilution series is indicated on the left.

All *E. coli* OD259 transformants could grow successfully at 30°C regardless of genotype (data not shown). *Agt* DnaJ therefore must be able to form a productive interaction with *E. coli* DnaK, as it was able to complement for the lack of *E. coli* DnaJ and CbpA, whereas *Agt* DnaJ-H33Q could not. This implies that specificity determinants necessary for the interaction of *E. coli* DnaJ with *E. coli* DnaK are potentially present in *Agt* DnaJ. *Agt* DnaJ could be detected by chemiluminescent-based immunodetection in *E. coli* OD259 (Figure 4.5).

3.3.8 Complementation by Type III DnaJ-like proteins in *E. coli* OD259

Plasmids encoding Type III DnaJ-like proteins from *A. tumefaciens* (pC1-30, pC2-30 and pC5-30) encoding *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5, as well as pRJ30 (encoding *Agt* DnaJ) were transformed into *E. coli* OD259. The ability of the proteins encoded by these plasmids to replace *E. coli* DnaJ was assessed. The results are shown in Figure 3.25.

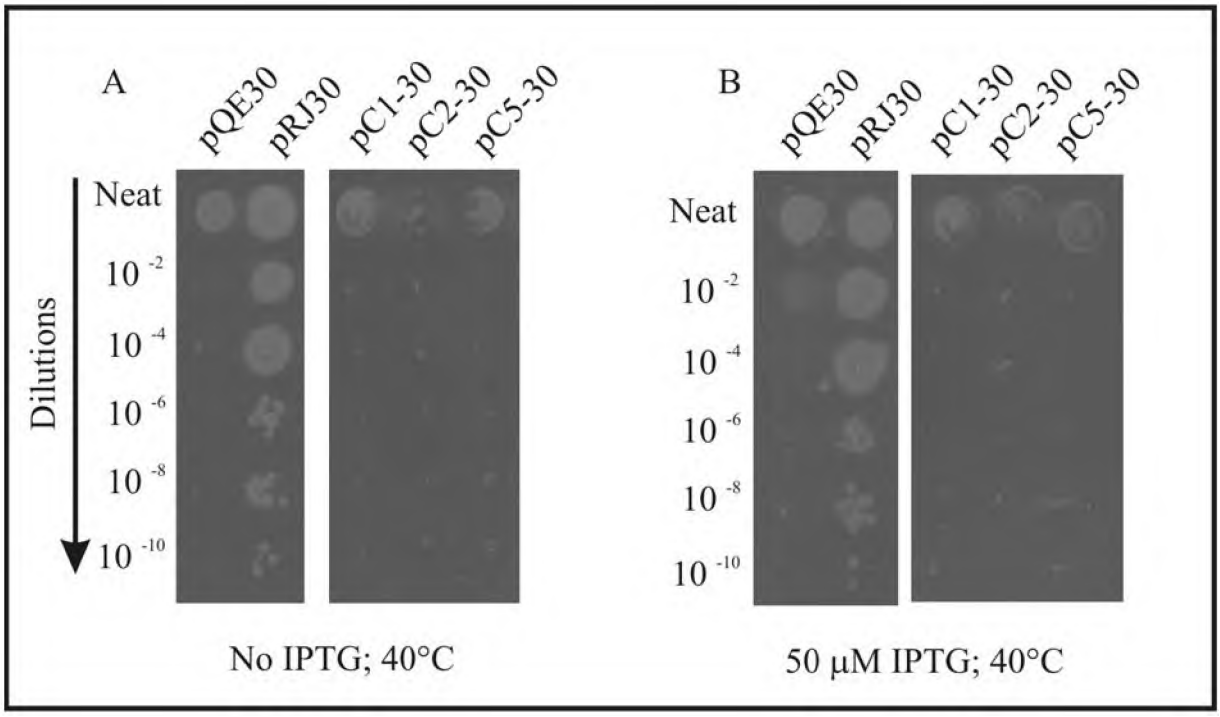

Figure 3.25: Type III DnaJ-like proteins from *A. tumefaciens* could not replace the lack of chromosomally encoded *E. coli* DnaJ and CbpA

Plasmids encoding Type III DnaJ-like proteins from *A. tumefaciens* (pC1-30, pC2-30 and pC5-30 encoding *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5 respectively) as well as pQE30 and pRJ30 encoding *Agt* DnaJ were transformed into *E. coli* OD259. The ability of the proteins produced on these plasmids to replace *E. coli* DnaJ was assessed under stress conditions (40°C). Complementation was performed in (A) the absence of IPTG and (B) the presence of 50 μM IPTG. The plasmids present in each *E. coli* OD259 transformant are indicated above the figure. The dilution series used is indicted to the left.

Agt DnaJ could replace the lack of *E. coli* DnaJ and CbpA, however the Type III DnaJ-like proteins *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5 were unable to do so, in the absence and presence of IPTG. The lack of complementation is therefore not due to an error in the pQE30 construct. *Agt* DjC1, *Agt* DjC2 and *Agt* DjC5 were also not detected using chemiluminescent-based immunodetection using an anti-His antibody (data not shown). Hence to determine if

the J domains from these proteins could interact productively with *E. coli* DnaK, implying that lack of complementation was due to an inability of the remainder of the protein to substitute for the C terminus of *E. coli* DnaJ, the J domains from these proteins were used to substitute for the J domain from *Agt* DnaJ.

3.3.9 Complementation by chimeric *Agt* DnaJ proteins in *E. co li* OD259

J domains from the Type III DnaJ-like proteins present in *A. tumefaciens* RUOR were successfully inserted into pRJ30-F74, replacing the original J domain. The resulting plasmids were transformed into *E. coli* OD259, and the ability of the chimeric proteins to substitute for the lack of *E. coli* DnaJ and CbpA was monitored. The results are shown in Figure 3.26.

Agt DnaJ containing the silent mutation in the codon for Phe^{74} expressed off plasmid pRJ30-F74 was able to complement successfully for the lack of *E. coli* DnaJ and CbpA. However, chimeric *Agt* DnaJ proteins containing J domains from other *Agt* DnaJ-like proteins were unable to complement for the lack of *E. coli* DnaJ and CbpA. This was therefore not due to any errors in the pRJ30-F74 construct, and the correct insertion of the J domain coding region into pRJ30-F74 was validated by sequencing. All *E. coli* OD259 transformants, regardless of genotype, were able to grow at 30°C. Again, none of these chimeric proteins were detectable using chemiluminescent-based immunodetection (data not shown) implying a low level of protein being produced, which would potentially be below the detection limits of the system used.

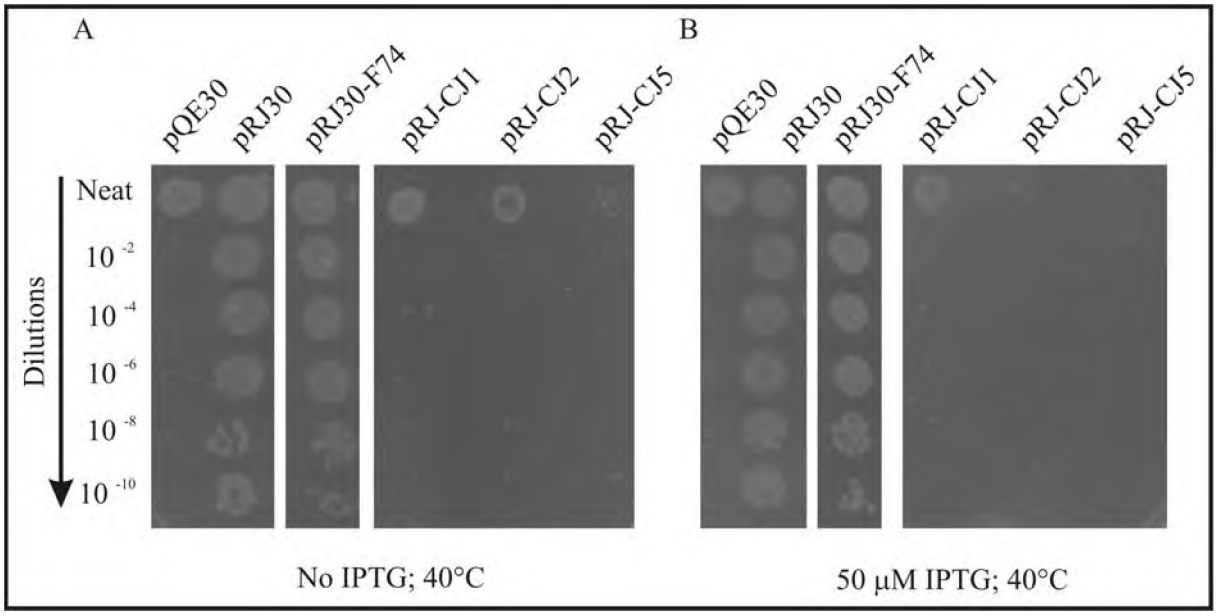

Figure 3.26: Chimeric *Agt* DnaJ proteins could not replace the lack of chromosomally encoded *E. coli* DnaJ and CbpA

Plasmids encoding chimeric *Agt* DnaJs containing Type III J domains from *A. tumefaciens* (pRJ-CJ1, pRJ-CJ2 and pRJ-CJ5 encoding *Agt* DnaJ-CJ1, *Agt* DnaJ-CJ2 and *Agt* DnaJ-CJ5 respectively) as well as pQE30, pRJ30 and pRJ30-F74 were transformed into *E. coli* OD259. The ability of the proteins produced on these plasmids to replace *E. coli* DnaJ was assessed. Complementation was performed in (A) the absence of IPTG and (B) the presence of 50 mM IPTG. The plasmids present in each *E. coli* OD259 are indicated above the figure. The dilution series used is indicted to the left.

3.3.10 Purification of Histidine-tagged *Agt* DnaJ, *Agt* DnaJ-H33Q and *Agt* DnaK

Agt DnaJ, *Agt* DnaJ-H33Q and *Agt* DnaK were produced and purified from *E. coli* XL1-Blue [pRJ30], *E. coli* XL1-Blue [pRJ30-H33Q] and *E. coli* XL1-Blue [pRK30], respectively. Induction levels were generally low for *Agt* DnaJ and *Agt* DnaJ-H33Q, but high for *Agt* DnaK, and purification was achieved using Nickel affinity chromotography. Purification products are shown below in Figure 3.27.

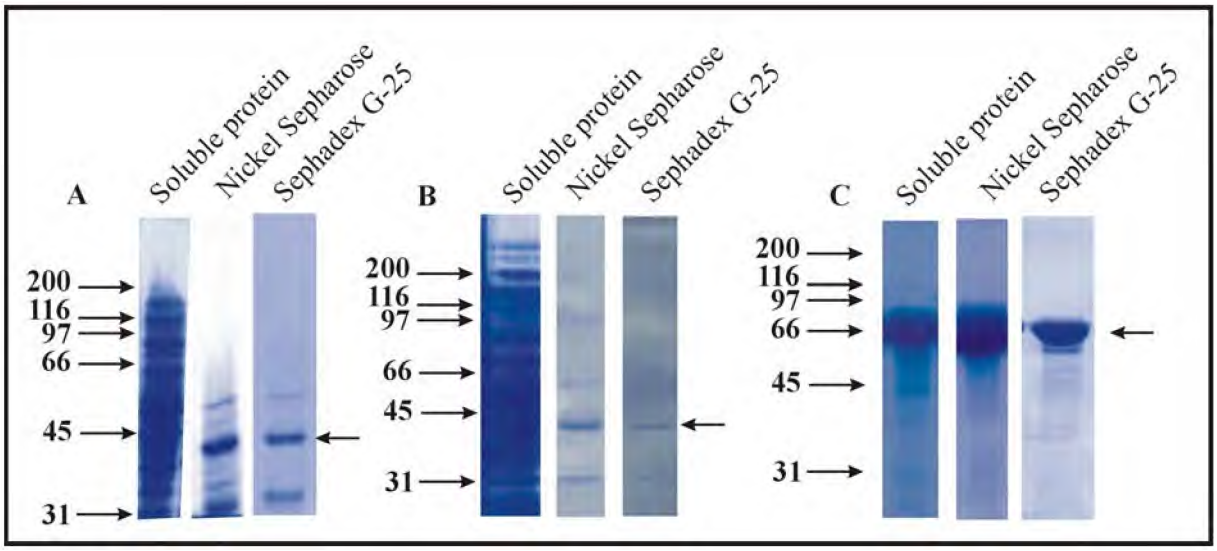

Figure 3.27: Purification of histidine-tagged *Agt* DnaJ, *Agt* DnaJ-H33Q and *Agt* DnaK using nickel affinity chromotography

(A) *Agt* DnaJ was purified from *E. coli* XL1-Blue [pRJ30], and the purification analysed using 12 % SDS-PAGE. Soluble protein represents the amount of protein loaded onto the nickel Sepharose column, nickel Sepharose shows the protein present in a typical elution, and Sephadex G-25 shows the protein recovered after buffer exchange. (B) *Agt* DnaJ-H33Q was purified from *E. coli* XL1-Blue [pRJ30-H33Q], and the purification analysed using 12 % SDS-PAGE. Soluble protein represents the amount of protein loaded onto the nickel Sepharose column, nickel Sepharose shows the protein present in a typical elution, and Sephadex G-25 shows the protein recovered after buffer exchange. (C) *Agt* DnaK was purified from *E. coli* XL1-Blue [pRK30], and the purification analysed using 12 % SDS-PAGE. Soluble protein represents the amount of protein loaded onto the nickel Sepharose column, nickel Sepharose shows the protein present in a typical elution, and Sephadex G-25 shows the protein recovered after buffer exchange. Molecular markers are indicated in kDa.

There was no obvious over induction band in the total protein sample for *Agt* DnaJ and *Agt* DnaJ-H33Q (Figure 3.27 A and B). In contrast to *Agt* DnaJ, *Agt* DnaK was produced to high levels in *E. coli* XL1-Blue, allowing for a relatively simple purification (Figure 3.27 C). Minimal *Agt* DnaJ and *Agt* DnaJ-H33Q was lost during the wash steps indicating a high level

of binding to the nickel Sepharose beads. However, a high proportion of DnaK appeared to be insoluble, although this could be due to the high levels of protein present in the cell causing some *Agt* DnaK to aggregate out of solution. Equally, a significant proportion of *Agt* DnaK did not bind to the column, and was washed off during the wash steps. Lower molecular mass products were seen during the elution steps for *Agt* DnaJ and *Agt* DnaJ-H33Q (Figure 3.27 A and B). These are likely to be break down products, containing a His tag, as they bind specifically to the nickel Sepharose, and are not released during the wash steps. Minimal protein remained bound to the beads. High levels of *Agt* DnaK were obtained after elution with imidazole, and higher levels remained bound to the beads as compared to the purifications for *Agt* DnaJ and *Agt* DnaJ-H33Q.

Subsequent to the purificati on using nickel Sepharose, imidazole was removed from the protein using buffer exchange on a Sephadex G-25 column. The protein typically eluted as a single sharp peak for *Agt* DnaJ and *Agt* DnaJ-H33Q, and in several peaks for *Agt* DnaK, with the imidazole eluting in subsequent elutions. This was confirmed by following the elutions using both Bradford determinations for protein, and $_{A280}$ readings for both imidazole and protein, as the imidazole ring absorbs at 280 nm. Protein samples were concentrated if necessary, although this was generally not required for *Agt* DnaK. Protein concentrations in the samples eluted from the Sephadex G-25 column for *Agt* DnaJ and *Agt* DnaJ-H33Q typically ranged between 2 to 4 μM, which was approximately 1 mg of *Agt* DnaJ from every litre of culture. Protein concentrations in the samples eluted from the Sephadex G-25 column for *Agt* DnaK typically ranged between 12 to 15 μM, which was approximately 3 mg of *Agt* DnaK from every litre of culture, with some purifications giving concentrations of between 40 to 50 μM in the samples eluted from the Sephadex G-25 column. .

3.3.11 Determination of kinetic constants for *Agt* DnaK

The basal specific ATPase activity of *Agt* DnaK was kinetically characterised for three independent batches of purified protein. The $_{Vmax}$ was determined to be 1.3 nmol phosphate released per minute per mg DnaK, and the K_m was determined to be 62.0 μM. Graphical analysis of the determination of the kinetic constants is shown in Figure 3.28.

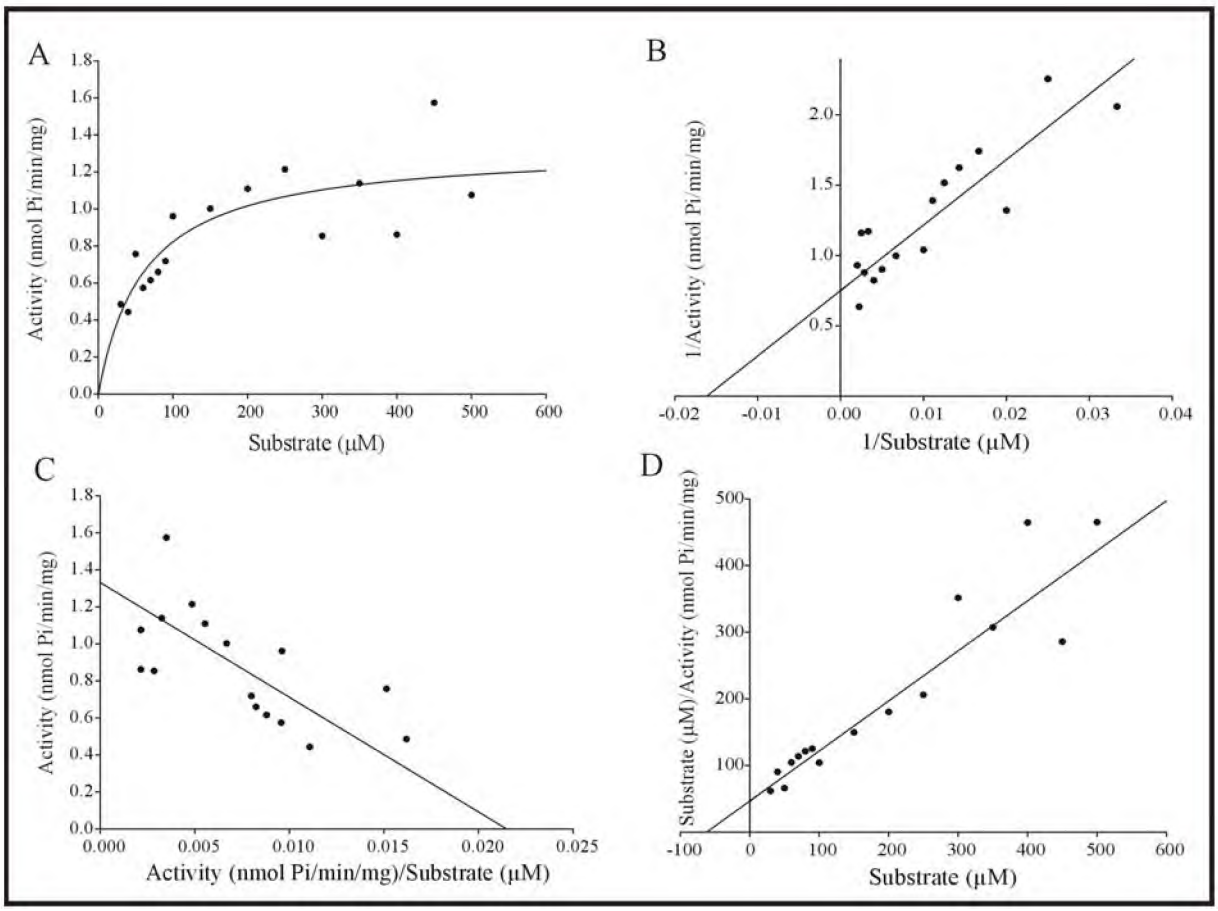

Figure 3.28: Enzyme kinetic plots for the determination of kinetic constants for *Agt* DnaK basal specific ATPase activity

(A) Michaelis-Menten, (B) Lineweaver-Burk, (C) Eadie-Hofstee and (D) Hanes-Wolf plots were derived from three independent batches of purified *Agt* DnaK (0.4 μM) by measuring the rate of phosphate release for varying ATP concentrations (0 to 500 μM). Each data point represents the average of three independent experiments.

3.3.12 Effect of *Agt* DnaJ and *Agt* DnaJ-H33Q on the stimulation of the basal ATPase activity of *Agt* DnaK

The stimulation of the basal ATPase activity of *Agt* DnaK by *Agt* DnaJ was monitored. Low levels of stimulation were observed, ranging from 1.5 times to 2 times the basal ATPase activity using 2 μM DnaJ. This was consistent with results obtained for the *E. coli* system, where low levels of DnaK stimulation by DnaJ were observed in the absence of GrpE or peptides.

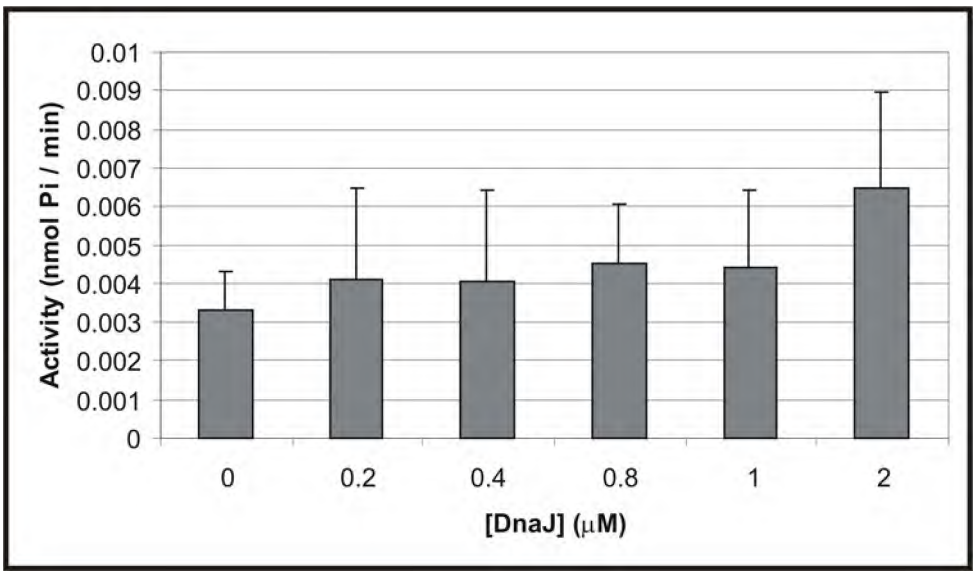

Figure 3.29: Stimulation of the ATPase activity of *Agt* DnaK by *Agt* DnaJ

Agt DnaK (0.4 μM) was incubated with varying concentrations of *Agt* DnaJ (0 to 2 μM) in the presence of 400 μM ATP, and the rate of phosphate release was monitored colourimetrically. The assay was performed in triplicate, and standard deviations are given.

The effect of *Agt* DnaJ-H33Q on the ATPase activity of *Agt* DnaK was monitored. The results are shown in Figure 3.30. *Agt* DnaJ-H33Q was incapable of stimulating the ATPase activity of *Agt* DnaK, whereas *Agt* DnaJ could stimulate by approximately 1.5 to 2 fold. This was consistent with results obtained for the *E. coli* system.

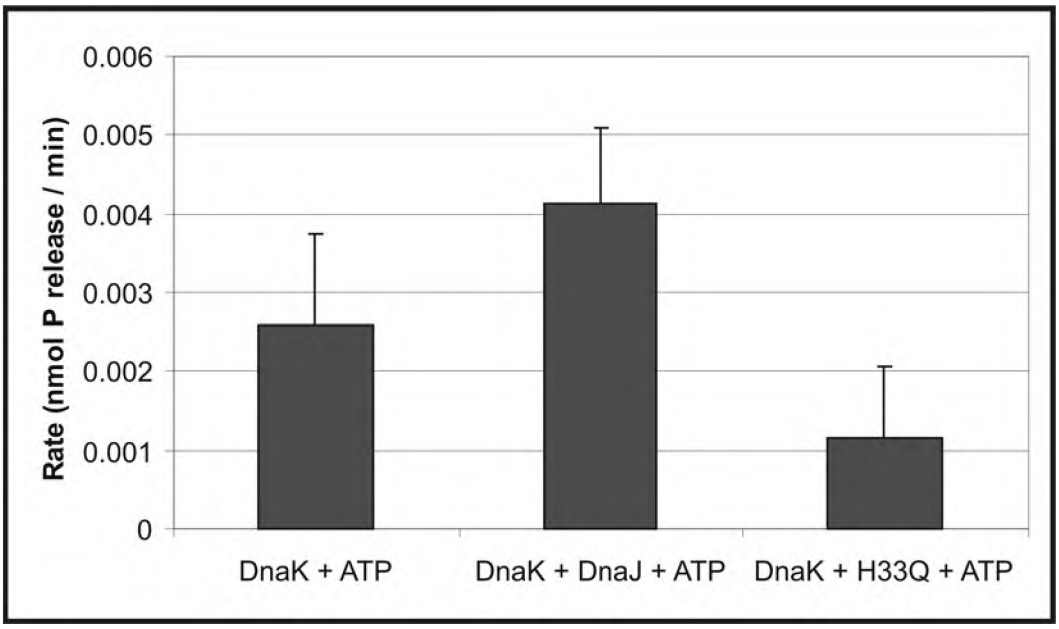

Figure 3.30: *Agt* DnaJ-H33Q was unable to stimulate the ATPase activity of *Agt* DnaK

Agt DnaK (0.4 μM) was incubated with *Agt* DnaJ (0.4 μM) and *Agt* DnaJ-H33Q (0.4 μm) in the presence of 400 μM ATP, and the rate of phosphate release was monitored colourimetrically. Three independent experiments were performed, and the average rate and standard deviation was calculated.

3.4 DISCUSSION

The coding region encoding the *Agt* DnaJ protein, as well as coding regions for three Type III DnaJ-like proteins from *A. tumefaciens* were successfully isolated and inserted into pQE30.

3.4.1 *Agt* DnaJ

Agt DnaJ was a Type I DnaJ-like protein, with a significant degree of identity to *E. coli* DnaJ (56.7 %). It was found in an operon with the *A. tumefaciens* DnaK (*Agt* DnaK), implying a close interaction between these two proteins similar to that found in *E. coli*. It appeared to be structurally similar to *E. coli* DnaJ, and was therefore likely to have similar functions, and act as a classical DnaJ protein.

3.4.2 *Agt* DjC1/DjlA

Agt DjC1 was the homologue to the DjlA protein found in *E. coli*, and other prokaryotes. It was similar to *E. coli* DjlA, with the exception of a poly glutamine rich stretch prior to the J domain that was not present in *Agt* DjC1. The encoded protein was likely to interact functionally with *Agt* DnaK, and have similar functions to *E. coli* DjlA. These included synthesis of colanic acid and synthesis of the colanic acid mucoid capsule (Clarke *et al.*, 1997; Kelley and Georgopoulos, 1997a), and aiding in *E. coli* growth, particularly in the absence of DnaJ (Genevaux *et al.*, 2001a).

3.4.3 *Agt* DjC2

Agt DjC2 was a 40 kDa DnaJ-like protein, but did not contain a Gly/Phe rich region or cysteine rich repeats. Its J domain bore the highest degree of identity and similarity to the J domain from *Agt* DnaJ, yet there was limited similarity to *Agt* DnaJ over the rest of the protein. The protein with the highest degree of similarity to *Agt* DjC2 was found in *S. melitoti* with few other proteins being identified with similarity to *Agt* DjC2. It could act as a CbpA-like protein, although there was low sequence identity between these two proteins (25.5 % identity) with the bulk of the similarity occurring in the N and C terminal portions of the proteins, with limited similarity occurring in the central portions of the proteins. The level of

similarity in the C terminal portions could allow *Agt* DjC2 to act as a replacement for *Agt* DnaJ under certain circumstances, and this role could possibly include limited chaperone activity.

3.4.4 *Agt* DjC5

Agt DjC5 was a small protein with a predicted molecular mass of approximately 23 kDa. Homologues to this protein were found in several organisms such as *S. melitoti*, as well as in *P. dentrificans*, and *B. melitensis*, but not in *E. coli*. The levels of conservation for these proteins were higher than seen for homologues for *Agt* DjC2, suggesting that *Agt* DjC5 homologues will be present in a variety of organisms, whereas *Agt* DjC2 homologues will specifically be found in *A. tumefaciens*, and a few related organisms. The type of DnaJ-like proteins in specific organisms may be related to the environment the organism is located in, and may provide a means of allowing specific organisms to function in specific environments, by targeting a partner Hsp70 to specific sites and functions.

3.4.5 *Agt* DjC3 and *Agt* DjC4

The coding regions for the DnaJ-like proteins *Agt* DjC3 and *Agt* DjC4 were not amplified. This may imply that the genes for these proteins are not present on the *A. tumefaciens* RUOR genome. This was likely the case for the *Agt djC4* gene. This was located on an *A. tumefaciens* C58 plasmid. The plasmid sequences varied considerably between *A. tumefaciens* strains, and proteins produced off the plasmid are often involved in the production of opines involved in cellular transformation (Goodner *et al.*, 2001; Wood *et al.*, 2001). Hence *Agt* DjC4 may be playing a role in cellular transformation, a function not present in the *A. tumefaciens* RUOR strain. Alternatively there may be additional DnaJ-like proteins present on the plasmids present in *A. tumefaciens* RUOR. The presence or absence of these genes could be detected using Southern analysis.

The lack of a coding region for *Agt* DjC3 was surprising. It was located on the circular chromosome of *A. tumefaciens*, and this chromosome seems relatively well conserved between *A. tumefaciens* C58 and *A. tumefaciens* RUOR. The likelihood was that *Agt djC3* was a pseudogene in *A. tumefaciens* C58, which has been lost from the *A. tumefaciens* RUOR

genome, or has been altered to such an extent that the primers designed would not anneal to the coding region under the conditions employed. The possibility also existed that the gene for *Agt djC4* has also been altered, and hence the primers designed would not anneal to the coding region under the conditions employed.

3.4.6 Sequence comparison of the J domains of the Type III *A. tumefaciens* DnaJ-like proteins

The J domains of the DnaJ-like proteins present in *A. tumefac iens* C58 were aligned, and the alignment is shown in Figure 3.31. There was a high degree of similarity and identity in the J domains between *Agt* DjC2 and *Agt* DjC3, and between *Agt* DjC4 and *Agt* DjC5, possibly indicating ancestral duplication events. The data are given in Table 3.3.

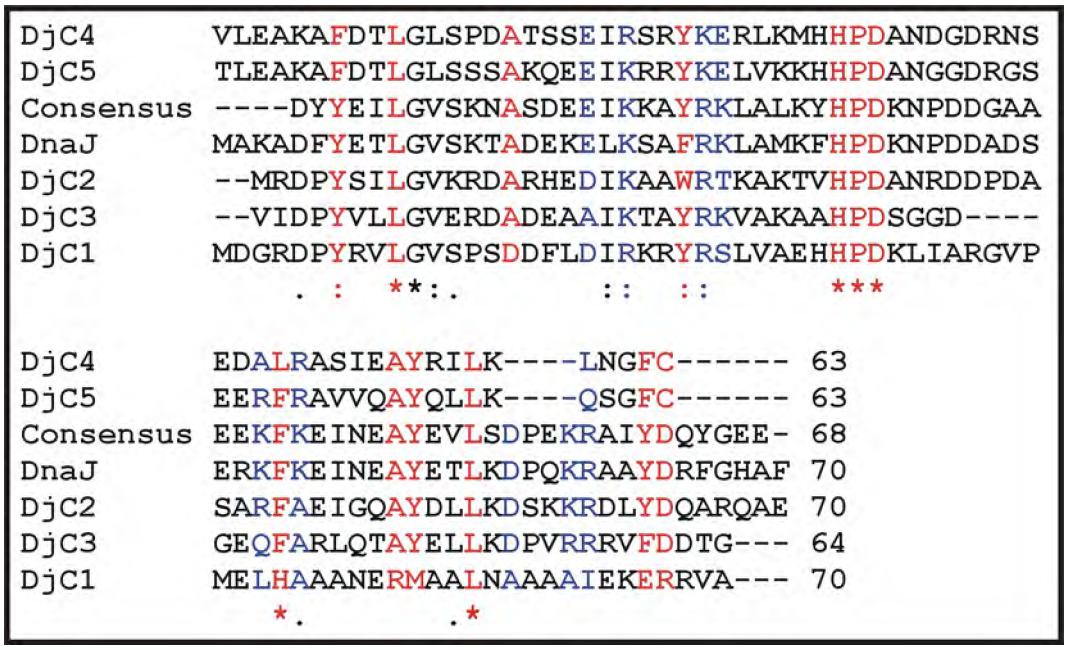

Figure 3.31: Alignment of the J domains from the DnaJ-like proteins found in *A. tumefaciens* C58

The alignment was performed using ClustalW (Thompson *et al.*, 1994), and includes the J domains from all the identified *A. tumefaciens* C58 DnaJ-like proteins, as well as the derived overall consensus sequence obtained previously. Amino acids in red correspond to those positions that are highly conserved in the overall consensus sequence; amino acids in blue correspond to those positions that contain charged, conserved amino acids in the overall consensus sequence. Asterisks indicate an identical match and colons and full stops indicate partial matches.

Table 3.3: Percentage similarity and identity between the J domains of the DnaJ-like proteins from *A. tumefaciens* C58

		Identity					
	Protein	**DnaJ**	**DjC1**	**DjC2**	**DjC3**	**DjC4**	**DjC5**
Similarity	**DnaJ**		27.3	43.1	40.0	30.0	32.9
	DjC1	58.4		31.6	27.3	26.7	29.6
	DjC2	63.9	49.4		41.4	22.2	30.6
	DjC3	62.9	51.9	71.4		23.2	30.0
	DjC4	52.9	44.0	47.2	56.5		63.5
	DjC5	61.4	50.7	52.8	55.7	88.3	

Percentage identity is given in the upper portion of the table; percentage similarity is given in the lower portion. The occurrences of greater than 40 % identity and greater than 70 % similarity between two sequences are highlighted.

However comparison of the full length *Agt* DjC2 and *Agt* DjC3 showed limited similarity over the rest of the protein (18.9 % identity). By contrast, *Agt* DjC4 and *Agt* DjC5 shared 48.5 % identity over the full length protein. Hence it was likely that an ancestral gene on the *A. tumefaciens* linear chromosome was duplicated to give rise to both *Agt* DjC4 on the plasmid, and *Agt* DjC5 on the linear chromosome. This may explain the apparent absence of *Agt* DjC4 from the *A. tumefaciens* RUOR genome. The similarity o f *Agt* DjC2 and *Agt* DjC3 in the J domain may be the result of a localised duplication of the J domain.

3.4.7 Ability of *Agt* DnaJ-like proteins to complement for lack of *E. coli* DnaJ and CbpA

Agt DnaJ was able to replace functionally DnaJ and CbpA in the mutant *E. coli* OD259 strain. This implies that *Agt* DnaJ must be able to interact with *E. coli* DnaK, and that differences between *E. coli* DnaJ and *Agt* DnaJ in the C terminal portion of these proteins do not hamper the ability of *Agt* DnaJ to function in *E. coli*. *Agt* DnaJ and *Agt* DnaK would also therefore interact similarly to *E. coli* DnaK and DnaJ.

Three Type III DnaJ-like proteins were isolated from *A. tumefaciens* RUOR, DjC1/DjlA, DjC2 and DjC5. DjC2 and DjC5 are novel proteins that have not been investigated prior to this. They potentially form a partnership with a second DnaK protein in *A. tumefaciens*.

None of these Type III proteins could replace DnaJ and CbpA in *E. coli* OD259. *Agt* DjC1 and DjC5 both contain C terminal J domains, and while *Agt* DjC2 contains an N terminal J domain, the remainder of its C terminus differs markedly from *E. coli* DnaJ and CbpA. Hence any regions in the C terminus of DnaJ and CbpA that are critical for normal cellular functioning are unlikely to be present in these proteins, which could explain their inability to replace DnaJ and CbpA in *E. coli* OD259. However, the presence of the trans-membrane region appeared to negatively impact on the ability of DjlA to complement for lack of DnaJ in a *E. coli dnaJ, djlA* strain (Genevaux *et al.*, 2001a). This region also impacted on the levels of produced *E. coli* DjlA. Removal of the trans-membrane region from *Agt* DjC1 may allow for production of higher levels of *Agt* DjC1 in *E. coli,* and therefore a potential ability to replace DnaJ and CbpA in *E. coli* OD259. The lack of the ability of these proteins to replace the lack of *E. coli* DnaJ and CbpA may be due to three possibilities. Firstly, the complete functioning of *E. coli* DnaJ requires residues in its C terminus, which these proteins lack, secondly, these proteins are unable to interact with *E. coli* DnaK, or thirdly, a combination of these two possibilities. It was known that full length *E. coli* DjlA can not complement for the lack of *E. coli* DnaJ and CbpA, so this result was not surprising (Genevaux *et al.*, 2001a; Genevaux *et al.*, 2001b). Equally *Agt* DjC5 was a small protein with a C terminal J domain, and was unlikely to contain determinants in the remainder of the protein to fulfil the *E. coli* DnaJs functions. *Agt* DjC2 was a similar size to *E. coli* DnaJ, but has minimal similarity to DnaJ in the non-J domain containing regions of the protein.

The level of specificity in their J domains was then also investigated, by replacing the J domain from *Agt* DnaJ, with the J domains from these proteins. These chimeric *Agt* DnaJ proteins were also unable to replace DnaJ and CbpA in *E. coli* OD259. This points to a high level of sequence specificity within the J domain. The *Agt* DnaJ J domain bears greater than 50 % similarity to all the J domains, yet this level of similarity, which implies a high degree of structural similarity, was not sufficient to give rise to functional chimeras under the system used here. The majority of the conserved amino acids highlighted in this thesis are conserved in all the J domains, with the exception of residues in Helix IV. The DjC1/DjlA J domain was the most disparate J domain when compared to the other J domains present in *A. tumefaciens* C58 (Figure 3.31 and Table 3.3). This low level of similarity may explain the inability of the chimeric protein to act like *Agt* DnaJ.

For the DjC1/DjlA J domain, this was a surprising finding. *E. coli* DjlA interacts with DnaK, and a chimeric *E. coli* DnaJ containing the J domain from DjlA was functional (Genevaux *et al.*, 2001b). This may imply a high degree of specificity in the DnaK-DjlA partnership, which will prevent DjlA proteins from interacting generically with DnaKs. The inability of the J domains from *Agt* DjC2 and *Agt* DjC5 to substitute for the *Agt* DnaJ J domain was not as surprising. These proteins potentially interact with an additional Hsp70 present in *A. tumefaciens*, and may not be able to interact with *E. coli* DnaK.

The non-detection of the Type III proteins and the chimeric *Agt* DnaJ proteins could imply that the proteins are unstable in *E. coli*, and produced at low levels. Hence levels of the proteins may have been below the detection limit of the Western analysis system employed. The technique used to detect the level of protein production in *E. coli* OD259 was of necessity an indirect technique, as there was no easy way of detecting levels of protein production in *E. coli* OD259 under stress conditions. *E. coli* OD259 containing plasmids that express proteins that do not successfully complement for the lack of *E. coli* DnaJ and CbpA will not grow successfully at 40°C. Hence obtaining representative cellular extracts of these *E. coli* OD259 strains would be problematic, in order to ascertain expression of the DnaJ-like protein under investigation. Alternatively performing colony Western analysis using colonies that do grow, and detecting Histidine-tagged proteins using immunodetection, may indicate levels of protein present; again however lack of growth at 40°C would prevent such an analysis. Analysis of protein production in another *E. coli* strain could also be performed. However, that would only imply that a specific protein was capable of being produced in that strain. It was also possible that some of the chimeric proteins are capable of being produced at high levels at 30°C, but not at 40°C. The fact that *Agt* DnaJ and *Agt* DnaK are produced at such varying levels from the same vector and promoter system type implies variability in protein production, despite consistency in experimental set up. DnaJ-like proteins appear to be consistently produced at low levels, and this may explain the lack of detection of the *Agt* Type III DnaJ-like proteins. Over expression of full length DnaJ in *E. coli* leads to a decrease in cell viability (Al-Herran and Ashraf, 1998), implying that high levels of DnaJ production has dire consequences for the cell. There were high differences between the amounts of *E. coli* DnaK and DnaJ produced, with there being approximately 10 fold more DnaK (Bardwell *et*

al., 1986), and there was an estimated ratio of 10:1:3 DnaK:DnaJ:GrpE (Pierpaoli *et al.*, 1998b).

An alternative explanation for the lack of detection was that a monoclonal anti-His antibody was used for immunodetection. It was possible that, due to varying conformations adopted by the different DnaJ-like proteins, the tag was inaccessible to the antibody, even after denaturing treatment with SDS and heat. Hence the antibody may not recognise the tag, and immunodetection would not be successful. However this antibody has been used successfully to detect *Agt* DnaJ, and additional Histidine tagged proteins.

3.4.8 Kinetic constants for the ATPase activity of *Agt* DnaK

E. coli DnaK has differing reported kinetic values including a $_{Vmax}$of 3.5 nmol Pi/min/mg DnaK, and a K_m of 20 μM (Liberek *et al.*, 1991), and an alternative K_m of 1 μM, and a calculated $_{Vmax}$ 0.43 nmol Pi/min/mg DnaK (Kamath-Loeb *et al.*, 1995). Hence, while the low $_{Vmax}$ that we have obtained for our DnaK was consistent with previous reports for *E. coli* DnaK, our K_m was higher. Differences between *Agt* DnaK and *E. coli* DnaK in terms of their kinetic constants can be attributed to two differing possibilities. Firstly, the histidine tag may be affecting the basal ATPase activity as has been seen for the murine protein BiP (Chevalier *et al.*, 1998). Secondly, *A. tumefac iens,* or specific strains of *A. tumefaciens,* may require a DnaK protein with a lower affinity for ATP than the *E. coli* DnaK. This could be important physiologically due to the differing en vironments in which *E. coli* and *A. tumefaciens* reside in. *A. tumefaciens* is a soil-base d organism, whereas *E. coli* resides predominately in the gut.

Agt DnaJ was able to stimulate the ATPase activity of *Agt* DnaK. The stimulation was approximately two fold at a molar excess of *Agt* DnaJ. This was consistent with previously obtained results for *E. coli* DnaK. In order to optimise stimulation, it will be necessary to isolate the *Agt* GrpE protein, and perform the assay with this present, and in the presence of substrate peptides. However, both *Agt* DnaJ and *Agt* DnaK appear to be functional, and appear to have a functional interaction. *Agt* DnaJ-H33Q was unable to stimulate the ATPase activity of *Agt* DnaK. This was consistent with previous results obtained for DnaJ proteins containing mutations in the HPD motif (Tsai and Douglas, 1996). However the low level of impurities detected in the purifications could impact on these results, and additional

purification steps could be attempted. These impurities were unlikely to be breakdown products as they were not detected by the anti-His antibody; however the inclusion of multiple protease inhibitors in future exp eriments may aid in reducing these contaminants if they are break down products.

3.4.9 Comparison of Hsp70s present in *A. tumefaciens* C58

Analysis of the genome sequence of *A. tumefaciens* indicated potential genes for two Hsp70s. The first was for the DnaK protein used in this thesis. The protein with the highest level of similarity to the *A. tumefaciens* DnaK protein was the DnaK from *S. meliloti*. These two proteins are 89.1 % identical (comparison performed using the DnaK from *A. tumefaciens* C58).

The second, termed for this thesis *Agt* DnaK2, has similarity to *Agt* DnaK over the ATPase domain, but low similarity over the rest of the pro tein. It was a 434 amino acid protein, which was markedly smaller than other Hsp70s. Its substrate binding region was smaller than other Hsp70s. Homologues are found in the closely related species *S. meliloti* and *Bradyrhizobium japonicum* (Figure 3.32). It was possible that DnaK2 will interact with *Agt* DjC2 and *Agt* DjC5, whereas DnaK will interact with DnaJ and *Agt* DjC1/DjlA.

There do not appear to be homologues of the *E. coli* proteins Hsc66 or HscC encoded in the *A. tumefaciens* C58 genome, which was consistent with there being no apparent homologues to the DnaJ-like proteins Hsc20 (partner to Hsc66), and DjlB and DjlC (partners to HscC). There also does not appear to be a Type II equivalent to CbpA, although this role could be performed by *Agt* DjC2. This could imply that a knock out of *Agt* DnaJ could be lethal to the *A. tumefac iens* cell, as CbpA can compensate for loss of *E. coli* DnaJ (Ueguchi *et al.*, 1994).

There also appears to be the formation of distinct partnerships within different organisms. The lack of the HscC – DjlB/DjlC and the Hsc66 – Hsc20 partnerships in *A. tumefaciens* implies that these partnerships play a specific role in *E. coli* functioning, that does not occur in *A. tumefaciens.* For example the Hsc66 – Hsc20 partnership was implicated in protection against cold shock in *E. coli* (Lelivelt and Kawula, 1995); however, one would expect a soil based organism also to require a means of protection against cold shock. Presumably other

systems, non-Hsp70 based, have evolved to provide this protection. The generation of a second type of partnerships in *A. tumefaciens* was of interest, as the *Agt* DnaK2 appears to be markedly different to other Hsp70s, as its substrate binding region was significantly shorter. This system may have evolved in response to the highly specialised nature of the *A. tumefaciens* life cycle.

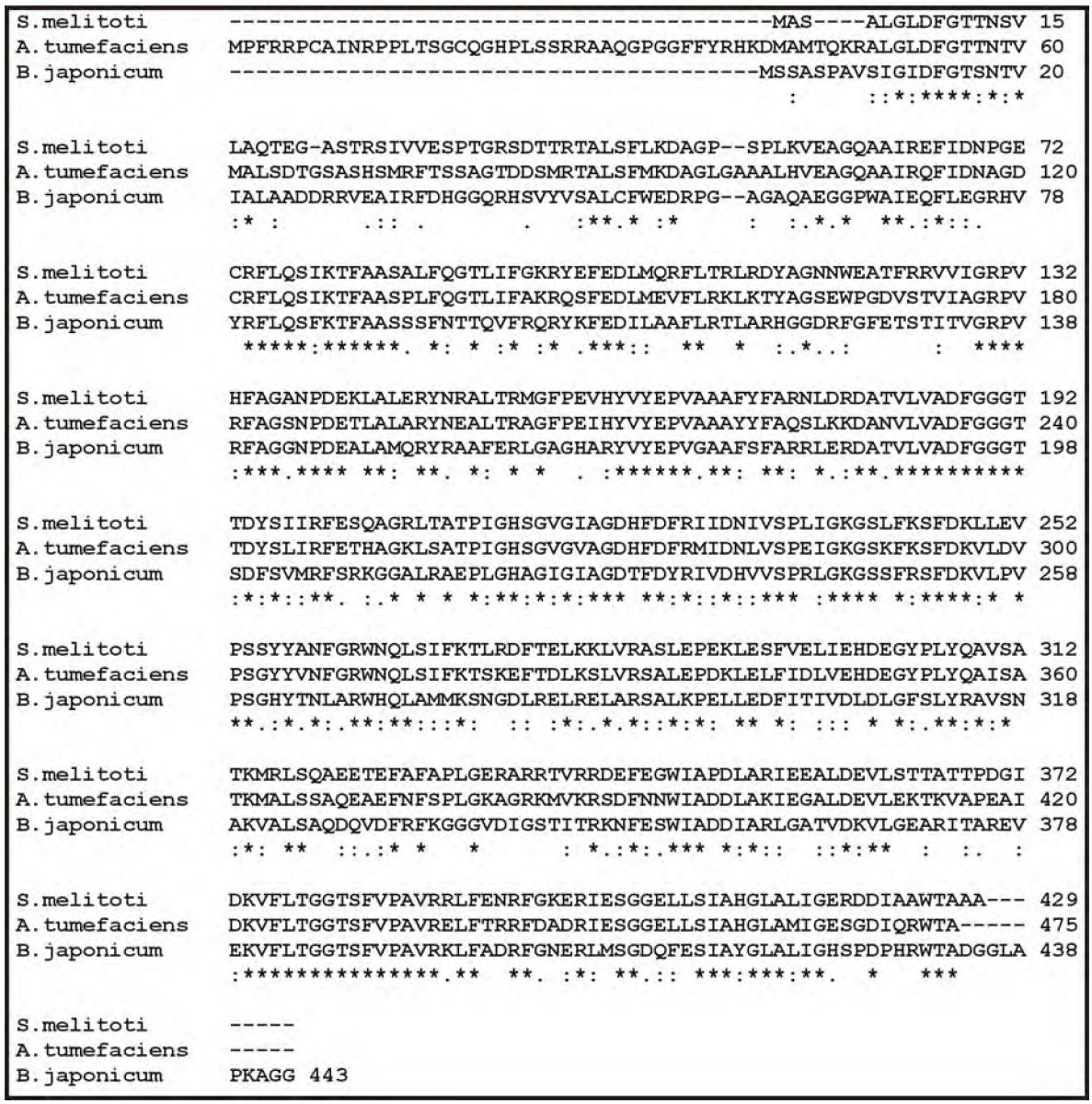

Figure 3.32: Alignment of the DnaK2 proteins from *A. tumefaciens, S. rhizobium* and *B. japonicum*

DnaK2 proteins were identified by a BLAST search of the non redundant database (NCBI) for similar proteins, using *Agt* DnaK2 as a query sequence. Homologous proteins from *A. tumefaciens* C58 (AD2698), *S. melitoti* (NP_385138) and *Bradyrhizobium japonicum* (NP_769792) were aligned. The alignment was performed using ClustalW (Thompson *et al.*, 1994).

3.4.10 Conclusion

In conclusion, the Dn aK and DnaJ partner proteins from *A. tumefaciens* have been isolated and sho wn to have a functional partnership. *Agt* DnaJ was also capable of interacting with *E. coli* DnaK, however additional DnaJ-like proteins from *A. tume faciens* could not do so, nor could chimeric *Agt* DnaJ proteins containing J domains from the Type III DnaJ-like proteins interact with *E. coli* Dna K. This implies a level of specificity with respect to the DnaK-DnaJ interaction. *A. tumefaciens* co ntained two DnaJ-like proteins with well characterised homologues, as well as several DnaJ-like proteins whose func tion had not been previously characterised. The presence of alternative DnaJ-like proteins may allow partner Hsp70s to perform very specific functions, with the conserved Type I protein DnaJ, interacting with DnaK to perform general protein folding functions in the cell.

A m odel system had now been established involving the interaction of *Agt* DnaJ with *E. coli* DnaK in an *in vivo* con text. This allowed for the analysis of conserved residues in the J domain of *Agt* DnaJ.

Chapter 4: Mutagenesis of the *A. tumefaciens* DnaJ J domain

4 MUTAGENESIS OF THE *A. TUMEFACIENS* DNAJ J DOMAIN

4.1 INTRODUCTION

4 .1.1 Effect of substitutions on the J domain

The majority of substitutions performed on the J domain have involved mutations in the HPD motif (Cheetham and Caplan, 1998). Recent work has investigated substitutions in other sections of the J domain (Fewell *et al.*, 2002; Genevaux *et al.*, 2002). A summary of mutations performed in the J domain is given in Table 4.1.

Table 4.1: Effect of substitutions performed in the J domain

Mutation	Helix	Protein	Effect	Ref *
L13M; L13I; L14V	I	T Ag	Functional	17
L13V	I	T Ag	Decrease in viral DNA synthesis	17
L17K	I	Large T antigen	Non-functional – reduction of levels of p130	19
A33R; Y34K	II	Large T antigen	Functional– reduction of levels of p130	19
RE(19,20)AA; RK(22,23)AA; R27A; L28A; A29G; A29G; MK(30,31)AA; Y32A	II	*E. coli* DnaJ	No effect	13
RK(22,23)AA/KR(26,27)AA; Y25A, K26A; KR(26,27)AA; K26E	II	*E. coli* DnaJ	Abolished J domain activity	13
W24R; L29P; M30T; Y34N; A37V; C38R	II	T Ag-Ydj1 chimera	Non-functional – complementation for loss of Ydj1 in yeast	16
Y24A; RR(25,26)AA	II	Hdj1-DnaJ chimera	Abolish function	13
YKR(25-27)AAA	II	*E. coli* DnaJ	Toxic	13
A181T	III	Sec63	Translocation defects	8
F45A	III	Hdj1-DnaJ chimera	Abolish function	13
F46L	III	Ydj1	Lack of complementation	18
F47A	III	*E. coli* DnaJ	Loss of complementation	13
K51E; K53R; M55T; N56D; Y59N	III	T Ag-Ydj1 chimera	Non-functional – complementation for loss of Ydj1 in yeast	16

Mutation	Helix	Protein	Effect	Ref *
KE(41,42)AA; E44A; K46A; KE(48,49)AA; KE(51,52)AA; Y54A; E55A; T58A; TD(58-59)AA	III	*E. coli* DnaJ	No effect	13
K62A; R63A; K62A,R63A	IV	*E. coli* DnaJ	Diminished interaction with DnaK SPR	11
SQ(60,61)AA; KR(62,63)AA; DQ(66,67)AA	IV	*E. coli* DnaJ	No effect	13
D35N	Loop	*E. coli* DnaJ J domain	Diminished interaction with DnaK fluorescence	12
D44N	Loop	T Ag	Non-functional	15
H233Q; D235N	Loop	DjlA-DnaJ chimera	Lack of complementation	5
H32Q	Loop	Hsp40	Inhibition of Hsp70 luciferase refolding	6
H33Q	Loop	Ydj1	No stimulation of ATPase	1
H33Q	Loop	*E. coli* DnaJ	Lack of complementation	4
H33Q	Loop	*E. coli* DnaJ	Lack of interaction with DnaK SPR	9
H33Q	Loop	*E. coli* DnaJ	Stimulation of DnaK lost	10
H33Q; ΔH33; ΔP34; P34F; ΔD35	Loop	*E. coli* DnaJ	Loss of complementation	13
H42Q; D44N	Loop	T Ag	Non-functional	14
H42Q; D44N	Loop	Large T antigen	Non-functional– reduction of levels of p130	19
H42R; D48G	Loop	T Ag-Ydj1 chimera	Non-functional – complementation for loss of Ydj1 in yeast	16
H43Q; D45A	Loop	Csp	No stimulation of Hsc70 ATPase	3
H89Q	Loop	J-MTJ1	No stimulation of BiP ATPase; Poor interaction	7
HPD-AAA	Loop	P58IPK	No stimulation of Hsc70 ATPase	2
K35G	Loop	Hdj1-DnaJ chimera	Marked effects on function	13
N36G	Loop	Hdj1-DnaJ chimera	No effect	13
R36G; N37G; Q38G	Loop	*E. coli* DnaJ	Partial complementation	13
RNQ (36-38) GGG	Loop	*E. coli* DnaJ	Loss of complementation	13

* References: 1 – (Tsai and Douglas, 1996); 2 – (Yan *et al.*, 2002); 3 – (Chamberlain and Burgoyne, 1997b); 4 – (Kelley and Georgopoulos, 1997b); 5 – (Genevaux *et al.*, 2001b); 6 – (Michels *et al.*, 1999); 7 – (Chevalier *et al.*, 2000); 8 – (Lyman and Schekman, 1995); 9 – (Mayer *et al.*, 1999); 10 – (Laufen *et al.*, 1999); 11 – (Suh *et al.*, 1999); 12 – (Wittung-Stafshede *et al.*, 2003); 13 – (Genevaux *et al.*, 2002); 14 – (Zalvide *et al.*, 1998); 15 – (Sullivan *et al.*, 2000a); 16 – (Fewell *et al.*, 2002); 17 – (Li *et al.*, 2001); 1 8 - (Johnson and Craig, 2000); 19 - (Stubdal *et al.*, 1997).

The majority of mutations performed have involved the HPD motif in the loop region, with minimal work having investigated residues in Helices I and IV. Hence, it appears that, despite the presence of highly conserved residues in these helices, they play an insignificant role in the interaction with partner Hsp70s. It is also apparent that the majority of mutations performed on residues outside the HPD motif employ the paradigmatic *E. coli* system. The assumption therefore is that the *E. coli* paradigm is relevant and applicable to all prokaryotic systems, whereas there may be significant differences in the manner of action of the DnaK system within differing organisms.

4.1.2 Aims and objectives

The aim of this section of work was to perform substitutions on conserved residues in the J domain of *Agt* DnaJ. The residues chosen, and the rationale are given below in Table 4.2. A significant number of the residues substituted in this work have only been substituted as part of a double substitution. Hence, it was decided to perform substitutions on these residues singly to ascertain their role in J domain functioning. Equally, few hydrophobic residues have been substituted in order to study their role in maintaining correct functioning of the J domain. Substitutions were performed on the *Agt* DnaJ J domain, and the ability of the derived proteins to complement for lack of *E. coli* DnaJ and CbpA was investigated.

Table 4.2: Rationale for amino acid substitutions performed in the *Agt* DnaJ J domain

No. *	aa	Nature of conservation	Change ^	Rationale	Altered previously?
7	Y	Aromatic	A	Potential structural residue	No
10	L	Hydrophobic	A	Potential structural residue	No
20	E	Negative	A	Potential functional residue	Double substitution
22	K	Positive	A	Potential functional residue	Double substitution
26	R	Positive	A	Potential functional residue	Yes
27	K	Positive	A	Potential functional residue	Yes
26, 27	RK	Positive	AA	Double mutation	Yes
32	Y	Aromatic	A	Potential structural residue	No
33	H	Critical; HPD	Q	Defined mutation; negative control	Yes
35	D	Critical; HPD	E	HPE motif; DjlC family of proteins	No
33, 35	HPD	Critical; HPD	YPY	YPY motif; RESA family of proteins	No
46	K	Positive	T	Potential functional residue	Ala scanning
47	F	Hydrophobic	L	Potential structural residue	Yes
48	K	Positive	T	Potential functional residue	Double substitution
53	A	Hydrophobic	S	Potential structural residue	Yes
57	L	Hydrophobic	S	Potential structural residue	No
59	D	Negative	A	Potential functional residue	Double substitution
59	D	Negative	N	Potential functional residue	Double substitution
62	K	Positive	A	Potential functional residue	Yes
63	R	Positive	A	Potential functional residue	Yes
66	Y	Aromatic	A	Potential structural residue	No
67	D	Negative	A	Potential functional residue	Double substitution
67	D	Negative	N	Potential functional residue	Double substitution

* - Numbering is for the *Agt* DnaJ numbering

^- Alanine scanning was performed for most substitutions, as inserting an alanine was unlikely to deleteriously impact on the structure, due to its lack of bulk and charge.

4.2 EXPERIMENTAL PROCEDURES

4.2.1 Site directed mutagenesis

Substitutions were performed on the codons encoding conserved residues identified in Chapter 2. Primers were designed using the software package GeneRunner (Hastings Software). Primer sequences are given in the Appendix (Section 6.2, Table 6.5). Primers were synthesised by IDT (Iowa, USA). All mutagenic primers allowed the insertion or deletion of a restriction endonuclease site to allow for confirmation of the insertion of the substitution. The position of these sites on the parental plasmid is shown in the Appendix (Section 6.2, Figure 6.3). Site directed mutagenesis was performed using the QuikChange site directed mutagenesis kit (Stratagene, USA) as described in the Appendix (Section 6.4.1.9).

Plasmid DNA from the resultant transformant cells was isolated and restricted with the appropriate restriction endonuclease to confirm the presence of the substitution. All mutant constructs were sequenced to confirm the presence of the desired substitution, and to ensure the absence of undesired substitutions.

4.2.2 Complementation Assays

Plasmids encoding *Agt* DnaJ mutant derivatives were transformed into *E. coli* OD259 (Appendix, Section 6.4.1.8). Complementation assays were performed as described previously (Section 3.2.9). Plasmids pQE30 and pRJ30 were also transformed into *E. coli* OD259 to give *E. coli* OD259 [pQE30] as a negative control and *E. coli* OD259 [pRJ30] as a positive control. All transformants were repeated twice with fresh transformants, and were reproducible with the exception of *Agt* DnaJ-R63A.

4.3 RESULTS

4.3.1 Investigation of the functional effect of substitutions on conserved, charged residues in the J domain

4.3.1.1 Effect of substitutions in the HPD motif

Substitutions in the *Agt* DnaJ J domain were investigated by following the success or lack of success in complementing for the lack of *E. coli* DnaJ. The first group of substitutions examined involved substitutions in the HPD motif. Three substitutions were performed. The first was the H33Q substitution; this would act as a negative control for future experiments. A D35E substitution was then performed to form a HPE motif, similar to that found in the *E. coli* proteins DjlB/DjlC (Kluck *et al.*, 2002). The final substitution was a H33Y, D35Y double substitution to give rise to a YPY motif, similar to that found in the RESA proteins of *P. falciparum* (Bork *et al.*, 1991). The results of complementation assays for these substitutions are given in Figure 4.1.

All *E. coli* OD259 transformants grew successfully at 30°C, but none of the strains containing plasmids with substitutions in the HPD motif were able to grow at 40°C, in the presence or absence of IPTG (Figure 4.1), although the *E. coli* OD259 [pRJ30] was able to do so (Figure 3.24). This was consistent with previous results that, no matter the nature of the substitution, changes in the HPD motif affect J domain functioning (Tsai and Douglas, 1996). Of interest was that even a conservative substitution from aspartic acid to glutamic acid leads to a non-functional protein. The addition of an extra methyl group at this position was therefore deleterious to J domain function. This could also mean that the HscC protein in *E. coli* has a modification that allows for the presence of a glutamic acid in its HPE containing DjlC partner. This could also partially explain the reason for the inability for DnaJ to interact with HscC, and for DjlC to interact with DnaK (Kluck *et al.*, 2002). Similarly, the presence of an alternative to the HPD motif in the RESA family of proteins could imply an altered Hsp70 either in *P. falciparum* or in humans.

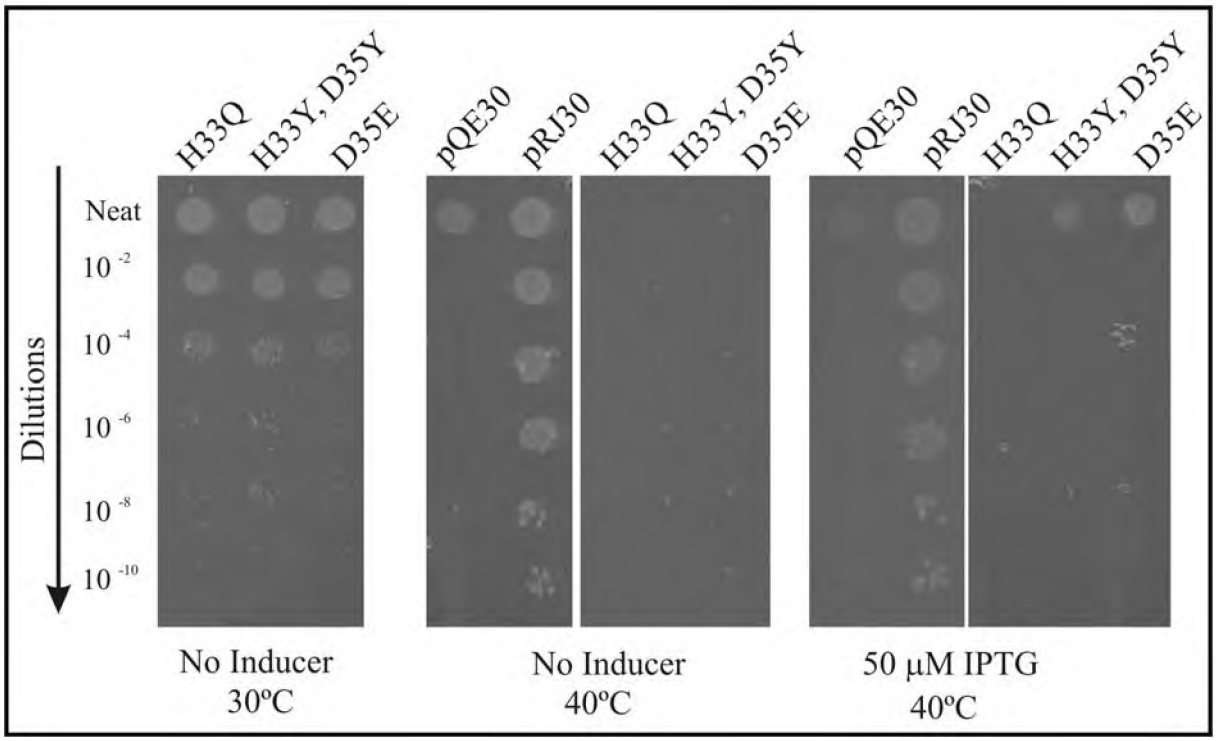

Figure 4.1: *Agt* DnaJ derivatives with substitutions in the HPD motif are incapable of complementing for the lack of *E. coli* DnaJ

Plasmid encoded *Agt* DnaJ with substitutions in the J domain were transformed into *E. coli* OD259. Growth, or lack of growth of the mutants was investigated at 30°C and 40°C (stress conditions) in the presence and absence of the inducer IPTG. The dilution series is indicated on the left hand side, and the nature of the substitution being investigated is indicated at the top of the figure. Growth of E. coli OD259 containing either pQE30 or pRJ30 at 40°C is also indicated.

4.3.1.2 Effect of substitutions in charged residues on Helix II

The second group of substitutions created were aimed at altering charged residues located in Helix II of the J domain. Substitutions performed were E20A, K22A, R26A, K27A and the double substitution R26A,K27A. The results are shown in Figure 4.2.

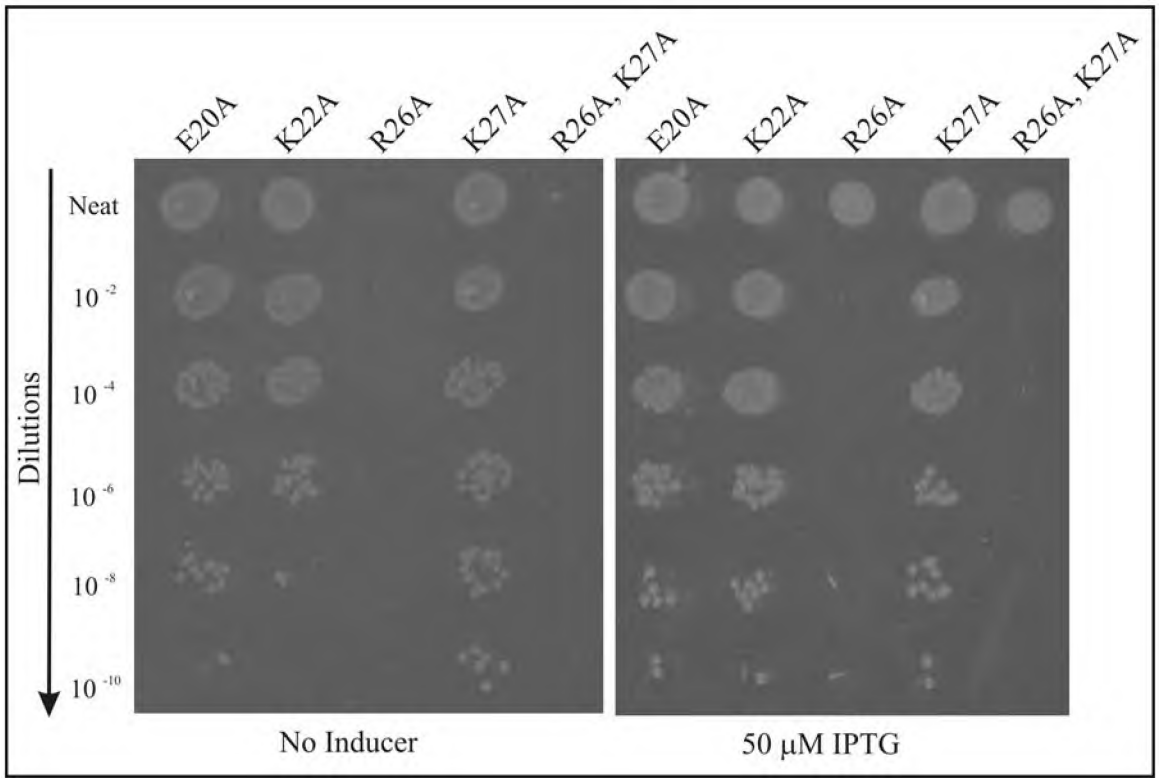

Figure 4.2: Analysis of the *in vivo* effect of substitutions on charged residues in Helix II of the J domain of *Agt* DnaJ using complementation assays

Conserved, charged residues in Helix II of the J domain of *Agt* DnaJ were substituted. *E. coli* OD259 was transformed with pRJ30-E20A, pRJ30-K22A, pRJ30-R26A, pRJ30-K27A and pRJ30-R26A, K27A. The effect of the substitution was investigated using these transformants under stress conditions (40°C). The dilutions used are indicated on the left, and the substitution being investigated is indicated at the top of the figure.

Agt DnaJ-E20A, *Agt* DnaJ-K2 2A and *Agt* DnaJ-K27A were all capable of complementing for the lack of *E. coli* DnaJ. *Agt* DnaJ-K22A had slightly abrogated complementation in the absence of IPTG, although the levels were similar to *Agt* DnaJ in the presence of IPTG (compare Figure 3.24). In accordance with previous results proteins containing either the *Agt* DnaJ-R26A or the *Agt* DnaJ-R26A, K27A substitutions were unable to complement for lack of *E. coli* DnaJ. As seen for *Agt* DnaJ, no inducer effect was apparent.

4.3.1.3 Effect of substitutions in charged residues in Helices III and IV

The next class of substitutions involved alterations of charged residues in Helices III and IV. The results are depicted in Figure 4.3 and Figure 4.4. The following substitutions were performed: K46T, K48T, D59A, D59N, K62A, R63A, D67A and D67N.

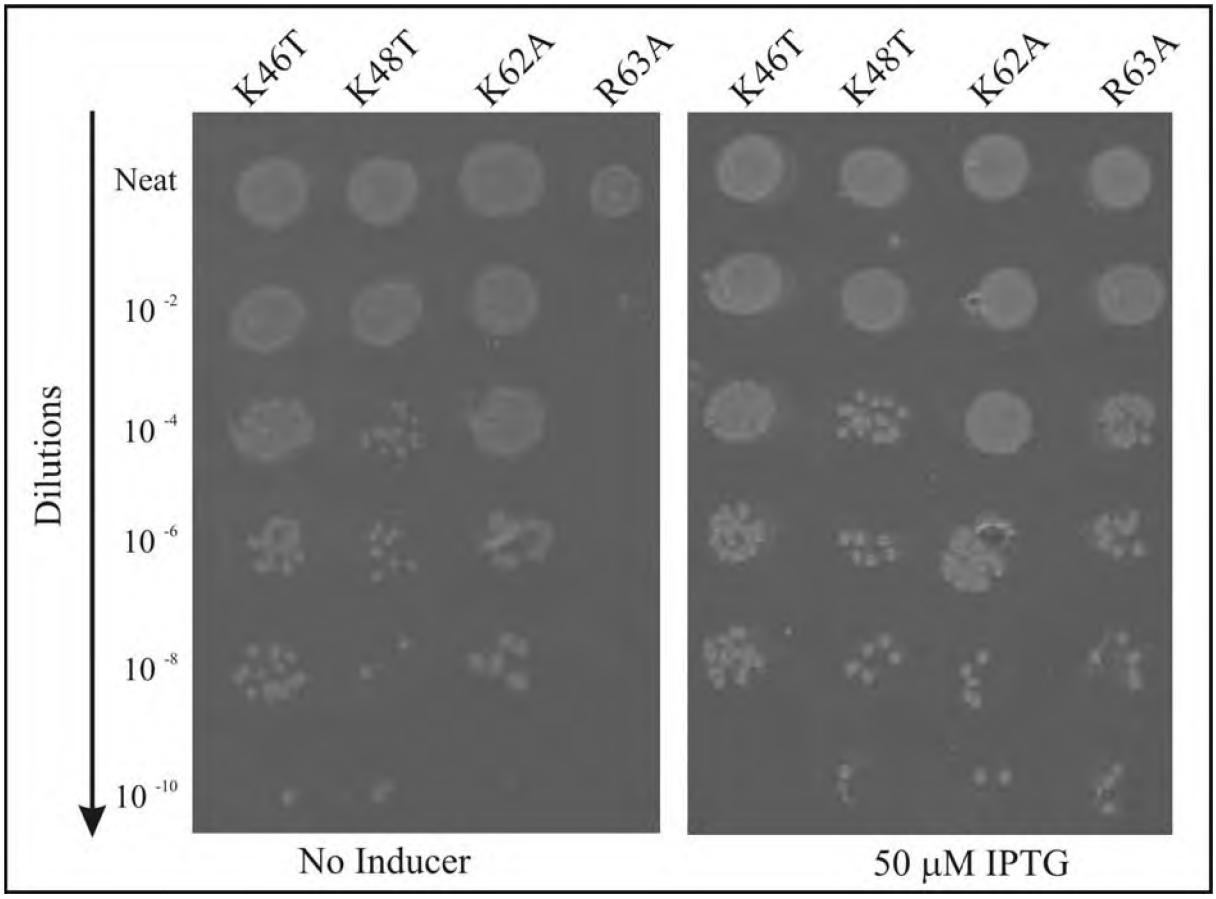

Figure 4.3: Analysis of *in vivo* effect of substitutions on conserved, positively charged residues in Helices III and IV of the J domain of *Agt* DnaJ using complementation assays

Conserved, positively charged residues in Helices III and IV of the J domain of *Agt* DnaJ were substituted. *E. coli* OD259 was transformed with pRJ30-K46T, pRJ30-K48T, pRJ30-K62A and pRJ30-R63A. The effect of the substitution was investigated using these transformants under stress conditions (40°C). The dilutions used are indicated on the left, and the substitution being investigated is indicated at the top of the figure.

Agt DnaJ-K46T and *Agt* DnaJ-K62A were able to successfully complement for the lack of *E. coli* DnaJ and CbpA. Lys^{48}, which was located in the previously described KFK motif (Hennessy *et al.*, 2000) appeared to have a slight role in J domain functioning, as substitution of this residue to a threonine (which was capable of being solvent exposed, while deleting the positive charge) led to a slight abrogation in efficiency of complementation. However, altering Arg^{63}, found in the QKRAA motif (Auger and Roudier, 1997) to an alanine lead to the only instance of an inducer-effect experienced in these experiments. No complementation was seen in the absence of IPTG, although complementation was successful in its presence.

Even in the presence of IPTG, the level of complementation was not as effective as by *Agt* DnaJ. This implied that a higher level of protein was required for successful interaction with DnaK than for the unmodified *Agt* DnaJ. This could mean that there was a lower chance of succes sful stimulation of the ATPase activity of *E. col i* DnaK by *Agt* DnaJ-R63A, than by *Agt* DnaJ. However, the phenotype of *E. coli* OD259 [pRJ30-R63A] was highly variable, and not reproducible, as in some instances *Agt* DnaJ-R63A was able to successfully complement for the lack of *E. coli* DnaJ and CbpA in both the presence and absence of IPTG, and in some instanc es it could not do so. This protein may be highly unstable, and high levels of phenotypic variance may be the net result of this substitution.

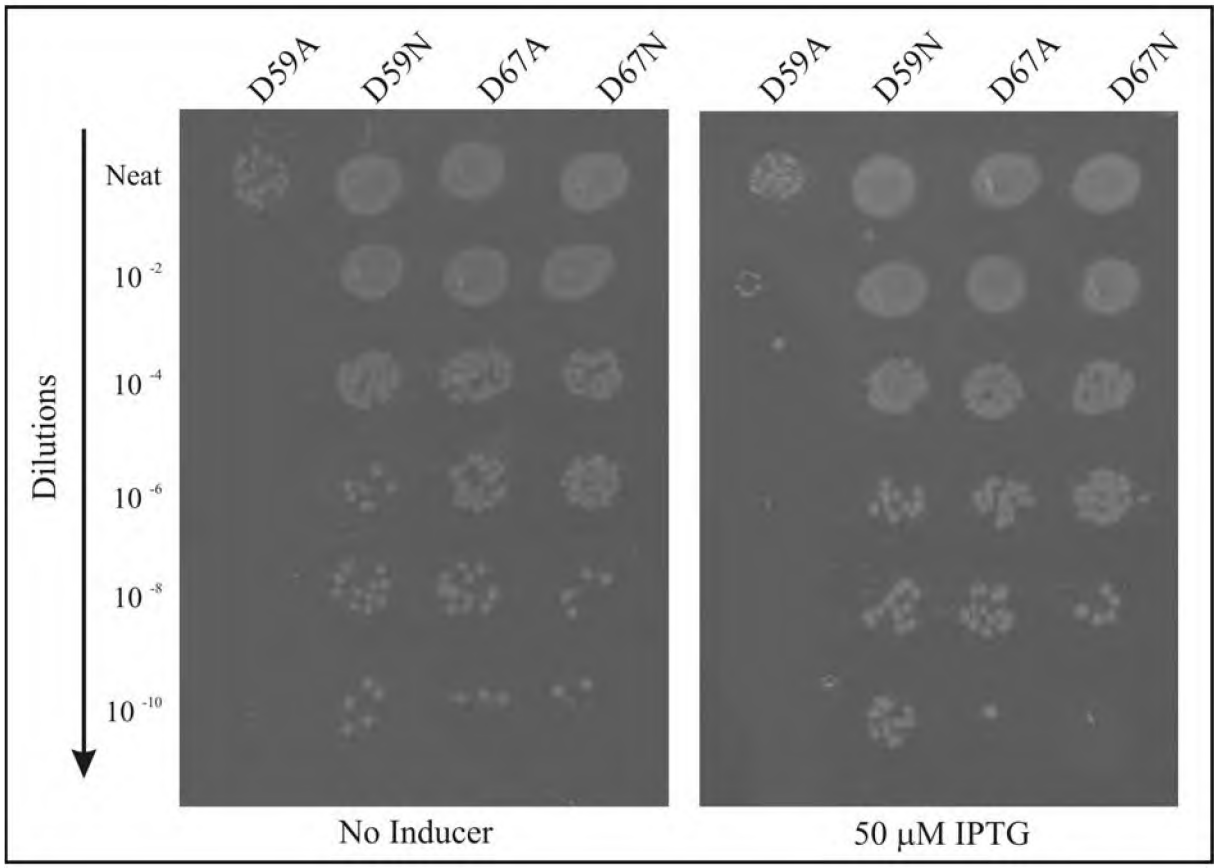

Figure 4.4: Analysis of *in vivo* effect of substitutions on conserved negatively charged residues on Helix IV of the J domain of *Agt* DnaJ using complementation assays

Conserved, negatively charged residues in Helix IV of the J domain of *Agt* DnaJ were substituted. *E. coli* OD259 was transformed with pRJ30-D59A, pRJ30-D59N, pRJ30-D67A and pRJ30-D67N. The effect of the substitution was investigated using these transformants under stress conditions (40°C). The dilutions used are indicated on the left, and the substitution being investigated is indicated at the top of the figure.

Substitutions in conserved, negatively charged residues also suggested a role for a conserved residue in Helix IV. A conservative substitution of Asp^{59} to Asn, resulted in the loss of the negative charge, but retention of the bulk and shape of the side chain, had a slight effect on the level of complementation (compare Figure 3.24 and Figure 4.4; *Agt* DnaJ was able to complement to a lower dilution of cells than *Agt* DnaJ-D59N). Substitution of this residue to Ala however, resulted in complete loss of complementation. This residue could potentially interact with Val^{56}, Glu^{55}, Leu^{57}, Lys^{62} and Arg^{63}, with the likely interaction of the carboxyl group being with Lys^{62}. Both *Agt* DnaJ-D67A and *Agt* DnaJ-D67N were able to complement for the lack of *E. coli* DnaJ and CbpA.

4.3.2 Analysis of levels of production of *Agt* DnaJ proteins with substitutions in charged residues in the J domain

An obvious reason for the inability of proteins to complement in *E. coli* OD259 was low protein production. Chemiluminescent based immunodetection of the histidine tagged *Agt* DnaJ and its derivatives was performed using a monoclonal anti-Histidine tag antibody on whole cell extracts of *E. coli* OD259 containing the relevant plasmids The results obtained are shown in Figure 4.5. *Agt* DnaJ, *Agt* DnaJ-K27A, *Agt* DnaJ-R26A, K27A, *Agt* DnaJ-H33Q, *Agt* DnaJ-D35E, *Agt* DnaJ-H33Y, D35Y, *Agt* DnaJ-K46T, *Agt* DnaJ-K48T, *Agt* DnaJ-D59N, *Agt* DnaJ-K62A and *Agt* DnaJ-D67N were all detected by chemiluminescent based immunodetection. However, *Agt* DnaJ-R26A, *Agt* DnaJ-D59A, *Agt* DnaJ-R63A and *Agt* DnaJ-D67A were not detected.

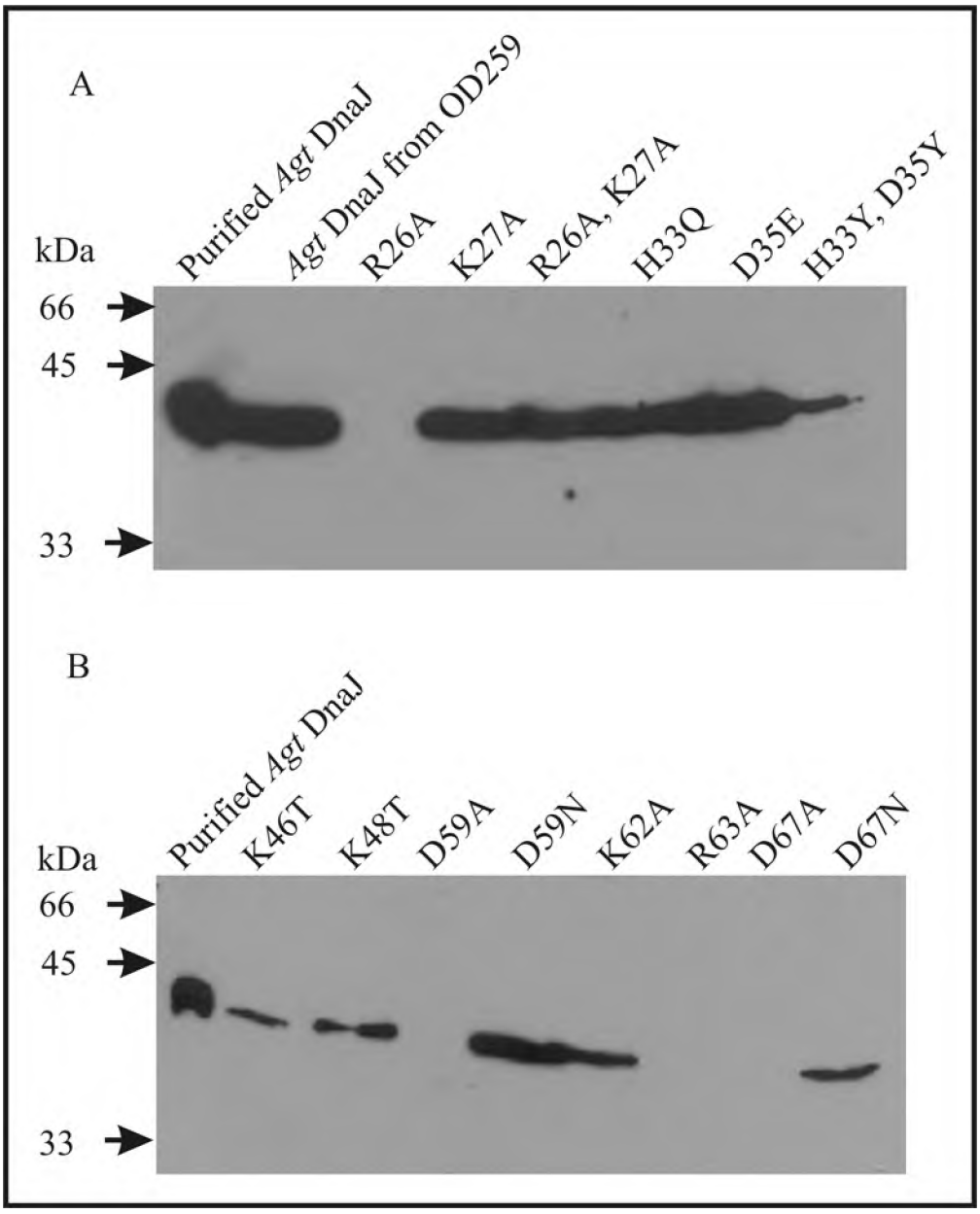

Figure 4.5: Western analysis of levels of protein production of *Agt* DnaJ proteins containing substitutions of charged residues in the J domain in *E. coli* OD259

Agt DnaJ proteins containing substitutions in the J domain were produced in *E. coli* OD259. Protein extracts from equal numbers of cells were run on a 12 % SDS-PAGE, and transferred to nitrocellulose. An anti-His antibody was used to detect the histidine-tagged proteins. Purified *Agt* DnaJ is shown in the first lane of both (A) and (B). *Agt* DnaJ produced in *E. coli* OD259 is in the lane labelled *Agt* DnaJ from OD259. The remaining lanes are labelled according to the substitution present in the *Agt* DnaJ J domain.

4.3.3 Investigation of the functional effect of substitutions on conserved, hydrophobic residues in the J domain

Amino acid substitutions were performed in conserved, hydrophobic residues in the J domain of *Agt* DnaJ. The following substitutions were performed: Y7A, Y32A, F47L, A53S, L57S and Y66A. The ability of these modified proteins to replace the lack of *E. coli* DnaJ and CbpA in *E. coli* OD259 was assessed by complementation assays. The results are shown in Figure 4.6.

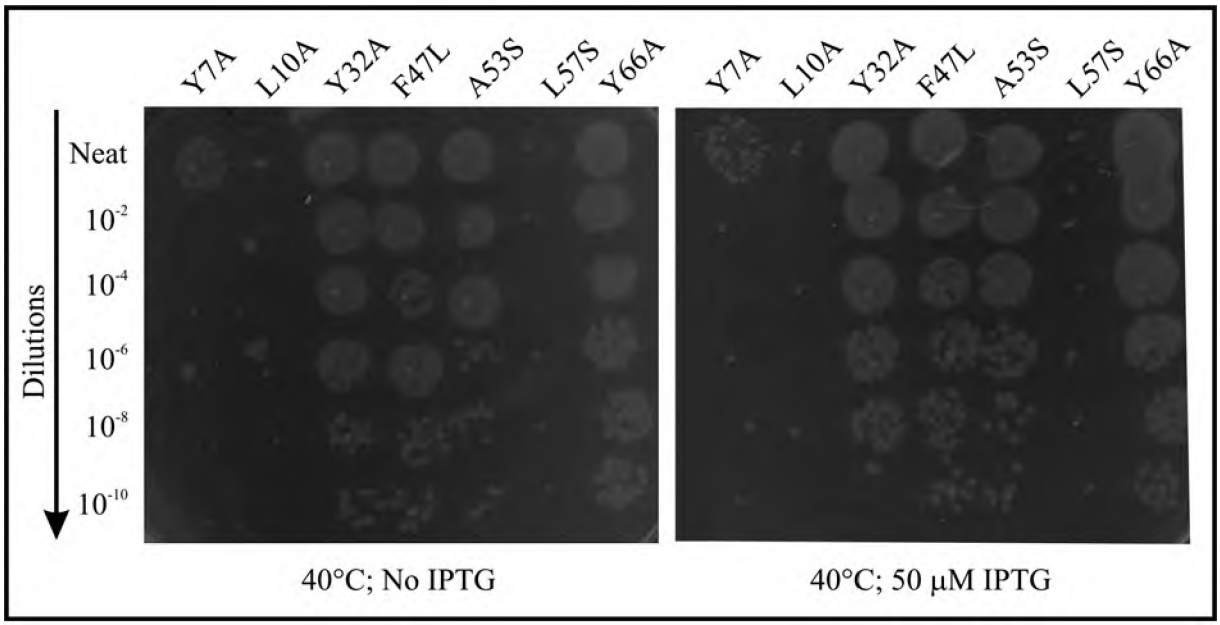

Figure 4.6: Analysis of *in vivo* effect of substitutions on conserved, structural residues in the J domain of *Agt* DnaJ using complementation assays

Conserved, hydrophobic residues in Helix IV of the J domain of *Agt* DnaJ were substituted. *E. coli* OD259 was transformed with pRJ30-Y7A, pRJ30-L10A, pRJ30-Y32A, pRJ30-F47L, pRJ30-A53S, pRJ30-L57S and pRJ30-Y66A. The effect of the substitution was investigated using these transformants under stress conditions (40°C), in the presence and absence of inducer. The dilutions used are indicated on the left, and the substitution being investigated is indicated at the top of the figure.

Agt DnaJ-Y32A, *Agt* DnaJ-F47L, *Agt* DnaJ-A53S and *Agt* DnaJ-Y66A were all able to complement for the lack of *E. coli* DnaJ and CbpA. *Agt* DnaJ-Y7A, *Agt* DnaJ-L10A and *Agt* DnaJ-L57S were unable to complement for the lack of *E. coli* DnaJ and CbpA.

4.3.4 Analysis of levels of production of *Agt* DnaJ proteins with substitutions in conserved, structural residues in the J domain

Western analysis was performed using whole cell extracts of *E. coli* OD259 containing the relevant plasmids and an antibody against the histidine tag. The results obtained are shown in Figure 4.7, and summarised in Table 4.3.

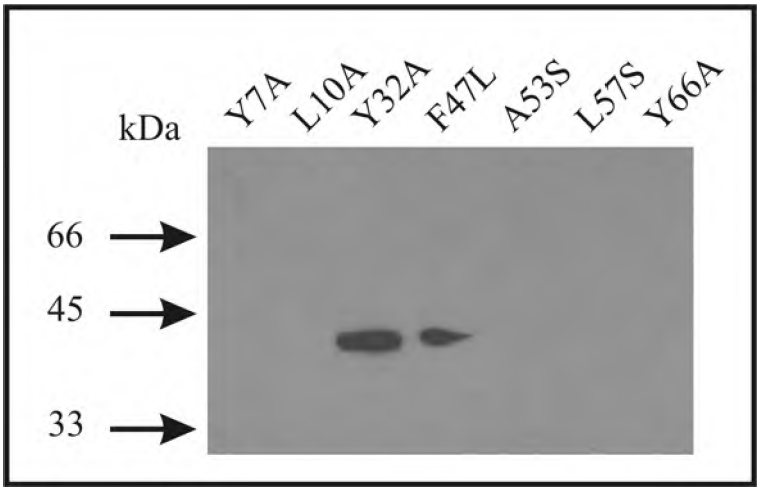

Figure 4.7: Western analysis of levels of protein production of *Agt* DnaJ proteins containing substitutions in structural residues in the J domain in *E. coli* OD259

Agt DnaJ proteins containing substitutions in the J domain were produced in *E. coli* OD259. Protein extracts from equal numbers of cells were run on a 12 % SDS-PAGE, and transferred to nitrocellulose. An anti-His antibody was used to detect the histidine-tagged proteins. The lanes are labelled according to the substitution present in the *Agt* DnaJ J domain.

The only proteins that were detectable were *Agt* DnaJ-Y32A and *Agt* DnaJ-F47L. Both of these were able to complement for the lack of *E. coli* DnaJ and CbpA.

4.4 DISCUSSION

4.4.1 Protein production

A comparison of the ability of the substituted proteins to complement, and their detection using chemiluminescent based immunodetection is given in Table 4.3.

It was found that there was not always a correlation between the levels of complementation and the levels of protein production. For example, both *Agt* DnaJ-R26A and *Agt* DnaJ-R26A,K27A do not complement, but *Agt* DnaJ-R26A,K27A can be detected *via* Western analysis, but *Agt* DnaJ-R26A can not. Similarly, there are examples of proteins that can not be detected *via* Western analysis, but can complement successfully. Presumably the levels of protein required for successful complementation was below the levels of detection using an anti-His antibody. Equally the possibility remains that the substitution has formed a protein with an unstable J domain, prone to premature degradation. Hence levels of protein in the cell would be low, and below the detection limit of the system employed. This may explain results where there was a lack of complementation as well as an inability to detect the protein.

The problems of detecting the proteins in *E. coli* OD259 was discussed previously (Section 3.4.7), but is worth re-iterating. The technique employed here was of necessity an indirect technique. There was no easy way of detecting levels of protein production in *E. coli* OD259 under stress conditions. *E. coli* OD259 containing plasmids that express proteins that do not successfully complement for the lack of *E. coli* DnaJ and CbpA will not grow successfully at 40°C. Hence obtaining representative cellular extracts of these *E. coli* OD259 strains would be problematic, in order to ascertain expression of the DnaJ-like protein under investigation. Alternatively performing colony Western analysis using colonies that do grow, and detecting Histidine-tagged proteins using immunodetection, may indicate levels of protein present; again however lack of growth at 40°C would prevent such an analysis. Analysis of protein production in another *E. coli* strain could also be performed. However, that would only imply that a specific protein was capable of being produced in that strain. It was also possible that some of the derived proteins are capable of being produced at 30°C, but not at 40°C.

Table 4.3: Comparison of complementation results by *Agt* DnaJ proteins containing substitutions in the J domain and levels of protein production of these proteins in *E. coli* OD259

Substitution	Complementation [1]	Protein Production [2]
Agt DnaJ	+++	++
K27A	++	++
Y32A	++	+
K46T	++	+
F47L	++	+
K48T	+	+
K62A	++	+
D59N	+++	++
D67N	++	+
A53S	++	-
R63A	+/-	-
Y66A	++	-
D67A	++	-
H33Q	-	++
H33Y, D35Y	-	++
D35E	-	+
R26A, K27A	-	++
Y7A	-	-
L10A	++	-
R26A	-	-
L57S	-	-
D59A	-	-
E20A	++	N/D
K22A	+++	N/D

[1] – The ability to complement for lack of *E. coli* DnaJ by *Agt* DnaJ proteins altered in the J domain was compared to *Agt* DnaJ. An ability to complement similarly to *Agt* DnaJ is indicated by +++, with lower abilities indicated by ++ and +. A +/- indicates complementation only under specific circumstances.

[2] – The levels of *Agt* DnaJ altered in the J domain detected by an anti-His antibody was compared to the level of *Agt* DnaJ \detected. Levels of protein similar to *Agt* DnaJ are indicated by +++, with lower levels indicated by ++ and +. A N/D indicates that detection of protein levels was not performed.

4.4.2 HPD motif

Work performed herein emphasised the importance of the HPD motif. Regardless of the substitution performed, alterations in that region abolished the interaction of *Agt* DnaJ with DnaK, as monitored by complementation assays. Similarly, *Agt* DnaJ-H33Q was unable to stimulate the ATPase activity of *Agt* DnaK. That the conservative substitution of glutamic acid for aspartic acid abolished complementation was striking, and provided a rational explanation for the inability of DjlC to interact with DnaK in *E. coli*. Presumably, the added bulk of the glutamic acid side chain was responsible for this result, possibly through additional interactions that prevented other, critical interactions from occurring. The rationale for the inability of the *Agt* DnaJ-H33Y, D35Y to complement was easier to explain. These were not conservative substitutions, although they do conserve bulk, and this result raises a query over the ability of RESA proteins to act as true DnaJ-like proteins, an ability that has not been demonstrated at present. This was the first time that HPD-like motifs found in DnaJ-like proteins have been shown not to be able to interact with an Hsp70. This in turn suggests that a potential subgroup of Hsp70s have evolved to compensate for the alterations in the HPD motif. This raised the question as to why proteins containing a J domain would have evolved by altering the HPD motif?

4.4.3 Helix II

Interestingly, Lys^{22} was more conserved in all J domains (except for Type III J domains) than Arg^{26}, yet Arg^{26} appeared to be the critical residue. This could mean that Lys^{22} was important for general interactions between DnaJ proteins and partner Hsp70s, yet its function could be successfully replaced by any of the other positively charged residues in Helix II, whereas Arg^{26} was important for a specific interaction between DnaJ proteins and DnaK proteins, and its role was not compensated for by the other positively charged residues.

The orientation of the conserved, charged residues in Helix II is shown below in Figure 4.8. As can be seen, all the charged residues project outwards from Helix II. They potentially interact with the underside cleft of the ATPase domain of a partner Hsp70, and the charged nature of Helix II may therefore serve as a mechanism for orientating the J domain with the ATPase domain.

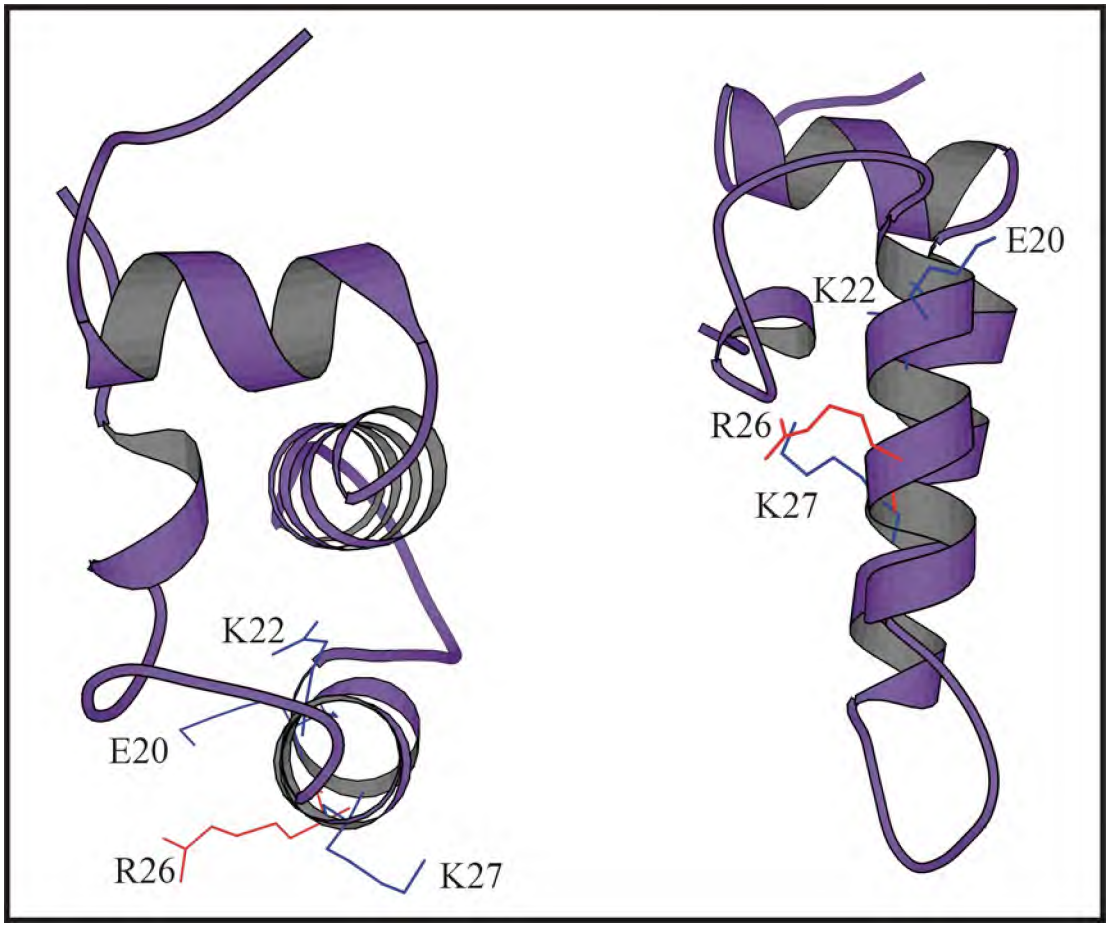

Figure 4.8: Orientation of residues substituted in Helix II of the J domain of *Agt* DnaJ

Conserved residues that were altered in this thesis are shown on a modelled structure of the *Agt* DnaJ J domain structure. Residues are shown in stick format and the critical residue Arg^{26} is depicted in red. The figure shows two different orientations to depict the outwards projection of these residues. Modelling was performed using the *E. coli* DnaJ structure (1XBL) (Pellechia *et al.*, 1996) as a template and using the modelling package in Whatif (Vriend, 1990). Figures were generated in Molscript (Kraulis, 1991).

Once again Helix II has been shown to be important in the interaction of DnaJ-like proteins with partner Hsp70s. The role of the Arg^{26} residue and its structural equivalents was obviously critical in correct functioning of the J domain. The residue was solvent exposed, hence substitutions are unlikely to have a structural role, particularly if polarity was maintained. What was interesting was that other positively charged residues (Lys^{22}) in Helix II retain higher levels of conservation, yet alterations at those positions had no detectable, functional effect. Why was the Arg^{26} residue so vital for the functioning of the J domain? It was unlikely to be a lack of tolerance of substitutions at that position. When examining the sequence alignment described previously, the positions equivalent to those residues (Lys^{22}

and Arg[26]) did contain other amino acid types. Additionally, there were fewer changes at the p osition corresponding to Lys[22], particularly in Type I J domains. The implication was that residues at the corresponding position to Arg[26] make a network of interactions with DnaK; and that, while different amino acids can in some circumstances be tolerated in this position, any residues located there has to have the ability to be involved in the network of interactions. Hence substituting Arg[26] with residues such as glutamic acid, methionine or histidine may be additional experiments that could provide interesting data. By contrast, Lys[22], while important, can potentially be substituted for without a serious deleterious affect being observed. Lys[22] may have been evolutionarily conserved for subtle reasons such as structural stability, such that engineered changes to this residue may slightly perturb the structure of the protein, but without observable functional consequences. A negatively charged residue at the equivalent position to Glu[20] was conserved, and yet substituting this residue does not appear to have a markedly deleterious affect on the functioning of *Agt* DnaJ, under the conditions employed. Hence the J domain appears to have several conserved residues, that can be substituted, and have their role potentially compensated for by additional residues in the J domain.

4.4.4 Helices III and IV

As for Helix II, all the residues investigated were solvent exposed (Figure 4.9). Helix III tended to have a more equal amount of negatively and positively charged residues than Helix II, and may play a different role in terms of interacting with a partner Hsp70.

There was partial abrogation of function by *Agt* DnaJ-K48T as compared to *Agt* DnaJ. This could be due to two effects. Firstly, a significant amount of bulk was removed from the side chain. This may prevent a significant interaction occurring due to the consequent inability of the Thr[48] side chain to interact with additional residues due to lack of bulk. Secondly, while polarity was retained, the positive charge was not. Hence a specific interaction that required the positive charge of the lysine to be created would not be formed. Significantly, *Agt* DnaJ-K46T did not appear to be significantly impaired in function. This was consistent with Lys[48] (54% in overall consensus; 88% in Type I consensus) being more conserved than Lys[46] (46% in overall consensus; 76% in Type I consensus).

Despite previous results (Genevaux *et al.*, 2002; Genevaux *et al.*, 2003), there does appear to be a role for Helix IV in both the structure and function of the J domain. Its relative lack of hydrophobicity meant it was likely to be solvent exposed, and may function in a hitherto unrecognised interaction with a partner Hsp70. The two residues in Helix IV implicated as being important are Arg^{63} (part of the QKRAA motif) and Asp^{59}.

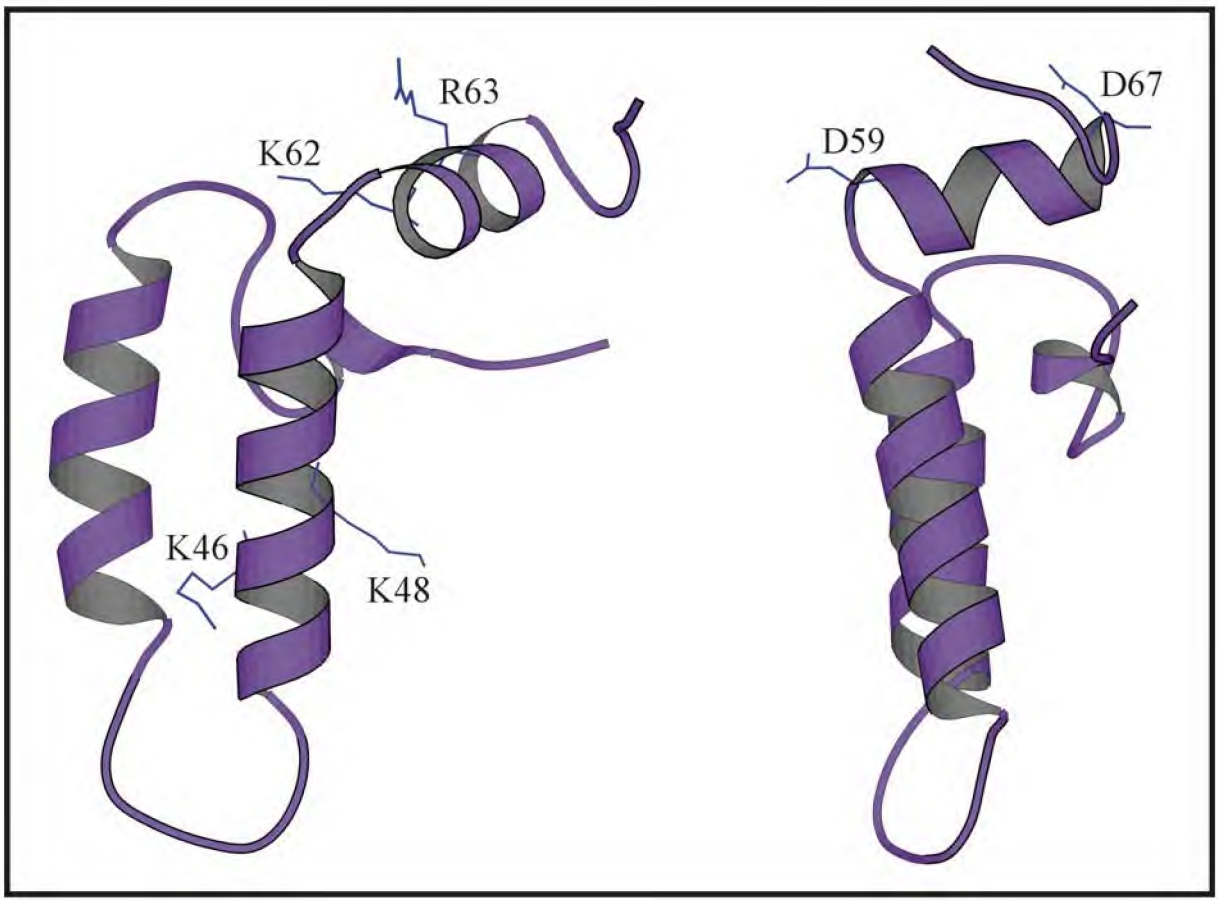

Figure 4.9: Orientation of conserved residues in Helices III and IV

Conserved residues that were altered in this thesis are shown on a modelled structure of the *Agt* DnaJ J domain structure. The left hand figure depicts the orientation of Lys^{46}, Lys^{48}, Lys^{62} and Arg^{63}. The right hand figure shows the orientation of Asp^{59} and Asp^{67}. Modelling was performed using the *E. coli* DnaJ structure (1XBL) (Pellechia *et al.*, 1996) as a template and using the modelling package in Whatif (Vriend, 1990). Figures were generated in Molscript (Kraulis, 1991).

Under some circumstances expression of *Agt* DnaJ-R63A was not capable of replacing the lack of *E. coli* DnaJ and CbpA in *E. coli* OD259 except in the presence of inducer; it therefore follows that higher levels of this protein are required for complementation than for other proteins. The probability of this helix being solvent exposed may imply that an additional hydrophobic residue may cause a conformational shift, leading to a more rigid structural

conformation, and a lower ability to interact with a partner Hsp70. This was potentially similar to the reason for the inability of *Agt* DnaJ-D59A to act similarly to *Agt* DnaJ. Asp^{59} was located at the beginning of Helix IV (Figure 4.9), and loss of the side chain of this amino acid may be deleterious in maintaining the correct structure of the J domain. Its location on the border between Helices III and IV make it structurally as well as functionally relevant, and substitution to an alanine residue may prevent conformational shifts in the labile J domain from occurring. This was the first instance of a residue at this position being shown to have a role in the functioning of the J domain. A previous *E. coli* DnaJ T58A, D59A double substitution was still active (Genevaux *et al.*, 2002), yet the *Agt* DnaJ-D59A protein was inactive. The activity of *E. coli* DnaJ-T58A, D59A may be due to a suppressor effect occurring as a result of the double mutation. Alternatively there may be slight structural differences in the J domains of *E. coli* DnaJ and *Agt* DnaJ, and a substitution in one may be more deleterious then in the other, as altering the structure of the J domain may impact on its functioning. It therefore appears that the interaction of *Agt* DnaJ with *E. coli* DnaK or *Agt* DnaK has some subtle differences when compared with the interaction of *E. coli* DnaJ and DnaK, as corresponding mutations have differing effects.

4.4.5 Structural residues

Three residues were implicated in being important for the structural integrity of the J domain, namely Tyr^{7}, Leu^{10} and Leu^{57}, however substitutions in other conserved, hydrophobic residues (Tyr^{32}, Phe^{47}, Ala^{53} and Tyr^{66}) still gave rise to functional *Agt* DnaJ. Tyr^{7} projects outwards from the J domain, yet can potentially make contacts with residues in Helices I, II and III, including Leu^{57}. Truncation of the side chain will prevent the majority of these interactions occurring, potentially destabilising the structure. Leu^{10} appears to interact predominately with residues in Helix II. It may be critical for ensuring the stability of that helix, such that the J domain scaffold was correctly orientated for interaction with Hsp70. Leu^{57}, which projects into the interior of the J domain, was likely to be a key residue that holds Helices II and III together. It appears to co-ordinate the network of interactions, including an interaction with the highly conserved Ala^{53}, and the substitution of serine for leucine may not be able to participate in that network. The most common substitutions at this position are the aromatic residues, and isoleucine and leucine, implying that bulk and hydrophobicity are key criteria at this position. The substitution of Ala^{53} in *Agt* DnaJ-A53S

did not appear to prevent complementation implying that polarity at that position may be compensated for. It was likely however that mutating that residue to a more charged and bulkier residue such as lysine may affect the J domain structure, as Ala^{53} was buried. The corresponding residue in *S. cerevisiae* Sec63p gave rise to translocation defects when substituted with threonine (Lyman and Schekman, 1995). This may imply that this residue has a role in specific J domain functioning, but a perturbed effect can not be detected under the conditions used.

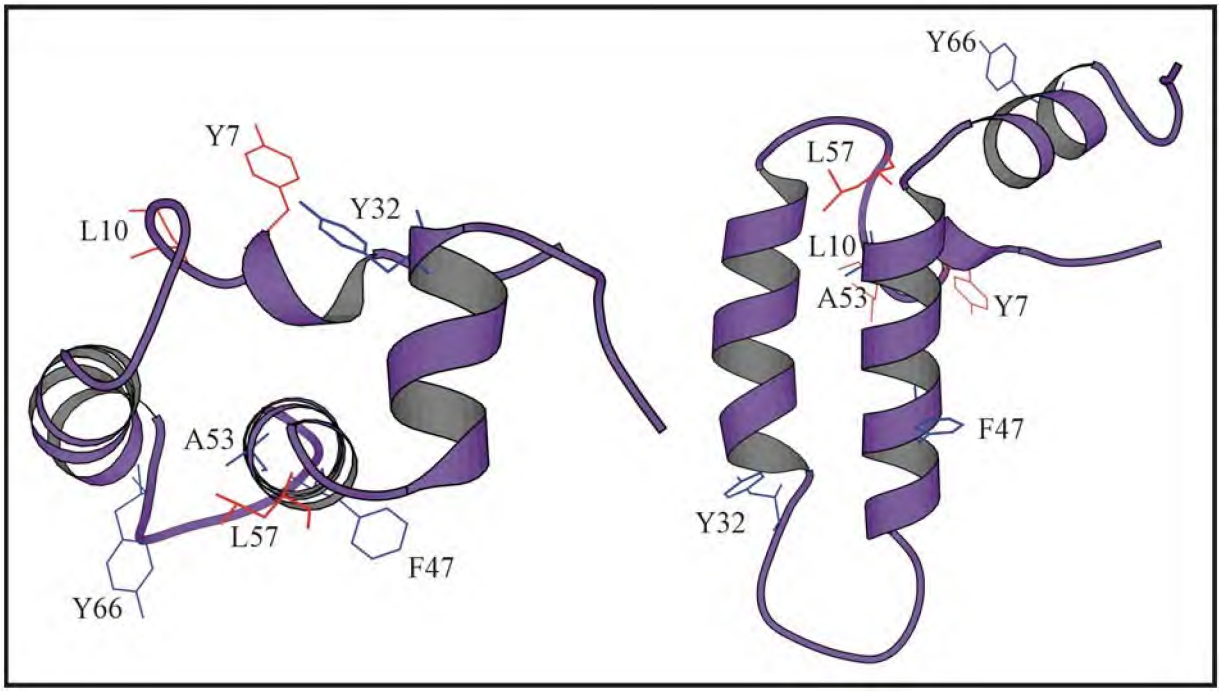

Figure 4.10: Orientation of conserved structural residues in the J domain

Conserved residues that were altered in this thesis are shown on a modelled structure of the *Agt* DnaJ J domain structure. Residues are shown in stick format, and the critical residues Tyr^{7}, Leu^{10} and Leu^{57} are shown in red. The figure shows two different orientations to depict the projection of these residues. Modelling was performed using the *E. coli* DnaJ structure (1XBL) (Pellechia *et al.*, 1996) as a template and using the modelling package in Whatif (Vriend, 1990). Figures were generated in Molscript (Kraulis, 1991).

In contrast to previous results (Johnson and Craig, 2000; Genevaux *et al.*, 2002) the substitution of Phe^{47} in *Agt* DnaJ-F47L had no obvious effect. Leucine was a conservative substitution, and leucine residues were often found in the equivalent position. The leucine side chain may have contained sufficient bulk to compensate for the loss of the remaining ring structure. Histidine tagged proteins expressed in *E. coli* OD259 were detected using chemiluminescent based immunodetection. Only *Agt* DnaJ-Y32A and *Agt* DnaJ-F47L were

detected. Both of these were able to complement, but other modified proteins that complemented for loss of *E. coli* DnaJ were not detectable. This was likely due to the levels of *Agt* DnaJ being produced being below the detection limits of the system used.

The fact that substitutions in Helix IV affect the complementation differs from results previously obtained (Genevaux *et al.*, 2002). This can be due to the slightly different systems used. However other work has indicated the loop between helices II and III has two preferred conformations that interact with Helix IV, which has alternate conformations (Pellechia *et al.*, 1996). Helix IV does appear to be highly mobile, and alterations of conserved residues in this helix may either negatively impact on the structural integrity of the J domain, thereby affecting the function of the J domain, or directly affect the interaction of DnaJ-like proteins with partner Hsp70s. Helix IV was postulated to affect the conformation of the adjacent helices depending on its orientation (Pellechia *et al.*, 1996). Conserved residues in Helix IV were well defined structurally (Lys^{62}, Arg^{63}, Tyr^{66}) in *E. coli* DnaJ (Pellechia *et al.*, 1996), but the orientation of the helix with respect to the remainder of the domain was not. Hence this thesis proposes that Helix IV has been overlooked in terms of the function of the J domain, and that conservation of residues in Helix IV are critical with respect to the correct functioning of the J domain.

4.4.6 Conclusions

Several conserved amino acids have been identified as being important in the structural integrity and the functional integrity of the J domain of *Agt* DnaJ. These include Arg^{26}, Arg^{63}, Asp^{59}, Tyr^{7}, Leu^{10} and Leu^{57}.

Chapter 5: Conclusions

5 CONCLUSIONS

5.1 GENERATION OF A MODEL SYSTEM

DnaJ-like proteins from *A. tumefaciens* RUOR were successfully isolated, including the Type I protein *Agt* DnaJ. This protein was shown to be active in both *in vivo* and *in vitro* assays. The histidine tagged *Agt* DnaK was also shown to be active *in vitro*, as it was capable of both hydrolysing ATP, and having this hydrolysis activity stimulated by *Agt* DnaJ. However, Type III DnaJ-like proteins isolated from *A. tumefaciens* RUOR did not appear capable of interacting with *E. coli* in an *in vivo* system, implying loss of amino acid determinants mediating the interacting of *Agt* DnaJ and *E. coli* DnaK. Hence amino acids in the *Agt* DnaJ J domain were targeted to elucidate which residues were necessary for the interaction of *Agt* DnaJ and DnaK.

5.2 AMINO ACIDS IMPORTANT IN THE STRUCTURE AND FUNCTION OF THE J DOMAIN

This thesis has identified several amino acids that are proposed to play a role in the cellular functioning of DnaJ-like partner proteins, specifically in the J domain. These are depicted below in Figure 5.1.

These residues are likely to be important for generic Hsp70 – DnaJ-like partnerships, and/or for establishing a correct J domain structure. The next stage of this work will be to determine what residues are important for the specific interaction of DnaJ and DnaK.

5.2.1 Helix I

Important residues in Helix I tended to be hydrophobic in nature, and include Tyr^{7} and Leu^{10} (*E. coli* DnaJ numbering). Both of these residues appeared to be important in the maintenance of J domain structure by making conserved interactions with conserved hydrophobic residue s on other Helices. There is also the possibility that these residues can be recognised by the substrate binding region of a partner Hsp70.

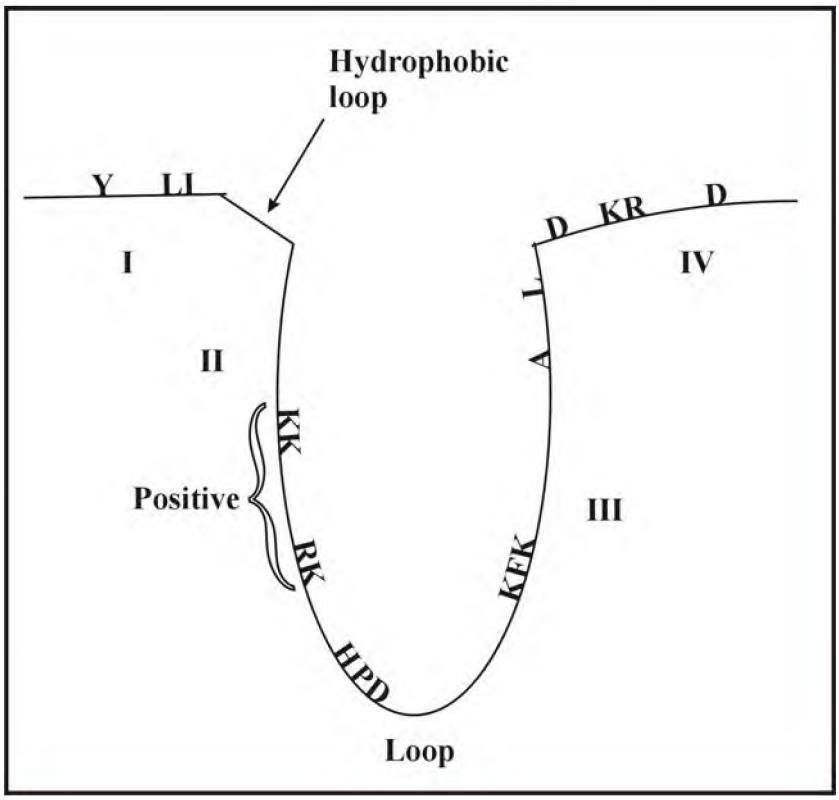

Figure 5.1: Schematic of the J domain showing the location of amino acids proposed to be involved in the interaction with partner Hsp70s

Residues shown to be important in J domain functioning either by *in vitro* and *in vivo*, or by *in silico* analysis are depicted schematically. Helices and the loop region are labelled (I to IV and loop). Amino acids are identified by single amino acid code. The hydrophobic loop between Helices I and II is labelled, as is the two positive dipeptides present on Helix II. I, II, III and IV refer to the J domain α-helices.

5.2.2 Helix II

Helix II appears to be the primary site of interaction with partner Hsp70s (Greene *et al.*, 1998; Lu and Cyr, 1998), and was enriched in positively charged residues, including four positively charged residues located in pairs. These residues are proposed to interact with negatively charged residues on the underside of the ATPase domain of partner Hsp70s (Suh *et al.*, 1998; Suh *et al.*, 1999). Substitution of these positively charged residues in *E. coli* DnaJ indicated the importance of Lys^{26} (Genevaux *et al.*, 2002), and the corresponding residue in *Agt* DnaJ (Arg^{26}) was also shown to be important in this thesis. However the other positively charged residues (Lys^{22}, Ser^{23}, Lys^{27} in *Agt* DnaJ) appeared to be less important individually when compared to Arg^{26}. These remaining residues may make separate interactions that as a group

are important, but the loss of one of these may be compensated for by the other residues, whereas Arg^{26} was required completely.

5.2.3 Loop region

As expected substitutions in the HPD motif completely abrogated interactions with partner Hsp70s. This applied even when the substitutions made were of known HPD-like motifs. The *E. coli* protein DjlC has a HPE motif, and interacts specifically with its partner protein, HscC. It was likely, given that there are far more HPD containing proteins than variants thereof, that the HPE motif evolved from the HPD motif, with a concurrent evolution in residues interacting with the HPE motif in HscC. DnaJ containing conservative substitutions of aspartic acid to glutamic acid (this work) or to asparagine (Wittung-Stafshede *et al.*, 2003) could not make a functional interaction with a partner Hsp70, indicating that both bulk and charge are important in this interaction. Hence the question that needs to be asked was why did a HPE motif develop? Additional work could focus on substituting the HPE motif to form a HPD motif in DjlC, and ascertain whether DnaJ-D35E could interact with HscC, and DjlC-E33D can interact with DnaK.

Similarly, why do the RESA family of proteins have such a wide variety of substitutions in the HPD motif. There must be a functional reason for such a shift in sequence, and this could prove to be a profitable area of investigation. It was possible that RESA proteins are fossilised, non-functional DnaJ-like proteins, as it does not appear that a functional interaction with a partner Hsp70 has been shown.

5.2.4 Helix III

Conserved charged residues in Helix III that were investigated were the lysine residues flanking the conserved phenylalanine (KFK motif) (Hennessy *et al.*, 2000). *Agt* DnaJ-K48T in particular showed a slight abrogation in function when replacing *E. coli* DnaJ and CbpA. This was the more conserved of the two lysines, and was presumably of functional importance. Three highly conserved structural residues are located in Helix III – Phe^{47}, Ala^{53} and Leu^{57}. Phe^{47} appears to be able to interact with His^{33}, and substitutions of this residue in previous experiments have given rise to DnaJ proteins with abrogated functions (Johnson and

Craig, 2000; Genevaux *et al.*, 2002). However the *Agt* DnaJ-F47L protein was capable of substituting for loss of *E. coli* DnaJ and CbpA. This may imply a slightly different interaction between *Agt* DnaJ and a partner Hsp70, than between *E. coli* DnaJ or yeast Ydj1 and their partner Hsp70s. Subtle differences such as these are important for determining correct, specific partnerships. *Agt* DnaJ-A53S was able to complement successfully for lack of *E. coli* DnaJ and CbpA, despite the presence of the hydroxyl group on the serine side chain. The increase of polarity was possibly compensated for by the rest of the J domain. Presumably, substitution of the alanine for bulkier polar and charged residues would negatively affect the J domain structure. *Agt* DnaJ-L57S was unable to complement successfully for the lack of *E. coli* DnaJ and CbpA. Presumably, either the loss of bulk, or addition of polarity has impacted negatively on the structural integrity of the J domain, leading to a non-functional protein.

5.2.5 H elix IV

Four conserved charged residues were investigated on Helix IV. The most conserved charged residue, Asp^{67}, was substituted for by alanine and asparagine. Both *Agt* DnaJ-D67A and *Agt* DnaJ-D6 7N were functional. This was surprising, as the residue was the only highly conserved, charged residue in the J domain. If it can be substituted for, without any apparent abrogation of function, why was it conserved? However, an additional conserved negatively charged residue, Asp^{59} was shown to be important. Substitution of this residue to alanine completely abrogated the ability of *Agt* DnaJ to replace the lack of *E. coli* DnaJ; however substitution to asparagine gave a functional protein. Hence the side chain of Asp^{59}, but not necessarily its negative charge, was important for J domain functioning. Arg^{63} was also shown to be important, as under some conditions *Agt* DnaJ-R63A was able to complement for lack of *E. coli* DnaJ, but not under others. This may imply that this mutation was destabilising the structure of the J domain, leading to varying species of *Agt* DnaJ-R63A, some of which are stable and functional, and some that are not. It does appear that residues in Helix IV are possibly required for J domain structure and function.

5.2.6 Conserved amino acids in the J domains of *A. tumefaciens* Type III DnaJ-like proteins

Performing a multiple sequence alignment of *E. coli* and *A. tumefaciens* proteins with either the ability or the inability to interact with *E. coli* DnaK shows some interesting trends developing (Figure 5.2). Firstly, assuming that the lack of a functional result for the chimeric *Agt* DnaJ proteins was due to sequence differences, the lack of an ability to replace *E. coli* DnaJ and CbpA in *E. coli* OD259 was not due to the loss of Arg^{26}. All sequences examined contained a positively charged residue at that point. Positions where the positive charge was lost are equivalent to Arg^{22}, Lys^{23} and Lys^{27} in *E.coli* DnaJ. However, no consistent results can be extrapolated, as some J domain sequences that could not functionally replace the *Agt* DnaJ J domain contain positively charged residues in the equivalent positions.

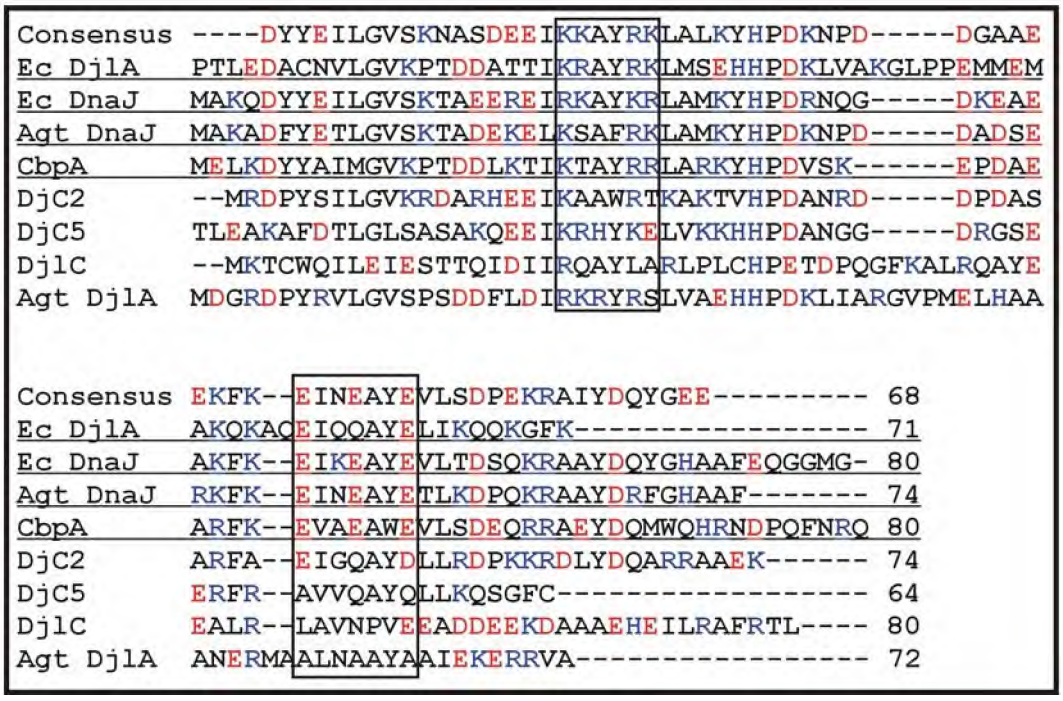

Figure 5.2: Multiple sequence alignment of J domains from DnaJ-like proteins in *E. coli* and *A. tumefaciens*

The *E. coli* proteins DnaJ (Ec DnaJ), DjlA (Ec DjlA), CbpA and DjlC, and the *A. tumefaciens* proteins *Agt* DnaJ, *Agt* DjC1 (*Agt* DjlA), *Agt* DjC2 and *Agt* DjC5, as well as the overall consensus sequence were aligned using ClustalW (Thompson *et al.*, 1994). J domains known to be able to interact with *E. coli* DnaK are underlined. Positively charged amino acids are in blue, and negatively charged amino acids are in red. Two possible regions interacting with DnaK are boxed.

A second group of residues was then noted. This was a group of three glutamic acid residues at positions 49, 52 and 55 in *E. coli* DnaJ. All J domains that could interact with *E. coli* DnaK contained these residues except for *E. coli* DjlA, which had a glutamine at the equivalent position to Glu^{52}. If this was a specificity motif, it would be expected that *Agrobacterium* DnaJ would bind to *E. coli* DnaK, as it contains glutamic acid, glutamine and aspartic acid at the three relevant positions.

However it was possible that a minimum of two glutamic acids in that region are required to interact with *E. coli* DnaK, and that aspartic acid can not successfully replace glutamic acid in this instance. Regardless, the one possible conclusion that could be drawn was that Helices III and IV provide the specificity determinants for the interaction with partner Hsp70s, and that Helix II provides the general binding determinants.

5.3 FUTURE WORK

5.3.1 Investigation of the interaction between substituted *Agt* DnaJ proteins, and *Agt* DnaK and *E. coli* DnaK

The next stage in the investigation of these residues will be to examine their effect of substitutions in these residues on the stimulation of the ATPase activity of both *Agt* DnaK and *E. coli* DnaK. There are potentially subtle differences in the interaction between *Agt* DnaJ and *Agt* Dna K, and between *Agt* DnaJ and *E. coli* DnaK. These subtleties may not be able to be recognised using the methods employed here and may require additional techniques such as SPR. This will allow an investigation of the levels of interaction, as an alternative to investigating stimulation of basal ATPase activity of a partner Hsp70. A DnaJ protein with substitutions in the J domain, may be able to bind to a partner Hsp70, but unable to stimulate the ATPase activity (Landry, 2003). Hence, employing other techniques will overcome this disadvantage. This level of subtlety may also be seen by showing that a substitution in one residue will abrogate the interaction with *E. coli* DnaK, but not necessarily with *Agt* DnaK. Stimulation of the ATPase activity will be the most appropriate start point in investigating this hypothesis.

5.3.2 Isolation of additional *A. tumefaciens* Hsp70 – DnaJ-partnerships

The discovery of an additional, potential Hsp70-DnaJ-like partnership in *A. tumefaciens* provides another avenue of exploration (*Agt* DnaK2 with *Agt* DjC2 and *Agt* DjC5). Isolation of that partnership, and characterising its interaction will possibly explain the inability of the *Agt* DjC2 and *Agt* DjC5 J domains to interact with *E. coli* DnaK.

There is also the obvious experiment of performing site-directed mutagenesis on the J domains of the chimeric *Agt* DnaJ proteins, to determine which residues are important for specific DnaK-DnaJ interactions, by looking for gain of function substitutions, as opposed to generic Hsp70– DnaJ-like interactions. A similar experiment was performed on Sec63-Sis1 chimeric protein, that allowed regain of a functional Sec63 protein (Schlenstedt *et al.*, 1995)

5.3.3 Generation of *A. tumefaciens dnaK* and *dnaJ* strains

Finally the last possibility is the generation of knockout *A. tumefaciens* strains to investigate the *Agt* DnaJ and DnaK in a homologous system as opposed to a heterologous system. The absence of a CbpA type gene could potentially imply that deletion of *Agt* DnaJ will be more deleterious to *A. tumefaciens* than the deletion of DnaJ was to *E. coli*.

5.3.4 Development of other model Hsp70 systems

Additional model systems from other organisms could also be isolated, and substitutions performed on the J domain of the DnaJ proteins. This would provide mechanism for validating results from this, and other work (Fewell *et al.*, 2002; Genevaux *et al.*, 2002).

5.3.5 Conclusion

In conclusion, residues have been identified that are important for the functioning of the J domain from *Agt* DnaJ. These include Arg^{26}, Asp^{59} and Arg^{63} at a functional level and Tyr^{7}, Leu^{10} and Leu^{57} at a structural level. Additional residues are potentially important, as formation of *Agt* DnaJ chimeric proteins gave rise to non-functional proteins. This was despite the retention of some conservation in the J domains used in the formation of chimeric *Agt* DnaJ proteins. Further work will be needed to determine which residues are important.

Appendices

6 APPENDICES

6.1 BIOINFORMATIC ANALYSIS OF THE J DOMAIN

Table 6.1: Universal Resource Locations (URL's)

Site	URL	Use
Align	http://xylian.igh.cnrs.fr/bin/align-guess.cgi	Pairwise alignments
BLAST	http://www.ncbi.nlm.nih.gov/BLAST	Determination of similar sequences
ClustalW	http://www2.ebi.ac.uk/clustalw/	Multiple alignments
Expasy	http://www.expasy.ch/tools/	
Genestream	http://xylian.igh.cnrs.fr/	
SWISS-MODEL	http://www.expasy.ch/swissmod/SWISS-MODEL.html	Molecular modelling
WhatIf	http://www.cmbi.nl/whatif/	Molecular modelling and analysis

The complete alignment and sequence databases that the alignment was generated from can be found on the attached CD-ROM.

Table 6.2: Distances between amino acids in the *E. coli* DnaJ J domain structure (1XBL)

Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond	Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond
No.	Type	Atom	No.	Type	Atom				No.	Type	Atom	No.	Type	Atom			
3	GLN	C	5	TYR	N	3.27		BB	33	PRO	C	36	ASN	N	3.58		BB
3	GLN	O	5	TYR	N	2.89	0.36	BB	33	PRO	O	36	ASN	N	2.86	0.54	BB
3	GLN	O	5	TYR	CD2	3.32		BS	33	PRO	O	36	ASN	CB	3.07		BS
3	GLN	CB	5	TYR	CE1	3.8		SS	33	PRO	O	36	ASN	CG	3.35		BS
4	ASP	C	6	TYR	N	3.03		BB	33	PRO	O	36	ASN	ND2	2.76	0.61	BS
4	ASP	O	6	TYR	N	3.05		BB	34	ASP	C	36	ASN	N	3.07		BB
4	ASP	OD1	6	TYR	N	2.75	0.54	SB	34	ASP	O	36	ASN	N	2.91		BB
4	ASP	OD1	6	TYR	CA	3.25		SB	34	ASP	O	36	ASN	ND2	3.08	0.34	BS
4	ASP	OD1	6	TYR	CB	3.18		SS	35	ARG	C	37	GLN	N	3		BB
4	ASP	O	7	GLU	N	3.12	0.23	BB	35	ARG	O	37	GLN	N	2.82	0.18	BB
4	ASP	OD1	7	GLU	N	2.79	0.58	SB	37	GLN	C	39	ASP	N	3.03		BB
4	ASP	O	8	ILE	N	3.04	0.64	BB	37	GLN	O	39	ASP	N	2.87	0.07	BB
4	ASP	CA	65	TYR	CZ	3.72		BS	38	GLY	C	40	LYS	N	3.4		BB
4	ASP	CB	65	TYR	OH	3.27		SS	39	ASP	C	41	GLU	N	3.08		BB
4	ASP	CG	65	TYR	OH	3.26		SS	39	ASP	O	41	GLU	N	2.9		BB
4	ASP	OD2	65	TYR	OH	2.71	0.62	SS	39	ASP	OD2	41	GLU	N	3.24		SB
5	TYR	C	7	GLU	N	3.67		BB	39	ASP	OD2	42	ALA	N	2.72	0.66	SB
5	TYR	CD1	8	ILE	CD1	3.77		SS	39	ASP	OD2	42	ALA	CB	3.01		SS
5	TYR	O	9	LEU	CD1	3.09		BS	40	LYS	C	42	ALA	N	3.38		BB
5	TYR	O	52	ALA	CB	3.33		BS	40	LYS	O	44	ALA	N	3.15	0.74	BB
5	TYR	CB	52	ALA	CA	3.67		SB	41	GLU	C	43	GLU	N	3.27		BB
5	TYR	CD1	52	ALA	CA	3.56		SB	41	GLU	C	44	ALA	N	3.44		BB
5	TYR	CD2	55	VAL	CB	3.71		SS	41	GLU	C	44	ALA	CB	3.8		BS
5	TYR	CD2	55	VAL	CG1	3.54		SS	41	GLU	O	44	ALA	N	2.84		BB
5	TYR	CD2	55	VAL	CG2	3.72		SS	41	GLU	O	44	ALA	CA	3.19		BB
5	TYR	CE2	55	VAL	CG2	3.71		SS	41	GLU	O	44	ALA	C	3.32		BB
5	TYR	O	56	LEU	CD1	3.34		BS	41	GLU	O	44	ALA	CB	3.1		BS
6	TYR	C	8	ILE	N	3.36		BB	41	GLU	O	45	LYS	N	2.69	0.5	BB
6	TYR	O	9	LEU	CD1	3.36		BS	41	GLU	CD	45	LYS	NZ	3.62		SS
6	TYR	O	10	GLY	N	2.85	0.58	BB	41	GLU	OE1	45	LYS	NZ	2.59	0.68	SS
6	TYR	CB	13	LYS	O	3.31		SB	42	ALA	C	44	ALA	N	3.25		BB
6	TYR	CG	14	THR	CG2	3.72		SS	42	ALA	C	45	LYS	N	3.61		BB
6	TYR	CD1	14	THR	CA	3.61		SB	42	ALA	O	45	LYS	N	3.08		BB

Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond	Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond
No.	Type	Atom	No.	Type	Atom				No.	Type	Atom	No.	Type	Atom			
6	TYR	CD2	14	THR	CG2	3.57		SS	42	ALA	O	45	LYS	CB	3.12		BS
6	TYR	CE1	14	THR	CA	3.73		SB	42	ALA	O	46	PHE	N	3.05	0.69	BB
6	TYR	CE1	14	THR	CB	3.57		SS	43	GLU	C	45	LYS	N	3.06		BB
6	TYR	CE2	14	THR	CG2	3.59		SS	43	GLU	O	45	LYS	N	3.08		BB
6	TYR	CZ	14	THR	CB	3.61		SS	43	GLU	C	46	PHE	N	3.6		BB
6	TYR	CZ	14	THR	CG2	3.79		SS	43	GLU	O	46	PHE	N	2.87		BB
6	TYR	CD1	56	LEU	CD1	3.46		SS	43	GLU	O	46	PHE	CA	3.27		BB
6	TYR	CE1	56	LEU	CD1	3.77		SS	43	GLU	O	46	PHE	CB	3.25		BS
6	TYR	CE1	56	LEU	CD2	3.41		SS	43	GLU	O	47	LYS	N	2.77	0.6	BB
6	TYR	CZ	56	LEU	CD2	3.7		SS	44	ALA	C	46	PHE	N	3.37		BB
6	TYR	OH	66	ASP	OD1	2.66	0.56	SS	44	ALA	C	47	LYS	N	3.43		BB
7	GLU	C	9	LEU	N	3.23		BB	44	ALA	O	47	LYS	N	3.05		BB
7	GLU	O	9	LEU	N	3.21		BB	44	ALA	O	47	LYS	CA	3.38		BB
7	GLU	O	9	LEU	C	3.17		BB	44	ALA	O	47	LYS	CB	3.26		BS
7	GLU	C	10	GLY	N	3.42		BB	44	ALA	O	48	GLU	N	2.86	0.67	BB
7	GLU	O	10	GLY	N	2.95		BB	45	LYS	C	47	LYS	N	3.17		BB
7	GLU	O	10	GLY	CA	3.27		BB	45	LYS	O	47	LYS	N	3.29		BB
8	ILE	C	10	GLY	N	3.64		BB	45	LYS	O	48	GLU	N	3.11		BB
8	ILE	CG2	48	GLU	C	3.72		SB	45	LYS	O	48	GLU	CB	3.37		BS
8	ILE	CG2	52	ALA	CB	3.74		SS	45	LYS	O	49	ILE	N	2.91	0.6	BB
9	LEU	CD2	20	ILE	CG2	3.64		SS	45	LYS	O	49	ILE	CG1	3.22		BS
9	LEU	CA	27	LEU	CD1	3.77		BS	45	LYS	O	49	ILE	CD1	3.33		BS
9	LEU	C	27	LEU	CD1	3.49		BS	46	PHE	C	48	GLU	N	3.35		BB
9	LEU	O	27	LEU	CD1	2.98		BS	46	PHE	O	49	ILE	N	3.27		BB
11	VAL	C	13	LYS	N	3.2		BB	46	PHE	O	50	LYS	N	2.86	0.72	BB
11	VAL	O	13	LYS	N	2.67	0.53	BB	46	PHE	O	50	LYS	CB	3.32		BS
11	VAL	CG2	19	GLU	CG	3.47		SS	47	LYS	C	49	ILE	N	3.28		BB
11	VAL	CG2	19	GLU	CD	3.69		SS	47	LYS	O	50	LYS	CB	3.38		BS
11	VAL	CG2	19	GLU	OE2	3.16		SS	47	LYS	O	51	GLU	N	2.75	0.61	BB
11	VAL	CG1	20	ILE	CD1	3.24		SS	47	LYS	O	51	GLU	CB	3.35		BS
12	SER	C	14	THR	N	3.56		BB	48	GLU	C	50	LYS	N	3.29		BB
12	SER	O	14	THR	N	3.21	0.19	BB	48	GLU	C	51	GLU	N	3.52		BB
13	LYS	C	15	ALA	N	3.53		BB	48	GLU	O	51	GLU	N	3.03		BB
16	GLU	C	18	ARG	N	2.91		BB	48	GLU	O	51	GLU	CB	3.21		BS

Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond	Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond
No.	Type	Atom	No.	Type	Atom				No.	Type	Atom	No.	Type	Atom			
16	GLU	O	18	ARG	N	2.88		BB	48	GLU	O	52	ALA	N	2.75	0.64	BB
16	GLU	CD	18	ARG	N	3.7		SB	48	GLU	O	52	ALA	CB	3.37		BS
16	GLU	CD	18	ARG	CB	3.65		SS	49	ILE	C	51	GLU	N	3.36		BB
16	GLU	OE2	18	ARG	N	3.18	0.12	SB	49	ILE	O	53	TYR	N	3.04	0.76	BB
16	GLU	O	19	GLU	N	2.83	0.49	BB	50	LYS	C	52	ALA	N	3.17		BB
16	GLU	O	20	ILE	N	3.06	0.8	BB	50	LYS	O	52	ALA	N	3.26		BB
17	GLU	C	19	GLU	N	3.19		BB	50	LYS	O	53	TYR	N	3.1		BB
17	GLU	O	19	GLU	N	3.15		BB	50	LYS	O	54	GLU	N	2.85	0.6	BB
17	GLU	C	20	ILE	N	3.6		BB	50	LYS	O	54	GLU	CB	3.14		BS
17	GLU	O	20	ILE	N	3.1		BB	51	GLU	C	53	TYR	N	3.38		BB
17	GLU	O	21	ARG	N	2.84	0.7	BB	51	GLU	C	54	GLU	N	3.69		BB
17	GLU	CG	53	TYR	CE1	3.64		SS	51	GLU	O	54	GLU	N	3.09		BB
17	GLU	CD	53	TYR	CE1	3.7		SS	51	GLU	O	54	GLU	CB	3.38		BS
17	GLU	OE2	53	TYR	CE1	3.39		SS	51	GLU	O	55	VAL	N	2.7 7	0.67	BB
17	GLU	OE2	53	TYR	OH	2.77	0.59	SS	51	GLU	O	55	VAL	CG2	3.34		BS
17	GLU	CG	57	THR	CG2	3.75		SS	52	ALA	C	54	GLU	N	3.1 4		BB
17	GLU	CD	57	THR	CG2	3.29		SS	52	ALA	O	54	GLU	N	3.14		BB
17	GLU	OE1	57	THR	CG2	3.1		SS	52	ALA	O	55	VAL	N	3.15		BB
18	ARG	C	20	ILE	N	3.57		BB	52	ALA	O	56	LEU	N	2.91	0.69	BB
18	ARG	O	22	LYS	CG	3.34		BS	52	ALA	O	56	LEU	CG	3.27		BS
18	ARG	O	22	LYS	NZ	2.71	0.73	BS	53	TYR	C	55	VAL	N	3.3 6		BB
19	GLU	C	21	ARG	N	3.13		BB	53	TYR	C	56	LEU	N	3.63		BB
19	GLU	O	21	ARG	N	3.13		BB	53	TYR	O	56	LEU	N	3.13		BB
19	GLU	C	22	LYS	N	3.56		BB	53	TYR	O	57	THR	N	3.29		BB
19	GLU	O	22	LYS	N	2.74	0.33	BB	53	TYR	CE1	57	THR	CG2	3.58		SS
19	GLU	O	22	LYS	CA	3.22		BB	53	TYR	CZ	57	THR	CG2	3.48		SS
19	GLU	O	22	LYS	CB	3.14		BS	54	GLU	C	56	LEU	N	3.19		BB
19	GLU	O	23	ALA	N	2.91	0.66	BB	54	GLU	C	57	THR	N	3.69		BB
20	ILE	C	22	LYS	N	3.11		BB	54	GLU	O	57	THR	N	3.07	0.44	BB
20	ILE	O	22	LYS	N	3.14		BB	54	GLU	O	58	ASP	N	3.22	0.25	BB
20	ILE	O	23	ALA	N	3.09		BB	55	VAL	C	57	THR	N	3.02		BB
20	ILE	O	24	TYR	N	2.75	0.72	BB	55	VAL	O	57	THR	N	2.96		BB
20	ILE	O	24	TYR	CB	3.3		BS	55	VAL	C	58	ASP	N	3.59		BB
20	ILE	CD1	56	LEU	CD1	3.42		SS	55	VAL	O	58	ASP	N	2.83	0.56	BB

Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond	Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond
No.	Type	Atom	No.	Type	Atom				No.	Type	Atom	No.	Type	Atom			
21	ARG	C	23	ALA	N	3.48		BB	55	VAL	O	58	ASP	CB	3.28		BS
21	ARG	NH1	24	TYR	CD2	3.28		SS	55	VAL	CG1	61	LYS	CB	3.6		SS
21	ARG	NH1	24	TYR	CE2	3.29		SS	55	VAL	CG2	61	LYS	CD	3.74		SS
21	ARG	O	25	LYS	N	2.83	0.58	BB	55	VAL	O	62	ARG	CB	3.24		BS
21	ARG	O	25	LYS	CB	3.36		BS	55	VAL	O	62	ARG	CG	3.01		BS
21	ARG	CD	53	TYR	CE1	3.27		SS	56	LEU	C	58	ASP	N	3.66		BB
21	ARG	CD	53	TYR	CZ	3.63		SS	56	LEU	CA	62	ARG	CG	3.71		BS
21	ARG	NE	53	TYR	CE1	3.57		SS	56	LEU	C	62	ARG	CZ	3.74		BS
21	ARG	NE	53	TYR	CZ	3.34		SS	56	LEU	O	62	ARG	CZ	3		BS
21	ARG	NE	53	TYR	OH	3.05		SS	56	LEU	O	62	ARG	NH2	2.73		BS
21	ARG	CZ	53	TYR	CE1	3.72		SS	58	ASP	CB	61	LYS	CE	3.67		SS
21	ARG	CZ	53	TYR	CE2	3.38		SS	58	ASP	CG	61	LYS	CE	3.33		SS
21	ARG	CZ	53	TYR	CZ	3.18		SS	58	ASP	CG	61	LYS	NZ	3.65		SS
21	ARG	CZ	53	TYR	OH	3.14		SS	58	ASP	OD2	61	LYS	CE	2.86		SS
21	ARG	NH1	53	TYR	CD2	3.36		SS	58	ASP	OD2	61	LYS	NZ	2.7	0.46	SS
21	ARG	NH1	53	TYR	CE1	3.55		SS	58	ASP	C	62	ARG	CB	3.8		BS
21	ARG	NH1	53	TYR	CE2	3.08		SS	58	ASP	O	62	ARG	CB	3.29		BS
21	ARG	NH1	53	TYR	CZ	3.22		SS	59	SER	C	61	LYS	N	3.66		BB
22	LYS	C	24	TYR	N	3.32		BB	60	GLN	C	62	ARG	N	2.97		BB
22	LYS	O	25	LYS	N	3.28		BB	60	GLN	O	62	ARG	N	2.95		BB
22	LYS	O	26	ARG	N	3.13	0.72	BB	60	GLN	O	63	ALA	N	2.99	0.48	BB
23	ALA	C	25	LYS	N	3.19		BB	60	GLN	O	64	ALA	N	2.88	0.81	BB
23	ALA	O	25	LYS	N	3.13		BB	60	GLN	NE2	73	PHE	O	2.91	0.35	SB
23	ALA	C	26	ARG	CG	3.76		BS	61	LYS	C	63	ALA	N	3.27		BB
23	ALA	O	26	ARG	N	3.07	0.12	BB	63	ALA	C	65	TYR	N	3.23		BB
23	ALA	O	26	ARG	CG	2.97		BS	63	ALA	O	65	TYR	N	3.25		BB
23	ALA	C	27	LEU	CD1	3.71		BS	63	ALA	C	66	ASP	N	3.59		BB
23	ALA	O	27	LEU	N	2.88	0.8	BB	63	ALA	O	66	ASP	N	2.91	0.05	BB
23	ALA	O	27	LEU	CB	3.35		BS	63	ALA	O	66	ASP	CB	3.31		BS
23	ALA	O	27	LEU	CD1	3.26		BS	63	ALA	O	67	GLN	N	2.89	0.76	BB
24	TYR	C	26	ARG	N	3.46		BB	64	ALA	C	66	ASP	N	3.28		BB
24	TYR	O	28	ALA	N	2.77	0.59	BB	64	ALA	O	68	TYR	N	2.79	0.56	BB
24	TYR	O	28	ALA	CB	2.96		BS	64	ALA	O	68	TYR	C	3.03		BB
24	TYR	OH	50	LYS	CD	3.35		SS	64	ALA	C	69	GLY	O	3.1		BB

Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond	Amino acid 1			Amino acid 2			Dist	H bond[1]	Bond
No.	Type	Atom	No.	Type	Atom				No.	Type	Atom	No.	Type	Atom			
25	LYS	C	27	LEU	N	3.29		BB	64	ALA	O	69	GLY	N	3.02		BB
26	ARG	C	28	ALA	N	3.36		BB	65	TYR	C	67	GLN	N	3.24		BB
26	ARG	O	29	MET	CG	3.07		BS	65	TYR	N	69	GLY	O	3.1		BB
26	ARG	O	30	LYS	N	2.98	0.7	BB	65	TYR	CA	69	GLY	N	3.49		BB
27	LEU	C	29	MET	N	3.3		BB	65	TYR	CA	69	GLY	CA	3.72		BB
27	LEU	C	30	LYS	N	3.69		BB	65	TYR	C	69	GLY	N	3.64		BB
27	LEU	O	30	LYS	N	3.11		BB	65	TYR	O	69	GLY	N	3.14	0.57	BB
27	LEU	O	31	TYR	N	2.8	0.6	BB	67	GLN	CB	72	ALA	CB	3.69		SS
27	LEU	O	31	TYR	CB	3.35		BS	67	GLN	CD	72	ALA	CB	3.53		SS
27	LEU	CD2	31	TYR	CE2	3.5		SS	67	GLN	OE1	73	PHE	CE2	3.23		SS
28	ALA	C	30	LYS	N	3.34		BB	67	GLN	OE1	73	PHE	CZ	2.89		SS
28	ALA	C	32	HIS	ND1	3.68		BS	68	TYR	C	70	HIS	N	3.45		BB
28	ALA	O	32	HIS	N	3.05	0.19	BB	68	TYR	O	70	HIS	N	2.79		BB
28	ALA	O	32	HIS	ND1	2.81	0.52	BS	68	TYR	O	71	ALA	N	2.79	0.66	BB
28	ALA	CB	46	PHE	CE1	3.56		SS	68	TYR	CB	71	ALA	O	3.32		SB
29	MET	C	31	TYR	N	3.26		BB	70	HIS	CA	74	GLU	CD	3.63		BS
29	MET	O	32	HIS	N	3.27	0.46	BB	70	HIS	CA	74	GLU	OE2	3.22		BS
30	LYS	C	32	HIS	N	3.26		BB	70	HIS	C	74	GLU	CD	3.23		BS
31	TYR	CE1	45	LYS	CG	3.59		SS	70	HIS	C	74	GLU	OE1	3.22		BS
31	TYR	CE1	45	LYS	CD	3.18		SS	70	HIS	C	74	GLU	OE2	3.33		BS
31	TYR	CZ	45	LYS	CG	3.68		SS	71	ALA	N	74	GLU	CD	3.58		BS
31	TYR	CZ	45	LYS	CD	3.54		SS	71	ALA	N	74	GLU	OE1	3.18		BS
31	TYR	CD2	49	ILE	CD1	3.14		SS	71	ALA	CA	74	GLU	OE1	2.98		BS
31	TYR	CE2	49	ILE	CD1	3.22		SS	71	ALA	C	74	GLU	OE1	3.28		BS
32	HIS	C	34	ASP	N	3.39		BB	72	ALA	N	74	GLU	CD	3.5		BS
32	HIS	CE1	46	PHE	CE2	3.33		SS	72	ALA	N	74	GLU	OE1	2.68	0.74	BS
32	HIS	CE1	46	PHE	CZ	3.52		SS	73	PHE	C	75	GLN	N	3.36		BB
33	PRO	C	35	ARG	N	3.53		BB									

Highlighted cells contain bonds that involve the HPD motif.
Numbering was done according to the numbers in the structure file, which lacks the initial Methionine.
[1] – This column gives an idea of the quality of the hydrogen bond, if appropriate. The nearer the value is to 1, the better the bond. No value means there is no hydrogen bond between those two atoms.

6.2 STRAINS OF *E. COLI*, PLASMIDS AND PRIMERS

Table 6.3: Genotypes of *E. coli* strains used in this thesis

E. coli strains [1]	Genotype
OD259 / WKG190	MC4100 *araD139* Δ*ara714* Δ*cbpA*::*ka*n *dnaJ*::Tn*10*-42
XL1-Blue	*recA1 endA1 gyrA96 th i-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*Z*Δ*M15* Tn*10* (Tetr)]

[1] – Deloche *et al.*, 1997; This strain is resistant to kanamycin, hence all cultures were grown in the presence of 25 μg/ml kanamycin.

Table 6.4: Plasmids used in this thesis

Plasmid	Function	Other
pQE30	T5 promoter-based expression vector used; allows insertion of a coding region in frame with the codons for six histidines	Qiagen (USA)
pRJ30	pQE30 containing *Agt dnaJ* coding region	Generated in this thesis
pBAD22A	Plasmid used as a negative control for complementation	Deloche *et al.*, 1997
pWKG90	pBAD22A containing *E. coli dnaJ* coding region	Deloche *et al.*, 1997
pRJ30-F74	*Agt* DnaJ containing a *Sfu* I site for domain swapping	Generated in this thesis
pGEM-T Easy	Cloning of PCR products	Promega (USA)

All plasmids contained the β-lactamase gene, which allows ampicillin resistance, and hence were selected for by including 100 μg/ml ampicillin in cultures

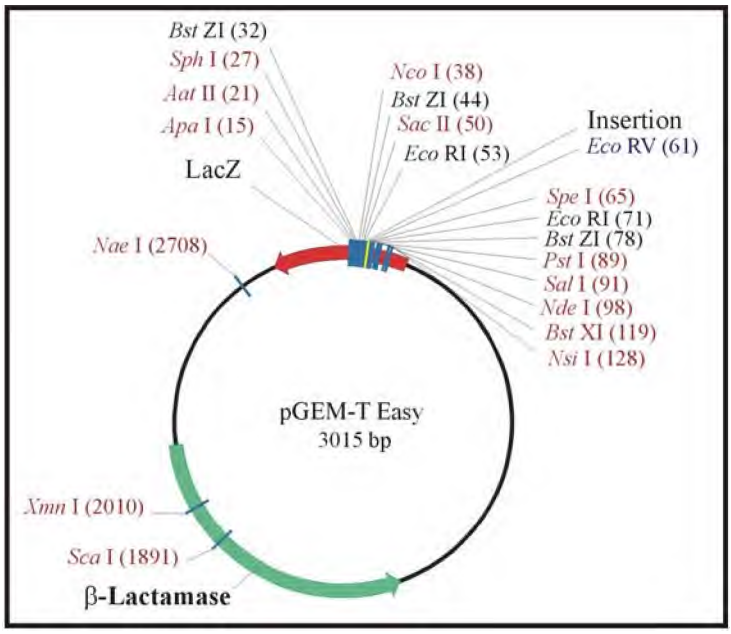

Figure 6.1: Plasmid map of pGEM-T Easy

Unique restriction sites are in red; multiple cut sites are in black. The *Eco* RV site which is the site of insertion by PCR fragments is in blue, and is indicated in yellow on the map. The ampicillin resistance gene encoding β-lactamase is in green, and the *lacZ* gene, which is interrupted during a succe ssful insertion of a PCR product, is in red.

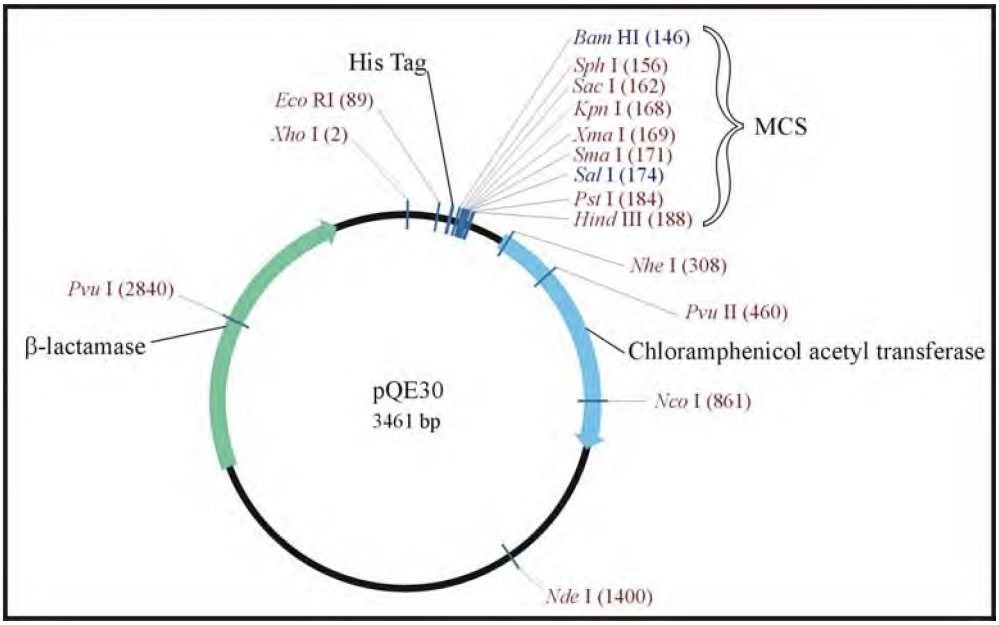

Figure 6.2: Plasmid map of pQE30

The *Bam* HI and *Sal* I sites used for the insertion of the *Agt dnaJ* coding region are shown in blue. Restriction sites present in the multiple cloning site (MCS) are also shown. The chloramphenicol acetyl transferase coding region is shown as a blue arrow, and the β-lactamase-encoding gene is shown in green

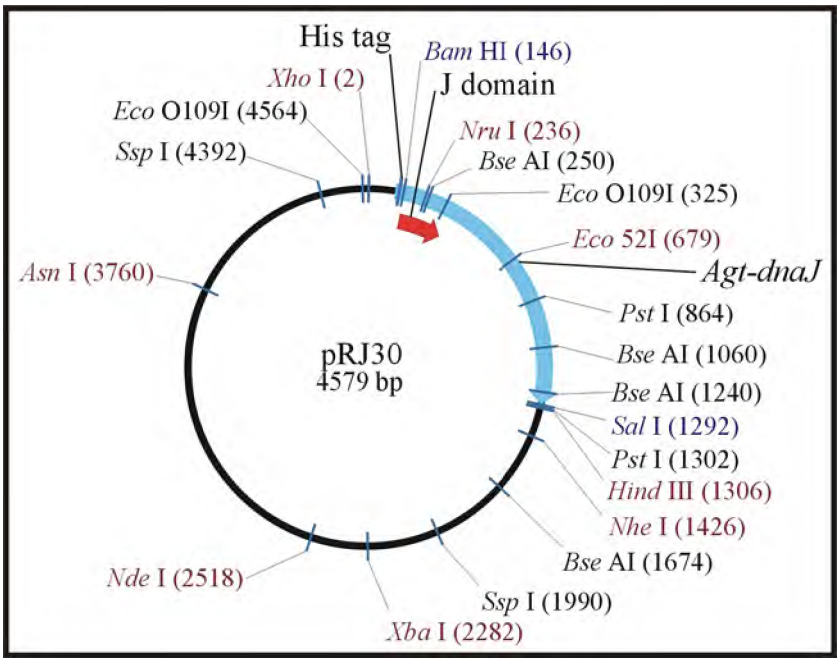

Figure 6.3: Plasmid map of pRJ30 showing diagnostic restriction endonuclease sites for confirmation of mutagenesis

The position of restriction sites used for confirmation of mutagenesis are shown. The position of *Bam* HI and *Sal* I are depicted in blue, unique restriction sites are shown in red, and multiple cutters in black.

Table 6.5: Primers used for PCR, sequencing and mutagenesis

All sequences are given from 5' to 3'.

PCR primers

Name	Sequence	Use	Restriction site
AtDnaJF	GGATCCATGGCGAAAGCAGACT	Amplifying *Agt dnaJ*	*Bam* HI
AtDnaJR	GTCGACTCAGCCSTCAAAGAATTYYTYCATCCG	Amplifying *Agt dnaJ*	*Sal* I
C1F	GGATCCATGCTTGCCTCTATGATGTTCG	Amplifying *Agt djC1*	*Bam* HI
C1R	CTCGAGTCATGCGACGCGGCGTTC	Amplifying *Agt djC1*	*Xho* I
C2F	AGATCTATGCGCGATCCTTATTCT	Amplifying *Agt djC2*	*Bgl* II
C2R	GTCGACTCACAAATACAGGCCTTC	Amplifying *Agt djC2*	*Sal* I
C3F	AGATCTATGATCGATCCCTATGTT	Amplifying *Agt djC3*	*Bgl* II
C3R	GTCGACTTATTCCGCCGCTTCGCC	Amplifying *Agt djC*3	*Sal* I
C4F	GGATCCATGGCATTTCAAAAGACC	Amplifying *Agt djC4*	*Bam* HI
C4R	GTCGACTCAGCAAAAACCGTTCAG	Amplifying *Agt djC4*	*Sal* I
C5F	AGATCTATGAAATTGGATTCGAAA	Amplifying *Agt djC5*	*Bam* HI
C5R	GTCGACTTAGCAGAAACCTGACTG	Amplifying *Agt djC5*	*Sal* I
CJ1F	GGATCCATGGATGGGCGCGATC	J domain of *Agt* DjC1	*Bam* HI
CJ1R	TTCGAATTGCGACGCGGCGTTC	J domain of *Agt* DjC1	*Sfu* I
CJ2R	TTCGAACTTCTCGGCCGCCCTGCG	J domain of *Agt* DjC2	*Sfu* I
CJ5F	GGATCCACGCTGGAGGCA	J domain of *Agt* DjC5	*Bam* HI
CJ5R	TTCGAAGCAGAAACCTGACTG	J domain of *Agt* DjC5	*Sfu* I

Sequencing primers

Name	Sequence	Use
DnaJseqF1	TGA TGA TGC CGA TTC CGA AC	DnaJ sequencing
DnaJseqF2	CGT GTG ACG GAA GAG CGT T	DnaJ sequencing
DnaJseqR1	AAA GAT CAG CAC CGT CAC GC	DnaJ sequencing
DnaJseqR2	GAA GAT ATC GGA GAA GCC	DnaJ sequencing
pQEforward	CCC GAA AAG TGC CAC CTG	Sequencing pQE vectors
pQEreverse	GTT CTG AGG TCA TTA CTG G	Sequencing pQE vectors
pUCF	CGC CAG GGT TTT CCC AGT CAC GAC	Sequencing pGEM clones
pUCR	TCA CAC AGG AAA CAG CTA TGA C	Sequencing pGEM clones

Mutagenic primers

Name *	Sequence	Restriction Enzyme
DY7LF	AAAGCAGACTTTCTAGAAACACTTGGC	*Xba* I
DY7LR	GCCAAGTGTTTCTAGAAAGTCTGCTTT	*Xba* I
DL10AF	TTTTACGAAACGGCCGGCGTCAGCAAG	*Eco* 52 I
DL10AR	CTTGCTGACGCCGGCCGTTTCGTAAAA	*Eco* 52 I
DE20AF	GCGGACGAAAAAGCTTTGAAAAGCGCC	*Hind* III
DE20AR	GGCGCTTTTCAAAGCTTTTTCGTCCGC	*Hind* III
DK22AF	GAAAAAGAGCTGGCTAGCGCCTTCCG	*Nhe* I
DK22AR	CGGAAGGCGCTAGCCAGCTCTTTTTC	*Nhe* I
DR26AF	AAAAGCGCCTTCGCGAAACTAGCAATGAAATACCAT	*Nru* I
DR26AR	ATGGTATTTCATTGCTAGTTTCGCGAAGGCGCTTTT	*Nru* I
DK27AF	GCCTTCCGCGCGCTAGCAATGAAATACCAT	*Nhe* I
DK27AR	ATGGTATTTCATTGCTAGCGCGCGGAAGGC	*Nhe* I
DR26A,K27AF	GCCTTCGCAGCGCTAGCAATGAAATACCAT	*Nhe* I
DR26A,K27AR	ATGGTATTTCATTGCTAGCGCTGCGAAGGC	*Nhe* I
DY32LF	CTCGCGATGAAGCTTCATCCGGACAAA	*Hind* III
DY32LR	TTTGTCCGGATGAAGCTTCATCGCGAG	*Hind* III
DH33QF	GCGATGAAATACCAACCGGACAAAAACCC	*Bse* AI
DH33QR	GGGTTTTTGTCCGGTTGGTATTTCATCGC	*Bse* AI
DH33Y, D35YF	CTCGCGATGAAATATTATCCTTATAAAAACCCTGATGAT	*Ssp* I
DH33Y, D35YR	ATCATCAGGGTTTTTATAAGGATAATATTTCATCGCGAG	*Ssp* I
DD35EF	AAATACCATCCAGAAAAAAACCCTGAT	*Bse* AI
DD35ER	ATCAGGGTTTTTTTCTGGATGGTATTT	*Bse* AI
DK46TF	GCCGATTCCGAACGTACGTTCAAAGAAATC	*Bsi* WI
DK46TR	GATTTCTTTGAACGTACGTTCGGAATCGGC	*Bsi* WI
DF47LF	TCCGAACGGAAACTTAAGGAAATCAACGAA	*Bst* 98 I
DF47LR	TTCGTTGATTTCCTTAAGTTTCCGTTCGGA	*Bst* 98 I
DK48TF	GAACGGAAATTCACAGAAATTAATGAGGCCTATGAAACG	*Asn* I
DK48TR	CGTTTCATAGGCCTCATTAATTTCTGTGAATTTCCGTTC	*Asn* I
DA53SF	GAAATCAACGAATCATATGAAACGCTG	*Nde* I
DA53SR	CAGCGTTTCATATGATTCGTTGATTTC	*Nde* I
DL57SF	GCCTATGAAACTTCGAAGGACCCGCAG	*Sfu* I
DL57SR	CTGCGGGTCCTTCGAAGTTTCATAGGC	*Sfu* I
DD59AF	GAAACGCTGAAGGCCCCGCAGAAG	*Eco* O 109 I
DD59AR	CTTCTGCGGGGCCTTCAGCGTTTC	*Eco* O 109 I

Name *	Sequence	Restriction Enzyme
DD59NF	GAAACGCTGAAGAACCCGCAGAAG	*Eco* O 109 I
DD59NR	CTTCTGCGGGTTCTTCAGCGTTTC	*Eco* O 109 I
DK62AF	TGAAGGACCCACAGGCTCGAGCGGC	*Xho* I
DK62AR	GCCGCTCGAGCCTGTGGGTCCTTCAG	*Xho* I
DR63AF	CTGAAGGACCCACAGAAAGCTGCAGCCTATGACCGT	*Pst* I
DR63AR	ACGGTCATAGGCTGCAGCTTTCTGTGGGTCCTTCAG	*Pst* I
DY66AF	CAGAAGCGCGCAGCTGCAGACCGTTTCGGC	*Pst* I
DY66AR	GCCGAAACGGTCTGCAGCTGCGCGCTTCTG	*Pst* I
DD67AF	CAGAAGCGCGCAGCATATGCACGTTTCGGCCAC	*Nde* I
DD67AR	GTGGCCGAAACGTGCATATGCTGCGCGCTTCTG	*Nde* I
DD67NF	GCGGCCTATAATCGATTCGG	*Nde* I
DD67NR	CCGAATCGATTATAGGGCGC	*Nde* I

- Naming of mutagenic primers was done as follows: an initial D for DnaJ followed by the amino acid in single letter code and its position, followed by the substituted amino acid. F refers to the forward primer, and R to the reverse primer.

6.3 COMPOUNDS AND SUPPLIERS

Compounds and suppliers used in this thesis are given in Table 6.6 below.

Table 6.6: Compounds and suppliers used in this thesis

Compound	Supplier	Country of Origin
Acrylamide	Sigma	Germany
Agar (Bacteriological)	Biolab (Merck)	South Africa
Agarose	WhiteSci	South Africa
Ammonium acetate	Saarchem	South Africa
Ammonium heptamolybdate trihydrate	Merck	Germany
Ammonium persulphate	Saarchem	South Africa
Ammonium sulphate	Saarchem	South Africa
Ampicillin	Roche	Germany
Arabinose, L (+)	Sigma	Germany
Ascorbic acid, L (+)	Merck	Germany
Bis-Acrylamide	Sigma	Germany
Boric Acid	Saarchem	South Africa
Bromophenol Blue	Sigma	Germany
BSA	Roche	Germany
Calcium Chloride	Saarchem	South Africa
Chloroform:Isoamyl alcohol (24:1)	Sigma	Germany
Coomassie Brilliant Blue G250	USB	USA
Coomassie Brilliant Blue R250	USB	USA
CTAB	Sigma	Germany
EDTA	Saarchem	South Africa
Ethidium Bromide	Sigma	Germany
Glucose	Saarchem	South Africa
Glycine	Sigma	Germany
Imidazole	Sigma	Germany
IPTG	Roche	Germany
Kanamycin sulphate	Roche	Germany
Lambda DNA	Promega	USA
Lysozyme	Roche	Germany
MacConkey Agar	Biolab (Merck)	South Africa
Magnesium chloride	Saarchem	South Africa

Compound	Supplier	Country of Origin
Mercaptoethanol	Merck	Germany
Nickel sulphate	Fluka	Germany
Nucleotide mix	Roche	Germany
Phenol:Chloroform:Isoamyl alcohol (25:24:1)	Sigma	Germany
PMSF	Sigma	Germany
Ponceau S	Sigma	Germany
Potassium acetate	Saarchem	South Africa
Potassium chloride	Saarchem	South Africa
Potassium hydroxide	BDH	Germany
Potassium phosphate; K_2H	Merck	Germany
Potassium phosphate; KH_2	Merck	Germany
Proteinase K	Roche	Germany
Sephadex G-25	Amersham	Sweden
Sepharose fast flow, Chelating	Amersham	Sweden
Sodium acetate	Merck	Germany
Sodium azide	Sigma	Germany
Sodium chloride	Saarchem	South Africa
Sodium dodecyl sulphate	BDH	Germany
Sodium hydroxide	Saarchem	South Africa
Sodium Phosphate; Na_2H	Saarchem	South Africa
Sodium Phosphate; NaH_2	BDH	Germany
Sucrose	Sigma	Germany
TEMED	Sigma	Germany
Tris	Sigma	Germany
Triton X-100	Sigma	Germany
Tryptone powder	Oxoid	England
Tween 20	Saarchem	South Africa
Yeast extract	Oxoid	England

Restriction endonucleases were purchased from Amersham (Sweden) or from Roche (Germany).

6.4 STANDARD PROCEDURES

Standard molecular biology techniques protocols are summarised below (Sambrook and Russell, 2001).

6.4.1 DNA isolation and usage

6.4.1.1 Preparation of plasmid DNA

Plasmid DNA was prepared using the QiaPrep kit (Qiagen, USA), a modified alkaline lysis method (Birnboim and Doly, 1979). Briefly, the protocol is as follows. *E. coli* XL1 Blue containing the plasmid of interest was grown overnight in five ml cultures YT broth (1.6 % Tryptone, 1 % yeast extract, 0.5 % NaCl) containing the appropriate antibiotic. The bacterial pellet was resuspended in the manufacturers buffer P1 (250 µl). The manufacturers buffer P2 was added (250 µl), followed by the manufacturers buffer P3 (350 µl). The samples were then centrifuged for 10 minutes. The supernatant was applied to a Qiaprep spin column placed in a 2 ml collection tube. The tubes were centrifuged for 60 seconds in a microcentrifuge at 16 000 x g, washed with 0.75 ml the manufacturers buffer PE, and spun for 60 seconds at maximum speed (16 000 x g) in a microcentrifuge. The tubes were centrifuged for an additional minute to remove any residual buffer at 16 000 x g in a microcentrifuge. The column was placed in a clean Eppendorf and 50 µl of sterile distilled water was added to the center of the column. This was left to stand for one minute, and thereafter centrifuged for one minute at 16 000 x g in a microcentrifuge to elute the DNA.

6.4.1.2 Restriction of plasmid DNA

Plasmid DNA was prepared as described (Appendix, Section 6.4.1.1). This DNA was restricted with restriction endonucleases for two hour at the appropriate temperature. The reaction mixture comprised plasmid DNA (300 ng), the appropriate restriction buffer (2 µl), and 2 U of the enzyme, up to a total volume of 20 µl. The restriction digest was resolved by gel electrophoresis using a 0.8 % agarose gel as described in the Appendix (Section 6.4.1.3).

6.4.1.3 Resolving of DNA on agarose gels

Molecular grade agarose was melted in 0.5 x Tris-Borate-EDTA (ethylene diamine tetra-acetic acid) buffer (TBE; 45 mM Tris, 45 mM Borate, 1 mM EDTA) to give a percentage of 0.8 % (w/v) agarose. The mixture was allowed to cool until hand hot and ethidium bromide was added to a final concentration of 0.5 μg/ml. This was left to set in a sealed gel platform. DNA gel loading buffer (6x loading buffer; 0.25 % bromophenol blue, 30 % glycerol) was added to the DNA samples at a ratio of 1 μl buffer to 5 μl DNA. The marker used was λ DNA restricted with *Pst* I. Agarose gels were run at 100 V for an hour, and the DNA was visualised using ultra-violet (UV) light. Smaller DNA fragments (< 300 bp) were resolved on 2 % (w/v) agarose gels.

6.4.1.4 Purification of DNA from an agarose gel

Purification of DNA from an agarose gel was performed using the Nucleospin kit (Clontech, USA). Restricted DNA was resolved using a 0.8 % agarose gel (Appendix, Section 6.4.1.3). The location of the band of interest was determined using long wave UV light. The band was excised and placed in a sterile Eppendorf tube. Buffer NT1 (manufacturers buffer) was added to each agarose gel fragment (300 μl buffer for every 100 mg). This was incubated at 50°C for 10 minutes. The liquid sample was loaded into a Nucleospin column and centrifuged for 60 seconds at maximum speed (16 000 x g) in a microcentrifuge. The column was washed with buffer NT3 (manufacturers buffer) (600 μl) and centrifuged for one minute at maximum speed (16 000 x g) in a microcentrifuge. A further wash step using buffer NT3 was performed (200 μl), and centrifugation was performed for two minutes in a microcentrifuge at maximum speed (16 000 x g) to remove any residual buffer. The column was placed into a clean Eppendorf tube, and elution buffer NE (manufacturers buffer) (50 μl) was added. This was left at room temperature for one minute, and the DNA was thereafter eluted by centrifugation for one minute at top speed (16 000 x g) in a microcentrifuge.

6.4.1.5 Alkaline Phosphatase Treatment of restricted DNA fragments

The plasmid to be treated with alkaline phosphatase was restricted with the appropriate restriction endonuclease as described in the Appendix (Section 6.4.1.2). The vector was

treated with shrimp alkaline phosphatase (SAP; Roche, Germany) to prevent recircularisation of the plasmid vector. The restriction endonuclease was inactivated by heating the restriction mixture (20 µl volume) to 65°C for 15 minutes. This was followed by the addition of 10 x dephosphorylation buffer (0.5 M Tris-HCl, 50 mM MgC l_2, pH 8.5) (2 µl), 1 U of the SAP and 2 µl water. This mixture was incubated for 15 minutes at 37°C, and for a further 15 minutes at 65°C. A further 1 µl of the SAP was added, and the reaction mixture was incubated for 15 minutes at 37°C, and thereafter for 15 minutes at 65°C. The DNA was then used for ligation reactions.

6.4.1.6 Ligation of DNA fragments

T4 DNA ligase (Roche, Germany) was used for ligation of DNA fragments. Appropriate amounts of the DNA fragments (500 ng) to be ligated were incubated with the 10 x ligase buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 10 mM ATP, pH 7.5) and 1 U of the ligase, to a final volume of 30 µl. All ligation reactions were performed overnight at 15°C, or at room temperature for four hours. Resultant products were transformed into competent *E. coli* XL1-Blue (Appendix, Section 6.4.1.8).

6.4.1.7 Ligation of PCR fragments into pGEM-T Easy

The standard protocol for inserting PCR products into pGEM-T (Promega, USA) was followed. Briefly, the PCR product (100 ng) was added to 2 x rapid ligation buffer (60 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP, 10 % polyethylene glycol) and pGEM-T Easy vector (50 ng). T4 DNA ligase was added (0.3 U). The mixture was incubated at 4°C overnight and transformed into competent *E. coli* XL1-Blue (Appendix, Section 6.4.1.8). Cells were plated on MacConkey agar and white colonies were selected for DNA isolation, as described previously. The isolated plasmid DNA was restricted with *Eco* RI or *Not* I, which release inserted fragments (Appendix, Figure 6.1).

6.4.1.8 Preparation of competent cells, and transformation of DNA

Competent *E. coli* were prepared using a modification of the calcium chloride method (Dagert and Ehrlich, 1979). The required strain was grown overnight with the appropriate antibiotic, if any. The overnight culture was diluted 1 in 200 in YT broth and grown in containers 10 x culture volume until early log phase was reached. The cells were centrifuged at 5 000 x g for 5 minutes at 4 °C. The cells were resuspended in one culture volume ice-cold 0.1 M $MgCl_2$ and left at 4°C for 1 minute. They were thereafter centrifuged at 5 000 x g for 5 minutes at 4°C. They were resuspended in 1/2 culture volume ice cold 0.1 M $CaCl_2$ and left at 4°C for 2 hours. The cells were centrifuged again at 5 000 x g for 5 minutes at 4°C and resuspended in 1/10 culture volume 0.1 M $CaCl_2$. The cells had one volume ice-cold, sterile 30 % glycerol added, and were stored at –70°C.

Transformations were performed as follows. Ligation reactions or plasmid DNA to be transformed were added to 100 µl of the competent cells. The cells were left on ice for 20 minutes. The mixture was heat shocked at 42°C for 2 minutes. The cells were returned to ice for one minute, and 0.9 ml YT broth was added. The cells were left to produce antibiotic resistance for 1 hour at 37°C and subsequently plated on YT-agar plates (15 g agar per 1 l broth) containing appropriate antibiotics. Selected colonies were grown in YT broth and plasmid DNA was recovered as described previously (Section 6.4.1.1).

The generation of competent *E. coli* OD259 required a change in these procedures. Cells were grown at 30°C, as opposed to 37°C during the initial growth steps. The transformation protocol was also altered to allow for an hour growth at 30°C as opposed to 37°C, and plates were incubated overnight at 30°C. This alteration was necessary to ensure that any partial heat stress did not negatively impact on the transformation results.

6.4.1.9 Site directed mutagenesis

Site directed mutagenesis was performed using the QuikChange site directed mutagenesis kit (Stratagene, USA). The reaction was performed according to the manufacturer's instructions, with minor modifications. Plasmid DNA was prepared using the Qiagen miniprep kit

(Qiagen, USA) (Section 6.4.1.1) and used in the reaction. A reaction mixture containing the reaction buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl, pH 8.0, 2 mM $MgSO_4$, 0.1 % Triton X-100, 0.1 mg/ml BSA), 1 µl of the manufacturers dNTPs, plasmid DNA (50 ng), forward and reverse primers (125 ng of each) and 2.5 U of the *Pfu* polymerase was used. The reaction was cycled as follows; one cycle of 95°C for 30 seconds, followed by 18 cycles comprising 95°C for 30 seconds, 52°C for 30 seconds and 68°C for 8 minutes. The reaction was thereafter cooled to 37°C. After the thermocycling the reaction had 10 U *Dpn* I added to it. This reaction was left to proceed at 37°C for 2 hours. Samples (10 µl) of the reaction before and after the addition of *Dpn* I were analysed on a 0.8 % agarose gel. The *Dpn* I treated DNA was used to transform supercompetent *E. coli* cells. Product from the reaction (1 µl) was added to 50 µl cells, and this was left on ice for 30 minutes. The cells were heat shocked at 42°C for 45 seconds, and returned to 4°C for two minutes. YT broth (500 µl) was added, and the cells were incubated at 37°C for one hour. The transformants were plated onto YT agar plates containing 100 µ/ml ampicillin, and incubated overnight at 37°C.

6.4.1.10 DNA sequencing

All constructs that were sequenced were prepared using the Qiagen Miniprep kit using the manufacturer's protocol (Qiagen, USA) (Section 6.4.1.1). The plasmid DNA was eluted in 50 µl sterile distilled water. The DNA was sequenced using the BigDye V3.1 Terminator cycle sequencing kit (Applied Bioystems, USA). Primers used for sequencing were the generic pUCR and pUCF primers when inserts were in pGEM-T based constructs, and the generic pQEforward and pQEreverse primers when inserts were in pQE30 based constructs. The internal, specific DnaJ primers DnaJF1, DnaJF2, DnaJR1 and DnaJR2 were used for sequencing *Agt dnaJ.* The sequences for all primers are given in the Appendix (Section 6.2, Table 6.4). Sequencing of coding regions was performed on three independent isolates.

Plasmid DNA (500 ng) and the appropriate primer (3.2 pmol) had 8 µl of the Big Dye added, in a total volume of 20 µl. The following thermal cycling protocol was performed for 25 cycles: 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. The extension products obtained were purified using Zymogen columns (Section 6.4.1.11) using the standard protocols. This allowed for the removal of unincorporated dye terminators. The

DNA was dried at 37°C, and then sequenced on a 3100 ABI Genetic Analyser. Sequence information was analysed using Vector NTi (Informax, USA).

6.4.1.11 Clean up sequencing products using Zymogen columns

Removal of unincorporated dye terminators after cycle sequencing was performed using Zymogen columns (Zymo research, USA). DNA binding buffer (100 µl) was added to the sequencing reaction. This sample was loaded onto a Zymo-Spin column, placed into a collection tube, and centrifuged at 16 000 x g for 30 seconds in a microcentrifuge. The flow through was discarded. Wash buffer was added to the column (200 µl). The c olumn was centrifuged at 16 000 x g for 30 seconds in a microcentrifuge. A further wash step was performed with the wash buffer (200 µl) followed by a centrifugation step at 16 000 x g for 30 seconds in a microcentrifuge. Sterile, distilled water was added to the column matrix (6 to 8 µl). The column was placed in a clean Eppendorf, and DNA was recovered by centrifugation at 16 000 x g for 30 seconds in a microcentrifuge.

6.4.2 Purification of Histidine-tagged proteins

6.4.2.1 Determination of optimal levels of induction

A 25 ml starter culture of *E. coli* XL1 Blue containing the appropriate plasmid was grown overnight at 37°C. This culture was diluted one in 10. The secondary culture was allowed to grow to an A_{600} of between 0.6 and 0.8. At this point induction of the production of His tagged protein was started using a final concentration of 1 mM IPTG. Samples (2 ml) were taken every hour from the initial dilution until the period induction had reached for 6 to 8 hours . The samples were centrifuged in a microcentrifuge at 16 000 x g, and the cell pellet was resuspended in 300 µl ice cold phosphate buffered saline (PBS; 150 mM NaCl, 16 mM NaH_2PO_4, 4 mM Na_2HPO_4, pH 7.5) for every 0.5 A_{600} unit. The protein samples were resolved using sodium dodecyl sulphate polyacrylamide gel electrophoresis.

6.4.2.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were resolved using a modification of the standard method (Laemmli, 1970). The BioRad minigel system was used for the running of all gels described. SDS-PAGE sample buffer (0.0625 M Tris, pH 6.8, 10 % glycerol, 2 % SDS, 5 % β-mercaptoethanol, 0.05 % bromophenol blue) was added to samples at a ratio of 1 parts sample to 4 parts buffer, after which samples were boiled at 95°C for 5 minutes. 12 % acrylamide separating gels (0.375 M Tris, pH 8.8, 0.1 % SDS, 12 % acrylamide) and 4 % stacking gels (0.125 M Tris, pH 6.8, 0. 1 % SDS, 4 % acrylamide) were prepared. Polymerisation of both separating and stacking gels was induced by the addition of 0.05 % ammonium persulphate (APS) and 0.005 % N,N,N'N' - tetramethylethylenediamine (TEMED). The SDS-PAGE running buffer was composed of 25 mM Tris, 192 mM glycine and 1 % SDS. Gels were resolved for one hour at 200 V. The staining of polyacrylamide gels was performed using Coomassie stain (40 % methanol, 0.7 % acetic acid, 0.075 % Coomassie dye) and the polyacrylamide gels were destained using Coomassie destain (40 % methanol, 0.7 % acetic acid).

6.4.2.3 Western Transfer of proteins

Transfer of proteins onto nitrocellulose was performed as described (Towbin *et al.*, 1979). Transfers were performed using the BioRad western apparatus using a methanol based transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol) prepared freshly and stored at 4°C. The transfer of protein from the polyacrylamide gel onto nitrocellulose was performed at 4°C for one hour at 100 V. Proteins transferred onto nitrocellulose were visualised using Ponceau S stain (0.5 % Ponceau S; 1 % glacial acetic acid). Destaining was performed using distilled water.

6.4.2.4 Chemiluminescent-based immunodetection of histidine-tagged proteins

Following transfer of proteins on to the nitrocellulose membrane, the membrane was washed twice with Tris buffered saline (TBS; 50 mM Tris, 150 mM NaCl; pH 7.5). The membrane was then left in blocking solution (5 % non fat milk powder in TBS) over night at 4°C. The membrane was incubated with a monoclonal anti-His antibody (Sigma, Germany) produced in a murine cell line, diluted 1 in 5000 in the blocking solution, for one hour at room

temperature. The membrane was thereafter washed in twice in Tris buffered saline-Tween 20 (TBST; 0.1 % Tween 20 in TBS) for 20 minutes each, and then twice with the blocking solution for 20 minute. The secondary antib ody (anti-mouse/anti-rabbit) was diluted to 1:5000 in the blocking solution, and the membrane was incubated with the antibody for half an hour. The membrane was thereafter washed four times with TBST for 15 minutes. The Roche chemiluminescence reagents (BM Chemiluminescence Western Blotting Kit Roche, Germany) were then prepared by adding the manufacturers Solution B to Solution A in a ratio of 1 to 100. This was left for 30 minutes before use. This reagent was added to the membrane, which was incubated for one minute. The membrane was then exposed to X-ray film for periods of time dependant on the strength of the chemiluminescence reaction. The film was developed according to the manufacturer's instructions. This was an incubation in developer for three minutes, passing through the stop solution, incubation in the fixer solution for 2 minutes, followed by a rinse with tap water before being left to dry. The developing and fixing solutions were purchased from Agfa, and the stop solution was 2 % acetic acid.

6.4.2.5 Large scale Purification of Histidine-tagged proteins

A 100 ml starter culture of *E. coli* XL1 Blue containing the plasmid of interest was grown overnight in YT broth at 37°C with shaking. This was diluted one in ten in YT broth and grown with shaking at 37°C until the A_{600} reached 0.8. Induction was performed using IPTG to a final concentration of 1 mM. The cells were left to grow for the optimum period of time as determined previously (Section 6.4.2.1). Cells were then centrifuged at 5 000 x g for 5 minutes at 4°C, and resuspended in lysis buffer (100 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF) in a final volume of 15 ml. The cells were frozen at -20°C overnight. They were thawed and lysed *via* the addition of ly sozyme (1 mg/ml final concentration). The reaction was left for 2 hours on ice. The mixture was thereafter centrifuged in a microcentrifuge at maximum speed (16 000 x g) for 30 minutes at 4°C. The supernatant was added to 1 ml of charged nickel Sepharose (Section 6.4.2.6). The lysate was left to bind to the column overnight. The flow through was collected, and the column was washed twice with 10 column volumes wash buffer I (100 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM PMSF) and then a further two times with 10 column volumes wash buffer II (100 mM Tris, pH 8.0, 300 mM NaCl, 80 mM imidazole, 1 mM PM SF). The sample

was eluted with 1 column volumes of elution buffer (100 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole, 1 mM PMSF) repeated seven times.

6.4.2.6 Preparation of the nickel Sepharose column

Sufficient Sepharose was removed for use in the purification. The gel was sedimented in a 20 ml glass column by gravity flow. The flow through was then stopped. Five gel volumes of distilled water was added, the gel was resuspended, and the water was allowed to flow through by gravity flow. Half a gel volume of 0.1 M $NiSO_4$ was added, the gel was resuspended, and the $NiSO_4$ was removed *via* gravity flow. The gel was washed three time s using five volumes distilled water. The gel was then resuspended in one volume of the lysi s buffer. After the column was used, it was stripped of nickel and recharged. Stripping was accomplished using 10 column volumes of the stripping buffer (0.05 M EDTA, 0.5 M NaCl) after which the EDTA was removed by washing the column with 3 column volumes 0.5 M NaCl. Ionically bound proteins were removed by washing the column with 0.5 column volumes 2 M NaCl for 15 minutes. Precipitated proteins, hydrophobically bound proteins and lipoproteins were removed by washing the column with 1 M NaOH for an hour. The column was washed with three column volumes of the lysis buffer and recharged as described above.

6.4.2.7 Bradford assay for determination of protein concentrations

The Bradford method was used for determination of protein concentration (Bradford, 1976). The Bradford reagent was made as follows. 100 mg Coomassie G250 was dissolved in 50 ml 95 % ethanol. 100 ml 85 % (w/v) phosphoric acid was then added, and the solution was made up to 1 litre. It was filtered and stored at 4°C in a dark bottle. The assay was performed in microtitre plates. Bradford reagent (200 µl) was added to the protein sample (10 µl). This was left for five minutes and the absorbance was read at 595 nm in a PowerWave microtitre plate reader (Biotek). A standard curve comprising 0 to 250 µg/ml bovine serum albumin (BSA) was used. Standards (10 µl) had the Bradford reagent added to them (200 µl), were left for five minutes, and again the absorbance was read at 595 nm in a microtitre plate reader. A sample standard curve is given in Figure 6.4.

6.5 STANDARD CURVES

6.5.1 Bradford standard curve

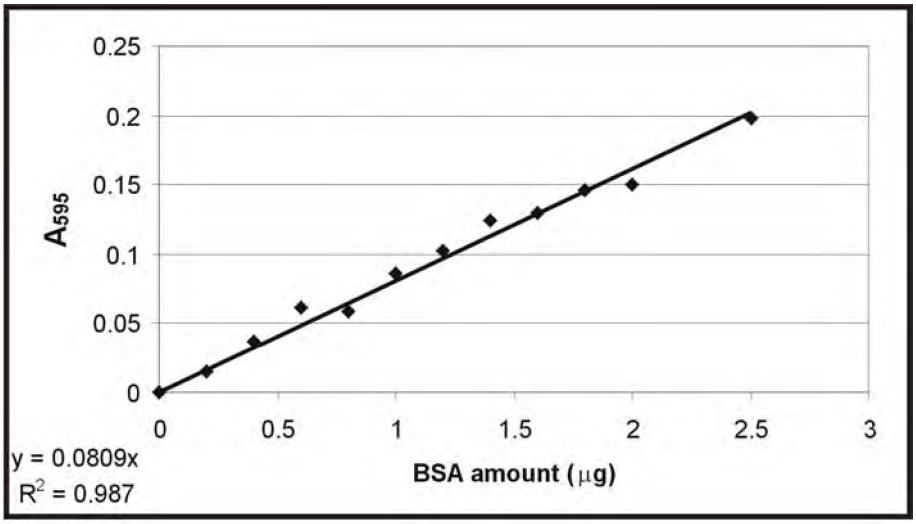

Figure 6.4: Sample standard curve for the Bradford protein dete rmination assay

The standard protein used was BSA.

6.5.2 Phosphate standard curve

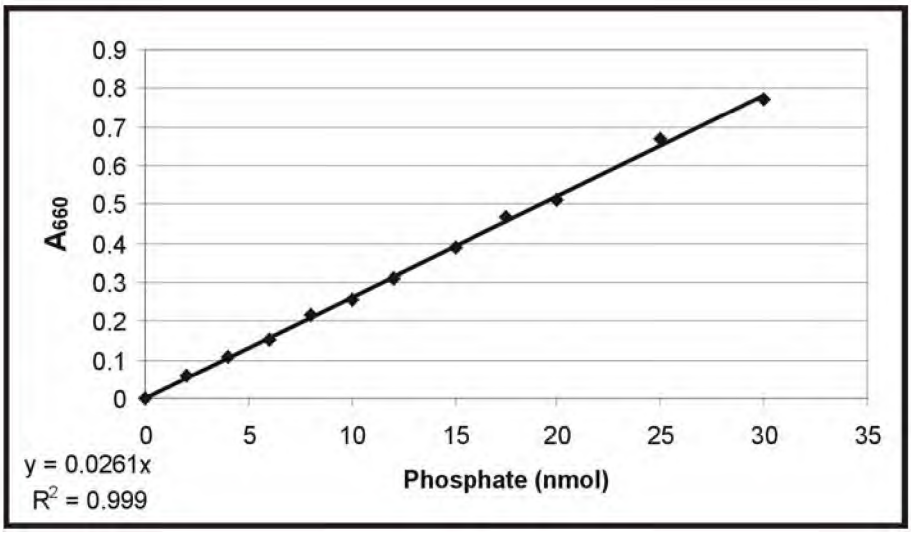

Figure 6.5: Sample standard curve for determination of phosphate release during ATPase assays

References

7 REFERENCES

Agarraberes, F.A., Dice, J.F. (2001). A molecular chaperone complex at the lysosomal membrane is required for protein translocation. Journal of Cell Science *114,* 2491-2499.

Al-Herran, S., Ashraf, W. (1998). Physiological consequences of the over-production of *E.coli* truncated molecular chaperone DnaJ. FEMS Microbiology Letters *162,* 117-122.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology *215,* 403-410.

Andrade, M.A., Sander, C. (1997). Bioinformatics: from genome data to biological knowledge. Current Opinion in Biotechnology *8,* 675-683.

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

Anfinsen, C.B., Haber, E., Sela, M. and White, F.H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proceedings of the National Academy of Sciences, USA *47,* 1309-1314.

Auger, I., Roudier, J. (1997). A function for the QKRAA amino acid motif: mediating binding of DnaJ to DnaK. Journal of Clinical Investigation *99,* 1818-1822.

Banecki, B., Liberek, K., Wall, D., Wawrzynów, A., Georgopoulos, C., Bertoli, E., Tanfani, F. and Zylicz, M. (1996). Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. Journal of Biological Chemistry *271,* 14840-14848.

Banecki, B., Zylicz, M. (1996). Real time kinetics of the DnaK/DnaJ/GrpE molecular chaperone machine action. Journal of Biological Chemistry *271,* 6137-6143.

Bardwell, J.C.A., Craig, E.A. (1984). Major heat shock gene of *Drosophila* and the *Escherichia coli* heat inducible *dnaK* gene are homologous. Proceedings of the National Academy of Sciences, USA *81,* 848-852.

Bardwell, J.C.A., Tilly, K., Craig, E., King, J., Zylicz, M. and Georgopoulos, C. (1986). The nucleotide sequence of the *Escherichia coli* K12 *dnaJ*+ gene. Journal of Biological Chemistry *261,* 1782-1785.

Baxevanis, A.D. (2001). The Molecular Biology Database Collection: an updated compilation of biological database resources. Nucleic Acids Research *29,* 1-10.

Baxevanis, A.D. (2002). The Molecular Biology Database Collection: 2002 update. Nucleic Acids Research *30,* 1-12.

Baxevanis, A.D. (2003). The Molecular Biology Database Collection: 2003 update. Nucleic Acids Research *31,* 1-12.

Becker, J., Craig, E.A. (1994). Heat shock proteins as molecular chaperones. European Journal of Biochemistry *219,* 11-23.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F.F., Rapp, B.A. and Wheeler, D.L. (1999). GenBank. Nucleic Acids Research *26,* 1-7.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2003). GenBank. Nucleic Acids Research *31,* 23-27.

Berg, J.M. (1990). Zinc fingers and other metal binding domains. Journal of Biological Chemistry *265,* 6513-6516.

Berjanskii, M.V., Riley, M.I. and Van Doren, S.R. (2002). Hsc70-interacting HPD loop of the J domain of polyomavirus T antigens fluctuates in ps to ns and μs to ms. Journal of Molecular Biology *321*, 503-516.

Berjanskii, M.V., Riley, M.I., Xie, A., Semenchenko, V., Folk, W.R. and Van Doren, S.R. (2000). NMR structure of the N-terminal J domain of murine polyomavirus T antigen s. Journal of Biological Chemistry *275*, 36094-36103.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Research *28*, 235-242.

Bessoule, J.-J., Testet, E. and Cassagne, C. (1994). Cloning of a new isoform of a DnaJ protein from *Allium porrum* epidermal cells. Plant Physiology and Biochemistry *32*, 723-727.

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

Bischofberger, P., Han, W., Feifel, B., Schönfeld, H.-J. and Christen, P. (2003). D-Peptides as inhibitors of the DnaK/DnaJ/GrpE chaperone system. Journal of Biological Chemistry *278*, 19044-19047.

Blond-Elguindi, S., Cwirla, S.E., Dower, W.J., Lipshutz, R.F., Sprang, S.R., Sambrook, J.F. and Gething, M.-J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. Cell *75*, 717-728.

Blumberg, H., Silver, P.A. (199 1). A homologue of the bacterial heat-shock gene *DnaJ* that alters protein sorting in yeast. Nature *349*, 627-630.

Boguski, M.S. (1998). Bioinformatics - A New Era. Trends Guide to Bioinformatics 1-3.

Bolliger, L., Deloche, O., Glick, B.S., Georgopoulos, C., Jenö, P., Kronidou, N., Horst, M., Morishima, N. and Schatz, G. (1994). A mitochondrial homologue of bacterial GrpE interacts with mitochondrial hsp70 and is essential for viability. EMBO Journal *13*, 1998-2006.

Bork, P., Sander, C. and Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proceedings of the N ational Academy of Sciences, USA *89*, 7290-7294.

Bork, P., Sander, C., Valencia, A. and Bukau, B. (1991). A module of the DnaJ heat shock proteins found in malaria paras ites. Trends in Biochemical Sciences *17*, 129.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the prinicple of protein-dye binding. Analytical Biochemistry *72*, 248-254.

Braun, J.E.A., Wilbanks, S.M. and Scheller, R.H. (1996). The cysteine string secretory vesicle protein activates Hsc70 ATPase. Journal of Biological Chemistry *271*, 2 5989-25993.

Brehmer, D., Rüdiger, S., Gässler, C.S., Klostermeier, D., Packschies, L., Reinstein, J., Mayer, M.P. and Bukau, B. (2001). Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. Nature Structural Biology *8*, 427-432.

Brodsky, J.L. (1996). Post-translational translocation: not all hsc70s are created equal. Trends in Biochemical Sciences *21*, 122-126.

Brodsky, J.L., Pipas, J.M. (1998). Polyomavirus T antigens: molecular chaperones for multiprotein complexes. Journal of Virology *72*, 5329-5334.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S.,

Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science *273*, 1058-1073.

Burton, S., Dorrington, R., Hartley, C., Kirchmann, S., Matcher, G. and Phehane, V. (1998). Production of enantiomerically pure amino acids: characterisation of South African hydantoinases and hydantoinase-producing bacteria. Journal of Molecular Catalysis B: Enzymatic *5*, 301-305.

Campbell, K.S., Mullane, K.P., Aksoy, I.A ., Stubdal, H., Zalvide, J., Pipas, J.M., Silver, P.A., Roberts, T.M., Schaffhausen, B.S. and DeCaprio, J.A. (1997). DnaJ/hsp40 chaperone domain of SV40 large T antigen promotes efficient viral DNA replication. Genes and Development *11*, 1098-1110.

Caplan, A.J., Cyr, D.M. and Douglas, M.G. (1992a). YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism. Cell *71*, 1143-1155.

Caplan, A.J., Douglas, M.G. (1991). Characterisation of YDJ1: a yeast homologue of the bacterial dnaJ protein. Journal of Cell Biology *114*, 609-621.

Caplan, A.J., Tsai, J., Casey , P.J. and Douglas, M.G. (1992b). Farnesylation of YDJ1p is required for function at elevated growth temperatures in *Saccharomyces cerevisiae*. Journal of Biological Chemistry *267*, 18890-18895.

Cappai, R., Kaslow, D.C., Peterson, M.G., Cowman, A.F., Anders, R.F. and Kemp, D.J. (1992). Cloning and analysis of the RESA-2 gene: a DNA homologue of the ring-infected erythrocyte surface antigen of *Plasmodium falciparum*. Molecular and Biochemical Parasitology *54*, 213-222.

Chai, Y., Koppenhafer, S.L., Boninim, N.M. and Paulson, H.L. (1999). Analysis of the role of heat shock protein (hsp) molecular chaperones in polyglutamine disease. Journal of Neuroscience *19*, 10338-10347.

Chamberlain, L.H., Burgoyne, R.D. (1997a). Activation of the ATPase activity of heat shock proteins Hsc70 / Hsp70 by cysteine-string protein. Biochemistry Journal *322*, 853-858.

Chamberlain, L.H., Burgoyne, R.D. (1997b). The molecular chaperone function of the secretory vesicle cysteine string protein. Journal of Biological Chemistry *272*, 31420-31426.

Chappell, T.G., Konforti, B.B., Schmid, S.L. and Rothman, J.E. (1987). The ATPase core of a clathrin uncoating protein. Journal of Biological Chemistry *262*, 746-751.

Chapple, J.P., Cheetham, M.E. (2003). The chaperone environment at the cytoplasmic face of the endoplasmic reticulum can modulate rhodopsin processing and inclusion formation. Journal of Biological Chemistry *278*, 19087-19094.

Cheetham, M.E., Brion, J.-P. and Anderton, B.H. (1992). Human homologues of the bacterial heat-shock protein DnaJ are preferentially expressed in neurons. Biochemistry Journal *284*, 469-476.

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

Chevalier, M., King, L., Wang, C., Gething, M.-J., Elguindi, E. and Blond, S. (1998). Substrate binding induces depolymerisation of the C-terminal peptide binding domain of murine Grp78/BiP. Journal of Biological Chemistry *273*, 26827-26835.

Chevalier, M., Rhee, H., Elguindi, E.C. and Blond, S.Y. (2000). Interaction of murine BiP/Grp78 with the DnaJ homologue MTJ1. Journal of Biological Chemistry *275*, 19620-19627.

Choglay, A.A., Chapple, J.P., Blatch, G.L. and Cheetham, M.E. (2001). Identification and characterisation of a human mitochondrial homologue of the bacterial co-chaperone GrpE. Gene *267*, 125-134.

Chung, K.T., Shen, Y. and Hendershot, L.M. (2002). BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. Journal of Biological Chemistry *277*, 47557-47563.

Clarke, D.J., Holland, I.B. and Jacq, A. (1996). A novel DnaJ-like protein in *Escherichia coli* inserts into the cytoplasmic membrane with a type III topology. Molecular Microbiology *20*, 1273-1286.

Clarke, D.J., Holland, I.B. and Jacq, A. (1997). Point mutations in the transmembrane domain of DjlA, a membrane-linked DnaJ-like protein, abolish its function in promoting colanic acid production via the Rcs signal transduction pathway. Molecular Microbiology *25*, 933-944.

Clarke, S. (1992). Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. Annual Review of Biochemistry *61*, 355-386.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537-544.

Conway De Macario, E., Macario, A.J.L. (1994). Heat shock response in archaea. Trends in Biotechnology *12*, 512-518.

Corsi, A.K., Schekman, R. (1997). The lumenal domain of Sec63p stimulates the ATPase activity of BiP and mediates BiP recruitment to the translocon in *Saccharomyces cerevisiae.* Journal of Cell Biology *137*, 1483-1493.

Cunnea, P.M., Miranda-Vizuete, A., Bertoli, G., Simmen, T., Damdimopoulos, A.E., Hermann, S., Leinonen, S., Huikko, M.P., Gustafsson, J.-Å., Sitia, R. and Spyrou, G. (2003). ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress. Journal of Biological Chemistry *278*, 1059-1066.

Cupp-Vickery, J.R., Vickery, L.E. (1997). Crystallisation and preliminary X-ray crystallographic properties of Hsc20, a J-motif co-chaperone protein from *Escherichia coli.* Protein Science *6*, 2028-2030.

Cupp-Vickery, J.R., Vickery, L.E. (2000). Crystal structure of Hsc20, a J-type co-chaperone from *Escherichia coli.* Journal of Molecular Biolo gy *304*, 835-845.

Cyr, D.M., Langer, T. and Douglas, M.G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends in Biochemical Sciences *19*, 176-181.

Dagert, M., Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene *6*, 23-28.

Davis, J.E., Voisine, C. and Craig, E.A. (1999). Intragenic suppressors of Hsp70 mutants: interplay between the ATPase- and peptide-binding domains. Proceedings of the National Academy of Sciences, USA *96*, 9269-9276.

De Crouy-Chanel, A., Kohiyama, M. and Richarme, G. (1995). A novel function of *Escherichia coli* chaperone DnaJ. Journal of Biological Chemistry *270*, 22669-22672.

De Crouy-Chanel, A., Kohiyama, M. and Richarme, G. (1996). Specificity of DnaK for arginine/lysine and effect of DnaJ on the amino acid specificity of DnaK. Journal of Biological Chemistry *271*, 15486-15490.

De Crouy-Chanel, A., Kohiyama, M. and Richarme, G. (1999). Interaction of DnaK with native proteins and membrane proteins correlates with their accessible hydrophobicity. Gene *230*, 163-170.

Deloche, O., Kelley, W.L. and Georgopoulos, C. (1997). Structure-function analyses of Ssc1p, Mdj1p and Mge1p *Saccharomyces cerevisiae* mitochondrial proteins in *Escherichia coli.* Journal of Bacteriology *179*, 6066-6075.

Demand, J., Lüders, J. and Höhfeld, J. (1998). The carboxy-terminal domain of Hsc70 provides binding sited for a distinct set of chaperone co-factors. Molecular and Cellular Biology *18*, 2023-2028.

Depre, C., Wang, L., Tomlinsin, J.E., Gaussin, V., Abdellatif, M., Topper, J.N. and Vatner, S.F. (2003). Characterisation of pDJA1, a cardiac-specific chaperone found by genomic profiling of the post-ischemic swine heart. Cardiovascular Research *58*, 126-135.

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

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

Ellis, R.J. (1990). Molecular Chaperones: the plant connection. Science *250*, 954-959.

Ellis, R.J., van der Vies, S.M. (1991). Molecular Chaperones. Annual Review of Biochemistry *60*, 321-347.

Federov, A.N., Baldwin, T.O. (1997). Cotransl ational protein folding. Journal of Biological Chemistry *272*, 32715-32718.

Feifel, B., Schönfeld, H.-J. and Christen, P. (1998). D-peptide ligands for the co-chaperone DnaJ. Journal of Biological Chemistry *273*, 11999-12002.

Feige, U., Mollenhauer, J. (1992). Heat shock proteins. Experientia *48*, 621-622.

Feldheim, D., Rothblatt, J. and Schekman, R. (1992). Topology and functional domains of Sec63p, an endoplasmic reticulum membrane protein required for secretory protein translocation. Molecular and Cellular Biology *12*, 3288-3296.

Fewell, S.W., Pipas, J.M. and Brodsky, J.L. (2002). Mutagenesis of a functional chimeric gene in yeast identifies mutations in the simian virus 40 large T antigen J domain. Proceedings of the National Academy of Sciences, USA *99*, 2002-2007.

Fink, A.L. (1999). Chaperone-mediated protein folding. Physiological reviews *79*, 425-449.

Flaherty, K.M., McKay, D.B., Kabsch, W. and Holmes, K.C. (1991). S imilarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. Proceedings of the National Academy of Sciences, USA *88*, 5041-5045.

Flaherty, K.M., Wilbanks, S.M., De Luca-Flaherty, C. and McKay, D.B. (1990). Three-dimensional structure of the ATPase fragment of a 70 kDa heat shock cognate protein. Nature *346*, 623-628.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott' J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrman, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496-512.

Flynn, G.C., Chappell, T.G. and Rothman, J.E. (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. Science *245*, 385-390.

Flynn, G.C., Pohl, J., Flocco, M .T. and Rothman, J.E. (1991). Peptide-binding specificity of the molecular chaperone BiP. Nature *353*, 726-730.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage , A.R., Quackenbush, J., Salzberg, S., Hanson, M., Van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature *390*, 580-586.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Furhmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson, C.A.I. and Venter, J.C. (1995). The minimal gene complement of *Mycoplasma genitalium.* Science *270*, 397-403.

Freeman, B.C., Myers, M.P., Schumacher, R. and Morimoto, R.I. (1995). Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. EMBO Journal *14*, 2281-2292.

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

Frydman, J., Nimmesgern, E., Ohtsuka, K. and Hartl, F.-U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. Nature *370*, 111-117.

Gamer, J., Multhaup, G., Tomoyasu, T., McCarty, J.S., Rüdiger, S., Schönfeld, J., Schirra, C., Bujard, H. and Bukau, B. (1996). A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor σ^{32}. EMBO Journal *15*, 607-617.

Gässler, C.S., Buchberger, A., Laufer, T., Mayer, M.P., Schröder, H., Valencia, A. and Bukau, B. (1998). Mutations in the DnaK chaperone affecting interaction with the DnaJ cochaperone. Proceedings of the National Academy of Sciences, USA *95*, 15229-15234.

Genevaux, P., Lang, F., Schwager, F., Vartikar, J.V., Rundell, K., Pipas, J.M., Georgopoulos, C. and Kelley, W.L. (2003). Simian virus 40 T antigens and J domains: analysis of Hsp40 cochaperone functions in *Escherichia coli.* Journal of Virology *77*, 10706-10713.

Genevaux, P., Schwager, F., Georgopoulos, C. and Kelley, W.L. (2001a). The *djlA* gene acts synegistically with *dnaJ* in promoting *Escherichia coli* growth. Journal of Bacteriology *183*, 5747-5750.

Genevaux, P., Schwager, F., Georgopoulos, C. and Kelley, W.L. (2002). Scanning mutagenesis identifies amino acid residues essential for the *in vivo* activity of the *Escherichia coli* DnaJ (Hsp40) J domain. Genetics *162*, 1045-1053.

Genevaux, P., Wawrzynów, A., Zylicz, M., Georgopoulos, C. and Kelley, W.L. (2001b). DjlA is a third DnaK co-chaperone of *Escherichia coli*, and DjlA-mediated induction of colanic acid capsule requires DjlA-DnaK interaction. Journal of Biological Chemistry *276*, 7906-7912.

Georgopoulos, C., Welch, W.J. (1993). Role of the major heat shock proteins as molecular chaperones. Annual Review of Cell Biology *9*, 601-634.

Gisler, S.M., Pierpaoli, E.V. and Christen, P. (1998). Catapult mechanism renders the chaperone action of Hsp7 0 unidirectional. Journal of Molecular Biology *279*, 833-840.

Goffin, L., Georgopoulos, C. (1998). Genetic and biochemical characterisation of mutations affecting the carboxy-terminal domain of the *Escherichia coli* molecular chaperone DnaJ. Molecular Microbiology *30*, 329-340.

Goldberger, R.F., Epstein, C.J. and Anfinsen, C.B. (1963). Acceleration of Reactivation of Reduced Bovine Pancreatic Ribonuclease by a Microsomal System from Rat Liver. Journal of Biological Chemistry *238*, 628-635.

Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C. and Slater, S. (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science *294*, 2323-2328.

Gragerov, A., Zeng, L., Zhao, X., Burkholder, W. and Gottesman, M.E. (1994). Specificity of DnaK-peptide binding. Journal of Molecular Biology *235*, 848-854.

Greene, M., Makos, K. and Landry, S.J. (1998). Role of the J-domain in the cooperation of Hsp40 with Hsp70. Proceedings of the National Academy of Sciences, USA *95*, 6108-6113.

Greene, M.K., Steede, N.K. and Landry, S.J. (2000). Domain-specific spectroscopy of 5-hydroxytryptophan-containing variants of *Escherichia coli* DnaJ. Biochimica et Biophysica Acta *1480*, 267-277.

Gross, J.D. (1957). Incorporation of Phosphorus-32 into salivary-type chromosomes which exhibit 'puffs'. Nature *180*, 440.

Guex, N., Peitsch, M. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Haas, IG., Wabl, M. (1983). Immunoglobulin heavy chain binding protein. Nature *306*, 387-389.

Haber, E., Anfinsen, C.B. (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. Journal of Biological Chemistry *237*, 1839-1844.

Han, W., Christen, P. (2001). Mutations in the interdomain linker region of DnaK abolish the chaperone action of the DnaK/DnaJ/GrpE system. FEBS Letters *497*, 55-58.

Han, W., Christen, P. (2003). Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system. Journal of Biological Chemistry *278*, 19038-19043.

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

Hartley, C., Kirchmann, S., Burton, S. and Dorrington, R. (1998). Production of D-amino acids from D,L-substituted hydantoins by an *Agrobacterium tumefaciens* strain and isolation of a mutant with inducer-independant expression of hydantoin hydrolysing activity. Biotechnology Letters *20*, 707-711.

Hartley, C., Manford, F., Burton, S. and Dorrington, R. (2001). Over-production of hydantoinase and N-carbomoylamino acids amidohydrolase enzymes by regulatory mutants of *Agrobacterium tumefaciens*. Applied Microbiology and Biotechnology *57*, 43-49.

Hennessy, F., Cheetham, M.E., Dirr, H.W. and Blatch, G.L. (2000). Analysis of the levels of conservation of the J domain among the various types of DnaJ-like proteins. Cell Stress and Chaperones *5*, 347-358.

Hodgkin, J., Plasterk, R.H.A. and Waterston, R.H. (1995). The nematode *Caenorhabditis elegans* and its genome. Science *27*, 410-414.

Hoff, K.G., Silberg, J.J. and Vickery, L.E. (2000). Interaction of the iron-sulfur cluster assembly protein IscU with Hsc66/Hsc20 molecular chaperone system of *Escherichia coli*. Proceedings of the National Academy of Sciences, USA *97*, 7790-7795.

Höhfeld, J. (1998). Regulation of the heat shock cognate Hsc70 in the mammalian cell: the characterisation o f the anti-apoptotic protein BAG-1 provides novel insights. Biological Chemistry *379*, 269-274.

Höhfeld, J., Cyr, D.M. and Patterson, C. (2001). From the cradle to the grave: molecular chaperones that may choose between folding and degradation. EMBO Reports *2*, 885-890.

Holstein, S.E.H., Ungewickell, H. and Ungewickell, E. (1996). Mechanism of clathrin basket dissociation: separate functions of protein domains of the DnaJ homologue Auxilin. Journal of Cell Biology *135*, 925-937.

Hosoda, A., Kimata, Y. and Kohno, K. (2003). JPD1, a novel endoplasmic reticulum-resident protein containing both a BiP-interacting J-domain and thioredoxin-like motifs. Journal of Biological Chemistry *278*, 2669-2676.

Huang, K., Flanagan, J.M. and Prestegard, J.H. (1998). The influence of C-terminal extension on the structure of the J domain in *E. coli* DnaJ. Protein Science *8*, 203-214.

Huang, K., Ghose, R., Flanagan, J.M. and Prestegard, J.H. (1999). Backbone dynamics of the N-terminal domain in *E. coli* DnaJ determined by ^{15}N- and ^{13}CO-relaxation measurements. Biochemistry *38*, 10567-10577.

Ingolia, T., Craig, E.A. (1982). *Drosophila* gene related to the major heat shock-induced gene is transcribed at normal temperatures and not induced by heat shock. Proceedings of the National Academy of Sciences, USA *79*, 525-529.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Iwahara, J., Clubb, R.T. (1999). Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). EMBO Journal *18*, 6084-6094.

Iwahara, J., Iwahara, M., Daughdrill, G.W., Ford, J. and Clubb, R.T. (2002). The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. EMBO Journal *21*, 1197-1209.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. Trends in Biochemical Sciences *23*, 403-405.

Jiang, J., Taylor, A.B., Prasad, K., Ishikawa-Brush, Y., Hart, P.J., Lafer, E.M. and Sousa, R. (2003). Structure-function analysis of the auxilin J-domain reveals an extended Hsc70 interaction interface. Biochemistry *42*, 5748-5753.

Jiang, R.-F., Greener, T., Baraouch, W., Greene, L. and Eisenberg, E. (1997). Interaction o f auxilin with the molecular chaperone, Hsc70. Journal of Biological Chemistry *272*, 6141-6145.

Johnson, J.L., Craig, E.A. (2000). A role for the Hsp40 YDJ1 in repression of basal steroid receptor activity in yeast. Molecular and Cellular B iology *20*, 3027-3036.

Johnson, J.L., Craig, E.A. (2001). An essential role for the substrate-binding region of Hsp40s in *Saccharomyces cerevisiae.* Journal of Cell Biology *152*, 851-856.

Jordan, R., McMacken, R. (1995). Modulation of the ATPase activity of the molecular chaperone DnaK by peptides and the DnaJ and GrpE heat shock proteins. Journal of Biological Chemistry *270*, 4563-4569.

Jubete, Y., Maurizi, M.R. and Gottesman, S. (1996). Role of the heat shock protein DnaJ in the Lon-dependant degradation of naturally unstable proteins. Journal of Biological Chemistry *271*, 30798-30803.

Kabani, M., Beckerich, J.-M. and Brodsky, J.L. (2002). Nucleotide exchange factor for the yeast Hsp70 molecular chaperone Ssa1p. Molecular and Cellular Biology *22*, 4677-4689.

Kabani, M., Beckerich, J.-M. and Brodsky, J.L. (2003). The yeast Sls1p and Fes1p proteins define a new family of Hsp70 nucleotide exchange factors. Current Genomics *4*, 273.

Kabani, M., Beckerich, J.-M. and Gaillardin, C. (2000). Sls1p stimulates Sec63p-mediated activation of Kar2p in a conformation-dependant manner in the yeast endoplasmic reticulum. Molecular and Cellular Biology *20*, 6923-6934.

Kamath-Loeb, A.S., Lu, C.Z., Suh, W.-C., Lonetto, M.A. and Gross, C.A. (1995). Analysis of three DnaK mutant proteins suggests that progression through the ATPase cycle requires conformational changes. Journal of Biological Chemistry *270*, 30051-30059.

Karzai, A.W., McMacken, R. (1996). A bipartite signalling mechanism involved in DnaJ-mediated activation of the *Escherichia coli* DnaK protein. Journal of Biol ogical Chemistry *271*, 11236-11246.

Kelley, W.L. (1998). The J-domain family and the recruitment of chaperone power. Trends in Biochemical Sciences *23*, 222-227.

Kelley, W.L., Georgopoulos, C. (1997a). Positive control of the two-component RcsC/B signal transduction network by DjlA: a member of the DnaJ family of molecular chaperones in *Escherichia coli*. Molecular Microbiology *25*, 913-931.

Kelley, W.L., Georgopoulos, C. (1997b). The T/t common exon of simian virus 40, JC, and BK polyomavirus T antigens can functionally replace the J domain of the *Escherichia coli* DnaJ molecular chaperone. Proceedings of the National Academy of Sciences, USA *94*, 3679-3684.

Kelley, W.L., Landry, S.J. (1994). Chaperone power in a virus? Trends in Biochemical Sciences *19*, 277-278.

Kim, H.-Y., Ahn, B.-Y. and Cho, Y. (2001). Structural basis for the inactivation of retinoblastoma tumor supressor by SV40 large T antigen. EMBO Journal *20*, 295-304.

Klapper, M.H. (1977). The independent distribution of amino acid near neighbor pairs into polypeptides. Biochemical and Biophysical Research Communications *78*, 1018-1024.

Kluck, C.J., Patzelt, H., Genevaux, P., Brehmer, D., Rist, W., Schneider-Mergener, J., Bukau, B. and Mayer, M.P. (2002). Structure-function analysis of HscC, the *Escherichia coli* member of a novel subfamily of specialised Hsp70 chaperones. Journal of Biological Chemistry *277*, 41060-41069.

Kraulis, P. (1991). Molscript: a programme to produce both detailed and schematic plots of protein structures. Journal of Applied Crystallography *24*, 946-950.

Kroczynska, B., Blond, S.Y. (2001). Cloning and characterisation of a new soluble murine J-domain protein that stimulates BiP, Hsc70 and DnaK ATPase activity with differrent efficiencies. Gene *273*, 267-274.

Kudlicki, W., Odom, O.W., Kramer, G. and Hardesty, B. (1996). Binding of a N-terminal Rhodanese peptide to DnaJ and to ribosomes. Journal of Biological Chemistry *271*, 31160-31165.

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

Laity, J.H., Lee, B.M. and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Current Opinion in Structural Biology *11*, 39-46.

Landry, S.J. (2003). Structure and energetics of an allele-specific interaction between *dnaJ* and *dnaK*: correlation of nuclear magnetic resonance chemical shift perturbations in the J-domain of Hsp40/DnaJ with binding affinity for the ATPase domain of Hsp70/DnaK. Biochemistry *42*, 4926-4936.

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

Lanzetta, P.A., Alvarez, L.J., Reinach, P.S. and Candia, O.A. (1979). An improved assay for nanomole amounts of inorganic phosphate. Analytical Biochemistry *100*, 95-97.

Laskey, R.A., Honda, B.M., Mills, A.D. and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature *275*, 416-420.

Laufen, T., Mayer, M.P., Beisle, C., Klostermeier, D., Mogk, A., Reinstein, J. and Bukau, B. (1999). Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. Proceedings of the National Academy of Sciences, USA *96*, 5452-5457.

Lelivelt, M.J., Kawula, T.H. (1995). Hsc66, an Hsp homologue in *Escherichia coli*, is induced by cold shock and not by heat shock. Journal of Bacteriology *177*, 4900-4907.

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

Li, H., Söderbärg, K., Houshmand, H., You, Z.-Y. and Magnusson, G. (2001). Effect on polyomavirus T-antigen function of mutations in a conserved leucine-rich segment of the DnaJ domain. Journal of Virology *75*, 2253-2261.

Li, J., Qian, X. and Sha, B. (2003). The crystal structure of the yeast hsp40 Ydj1 complexed with its peptide substrate. Structure *11*, 1475-1483.

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C. and Zylicz, M. (1991). *Escherichia coli* DnaJ and GrpE heat proteins jointly stimulate ATPase activity of DnaK. Proceedings of the National Academy of Sciences, USA *88*, 2874-2878.

Linke, K., Wolfram, T., Bussemer, J. and Jakob, U. (2003). The roles of the two zinc binding sites in DnaJ. Journal of Biological Chemistry *278*, 44457-44466.

Lu, Z., Cyr, D.M. (1998). The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone YDJ1 assist Hsp70 in protein folding. Journal of Biological Chemistry *273*, 5970-5978.

Lyman, S.K., Schekman, R. (1995). Interaction between BiP and Sec63p is required for the completion of protein translocation into the ER of *Saccharomyces cerevisiae.* Journal of Cell Biology *131*, 1163-1171.

Ma, Y., Greener, T., Pacold, M.E., Kaushal, S., Greene, L.E. and Eisenberg, E. (2002). Identification of domain required for catalytic activity of auxilin in supporting clathrin uncoating by Hsc70. Journal of Biological Chemistry *277*, 49267-49274.

Macario, A.J.L., Lange, M., Ahring, B.K. and Conway De Macario, E. (1999). Stress genes and proteins in the Archaea. Microbiology and Molecular Biology Reviews *63*, 923-967.

Marshall, C.J. (1993). Protein Prenylation: a mediator of protein-protein interactions. Science *259*, 1865-1866.

Martinez-Yamout, M., Legge, G.B., Zhang, O., Wright, P.E. and Dyson, H.J. (2000). Solution structure of the cysteine-rich domain of the *Escherichia coli* chaperone protein DnaJ. Journal of Molecular Biology *300*, 805-818.

Mayer, M., Reinstein, J. and Buchner, J. (2003). Modulation of the ATPase cycle of BiP by peptides and proteins. Journal of Molecular Biology *330*, 137-144.

Mayer, M.P., Laufen, T., Paal, K., McCarty, J.S. and Bukau, B. (1999). Investigation of the interaction between DnaK and DnaJ by surface plasmon resonance spectroscopy. Journal of Molecular Biology *289*, 1131-1144.

McCarty, J.S., Buchberger, A., Reinstein, J. and Bukau, B. (1995). The role of ATP in the functional cycle of the DnaK chaperone system. Journal of Molecular Biology *249*, 126-137.

Michels, A.M., Kanon, B., Bensaude, O. and Kampinga, H.H. (1999). Heat shock proteins (Hsp) 40 mutants inhibit Hsc70 in mammalian cells. Journal of Biological Chemistry *274*, 36757-36763.

Miernyk, J.A. (2001). The J domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. Cell Stress and Chaperones *6*, 209-218.

Minami, Y., Höhfeld, J., Ohtsuka, K. and Hartl, F.-U. (1996). Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homologue Hsp40. Journal of Biological Chemistry *271*, 19617-19624.

Misselwitz, B., Staeck, O. and Rapoport, T.A. (1998). J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences. Molecular Cell *2*, 593-603.

Montgomery, D.L., Morimoto, R.I. and Gierasch, L.M. (1999). Mutations in the substrate binding domain of the *Escherichia coli* 70 kDa molecular chaperone, DnaK, which alter substrate affinity or interdomain coupling. Journal of Molecular Biology *286*, 915-932.

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

Moro, F., Fernández, V. and Muga, A. (2003). Interdomain interaction through helices A and B of DnaK peptide binding helices. FEBS Letters *533*, 119-123.

Morshauser, R.C., Hu, W., Wang, H., Pang, Y., Flynn, G.C. and Zuiderweg, E.R. (1999). High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. Journal of Molecular Biology *289*, 1387-1403.

Motohashi, K., Taguchi, H., Ishii, N. and Yoshida, M. (1994). Isolation of the stable hexameric DnaK-DnaJ complex from *Thermus thermophilus.* Journal of Biological Chemistry *269*, 27074-27079.

Motohashi, K., Yohda, M., Endo, I. and Yoshida, M. (1996). A novel factor required for the assembly of the DnaK and DnaJ chaperones of *Thermus thermophilus*. Journal of Biological Chemistry *271*, 17343-17348.

Nagata, H., Hansen, W.J., Freeman, B. and Welch, W.J. (1998). Mammalian cytosolic DnaJ homologues affect the hsp70 chaperone-substrate reaction cycle, but do not interact directly with nascent or newly synthesised proteins. Biochemistry *37*, 6924-6938.

Naylor, D.J., Stines, A.P., Hoogenraad, N.J. and Høj, P.B. (1998). Evidence for the existence of distinct mammalian cytosolic, microsomal, and two mitochondrial GrpE-like proteins, the co-chaperones of specific Hsp70 members. Journal of Biological Chemistry *273*, 21169-21177.

Ohki, M., Tamura, F., Nishimura, S. and Uchida, H. (1986). Nucleotide sequence of the *Escherichia coli dnaJ* gene and purification of the gene product. Journal of Biological Chemistry *261*, 1778-1781.

Ohtsuka, K., Hata, M. (2000). Mammalian Hsp40/DnaJ homologs: cloning of novel cDNAs and a proposal for their classification and nomenclature. Cell Stress and Chaperones *5*, 98-112.

Osipiuk, J., Walsh, M.A., Freeman, B.C., Morimoto, R.I. and Joachimiak, A. (1999). Structure of a new crystal form of human Hsp70 ATPase domain. Acta Crystallographica *D55*, 1105-1107.

Pauli, D., Arrigo, A.-P. and Tissières, A. (1992). Heat shock response in *Drosophila.* Experientia *48*, 623-629.

Peitsch, M. (1996). ProMod and Swiss-Model: Internet based tools for automated comparative protein modeling. Biochemical Society Transactions *24*, 274-279.

Pellechia, M., Montgomery, D.L., Stevens, S.Y., Vander Kooi, C.W., Feng, H.-P., Gierasch, L.M. and Zuiderweg, E.R.P. (2000). Structural insights into substrate binding by the molecular chaperone DnaK. Nature Structural Biology *7*, 298-303.

Pellechia, M., Szyperski, T., Wall, D., Georgopoulos, C. and Wüthrich, K. (1996). NMR structure of the J-domain and the Gly/Phe-rich region of the *Escherichia coli* DnaJ chaperone. Journal of Molecular Biology *260*, 236-250.

Pelling, G. (1959). Chromosomal synthesis of ribonucleic acid as shown by incorporation of uridine labelled with tritium. Nature *184*, 655-656.

Pierpaoli, E.V., Gisler, S. and Christen, P. (1998a). Sequence-specific rates of interaction of target peptides with the molecular chaperones DnaK and DnaJ. Biochemistry *37*, 16741-16748.

Pierpaoli, E.V., Sandmeier, E., Baici, A., Schönfeld, H.-J., Gisler, S. and Christen, P. (1997). The power stroke of the DnaK/DnaJ/GrpE molecular chaperone system. Journal of Molecular Biology *269*, 757-768.

Pierpaoli, E.V., Sandmeier, E., Schönfeld, H.-J. and Christen, P. (1998b). Control of the DnaK chaperone cycle by substoichiometric concentrations of the co-chaperones DnaJ and GrpE. Journal of Biological Chemistry *273*, 6643-6649.

Qian, Y.Q., Patel, D., Hartl, F.-U. and McColl, D.J. (1996). Nucleic Magnetic Resonance solution structure of the human Hsp40 (HDJ-1) J domain. Journal of Molecular Biology *260*, 224-235.

Reid, K.L., Fink, A.L. (1996). Physical interactions between members of the DnaK chaperone machinery: characterisation of the DnaK.GrpE complex. Cell Stress and Chaperones *1*, 127-137.

Richarme, G., Kohiyama, M. (1993). Specificity of the *Escherichia coli* chaperone DnaK (70 kDa heat shock protein) for hydrophobic amino acids. Journal of Biological Chemistry *268*, 24074-24077.

Riley, M.I., Yoo, W., Mda, N.Y. and Folk, W.R. (1997). Tiny T antigen: an autonomous polyomavirus T antigen amino-terminal domain. Journal of Virology *71*, 6068-6074.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in *Drosophila.* Experientia *18*, 571-573.

Ritossa, F. (1996). Discovery of the heat shock response. Cell Stress and Chaperones *1*, 97-98.

Ruddon, R.W., Bedows, E. (1997). Assisted protein folding. Journal of Biological Chemistry *272*, 3125-3128.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J. and Bukau, B. (1997). Substrate-specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. EMBO Journal *16*, 1501-1507.

Rüdiger, S., Schneider-Mergener, J. and Bukau, B. (2001). Its substrate specificity characterises the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. EMBO Journal *20*, 1042-1050.

Russell, R., Jordan, R. and McMacken, R. (1998). Kinetic characterisation of the ATPase cycle of the DnaK molecular chaperone. Biochemistry *37*, 596-607.

Russell, R., Karzai, A.W., Mehl, A.F. and McMacken, R. (1999). DnaJ dramatically stimulates ATP hydrolysis by DnaK: insight into targeting of Hsp70 proteins to polypeptide substrates. Biochemistry *38*, 4165-4176.

Sakahira, H., Breuer, P., Hayer-Hartl, M. and Hartl, F.-U. (2002). Molecular chaperones as modulators of polyglutamine protein aggregation. Proceedings of the National Academy of Sciences, USA *99*, 16412-16418.

Sambrook,J., Russell,D.W. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory Press.

Sayle, R.A., Milner-White, E.J. (1995). RasMol: Biomolecular graphics for all. Trends in Biochemical Sciences *20*, 374-376.

Schiene-Fischer, C., Habazettl, J., Schmid, F.X. and Fischer, G. (2002). The Hsp70 chaperone DnaK is a secondary amide peptide bond *cis-trans* isomerase. Nature Structural Biology *9*, 419-424.

Schlenstedt, G., Harris, S., Risse, B., Lill, R. and Silver, P.A. (1995). A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum with BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s. Journal of Cell Biology *129*, 979-988.

Schumacher, R.J., Hansen, W.J., Freeman, B.C., Alnemri, E., Litwack, G. and Toft, D.O. (1996). Cooperative action of Hsp70, Hsp90 and DnaJ proteins in protein renaturation. Biochemistry *35*, 14889-14898.

Segal, G., Ron, E.Z. (1995). The *dnaKJ* operon of *Agrobacterium tumefaciens*: transcriptional analysis and evidence for a new heat shock promoter. Journal of Bacteriology *177*, 5952-5958.

Sell, S.M., Eisen, C., Ang, D., Zylicz, M. and Georgopoulos, C. (1990). Isolation and characterisation of *dnaJ* null mutants of *Escherichia coli.* Journal of Bacteriology *172*, 4827-4835.

Sha, B., Lee, S. and Cyr, D. (2000). The crystal structure of the peptide binding fragment from the yeast Hsp40 protein Sis1. Structure with Folding and Design *8*, 299-807.

Silberg, J.J., Hoff, K.G. and Vickery, L.E. (1998). The Hsc66-Hsc20 chaperone system in *Escherichia coli*: chaperone activity and interactions with the DnaK-DnaJ-GrpE system. Journal of Bacteriology *180*, 6617-6624.

Silberg, J.J., Vickery, L.E. (2000). Kinetic characterisation of the ATPase cycle of the molecular chaperone Hsc66 from *Escherichia coli.* Journal of Biological Chemistry *275*, 7779-7786.

Smith, D.R., Doucetter-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Chuech, G.M., Daniels, C.J., Mao, J.-I., Rice, P., Nölling, J. and Reeve, J.N. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. Journal of Bacteriology ***179***, 7135-7155.

Sondermann, H., Scheufler, C., Schneider, C., Höhfeld, J., Hartl, F.-U. and Moarefi, I. (2001). Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsc70 nucleotide exchange factors. Science *291*, 1553-1557.

Stubdal, H., Zalvide, J., Campbell, K.S., Schweitzer, C., Roberts, T.M. and DeCaprio, J.A. (1997). Inactivation of pRB-related proteins p130 and p107 mediated by the J domain of Simian Virus 40 Large T antigen. Molecular and Cellular Biology *17*, 4979-4990.

Suh, W.-C., Burkholder, W.F., Lu, C.Z., Zhao, X., Gottesman, M.E. and Gross, C.A. (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. Proceedings of the National Academy of Sciences, USA *95*, 15223-15228.

Suh, W.-C., Lu, C.Z. and Gross, C.A. (1999). Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its co-chaperone DnaJ. Journal of Biological Chemistry *274*, 30534-30539.

Sullivan, C.S., Cantalupo, P. and Pipas, J.M. (2000a). The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP dependant mechanism. Molecular and Cellular Biology *20*, 6233-6243.

Sullivan, C.S., Tremblay, J.D., Fewell, S.W., Lewis, J.A., Brodsky, J.L. and Pipas, J. (2000b). Species-specific elements in the large T antigen J domain are required for cellular transformation and DNA replication by Simian virus 40. Molecular and Cellular Biology *20*, 5749-5757.

Szabo, A., Korszun, R., Hartl, F.-U. and Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured substrates. EMBO Journal *15*, 408-417.

Szabo, A., Langer, T., Schröder, H., Flanagan, J., Flanagan, J., Bukau, B. and Hartl, F.-U. (1994). The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system-DnaK, DnaJ and GrpE. Proceedings of the National Academy of Sciences, USA *91*, 10345-10349.

Takayama, S., Bimston, D.N., Matsuzawa, S., Freeman, B.C., Aime-Sempe, C., Xie, Z., Morimoto, R. and Reed, J.C. (1997). BAG-1 modulates the chaperone activity of Hsp70/Hsc70. EMBO Journal *16*, 4887-4896.

Tatsuta, T., Joo, D.M., Calendar, R., Akiyama, Y. and Ogura, T. (2000). Evidence for an active role of the DnaK chaperone systen in the degradation of σ^{32}. FEBS Letters *478*, 271-275.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796-815.

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

Thulasiraman, V., Yang, C.F. and Frydman, J. (1999). *In vivo* newly translated polypeptides are sequestered in a protected folding environment. EMBO Journal *18*, 85-95.

Tilly, K., Hauser, R., Campbell, J. and Ostheimer, G.J. (1993). Isolation of *dnaJ, dnaK* and *grpE* homologues from *Borrelia burgdorferi* and complementation of *Escherichia coli* mutants. Molecular Microbiology *7*, 359-369.

Tissières, A., Mitchell, H.K. and Tracey, U. (1974). Protein synthesis in the salivary glands of *Drosophila melanogaster*: relation to chromosome puffs. Journal of Molecular Biology *84*, 389-398.

Towbin, H., Staehelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proceedings of the National Academy of Sciences, USA *76*, 4350-4354.

Tsai, J., Douglas, M.G. (1996). A conserved HPD sequence of the J-domain is neccessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. Journal of Biological Chemistry *271*, 9347-9354.

Tsugeki, R., Nishimura, M. (1993). Interaction of homologues of Hsp70 and Cpn60 with ferredoxin-NADP+ reductase upon its import into chloroplasts. FEBS Letters *320*, 198-202.

Ueguchi, C., Kakeda, M., Yamada, H. and Mizuno, T. (1994). An analogue of the DnaJ molecular chaperone in *Escherichia coli*. Proceedings of the National Academy of Sciences, USA *91*, 1054-1058.

Ueguchi, C., Shiozawa, T., Kakeda, M., Yamada, H. and Mizuno, T. (1995). A study of the double mutation of *dnaJ* and *cbpA*, whose gene products function as molecular chaperones in *Escherichia coli*. Journal of Bacteriology *177*, 3894-3896.

Ungewickell, E., Ungewickell, H. and Holstein, S.E.H. (1997). Functional interaction of the auxilin J domain with the nucleotide- and substrate-binding modules of Hsc70. Journal of Biological Chemistry *272*, 19594-19600.

Vickery, L.E., Silberg, J.J. and Ta, D.T. (1997). Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from *Escherichia coli*. Protein Science *6*, 1047-1056.

Vriend, G. (1990). Whatif: a molecular modelling and drug design programme. Journal of Molecular Graphics *8*, 52-56.

Wall, D., Zylicz, M. and Georgopoulos, C. (1995). The conserved G/F motif of the DnaJ chaperone is necessary for the activation of the substrate binding properties of the DnaK chaperone. Journal of Biological Chemistry *270*, 2139-2144.

Wang, H., Kurochkin, A.V., Pang, Y., Hu, W., Flynn, G.C. and Zuiderweg, E.R.P. (1998). NMR solution structure of the 21kDa chaperone protein DnaK substrate binding domain: a preview of chaperone-protein interaction. Biochemistry *37*, 7929-7940.

Wawrzynów, A., Zylicz, M. (1995). Divergent effects of ATP on the binding of the DnaK and DnaJ chaperones to each other, or to their various native and denatured substrates. Journal of Biological Chemistry *270*, 19300-19306.

Wegrzyn, A., Taylor, K. and Wegrzyn, G. (1996). The *cbpA* chaperone gene function compensates for *dnaJ* in plasmid replication during amino acid starvation of *Escherichia coli.* Journal of Bacteriology *178*, 5847-5849.

Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H.M. (2003). The Protein Data Bank and structural genomics. Nucleic Acids Research *31*, 489-491.

Westermann, B., Gaume, B., Herrmann, J.M., Neupert, W. and Schwarz, E. (1996). Role of the mitochondrial DnaJ homolog Mdj1p as a chaperone for mitochondrially synthesised and imported proteins. Molecular and Cellular Biology *16*, 7063-7071.

Wittung-Stafshede, P., Guidry, J., Horne, B.E. and Landry, S.J. (2003). The J domain of Hsp40 couples ATP hydrolysis to substrate capture in Hsp70. Biochemistry *42*, 4937-4944.

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F.J., Woo, L., Chen, Y., Paulsen, I.T., Eisen, J.A., Karp, P.D., Bovee, D.S., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M.J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z.Y., Dolan, M., Chumley, F., Tingey, S.V., Tomb, J.F., Gordon, M.P., Olson, M.V. and Nester, E.W. (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science *294*, 2317-2323.

Yan, W., Craig, E.A. (1999). The glycine-phenylalanine-rich region determines the specificity of the yeast Hsp40 Sis1. Molecular and Cellular Biology *19*, 7751-7758.

Yan, W., Gale, M.J.J., Tan, S.-L. and Katze, M.G. (2002). Inactivation of the PKR protein kinase and stimulation of mRNA translation by the cellular co-chaperone P58^{IPK} does not require J domain function. Biochemistry *41*, 4938-4945.

Yoshimune, K., Yoshimura, T., Nakayama, T., Nishino, T. and Esaki, N. (2002). Hsc62, Hsc56 and GrpE, the third Hsp70 chaperone system of *Escherichia coli.* Biochemical and Biophysical Research Communications *293*, 1389-1395.

Zalvide, J., Stubdal, H. and DeCaprio, J.A. (1998). The J domain of Simian Virus 40 Large T Antigen is required to functionally inactivate RB family proteins. Molecular and Cellular Biology *18*, 1408-1415.

Zhang, F.L., Casey, P.J. (1996). Protein prenylation: Molecular mechanisms and functional consequences. Annual Review of Biochemistry *65*, 241-269.

Zhang, X.-P., Elofsson, A., Andreu, D. and Glaser, E. (1999). Interaction of mitochondrial presequences with DnaK and mitochondrial hsp70. Journal of Molecular Biology *288*, 177-190.

Zhu, J.-K., Bressan, R.A. and Hasegawa, P.M. (1993). Isoprenylation of the plant molecular chaperone ANJ1 facilitates membrane association and function at high temperature. Proceedings of the National Academy of Sciences, USA *90*, 8557-8561.

Zhu, X., Zhao, X., Burkholder, W., Gragerov, A., Ogata, C.M., Gottesman, M.E. and Hendrickson, W.A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. Science *272*, 1606-1614.

Zuber, M., Hoover, T.A., Dertzbaugh, M.T. and Court, D.L. (1995). Analysis of the DnaK molecular chaperone system of *Francisella tularensis.* Gene *164*, 149-152.

Zwanzig, R., Szabo, A. and Bagchi, B. (1992). Levinthal's Paradox. Proceedings of the National Academy of Sciences, USA *89,* 20-22.

Zylicz, M., Ang, D. and Georgopoulos, C. (1987). The grpE protein of *Escherichia coli.* Journal of Biological Chemistry *262,* 17437-17442.

Zylicz, M., Yamamoto, T., McKittrick, N., Sell, S. and Georgopoulos, C. (1985). Purification and properties of the DnaJ replication protein of *Escherichia coli.* Journal of Biological Chemistry *260,* 7591-7598.