Structural and biochemical studies of the *Caenorhabditis elegans* Hsp70/Hsp90 chaperone system

A Thesis

Submitted for the Degree of

Doctor of Philosophy

by

Liam Worrall, B.Sc., MRes.

Structural Biochemistry Group

The Institute of Structural and Molecular Biology

University of Edinburgh

March 2007

Summary of contents

Abstract

In the crowded cellular environment, folding of newly transcribed polypeptides and maintenance of the correct folded state of proteins presents a significant problem. The cell has thus developed a sophisticated chaperone system to regulate protein quality control. Two of the major players are the 70 kDa and 90 kDa heat shock protein families. Hsp70 is predominantly involved in the folding of newly transcribed polypeptides and partially unfolded proteins whilst Hsp90 is involved at a later stage, regulating the functional maturation of a sub-set of client proteins. The actions of Hsp70 and Hsp90 are regulated by a multitude of co-chaperones; a major family common to both pathways is the tetratricopeptide repeat (TPR) domain containing co-chaperones. The TPR domain is a 34 amino acid helix-loop-helix that occurs in tandem arrays and commonly participates in protein-protein interactions. TPR co-chaperones provide a diverse array of functionality to the Hsp70 and Hsp90 machinery including facilitating communication between the two pathways, protein transport, mitochondrial/chloroplast protein import and providing a link to the protein degradation system. This thesis describes the structural and biochemical studies of members of the Hsp70/Hp90 chaperone machinery in the nematode worm *Caenorhabditis elegans*.

The crystal structure of the C-terminal helical lid domain from C. elegans Hsp70

Hsp70 proteins are composed of two functionally distinct domains; the 40 kDa N-terminal nucleotide binding domain (NBD) and the 30 kDa C-terminal substrate binding domain (SBD). The SBD can be further divided into an 18 kDa β-sandwich sub-domain which forms the hydrophobic binding pocket and a 10 kDa helical-bundle sub-domain which forms a lid over the binding pocket. Structures of the helical sub-domain are limited to *E. coli* homologues DnaK and HscA, and rat Hsc70. Despite evolutionary structural conservation in the NBD and β-sandwich, the lid was shown to adopt alternate conformations in prokaryotes and eukaryotes. This work presents the crystal structure of the C-terminal 10 kDa sub-domain from *C. elegans* Hsp70. Despite a high degree of sequence identity, the *C. elegans* domain is shown to adopt a conformation distinct from the rat crystal structure, consistent with the more distantly related bacterial homologues. Comparison with the rat structure reveals an intriguing putative domain-swap dimerisation mechanism though the isolated *C. elegans* domain was found to exist exclusively as a monomer in solution.

Biochemical characterisation of two putative Hsp70/Hsp90 interacting TPR co-chaperones

A previous study identified two TPR domain containing *C. elegans* putative proteins predicted to interact with Hsp90. These proteins were identified as the *C. elegans* homologues for small glutamine-rich TPR containing protein (SGT) and Hsp70/Hsp90 organising protein (HOP). These proteins have been successfully cloned, expressed and purified. Characterisation of purified SGT by mass spectrometry, cross-linking and gel filtration experiments provides unambiguous evidence that SGT forms homo-dimers in solution. Its hydrodynamic dimensions in relation to its molecular weight suggest a protein with a low level of compactness and an extended conformation. Further, it has been demonstrated that SGT interacts with the C-terminal peptides from both Hsp70 and Hsp90 with equal affinities. Crystals were obtained for full-length SGT and its isolated TPR domain but were of insufficient quality for X-ray data analysis. Studies on *C. elegans* HOP suggested it might exist as a dimer in solution. In addition, a tight binding interaction was demonstrated with human and *C. elegans* Hsp90 homologues.

Identification of the complete repertoire of C. elegans TPR co-chaperones

A thorough search of the complete *C. elegans* proteome and genome was performed to identify the complete repertoire of TPR domain containing proteins likely to interact with Hsp70 or Hsp90. Hsp70/90 interacting TPR motifs have a well-defined domain architecture and a highly conserved consensus carboxylate-clamp motif. Profile hidden Markov models (HMMs) provide a means of representing the amino acid probability distribution of sequence alignments and are powerful stochastic models of protein families. A profile HMM based search of the published *C. elegans* protein and DNA databases identified 11 proteins; eight of which are homologues of proteins known to interact with Hsp70 or Hsp90. The remaining three are uncharacterised putative proteins and represent targets for further study.

Declaration

The work presented in this thesis is the original work of the author unless otherwise acknowledged. This thesis has been composed by the author and has not been submitted in whole or in part for any other degree.

Acknowledgements

I would like to thank my supervisor, Professor Malcolm D. Walkinshaw for providing with the opportunity to study for the degree of Doctor of Philosophy and also for encouraging me to venture back into the world of experimental science.

I am also grateful to Dr. Anthony P. Page, University of Glasgow, for allowing me to visit his lab for four weeks in August 2004 and his group members for teaching me the technique of cloning.

I must make special acknowledgement to the members of the Structural Biochemistry Group who have provided significant help and assistance: Sandra Bruce, Dr. Martin Weir, Dr. Iain McNae, Dr. Amir Rabu, Dr. Jacqueline and Dr. Paul.

In addition I would like to mention Dr. Lindsay Tulloch, Dr. Chris Brown, Hugh Morgan, Dinesh Soares, Dr. Julia Richardson, Prof. Lindsay Sawyer, Dr. Marjorie Harding, Daphne Kan, Simon Harding and Connie Ludwig for help in a variety of things during the course of my studies. I would also like to thank everyone on the third floor of the Michael Swann Building for providing a friendly, stimulating work environment and for their past and continued friendship.

Lastly I would like to thank my family - Christine, Gary and Erin - for their constant support and encouragement. And Ellie, for proof reading this thesis and putting up with me over the last four years.

Abbreviation list

ADP	Adenine diphosphate
ATP	Adenine triphosphate
CCP4	Collaborative Computational Project Number 4
CD	Circular dichroism
cDNA	Complementary deoxyribonucleoide acid
Chip	C-terminal of Hsp70/Hsp90 interacting protein
CsA	Cyclosporine A
Cyp	Cyclophilin
DAF	Dauer larva formation
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FKBP	FK506 binding protein
FPLC	Fast protein liquid chromatography
HEPES	N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulphonic acid)
Hip	Hsp70/Hsp90 interacting protein
Hop	Hsp70/Hsp90 organising protein
Hsp	Heat shock protein
IPTG	Isopropyl β-D-thiolgalactopyranoside
K_d	The dissociation constant
kDa	Kilodalton
K_m	The Michaelis-Menten constant
LB	Luria-Bertani broth
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation Time-of-Flight
NBD	Nucleotide-binding domain
NTA	Nitrolotriacetic acid
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethyleneglycol
P_i	Inorganic phosphate
pI	Isoelectric point
PP5	Protein phosphatase 5
PPIase	Peptidyl prolyl cis-trans isomerase
RMSD	Root mean square deviation
SBD	Substrate-binding domain
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SGT	Small glutamine-rich tetratricopeptide
SGT	Small glutamine-rich tetratricopeptide
SMART	Simple Modular Architecture Research Tool
SPR	Surface Plasmon Resonance
STI1	Stress-induced 1
T_m	Transition temperature
TOM	Translocase of outer membrane
TPR	Tetratricopeptide repeat
UV	Ultraviolet
V_m	Matthew's coefficient

Table of Contents

List of Figures

List of Tables

1. Introduction - the Hsp70/Hsp90 chaperone machinery

The central dogma of biology describes the conversion of genetic information to functional proteins; a process that begins in the nucleus with the transcription of coding genes to mRNA, which are then exported from the nucleus and translated into polypeptides by the ribosome. There is, however, a further step - the folding of newly formed polypeptides into specific three-dimensional structures. A class of proteins known as molecular chaperones are vital for this final step, aiding the process of folding newly transcribed polypeptides and supervising the structural fidelity of existing proteins.

The native three-dimensional structure of proteins is encoded in their linear sequence and governed by the noncovalent interactions of their amino acid side chains. This process is spontaneous; however, in the busy milieu of the cell proteins sometimes require assistance to achieve or maintain their correct structure. The cell is a crowded environment with high concentrations of proteins, nucleic acids and other macromolecules. This results in a so called excluded volume effect that can favour intermolecular over intramolecular interactions, and protein aggregation competes with folding of newly formed polypeptides. This problem is exacerbated by the fact that a protein domain cannot adopt its native structure until it has exited the ribosome. In the meantime, the exposure of hydrophobic regions normally located in the core of proteins can lead to unwanted aggregation. To address this issue the cell has developed a complex system of molecular chaperones involved in supervising the correct folding of newly synthesised polypeptides, maintenance of the folded state of existing proteins and, in some cases, required for correct protein function. For an excellent review of the chaperone pathways see Young et al., 2004.

Two of the major chaperone pathways involve the heat shock proteins Hsp70 and Hsp90 (Figure 1-1). These were first observed as proteins upregulated in response to elevated temperatures and subsequently identified as protein chaperones (Welch and Feramisco, 1982). The Hsp70 chaperone pathway is the most common folding pathway with Hsp70 homologues ubiquitously expressed and present in virtually all living organisms (Wegele et al., 2004). It is involved in numerous protein folding processes including folding of nascent polypeptides, refolding of misfolded and aggregated proteins, and transmembrane protein transport. The Hsp90 chaperone machinery is somewhat different. It is mainly involved in the maintenance of the functional viability of a sub-set of client proteins which require Hsp90 to adopt their functionally active conformations. There is communication between the two pathways with some proteins processed first by the Hsp70 machinery prior to passing to

the Hsp90 machinery. Importantly, both chaperones are involved in protein turnover providing a direct link to the protein degradation machinery.
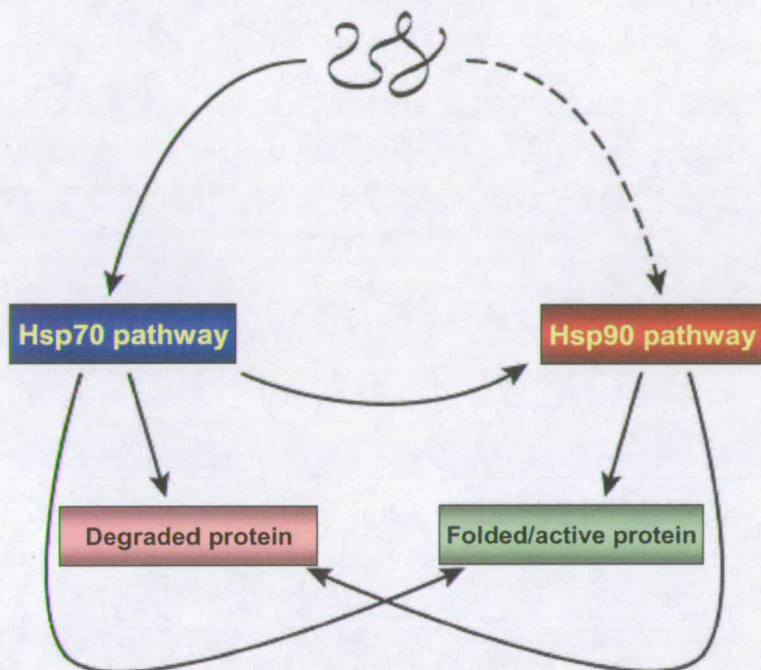
Figure 1-1 Overview of the Hsp70 and Hsp90 chaperone pathways. Hsp70 is involved in the folding of nascent and unfolded polypeptides whilst Hsp90 works together with Hsp70 in the functional maturation of a subset of client proteins. In addition, both chaperones are involved in protein quality control and are capable of targeting proteins for degradation. Figure adapted from Wegele et al., 2004.

1.1.　The Hsp70 chaperone machinery

The 70 kDa heat-shock proteins (Hsp70s) comprise a family of conserved chaperones that regulate a wide variety of cellular processes during normal and stress conditions (Boorstein et al., 1994). Hsp70 is one of the most abundant of these proteins, accounting for as much as 1-2% of total cellular protein (Herendeen et al., 1979). Humans possess at least 11 distinct genes that code for Hsp70 isoforms with homologues found in the cytoplasm (Hsp70 and Hsc70), endoplasmic reticulum (BiP/Grp78) and mitochondria (mtHsp70/Grp75) (Tavaria et al., 1996). Hsc70 is the major constitutively expressed isoform whilst Hsp70 is an inducible form upregulated in response to stress in addition to a variety of physiological processes such as cell cycle control, proliferation and differentiation. In *E. coli*, there are at least three cytosolic isoforms with the most common being DnaK whilst yeast has 14 isoforms, 9 of

which are found in the cytosol (Table 1-1) (Wegele et al., 2004). The evolution of multiple isoforms with varying sub-cellular localisation along with a large diverse collection of co-chaperones has facilitated the broad spectrum of activities with which Hsp70 is implicated. For clarity, Hsp70 will be used to refer to the Hsp70 family as a whole unless otherwise specified.

	Archaea	*Eubacteria*	*Yeast*	*Plants*	*Mammals*
Cytosol	DnaK[1]	DnaK	Ssa1	Hsp70	Hsp70
	-	Hsc66	Ssa2	Hsc70	Hsc70
	-	Hsc62	Ssa3	-	-
	-	-	Ssa4	-	-
	-	-	Ssb1	-	-
	-	-	Ssb2	-	-
	-	-	Ssz1	-	-
ER	-	-	Grp78/Bip	Grp78/Bip	Grp78/Bip
Mitochondria	-	-	Ssc1	mtHsp70	mtHsp70
	-	-	Ssc2	-	-
Chloroplasts	-	-	-	Com70	-
	-	-	-	IAP70	-
	-	-	-	sHsp70/CSS1	-
	-	-	-	sHsp70/S78	-

[1]not all archaea have Hsp70 homologues
Table 1-1 Species and organelle Hsp70 isoforms.

Hsp70 is involved in many processes including traditional chaperone roles of folding nascent polypeptides, the prevention of aggregation of unfolded proteins and the solubilisation and refolding of aggregated proteins. In addition, Hsp70 plays important roles in protein translocation across membranes and the disassembly of protein complexes including the clathrin cage, viral capsids and the nucleoprotein complex (Sousa and Lafer, 2006). These diverse functions are achieved via the repetitive transient association of Hsp70 with exposed hydrophobic patches in client proteins in an ATP-dependent manner.

It is estimated that 5-18% of bacterial proteins require Hsp70 for correct folding (Bukau et al., 2000) with this figure likely higher in eukaryotes due to the larger average protein size. Hsp70 is thought to assist protein folding in a passive manner, binding exposed hydrophobic patches on unfolded proteins thereby preventing them from aggregation and providing an amenable environment for correct folding. An additional theory is that Hsp70 uses energy

derived from its intrinsic ATPase activity to provide a "power stroke" which can overcome kinetic barriers for folding (Slepenkov and Witt, 2002).
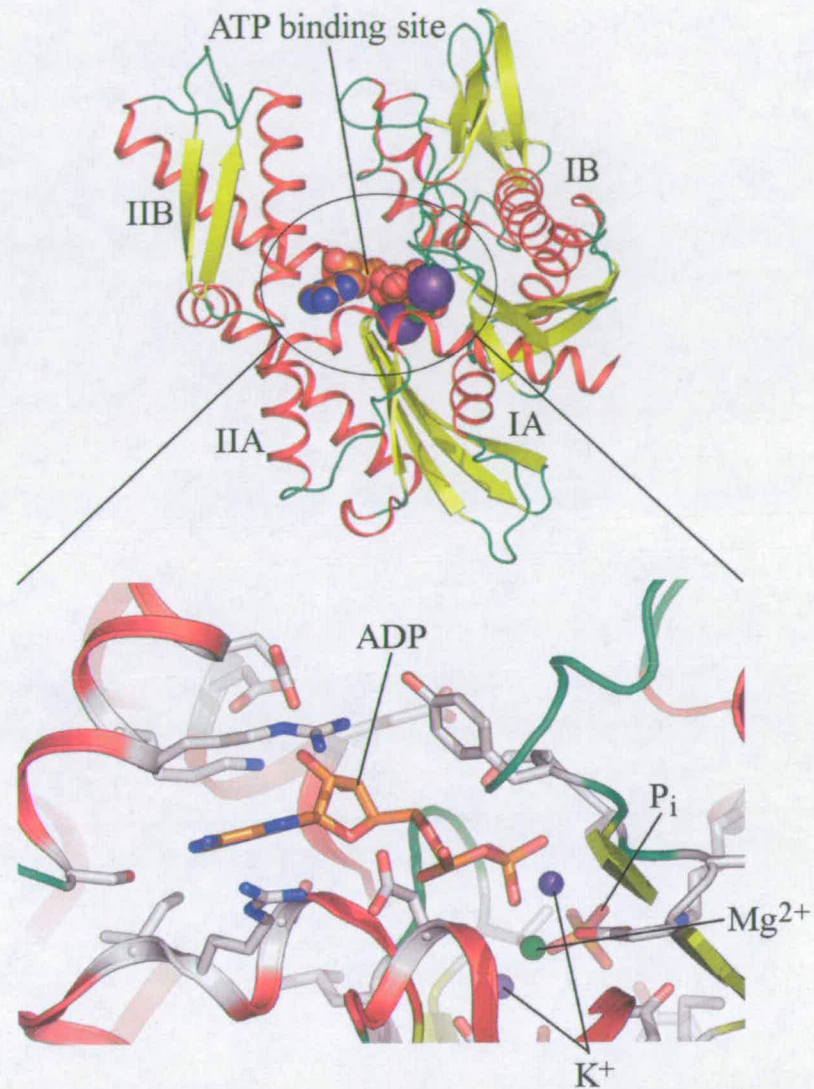
Figure 1-2 N-terminal Hsp70 nucleotide binding domain. The NBD is composed of two lobes (I and II) which are further divided into sub-domains A and B. ATP binds in a cleft between these lobes and makes contacts with residues from all four subdomains. This figure and all subsequent molecular graphics figures produced with PyMol (http://www.pymol.org).

1.1.1. The Hsp70 structure

Hsp70 is composed of two intimately related but functionally distinct domains; the 40 kDa N-terminal nucleotide binding domain (NBD), which both binds and hydrolyses ATP and the 30 kDa C-terminal substrate binding domain (SBD) (Chappell et al., 1987). The SBD can be

further divided into an 18 kDa β-sandwich subdomain which forms the hydrophobic binding pocket and a 10 kDa helical-bundle subdomain which forms a lid over the binding pocket (Zhu et al., 1996). Substrate binding and release is an allosteric process with ATP binding and hydrolysis in the NBD regulating client binding and release in the SBD (Flynn et al., 1989; Takeda and McKay, 1996).

The structure of the Hsp70 ATPase domain was first solved in 1990 revealing an actin-like fold (Flaherty et al., 1990). It comprises two subdomains (I and II) (Figure 1-2) which are in turn divided into two small subdomains (A and B). The nucleotide binding site is located in a cleft between the two lobes and nucleotide binding occurs in conjunction with one magnesium and two potassium ions.

The C-terminal SBD can be divided into an 18 kDa β-sandwich subdomain and a 10 kDa helical subdomain. Hendrickson and colleagues solved the first structure of the SBD from the *E. coli* Hsp70 homologue DnaK in complex with the heptapeptide NRLLLTG (Zhu et al., 1996). This revealed an 8-stranded anti-parallel β-sandwich, which contained the hydrophobic binding groove, and a helical-bundle which appeared to form a lid over the bound peptide (Figure 1-3).

The peptide binding groove is formed by pairs of inner and outer loops connecting the β-sheets. Conserved aliphatic and aromatic residues form a hydrophobic cavity that accommodates the central leucine of the bound peptide (Figure 1-3). The preferred substrate binding motif is characterised by a core of four or five consecutive amino acids enriched in hydrophobic residues, especially leucine, and flanking regions enriched in basic residues and such sequence bind with affinities in the range 5 nM to 5 μM (Bukau and Horwich, 1998).

The helical subdomain, composed of five α-helices (αA-αE), sits over the binding groove. Only helices αA and αB are in direct contact with the β-sandwich, with the long helix αB extending over the binding groove and forming contacts with the loop regions connecting the β-sheets (Figure 1-3). Of particular importance are conserved salt bridges between the C-terminal region of helix αB and the outer loops of the β-sandwich, termed the latch, with disruption shown to affect peptide binding (Figure 1-3) (Fernandez-Saiz et al., 2006). Helices αC-αE, together with the C-terminal half of helix αB, form an anti-parallel three-helix bundle. It has been proposed that the helical subdomain acts as a lid, regulating access to the substrate binding pocket (Zhu et al., 1996). The SBD terminates in a 20-30 residue flexible loop. The precise function of this is still unclear but it does interact with several co-chaperones and eukaryotic cytosolic isoforms terminate in a conserved GPTIEEVD motif

important in binding to Hsp40 (section 1.1.3.1) and TPR domain containing co-chaperones (section 1.4)
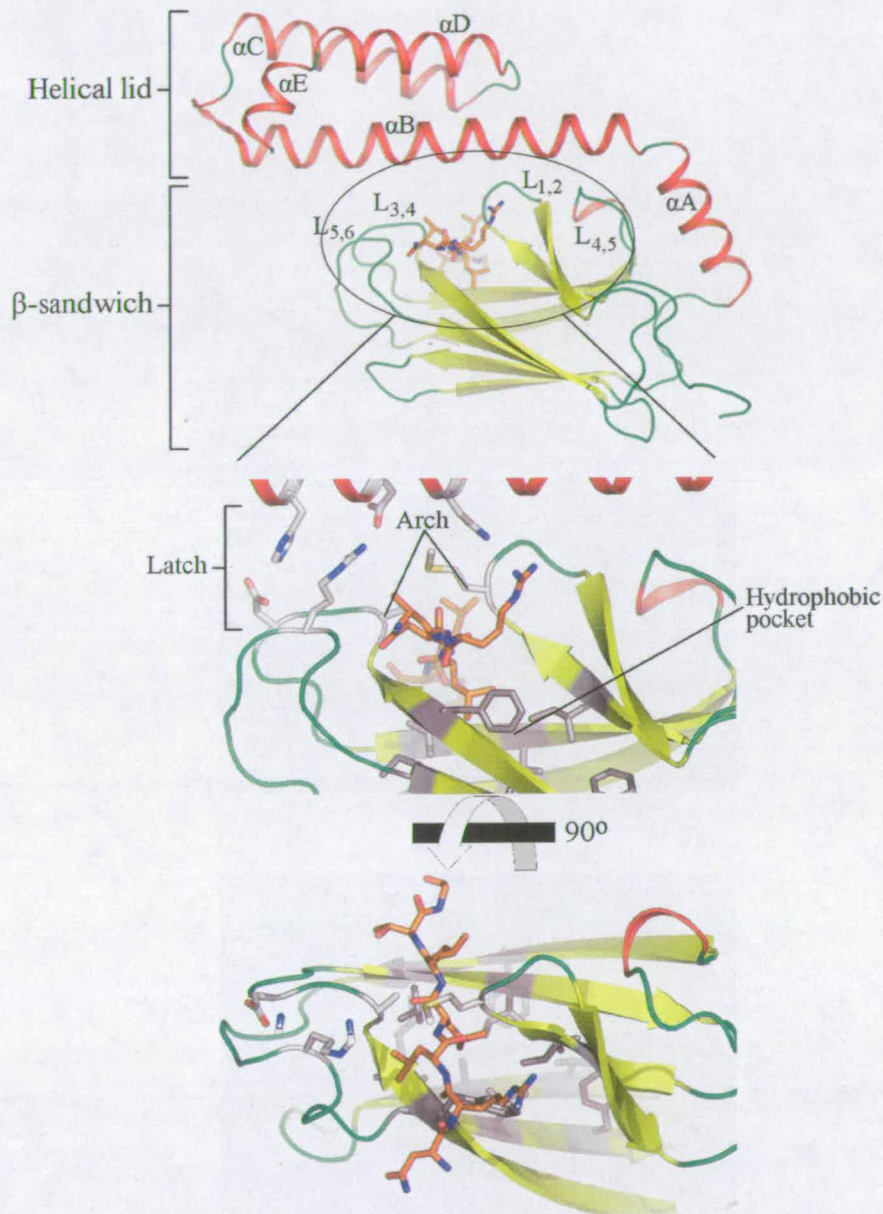
Figure 1-3 Hsp70 substrate binding domain. The SBD is composed of a β-sandwich subdomain and a helical lid subdomain. The loops connecting the β-sheets form a hydrophobic peptide binding groove which is covered by helix αB of the helical subdomain. The central leucine of the bound peptide is accommodated in a pocket lined with conserved hydrophobic residues and encapsulated by a pair of little and large conserved hydrophobic amino acids termed the arch. The peptide bound conformation is stabilised by "latch" interactions between the outer-loops and the C-terminal end of helix αB.
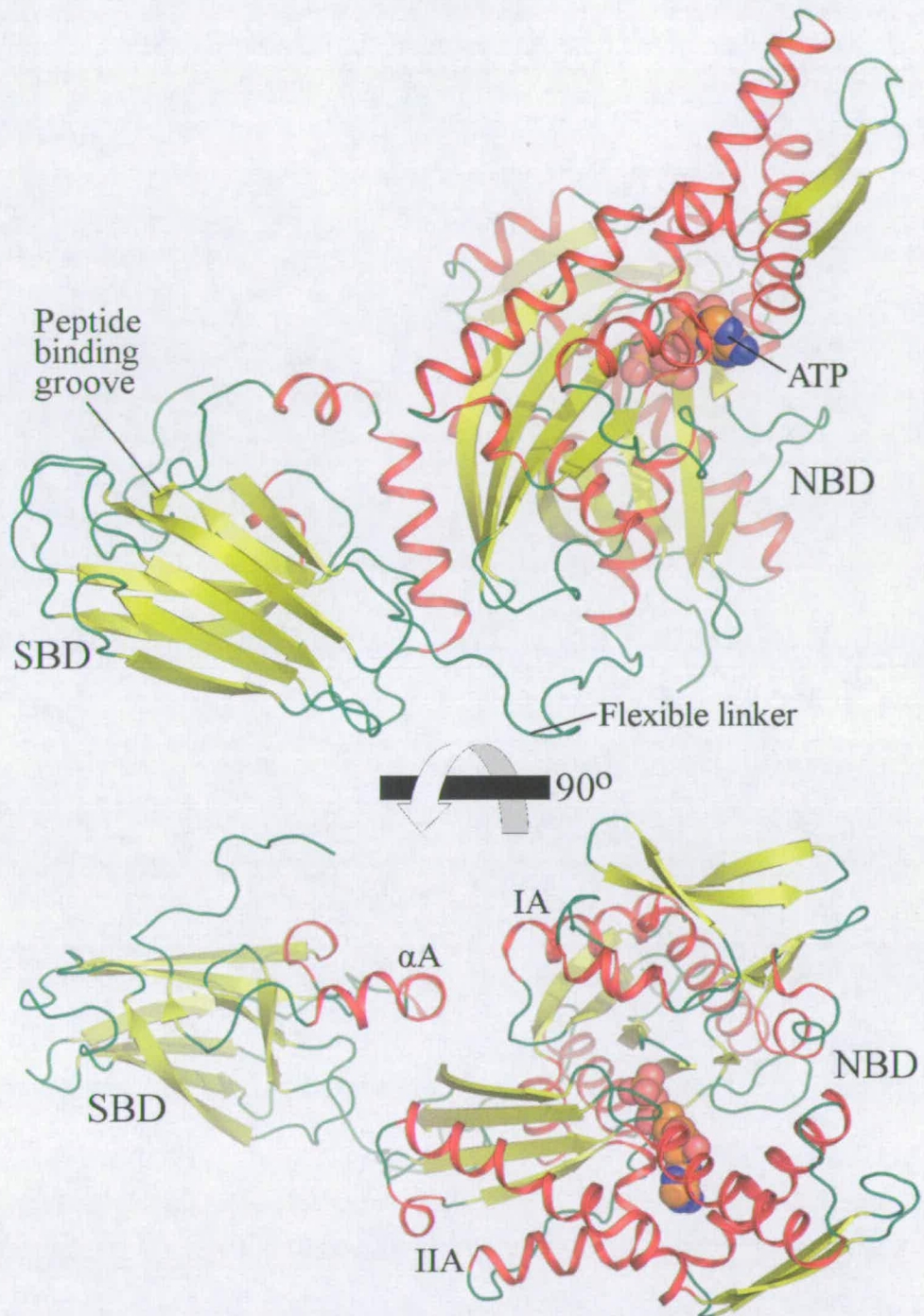
Figure 1-4 Two domain structure of bovine Hsc70. Bovine Hsc70 (residues 1-554) (Jiang et al., 2005) was solved in the absence of most of the C-terminal helical subdomain. The two domains are connected by an exposed flexible linker previously implicated in the allosteric control of substrate binding. The interdomain interface is formed by helix αA of the SBD and the cleft between lobes IA and IIA of the NBD. The C-terminal region of helix αB was found to have unwound and bound in the peptide binding groove.

The structure of the complete SBD from *E. coli* HscA, a specialised Hsp70 chaperone, has also been solved (Cupp-Vickery et al., 2004). Despite low sequence conservation with *E. coli* DnaK (<20 % sequence identity), HscA was shown to adopt an identical structure with the same mechanisms of peptide binding. Further β-sandwich subdomain structures have since been published from *E. coli* (Stevens et al., 2003) and rat (Morshauser et al., 1999); however, none together with the helical subdomain. These structures demonstrate evolutionary conservation from prokaryotes to eukaryotes with a near identical overall topology of the β-sandwich and good conservation of the residues important in substrate binding. In contrast, the only published eukaryotic structure of the helical subdomain is from rat Hsc70 (Chou et al., 2003). Despite structural conservation of the NBD and β-sandwich subdomain between *E. coli* and eukaryotic homologues, the helical subdomain of rat Hsc70 formed a helix-loop-helix that dimerised via a coiled-coil like interaction.

The structure of a full-length Hsp70 protein remains elusive with the most complete structure published to date being bovine Hsc70 (residues 1 - 554) (Jiang et al., 2005). As with many of the SBD structures, successful crystallisation was achieved by the removal of most of the helical subdomain, although the remaining protein has been shown to be functionally active. The structure provided the first direct insight into the relative spatial orientation of the NBD and SBD, showing a bi-lobal conformation with helix αA of the helical subdomain resting in a groove between lobes IA and IIA of the NBD (Figure 1-4). The structure also supported the hypothesis of an exposed linker connecting the two domains with important roles in the allosteric regulation of substrate binding (see below).

1.1.2. The allosteric regulation of substrate binding and release

Hsp70 achieves its multitude of functions via the repetitive transient association with exposed hydrophobic patches on client proteins. This process is allosteric, with ATP binding and hydrolysis in the NBD controlling substrate binding and release in the SBD. Structural and biochemical evidence has demonstrated that Hsp70 proteins exist in equilibrium between two conformations; a low-affinity "open" conformation characterised by rapid substrate binding and release and a high-affinity "closed" conformation with slow substrate association and dissociation (Mayer et al., 2000). There is a high energy barrier separating these two states; spontaneous transition occurs on the time scale of tens of minutes (Vogel et al., 2006a) and the functional interconversion is controlled by cycles of ATP binding,

substrate-binding, ATP hydrolysis and substrate release. This cycle, described in detail below, is graphically illustrated in figures 1-5 and 1-6.
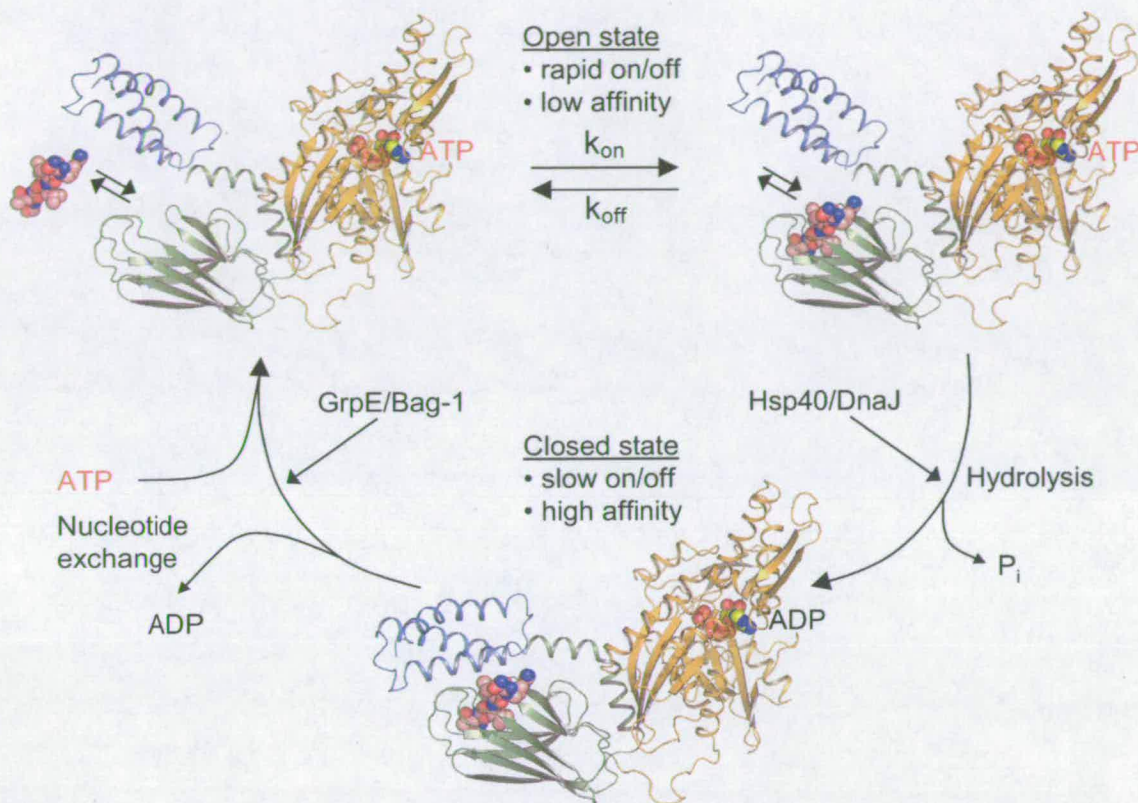
Figure 1-5 Hsp70 chaperone cycle. Hsp70 exists in equilibrium between two conformations; an open low-affinity state with rapid substrate binding and release and a high-affinity closed conformation with slow association and dissociation. ATP binding stabilises the open conformation. Substrate binding, in synergy with J domain co-chaperones, triggers ATP hydrolysis and a conformational change to the high-affinity closed state trapping the peptide. Exchange of ADP for ATP, facilitated by nucleotide exchange factors leads to opening of the binding pocket and substrate release. Open and closed structures are models and purely for illustrative purpose.

The cycle begins with ATP binding to the NBD inducing the adoption of the low-affinity open conformation. To do this, a signal must be transduced from the NBD to the substrate binding groove, some 50 Å in distance; the exact mechanisms of which are only gradually becoming clearer. This is thought to begin with the sensing of ATP binding by two residues (Lys[71] and Glu[176]; numbering refers to bovine Hsc70) at the bottom of the ATP binding pocket which interact with the γ-phosphate (Figure 1-6) (Vogel et al., 2006a). These residues, on lobes IB and IA respectively, effect a change in spatially adjacent Pro[148], located on a conserved strand of lobe IA (Vogel et al., 2006a). This residue has been proposed to be

the switch for the conformational change, possibly via a cis-trans isomerisation, with a high activation energy for transition between conformations. This switch is relayed, via Arg[156], to two positively charged residues on the surface of lobe IA (Lys[160] and Arg[171]) in the region of the inter-domain interface seen in the crystal structure (Vogel et al., 2006b). Although the exact mechanism is unclear, this in turn triggers a conformational change in the exposed inter-domain hydrophobic linker which is proposed to invade the inter-domain interface at the bottom of helix αA of the SBD (Figure 1-6) (Jiang et al., 2005; Vogel et al., 2006b).

390-DLLLLD-396
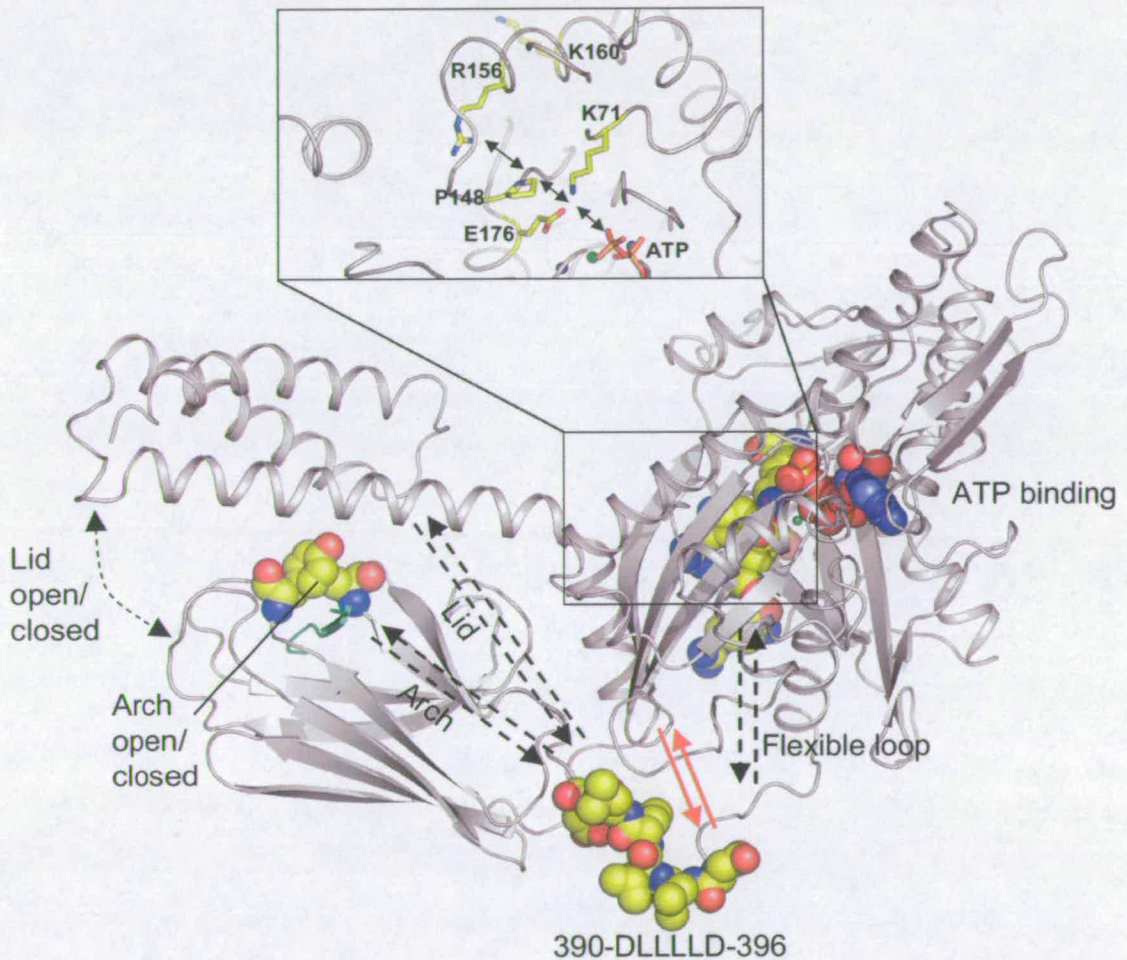
Figure 1-6 Allosteric changes involved in substrate binding and release. ATP binding is sensed by two residues at the base of the ATP binding cleft; K71 and E176. This is transduced to a surface cluster of positively charged residues (R156, K160 and R171 (not shown)) via P148. This triggers a change in the position of the flexible loop around D396 with the conserved hydrophobic motif (LLLL) inserting in a hydrophobic pocket at the base of helix αA. This causes changes in the SBD resulting in opening of the arch residues and opening of the lid subdomain. Substrate binding, along with J domain co-chaperones stimulates the reverse pathway.

The secondary phase is the allosteric changes in the SBD that contribute to the opening of the substrate binding pocket, proposed to involve two main parts: opening of the helical lid and opening of the β-sandwich (Figure 1-6). The crystal structure of the SBD of *E. coli* DnaK showed that a conformational change in the lid would be necessary for substrate binding or release to occur (Zhu et al., 1996). A second crystal form from the same study revealed a conformation whereby the helical lid was bent upwards by 11° half-way along helix αB, breaking the bonds of the latch residues, and this was proposed to be the means of the allosteric regulation. Evidence from additional structures has further suggested that the hinge could be located at the bend between helices αA and αB or could be affected by a complete rotation of the helical subdomain (Morshauser et al., 1999; Wang et al., 1998). The second, and perhaps more important, effector region is around the substrate binding groove itself. Experiments with lidless mutants show they are functionally active under physiological conditions (Ungewickell et al., 1997; Wilbanks et al., 1995) implying regions outwith the lid in the allosteric control. Changes in the residues lining the binding pocket have been implicated with the most important being the conserved complementary little and large hydrophobic residues that form the arch over the bound peptide (Ala^{406} and Tyr^{431} in bovine Hsc70) (Figure 1-6) (Mayer et al., 2000).

The net result of ATP binding is the opening of the SBD. This conformation has low substrate affinity and rapid association and dissociation kinetics. Substrate binding triggers the cascade of events that result in ATP hydrolysis and the transition to the high-affinity closed conformation. This is thought to involve the reverse pathway outlined above with the conformational change of the hydrophobic linker vital in positioning the ATPase active site into a catalytically favourable conformation (Vogel et al., 2006b). Substrate binding alone, however, is not sufficient to stimulate the ATPase rate enough for Hsp70 function and the action of a family of J domain containing co-chaperones is required (see section 1.1.3.1.). These bind the NBD and work in synergy with substrate binding to stimulate ATP hydrolysis by several orders of magnitude (Laufen et al., 1999). ATP hydrolysis stabilises the closed high-affinity conformation enclosing the client in the substrate binding groove. The cycle is completed by the exchange of ADP for ATP. ADP has relatively slow dissociation constant from the NBD and can represent the rate limiting step in substrate release. The regulation of nucleotide exchange comes in the form of nucleotide exchange factors (NEFs) GrpE, Bag-1 and HspBP1. These bind the mouth of the nucleotide binding cleft, inducing ADP dissociation by opening the NBD (Mayer and Bukau, 2005). The high physiological concentration of ATP leads to rapid association and the cycle continues.

1.1.3. Hsp70 co-chaperones

Hsp70 alone is poorly active and non-specific. As mentioned above, its basal rate of ATP hydrolysis is low and sequences suitable for Hsp70 binding occur approximately every 40 residues in proteins. In addition to the numerous isoforms with different expression patterns and sub-cellular locations, specificity is achieved by the interactions with numerous different co-chaperones which recruit Hsp70 proteins to carry out a variety of specific tasks. Many of these have a modular architecture in which a chaperone-interacting domain is fused to other domains of different function. These serve to regulate the activity of Hsp70, target Hsp70 to specific locations or bring Hsp70 together with specific binding partners. Co-chaperones include the J domain chaperones and NEFs mentioned above and also a diverse family of tetratricopeptide repeat (TPR) containing proteins.
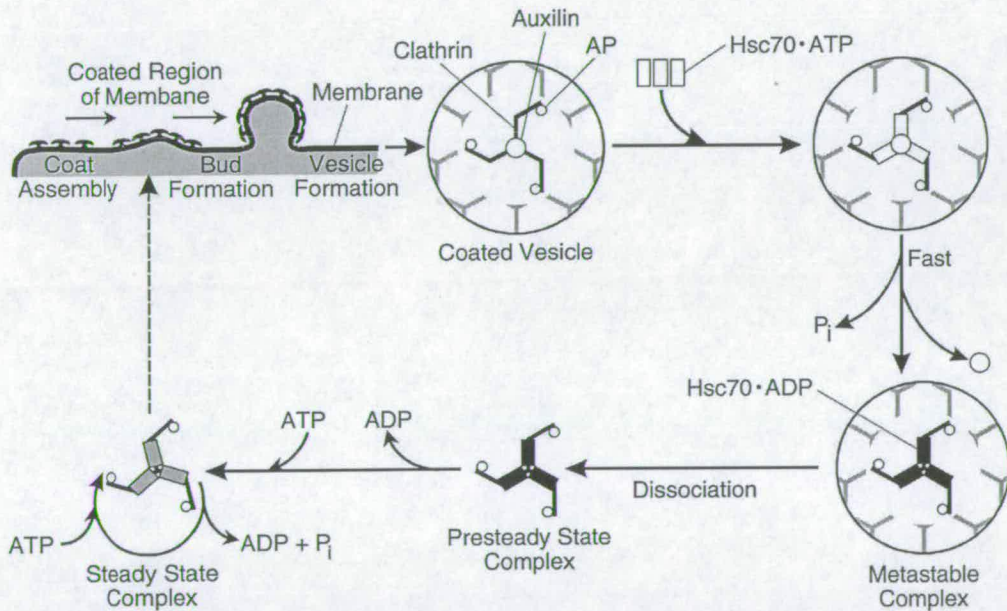
Figure 1-7 The role of Hsc70 in uncoating of clathrin-coated vesicles. Auxilin recruits Hsc70 to endocytosed clathrin-coated vesicles. ATP hydrolysis leads to formation of a metastable complex and ultimately dissociation. Free clathrin is recycled. Figure taken from Jiang et al., 2000.

1.1.3.1. The J domain family of Hsp70 co-chaperones

The J domain, or Hsp40/DnaJ, family of co-chaperones are modular adaptor proteins tailored to a specific function through the presence of target-specific domains in addition to a J domain capable of interacting with Hsp70. The J domain regulates the chaperone cycle by binding to Hsp70 in the closed conformation and potentiating the rate of ATP turnover (Hennessy et al., 2005). Hsp40-like co-chaperones are a diverse family of proteins with the number of isoforms from species to species exceeding the number of Hsp70 isoforms; 6 in *E. coli*, 20 in *S. cerevisiae*, 33 in *C. elegans* and 44 in human (Mayer and Bukau, 2005). Hsp40-like proteins are defined by the presence of the J domain, a 70-amino-acid domain with similarity to the initial 73 amino acids of the archetypal Hsp40, *E. coli* DnaJ (Pellecchia et al., 1996). The domain forms a four helix structure with a loop region containing a highly conserved histidine, proline, and aspartic acid (HPD) motif. This motif is present in virtually every J domain and is integral in the interaction with Hsp70 (Tsai and Douglas, 1996).

Hsp40/DnaJ is the archetypal family member. It has broad specificity and is thought to participate in the general protein folding pathway together with Hsp70. Hsp40/DnaJ is capable of binding unfolded proteins directly and is thought to bind to unfolded or nascent polypeptides and present them to Hsp70 (Hennessy et al., 2005).

Further Hsp40-like proteins have more specialised roles. One well studied example is auxilin, a J domain protein linking Hsp70 to the endocytic pathway (Lemmon, 2001). Receptor mediated endocytosis is an important cellular function for the rapid import of membrane bound receptors into cells. This is facilitated by the protein clathrin, which forms large cage-like vesicles encapsulating the imported proteins. On import, Hsp70 is recruited to the clathrin-coated vesicles by the J domain containing clatherin assembly protein auxilin (Pishvaee et al., 2000) and stimulates disassembly in an ATP dependent reaction (Figure 1-7). Of interest, the yeast auxilin homologue Swa2p contains both a J domain and a TPR domain (see section 1.4.), both of which are necessary for the interaction with Hsp70 (Xiao et al., 2006).

Although no structural evidence is available for a J domain-Hsp70 complex, an experimentally based model has been proposed with the auxilin J domain docked into the cleft between lobes IA and IIA (Figure 1-8) (Gruschus et al., 2004). This region is vital in the interdomain allosteric communication (section 1.1.2) and is also the site for the interdomain interface in the bovine Hsc70 two domain structure (section 1.1.1; Figure 1.4).

Figure 1-8 Experimentally based model of auxilin J domain-Hsp70 NDB complex. Auxilin (green) is predicted to bind in the cleft between lobes IA and IIA of the NBD (orange) in a position marginally overlapping with helix αA of the lid subdomain. The conserved J domain HPD motif is shown in spheres as is the bound ATP.

1.1.3.2. Nucleotide exchange factors

Substrate release requires the dissociation of ADP and the association of ATP. Spontaneous ADP dissociation is slow and can represent the rate limiting step in substrate release. Several unrelated proteins have evolved to facilitate the process of nucleotide exchange, namely GrpE, Bag-1 and HspBP1. These three structurally unrelated proteins function in a similar manner, binding to the mouth of the nucleotide binding cleft inducing a conformation incompatible with nucleotide binding (Figure 1-9).

1.1.3.3. The TPR family

A family of co-chaperones interact with Hsp70 via a TPR domain, a helical motif commonly implicated in protein-protein interactions. These provide a wide range of additional functionality encompassing communication with the Hsp90 machinery, protein degradation, protein transport, regulation of signal transduction pathways and neurotransmission amongst

many other roles. The TPR domain and the family of TPR domain containing co-chaperone is discussed in detail below (section 1.3).
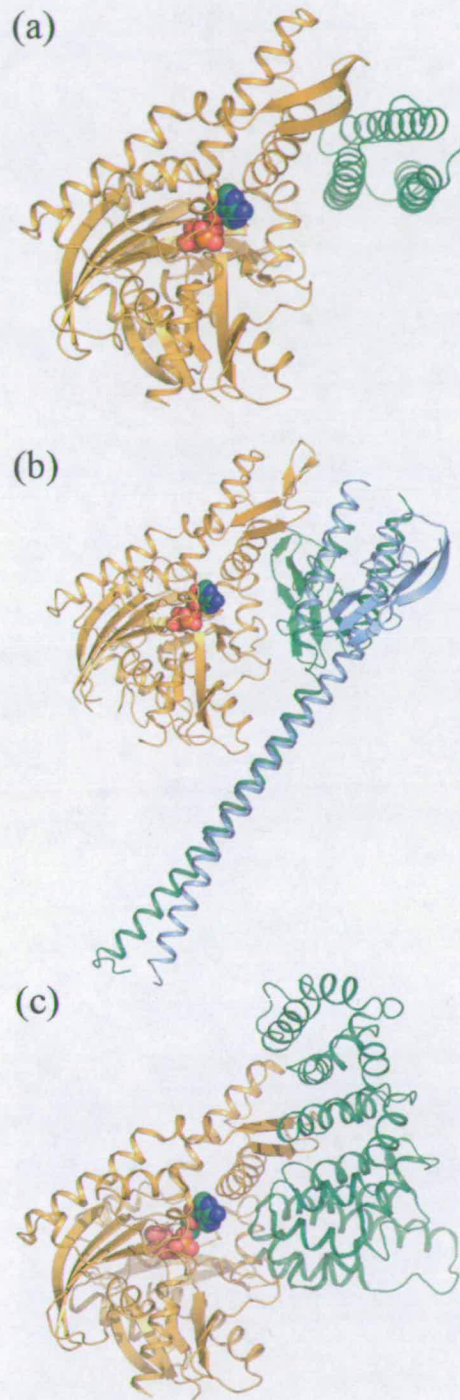
Figure 1-9 Crystal structures of the Hsp70 NBD complexed with nucleotide exchange factors. (a) Bag-1 (Sondermann et al., 2001), (b) GrpE (Harrison et al., 1997) and (c) HspBP1 (Shomura et al., 2005). All NEFs bind in a similar manner at the mouth of the ATP binding cleft, opening the nucleotide binding pocket and triggering ADP release.

1.2.   The Hsp90 chaperone machinery

Like Hsp70, the 90 kDa heat shock protein Hsp90 is one of the most abundant proteins constituting 1-2 % of total soluble protein (Lai et al., 1984). Homologues are found in all branches of life except archaea with an essential function in eukaryotes (Wegele et al., 2004). As with the Hsp70 family, different members are localised in different cellular compartments with inducible Hsp90α and constitutive Hsp90β (85% sequence identity) found in the cytosol (Hickey et al., 1989), the 94 kDa glucose regulated protein (Grp94) located in the endoplasmic reticulum (Little et al., 1994) and TNF receptor-associated protein 1 (Trap1/Hsp75) found in the mitochondrion (Song et al., 1995). The family is well conserved throughout evolution with yeast and *E. coli* homologues sharing 60% and 40% sequence identity respectively with human Hsp90 (Table 1-2; Figure 1-10).
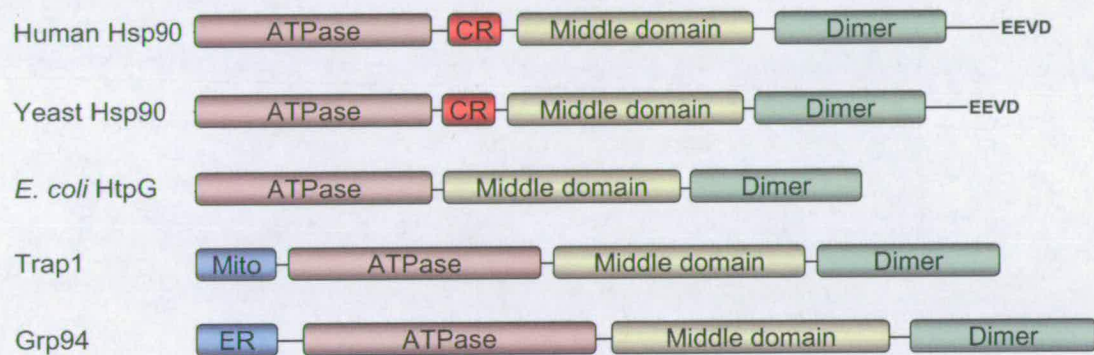
Figure 1-10 Domain architecture of Hsp90 homologues from human, yeast, *E. coli*, the mitochondrion and the endoplasmic reticulum. CR - charged region, Mito - mitochondrial signal peptide, ER - endoplasmic reticulum signal peptide.

Unlike the Hsp70 chaperone machinery, Hsp90s do not participate in the folding of newly translated polypeptides (Nathan et al., 1997) and are instead involved in the functional maturation of a sub-set of client proteins at a late stage of their folding process (Jakob et al., 1995). These proteins are commonly involved in signal transduction and include kinases e.g. Cdk2, Chk1 and ErbB2; transcription factors e.g. all steroid receptors and p53; enzymes e.g. DNA polymerase α, telomerase and nitric oxide synthase; and cytoskeletal proteins e.g. actin, tubulin and myosin (Figure 1-11). An updated list of ~120 proteins with direct biochemical evidence of interacting with Hsp90 is maintained by Cyril Picard (Appendix A.1). In addition, a recent survey of yeast proteomic and genomic data predicted 198 physical interactions (Zhao et al., 2005).

	Archea	Eubacteria	Yeast	Plants	Mammals
Cytosol	-	HtpG	Hsc83p	Hsp90	Hsp90a
			Hsp83	-	Hsp90b
ER	-	-	-	Grp94	Grp94/GP96
Mitochondria	-	-	-	-	Trap1/Hsp75
Chloroplasts	-	-	-	cpHsp82	-

Table 1-2 Species and organelle Hsp90 isoforms.
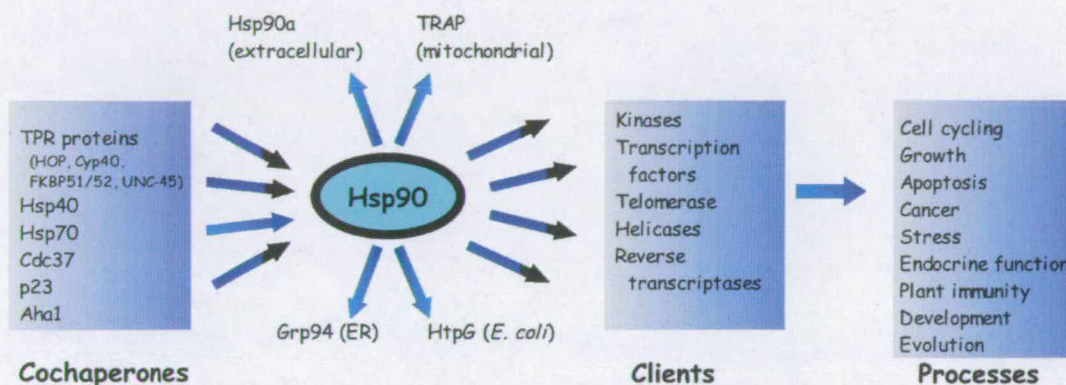
Figure 1-11 Overview of Hsp90 chaperone system. Hsp90 homologues, in conjunction with numerous co-chaperones, act on a wide range of client proteins and affect multiple cellular processes. Figure taken from Jackson et al., 2004.

1.2.1. The Hsp90 structure

Hsp90 exists predominantly as an elongated dimer. Each monomer is composed of three domains; an N-terminal ATPase domain, a middle domain and a C-terminal domain that is responsible for dimerisation (Figure 1-10). A highly charged, proteolytically-sensitive, linker region connects the N-terminal and middle domains; the length and composition of which varies both between species and isoforms (Wegele et al., 2004).

The 25 kDa N-terminal domain is well conserved across species and both binds and hydrolyses ATP. Crystal structures of this domain from human (Stebbins et al., 1997) and yeast (Prodromou et al., 1997) homologues revealed a two-layer α/β sandwich structure that forms an unusual ATP binding pocket known as the Bergerat fold (Figure 1-12) (Bergerat et al., 1997), an atypical ATP binding domain unlike others from kinases or Hsp70. The ATP binding site is also the target for N-terminal binding small-molecule Hsp90 inhibitors such as geldanamycin and radicicol. Despite poor sequence conservation, the structure was found to be similar to several DNA manipulating proteins, namely type II and type IV DNA topoisomerases, the MutL mismatch repair protein and bacterial DNA gyrase B, leading to the superfamily classification of the GHKL ATPases (Dutta and Inouye, 2000).
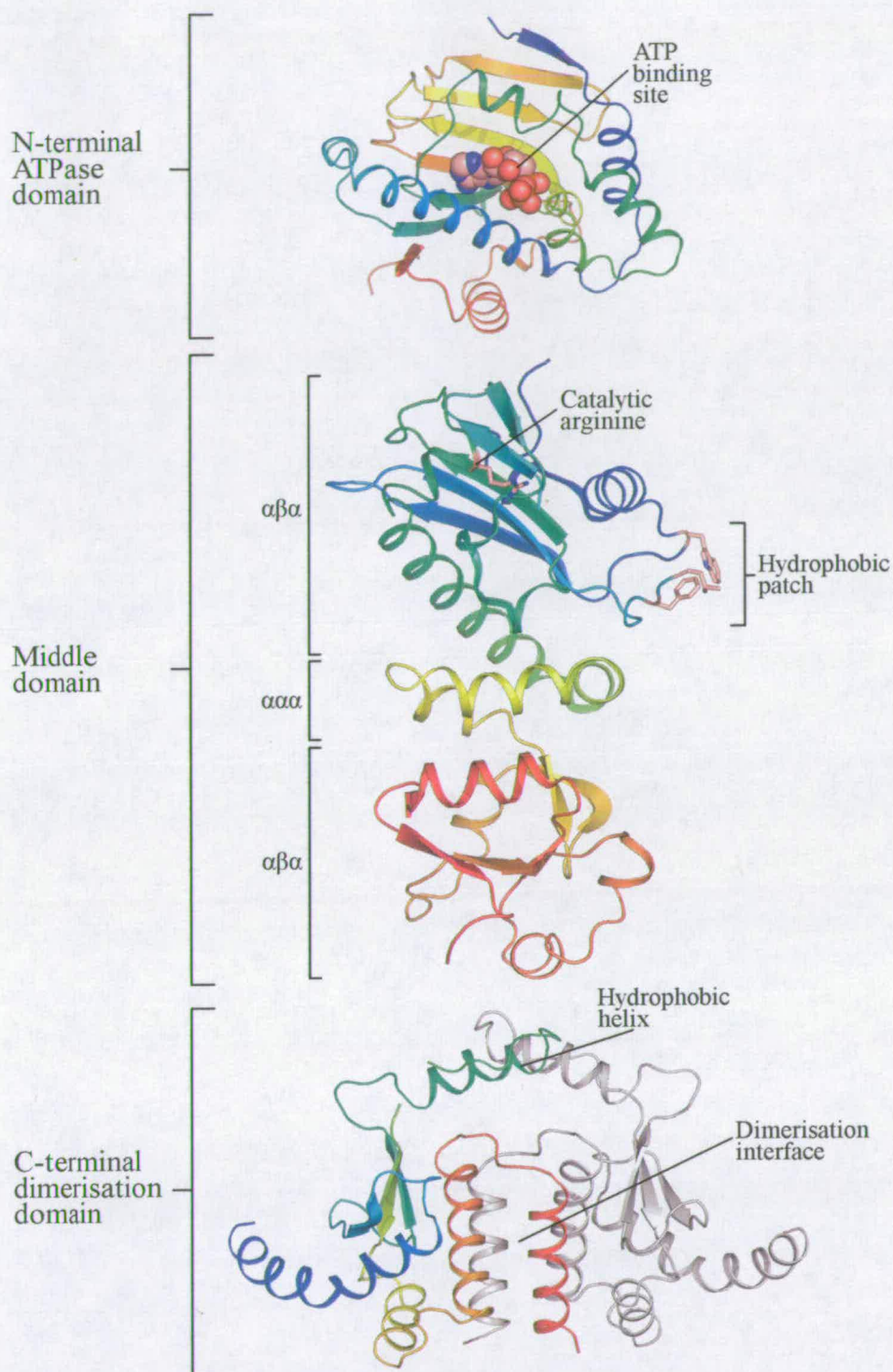
Figure 1-12 Structures of the N-terminal, middle-domain and C-terminal domain from Hsp90. The N-terminal domain forms an atypical ATP binding site called the Bergerat fold common in the GHKL family. The middle-domain is composed of three smaller subdomains, two terminal αβα subdomains and a linking small three helix domain. The N-terminal αβα subdomain contains a conserved catalytic arginine and a cluster of hydrophobic residues involved in substrate binding. The C-terminal domain is responsible for the dimerisation via a four-helix bundle. It also contains an amphiphilic helix which is predicted to be involved in client binding.

The crystal structure of the middle domain revealed three distinct parts - two terminal αβα domains connected by a three helix linker (Figure 1-12) (Meyer et al., 2003). The first subdomain consists of a five-stranded β-sheet sandwiched by an N-terminal small helix and a C-terminal three-turn helix. This leads into a short three helix linker followed by the final structural unit that consists of an unusual αβα fold different to previously described domains of similar architecture (Wegele et al., 2004). Extending the evolutionary relationship of the GHKL superfamily, the larger N-terminal αβα unit is structurally homologous to domains from DNA gyrase and MutL. Comparison with these homologues, which contribute a catalytic lysine that interacts with the γ-phosphate of the N-terminally bound nucleotide, highlighted a conserved arginine essential for ATPase activity (Meyer et al., 2003). Furthermore, a conserved hydrophobic patch and an amphiphilic protrusion have been implicated as a major site in the binding of client proteins.

The C-terminal domain is a small ~12 kDa domain responsible for the dimerisation of Hsp90 (Minami et al., 1994) and also the interaction with the family of TPR domain containing co-chaperones (Owens-Grillo et al., 1996). Despite lower sequence conservation compared to the N-terminal and middle-domains, structures from *E. coli* HtpG (Harris et al., 2004) and yeast Hsp90 (Ali et al., 2006) reveal a homologous mixed α/β domain with a core four-helix bundle, two pairs from each monomer, constituting the dimerisation interface (Figure 1-12). In both structures, a small amphiphilic helix caps the dimerisation interface and is proposed to participate in client binding (Harris et al., 2004). The extreme C-terminal >35 residues constitute a flexible loop, absent in HtpG, terminating in the highly conserved MEEVD-COOH motif implicated in the binding of the TPR co-chaperones (see below).

The intrinsic conformational flexibility of Hsp90 has hampered efforts to obtain a full-length atomic resolution structure with structures available only for the individual domains. In 2006, however, two groups presented full-length medium resolution Hsp90 structures; Pearl and colleagues first describing the structure of yeast Hsp90 in complex with co-chaperone p23/Sba1 (Ali et al., 2006) followed by the publication of the full-length *E. coli* homologue HtpG by Agard and co-workers (Shiau et al., 2006) (Figure 1-13). These structures not only provide direct evidence regarding the domain organisation, confirming predictions of parallel elongated dimers with proximal N-terminal domains, but also give valuable insight into the massive conformational changes involved in the chaperone cycle.
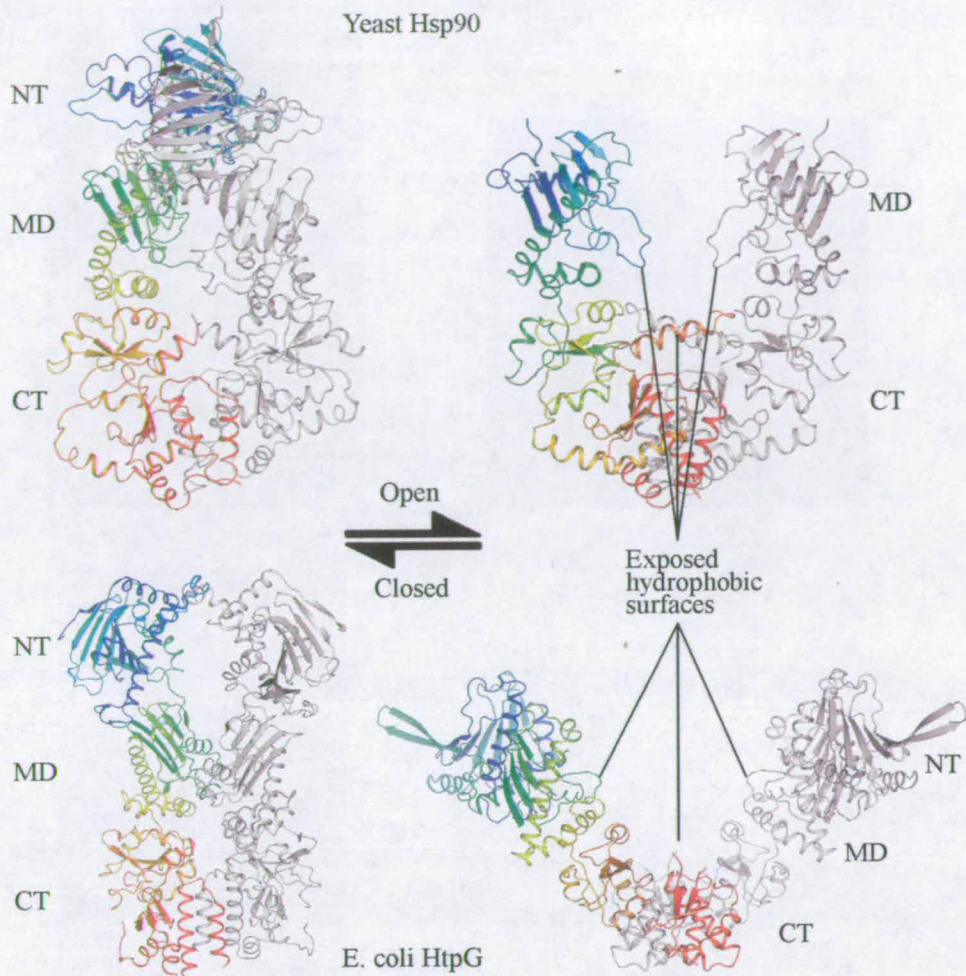
Figure 1-13 Structures of full-length Hsp90 from yeast and *E. coli* in ATP bound closed conformations and nucleotide free open conformations. In both structures, dramatic conformational changes are witnessed between the two states with the exposure of a series of hydrophobic surfaces implicated in client binding in the open state.

1.2.2. Structural aspects of the ATPase and chaperone cycle

The ATPase activity of the N-terminal domain is vital for Hsp90 function (Obermann et al., 1998; Panaretou et al., 1998). Hsp90 has a weak affinity for ATP with dissociation constants in the high micromolar range. It also possesses a very weak intrinsic ATPase activity with yeast Hsp90 hydrolysing one ATP approximately every three minutes and the human homologue about 10 times slower (Wegele et al., 2004). The kinetics are, in part, due to slow conformational changes that occur upon nucleotide binding. In the nucleotide free state, Hsp90 exists in an open, flexible conformation (Figure 1-14). ATP binding, along with client binding, triggers a major rearrangement leading to a more compact structure and also dimerisation of the N-terminal domains in a so called "molecular clamp" mechanism (Figure

1-14) (Prodromou et al., 2000; Prodromou et al., 1997). A domain-swap involving a single strand between the dimerised N-terminal domains is necessary to form a catalytically active conformation (Prodromou et al., 1997). The closed conformation of Hsp90 is reliant on the presence of a γ-phosphate and hydrolysis destabilises the N-terminal dimerisation interface leading to the relaxation to the open state and client release.
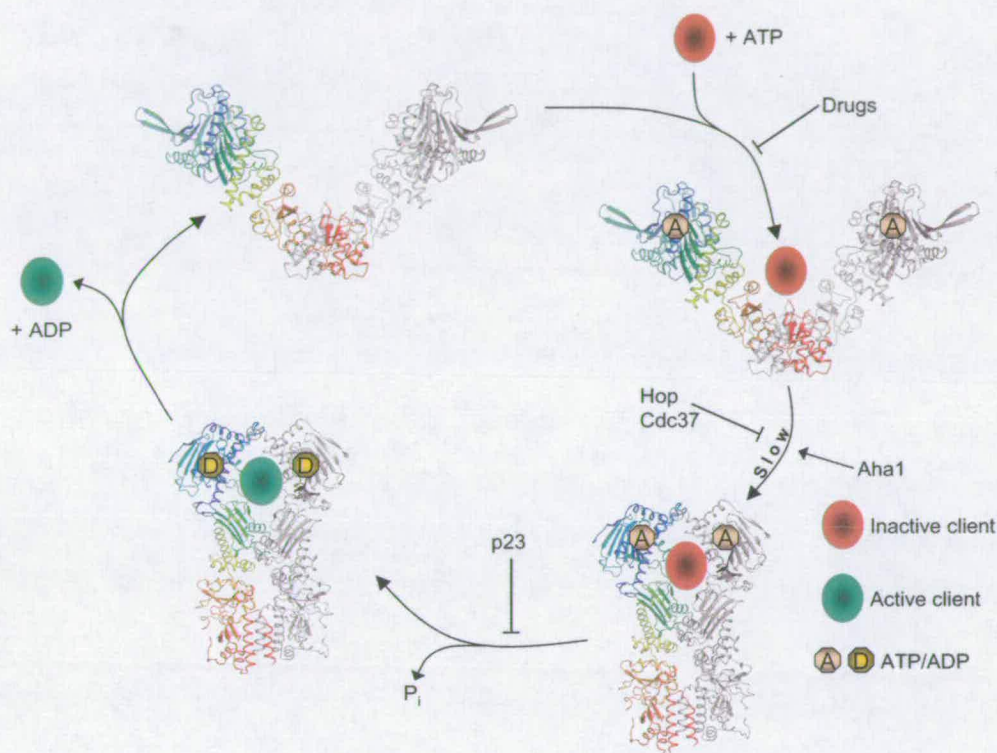
Figure 1-14 Hsp90 chaperone cycle. In the nucleotide free state Hsp90 exists in a highly flexible open conformation primed for client binding. ATP binding induces a conformational change resulting in dimerisation of the N-terminal domains. This step is slow and accounts for the low rate of ATP hydrolysis. N-terminal dimerisation is also inhibited by co-chaperones Hop and Cdc37 and facilitated by Aha1. ATP hydrolysis results in client activation and, due to destabilisation of the N-terminal dimerisation interface, client release. Client protein is shown purely for illustrative purposes.

The client binding site on Hsp90 remains poorly defined. Chaperones, as discussed with Hsp70, bind proteins via hydrophobic binding interfaces. Both complete open-state Hsp90 structures show exposed hydrophobic patches projecting into the cavity between the two monomers. These hydrophobic segments, from the middle and C-terminal domains, have been proposed to be client binding sites (Figure 1-13) (Harris et al., 2004; Meyer et al., 2003; Shiau et al., 2006). In both yeast and *E. coli*, ATP binding induces larges rearrangements in

these subdomains, changes that may be passed onto a bound client protein (Ali et al., 2006; Shiau et al., 2006).

The first insight into the structure of a client loaded Hsp90 complex comes from the electron microscopy single-particle reconstruction of Hsp90 complexed with kinase Cdk4 and co-chaperone Cdc37 (Vaughan et al., 2006). Although low resolution, this shows an asymmetric complex with Cdk4 bound to the N-terminal and middle-domain of one monomer, in the region of the middle-domain hydrophobic loop, and Cdc37 bound to the N-terminal domain of the other monomer, blocking the ATP induced dimerisation (Figure 1-15). The complex is proposed to represent an early complex in the chaperone cycle, prior to dissociation of Cdc37 and dimerisation of the N-terminal domains.
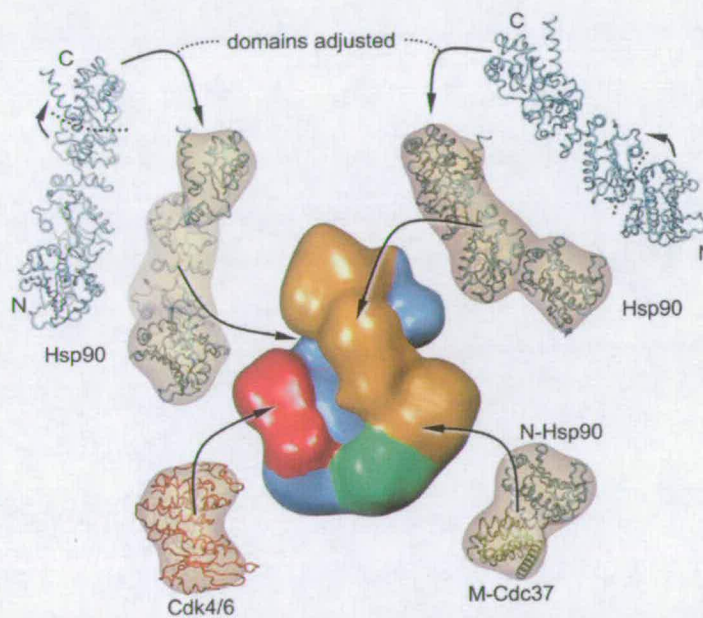
Figure 1-15 EM model of a Hsp90-Cdc37-Cdk4 complex. Cdc37 (green) and Cdk4 (red) bind asymmetrically with Cdc37 binding the N-terminal domain of one Hsp90 monomer (orange) and Cdk4 binding around the middle-domain of the other (blue). This is thought to represent an intermediate client complex, prior to Cdc37 dissociation and ATP hydrolysis. Figure taken from Vaughan et al., 2006. NB. Hsp90 inverted 180°, with C-terminal at top, compared to other figures.

1.2.3. Hsp90 co-chaperones

In eukaryotes, the function of Hsp90 is tightly regulated and fine-tuned by a multitude of co-chaperones. These can be loosely divided into those that contain TPR domains and those that do not. Similar to Hsp70, the largest family of these is the TPR domain containing co-chaperones, which are discussed below (section 1.4). Co-chaperones not containing TPR

domains include p23, Cdc37 and Aha1, which all regulate the Hsp90 chaperone cycle by controlling its rate of ATP turnover (Figure 1-14).

(a)

p23
- Binds dimerised N-terminal
- Slows ATPase

Hsp90-NT   ATP

(b)

Cdc37
- Binds open N-terminal
- Inhibits ATPase

(c) Catalytic loop

Hsp90-MD
- Binds middle-domain
- Stimulates ATPase
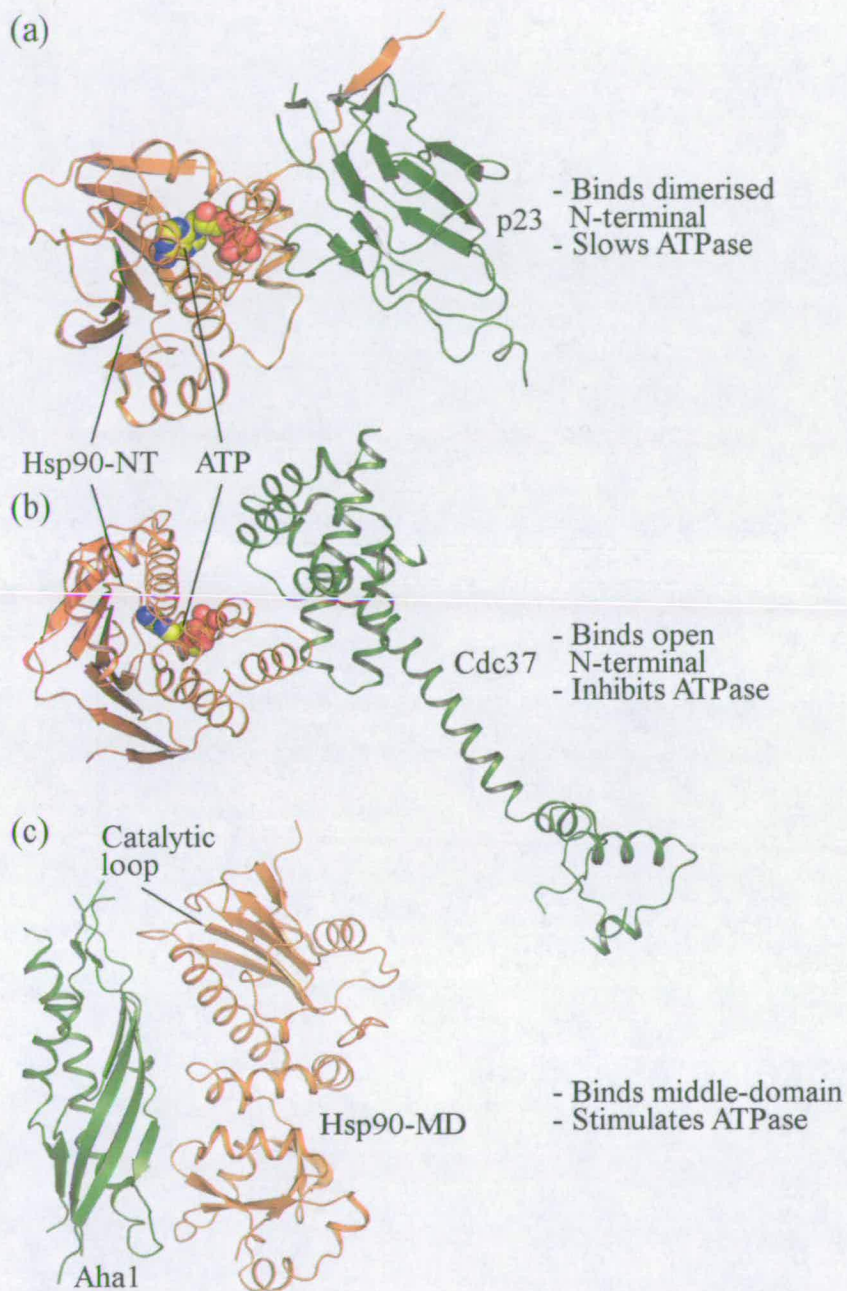
Aha1

Figure 1-16 The interaction of Hsp90 and co-chaperones p23, Cdc37 and Aha1. (a) p23 binds to the N-terminal domain in the ATP bound dimerised state slowing the rate of ATP turnover (Ali et al., 2006). (b) Cdc37 binds to the monomeric ATP bound N-terminal inhibiting dimerisation and consequently ATP hydrolysis (Roe et al., 2004). (c) Aha1 binds to the middle-domain and stimulates ATP hydrolysis (Meyer et al., 2004). It is thought to achieve this by positioning the conserved catalytic loop in a catalytically favourable conformation.

p23 is a small protein with chaperone activity (Freeman et al., 1996). It is found in a wide range of Hsp90-client complexes and is proposed to be involved at a late stage in the chaperone cycle, enhancing the release of active client proteins (Young and Hartl, 2000). The crystal structure of the Hsp90-p23 complex shows it binds to the N-terminal of Hsp90, in the ATP induced dimerised conformation (Figure 1-16a), slowing the rate of ATP turnover (Ali et al., 2006).

Cdc37 is required by many kinases for Hsp90 mediated functional maturation (Pearl, 2005). It acts as an adaptor protein, binding kinases via its N-terminal domain and Hsp90 via its C-terminal domain. Cdc37 binds to the N-terminal domain of Hsp90 and the crystal structure of the core Cdc37-Hsp90 interacting complex shows the interaction to be located around the lid to the ATP binding site (Figure 1-16b) (Roe et al., 2004). This arrests the Hsp90 ATPase cycle by blocking the ATP triggered N-terminal dimerisation (Roe et al., 2004; Siligardi et al., 2002).

Aha1 is involved in client activation and was shown to stimulate the Hsp90 ATPase rate to about 12 times the basal level (Wegele et al., 2004). Biochemical experiments showed N-terminal domain of Aha1 interacted with the middle-domain of Hsp90 (Meyer et al., 2003). This was confirmed in a crystal structure of the complex (Figure 1-16c), demonstrating that the N-terminal domain of Aha1 increases the ATPase rate by facilitating a change of the middle-domain catalytic loop into a more active conformation (Meyer et al., 2004).

1.3. The Hsp70/Hsp90 multichaperone machinery

Far from occurring in isolation, there is also a great deal of communication between the Hsp70 and Hsp90 pathways. A sub-set of Hsp90 clients are first processed by the Hsp70 chaperone pathway prior to passing to the Hsp90 pathway for final maturation. The most extensively studied example of this is in the activation of steroid hormone receptors (SHRs).

SHRs are cytosolic ligand activated transcription factors that translocate to the nucleus upon ligand binding, augmenting or suppressing the expression of certain genes. They require both Hsp70 and Hsp90 in addition to several further co-chaperones in order adopt a functionally active conformation. The identification of participating proteins and knowledge of the precise order of events have progressed significantly over the last 20 years and are now well understood. The first step involves the recognition of newly synthesised SHR by Hsp40 (section 1.1.3.1), which delivers it to Hsp70 (Figure 1-17). This then interacts with an Hsp90-Hop complex, forming a multi-protein intermediate complex consisting of Hsp70,

SHR, Hsp40, Hsp90 and Hop. Hop is an adaptor protein with two TPR-clamp domains (see section 1.4.2.1.) capable of simultaneously binding Hsp70 and Hsp90. Hsp70 and Hop then dissociate and are replaced by p23 (section 1.2.3) and a TPR-clamp domain containing immunophilin such as FKBP52 or Cyp40 (see section 1.4.2.2) to form the final complex. ATP hydrolysis by Hsp90 then triggers the release of the active SHR from the final complex. For a review see Pratt and Toft, 2003.

Although less well understood, Hsp70 and Hsp90 also cooperate in the activation of protein kinases. As discussed in section 1.2.3, Cdc37 is an adaptor protein linking kinase clients to Hsp90. Hsp70, Hop and p23 have been detected in Cdc37 complexes, although little is known about the sequence of chaperone interactions (Pearl and Prodromou, 2006).
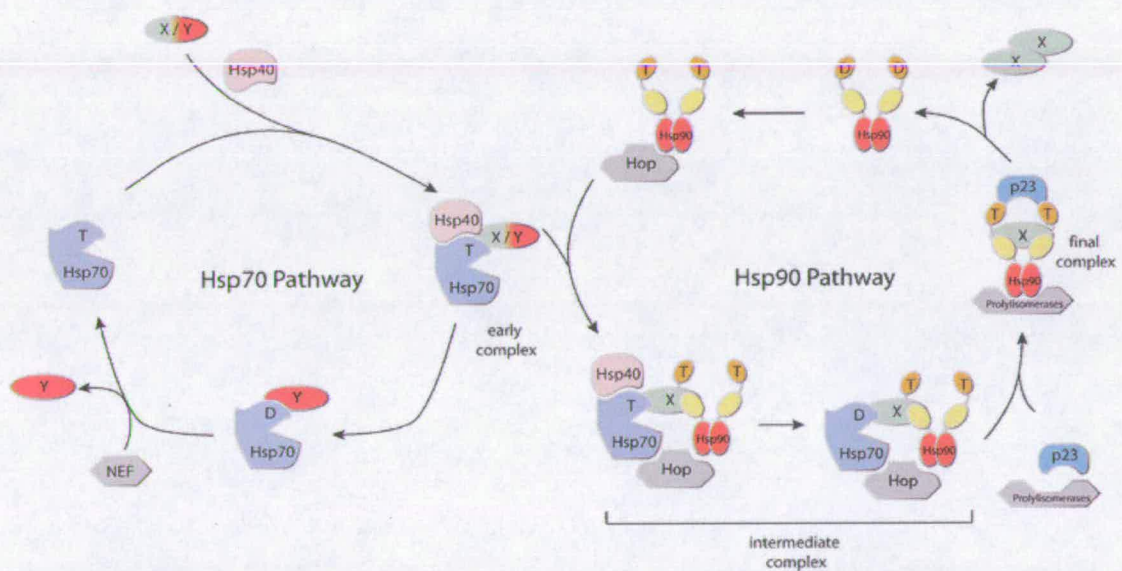
Figure 1-17 The Hsp70 and Hsp90 chaperone pathways. Nascent or unfolded polypeptides are presented to Hsp70 by Hsp40 forming the early complex. Non Hsp90 clients (labelled Y) are processed by the Hsp70 cycle alone. For a sub-set of Hsp90 substrates (labelled X), Hsp90 is recruited to the early complex by Hop, which is capable of interacting with both Hsp70 and Hsp90, forming the intermediate complex. Hsp70 and Hop are replaced by a prolylisomerase, i.e. the immunophilins, and p23 in the final complex. ATP hydrolysis triggers the release of active client. T - ATP bound form, D - ADP bound form. Figure taken from Wegele et al., 2004.

1.4. TPR domain containing co-chaperones

Central to the functional interplay between the Hsp70 and Hsp90 pathways are the diverse family of TPR domain containing co-chaperones. Fascinatingly, many TPR co-chaperones are capable of interacting with both Hsp70 and Hsp90 via their common extreme C-terminal EEVD peptide motifs. Some TPR proteins, such as Hop and Tpr2, contain multiple TPR motifs which may differentially recognise Hsp70 and Hsp90 whilst others contain a single TPR motif capable of binding both chaperones.

1.4.1. The TPR repeat

The tetratricopeptide repeat (TPR) is a short helical repeat belonging to the family of non-globular repeats, widely used to mediate protein-protein interactions, examples of which include ankyrins, armadillo repeats, HEAT repeats, hexapeptide repeats, leucine-rich repeats and WD-40 repeats. TPR proteins are ubiquitously distributed, found in eukaryotes, prokaryotes and archea; with 14,133 TPR containing proteins in the InterPro Tetratricopeptide-like helical family (IPR011990).

The TPR domain was first described in the early 1990s as a modular domain in the cell division cycle proteins in yeast (Hirano et al., 1990; Sikorski et al., 1990). The domain was named due its 34 amino acid periodicity with repeats showing a great diversity in primary sequence. The small domain is all helical, forming a tight helix-loop-helix structure, and is found to occurr in multiple tandem repeats. TPR motifs ranging from 1-16 repeats have been found, with three tandem repeats being the most prevalent (D'Andrea and Regan, 2003). The motif is commonly found in signaling proteins as a mediator of protein-protein interactions.

1.4.1.1. Sequence and structure

The TPR is a 34 amino acid degenerate repeat. Analysis of the first described motifs revealed a largely conserved pattern of amino acid type, size and physiochemical nature (Sikorski et al., 1990). Eight positions, in particular, were found to adhere to a consensus: helix A - W4, L7, G8, Y11; helix B - A20, F24, A27, P32 (Figure 1-18a). Out with these positions there is no sequence conservation amongst the TPR family as a whole; within families, such as the Hsp70/Hsp90 interacting TPR domains, there is conservation of key residues implicated in the interaction (Owens-Grillo et al., 1996).

(a)

(b)

(c)
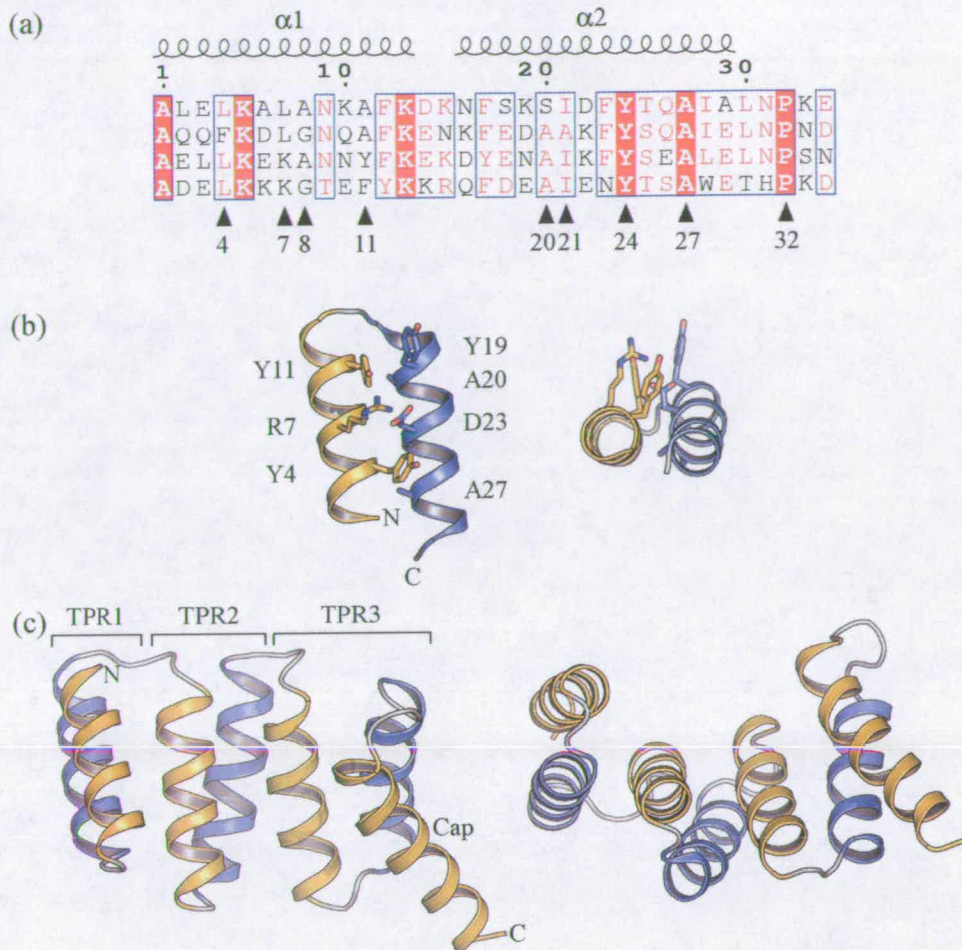
Figure 1-18 (a) Alignment of four 34 residue TPR repeats showing conserved positions. (b) The structure of one TPR domain; conserved small hydrophobic residues are located in positions of close contact and large residues form the helical interface. (c) The crystal structure of the TPR domain from PP5; three consecutive TPR motifs form a concave channel capped by a C-terminal helix.

The TPR domain was predicted to have a high propensity for forming amphiphilic helices forming a coiled-coil like interaction with "knobs into holes" packing (Hirano et al., 1990). Circular dichroism spectroscopy confirmed the high helical content and secondary structure prediction suggested each repeat to form two short stretches of α-helix. These predictions were confirmed with the first crystal structure of a TPR domain from protein phosphatase 5 (PP5) showing a completely novel helical array (Figure 1-18c) (Das et al., 1998). The TPR motif was shown to form a pair of anti-parallel α-helices, with three successive repeats packing in a parallel fashion with helix A_i of one repeat interacting with helix A_{i+1} of the following repeat. The packing generates a right-handed super-helical architecture with a concave channel defined by the side-chains of residues belonging to helix A, and a convex surface formed by residues from helices A and B. This packing has been predicted to form a

right handed super-helix with a helical repeat of approximately seven TPR motifs (Das et al., 1998). The PP5 structure contained an additional C-terminal capping helix; an equivalent is present in nearly all TPR structures solved to date and it has been suggested that this could be important in the solubility and stability of the fold (D'Andrea and Regan, 2003).

The structure of PP5 demonstrated the conserved consensus residues are important for the structural integrity of the fold. The small hydrophobic residues (8, 20 and 27) are located in the positions of the closest contacts between the two helices and the large hydrophobic residues (4, 24) form the intra-domain helical interface (Figure 1-18b).

Name	Domain organisation	Description
Hop	TPR — TPR — TPR	Hsp70/Hsp90 adaptor
Tpr2	TPR — TPR — DnaJ	Hsp70/Hsp90 adaptor
FKBP51	FK1 — FK2 — TPR	SHR maturation
FKBP52	FK1 — FK2 — TPR	SHR maturation
Cyp40	PPIase — TPR	SHR maturation
FKBP38	FK — TPR	FKBP38
AIP	FK — TPR	AhR maturation
PP5	TPR — Phosphatase	Protein phosphatase
Tom70	TPR — TPR — TPR	Mitochondrial import
Tom34	TPR — TPR	Mitochondrial import
Chip	TPR — U-box	Protein degradation
Unc-45	TPR — ARM	Myosin assembly
Wisp39	TPR	Cell-cycle control
Cns-1	TPR	Unknown, essential in yeast
SGT	TPR	Cell-cycle/neurotransmission

Table 1-3 Domain organisation of all TPR domain co-chaperones shown to interact with Hsp70 or Hsp90.

1.4.2. Hsp70/Hsp90 co-chaperones

TPR domain containing proteins are involved in numerous, diverse biological processes including cell cycle regulation, transcriptional control, mitochondrial transport, neurotransmission and protein folding. Biochemical evidence and structures of TPR-peptide complexes demonstrate that the TPR domain is involved in mediating protein-protein interactions and functions as an adaptor or scaffold domain to bring together components of multi-protein molecular complexes. As discussed above, one major class of TPR domain containing proteins is those interacting with chaperones Hsp70 and/or Hsp90. These proteins provide additional functionality and fine-tune the chaperone activity of Hsp70 or Hsp90 by targeting them to specific complexes and regulating their chaperone cycles.

Approximately 15 TPR domain containing proteins, from multiple species, have been shown to interact with Hsp70 and/or Hsp90. These include the Hsp70/Hsp90 adaptor protein Hop, several immunophilins with petidylprolyl isomerase activity, a DnaJ domain containing protein capable of regulating the ATPase activities of both Hsp70 and Hsp90, a protein phosphatase, and several members of the mitochondrial import machinery (Table 1-3). To illustrate the diverse functionality these co-chaperones impose, four examples are discussed in more detail, figure 1-20 provides an overview of these functions.

1.4.2.1. Cross-talk between the Hsp70 and Hsp90 pathways - Hop and TPR2

Hop and TPR2 both contain two TPR-clamp domains and have been shown to interact with and regulate the behaviour of both Hsp70 and Hsp90.

Hop was first identified in yeast and named Sti1 for stress inducible protein 1 (Nicolet and Craig, 1989). As discussed in section 1.3, Hop functions as an adaptor protein in the functional maturation of a sub-set of client protein that are first handled by the Hsp70 pathway prior to final processing by Hsp90 (Figure 1-17). Hop contains nine TPR repeats evenly clustered into three TPR motifs. The first of these, TPR1, is responsible the interaction with Hsp70 and the second, TPR2A is the interaction site for Hsp90 (Figure 1-19) (Scheufler et al., 2000). Hop inhibits the ATPase activity of Hsp90 (Prodromou et al., 1999), maintaining it in a state primed for client binding. Client loaded Hsp70 complexes are recruited to the Hop-Hsp90 complex, facilitating transfer of the substrate to Hsp90. Subsequent dissociation of Hsp70 and Hop ultimately leads to ATP hydrolysis and activation of the client.
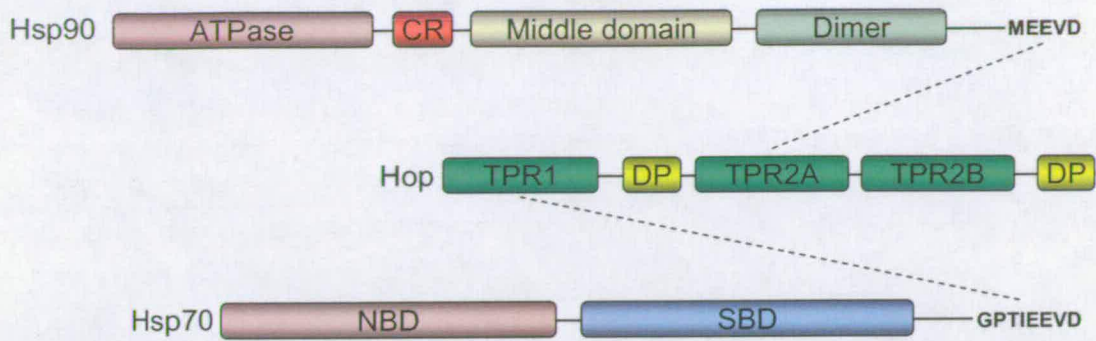
Figure 1-19 Hop can interact with both Hsp70 and Hsp90 via distinct TPR domains. TPR1 interacts with Hsp70 and TPR2A interacts with Hsp90. The ligand for TPR2B is unknown although it has been implicated in the interaction with Hsp70 and Hsp90.

The human protein Tpr2 was identified from yeast two-hybrid screens using the C-terminal domain of Hsp90 as bait (Brychzy et al., 2003). It contains two three-repeat TPR arrays in addition to a J domain homologous to the Hsp40 subdomain. Experiments showed it was capable of interacting with both Hsp70 and Hsp90, stimulating the ATPase activity of Hsp70 via its J domain and inducing the nucleotide-independent release of substrate from Hsp90. Whereas Hop interacts with Hsp70 and Hsp90 by distinct TPR motifs, the TPR domains of Tpr2 seem to bind both chaperones. Based on the results it was postulated that Tpr2 works in the opposite direction of Hop, allowing the recycling of polypeptides not fully folded after a single cycle through the Hsp70/Hsp90 system.

1.4.2.2. Protein transport - the peptidylprolyl isomerase TPR co-chaperones

A sub-set of the large immunophilin family, target proteins of a number of immunosuppressive agents including FK-506 and cyclosporin, contain TPR domains and are capable of binding Hsp70 and/or Hsp90 (Galat, 2003). These include the large immunophilins FKBP51, FKBP52 and cyclophilin 40; proteins characterised by peptidylprolyl isomerase (PPIase) activity. In addition, the protein phosphatase PP5 binds FK506 with low affinity and shares some sequence homology with the FKBP PPIase domain (Silverstein et al., 1997). These proteins play a critical role in SHR assembly (see section 1.3) with different immunophilins having preference for different SHRs.
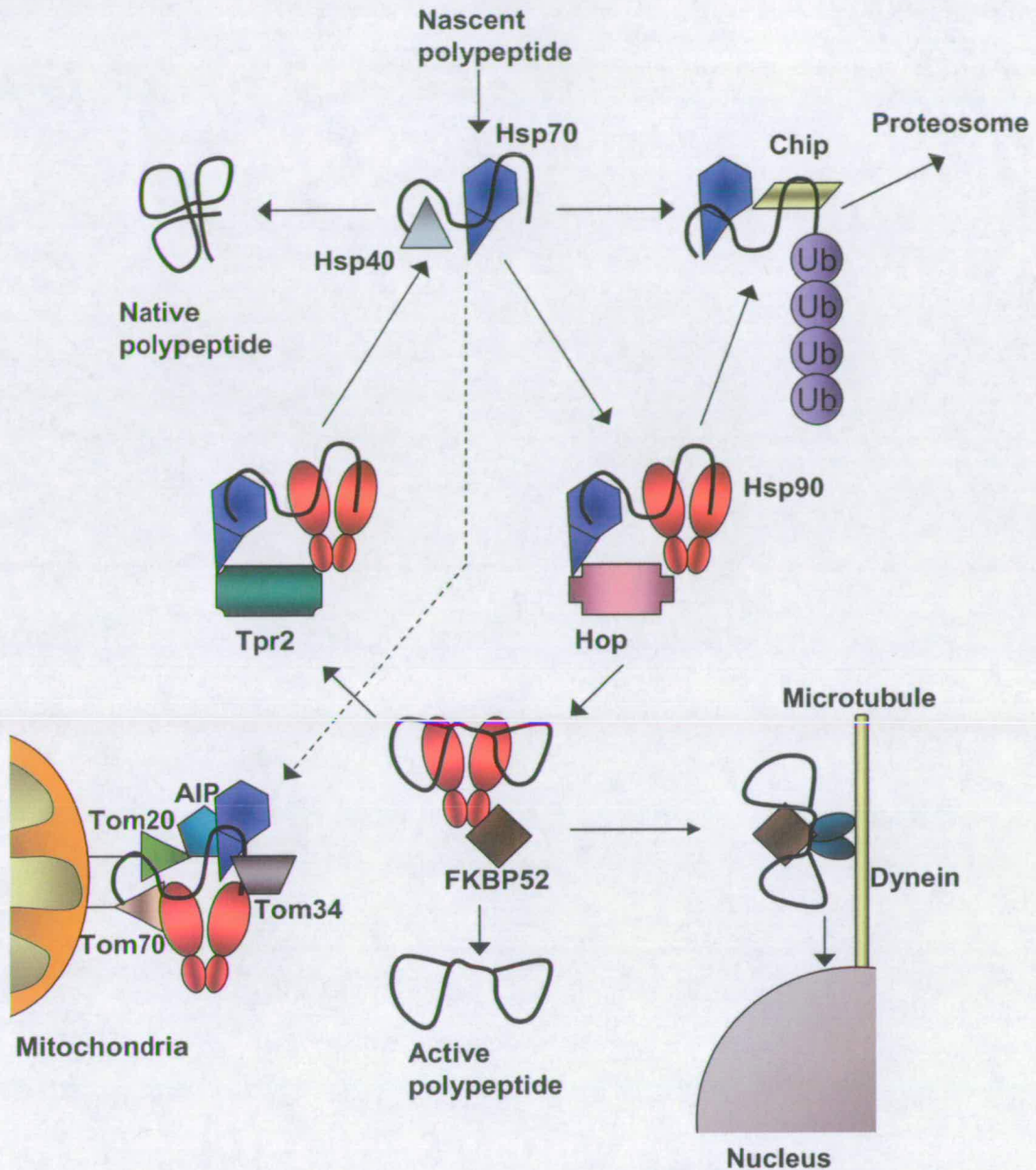
Figure 1-20 Multiple functions of TPR domain containing co-chaperones. TPR domain co-chaperones provide specificity and selectivity to the Hsp70/Hsp90 chaperone machinery. Examples discussed in section 1.4.2 include the cycling of proteins between the Hsp70 and Hsp90 systems by adaptor proteins Hop, involved in the transfer of clients from Hsp70 to Hsp90 and Tpr2, involved in the recycling of incompletely folded proteins back to Hsp70. Client proteins are matured in complexes containing imunophilins e.g. FKBP51, FKBP52 (shown), Cyp40, AIP and PP5. These have been shown to prime SHRs for ligand binding and also to bind the motor protein dynein and facilitate nuclear import. Several TPR proteins are also involved in import of nuclear encoded mitochondrial proteins including TOM34, TOM70 and AIP. Finally, both Hsp70 and Hsp90 also provide a link to protein degradation via the E3-ligase Chip. Figure adapted from Young et al., 2004.

In addition to selectively modulating hormone-binding affinity of some steroid receptors (Denny et al., 2000; Reynolds et al., 1999; Scammell et al., 2001), the TPR-containing immunophilins, excluding FKBP51, have been shown to interact with cytoplasmic dynein, one of the major microtubule-associated motor proteins (Pratt et al., 2004). The interaction with dynein is mediated by PPIase domain but is independent of its enzymatic activity. These co-chaperones are thought to link Hsp90-bound steroid hormone receptors to the microtubule cytoskeleton, facilitating import to the nucleus. In addition to SHRs, Hsp90 has also been proposed to regulate the trafficking of p53 in a similar manner (Pratt et al., 2004).

1.4.2.3. Mitochondrial protein import

Mitochondria contain roughly 1000 different proteins, only eight of which are encoded by the mitochondrial genome. The remaining 99% are nuclear encoded and synthesised by cytosolic ribosomes as preproteins with positively charged N-terminal or internal targeting sequences for import into the mitochondria. The translocation machinery of the mitochondrial outer membrane (TOM) includes the Tom70, Tom22 and Tom20 preprotein receptors (for review see van der Laan et al., 2006). Before import, many mitochondrial preproteins are bound by Hsp70 and/or Hsp90, which are proposed to maintain them in an unfolded state. The preprotein is delivered to the TOM, docking with Tom70 or Tom20 prior to passage though the Tom20/Tom22/Tom40 pore.

Several of the constituents of the TOM contain TPR domains, namely Tom70, Tom20 and Tom34. Tom70 contains nine TPR domains arranged as three three-repeat arrays (Table 1-3). Additionally, an N-terminal transmembrane domain serves to anchor it to the mitochondrial outer membrane. The first TPR domain is capable of binding Hsp70 and Hsp90 (only Hsp70 in yeast) and the second TPR domain recognises internal preprotein sequences (Young et al., 2003). Upon interaction of chaperone-preprotein complexes with Tom70, the ATPase activities of Hsp70 and Hsp90 have been proposed to facilitate protein translocation through the mitochondrial pore (Young et al., 2003).

The TPR domain of Tom20 is incapable of binding Hsp70 or Hsp90; instead it is the target for N-terminal preprotein sequences. Intriguingly, the C-terminal region of Tom20, ending EDDVE, was found to interact with arylhydrocarbon receptor interacting protein (AIP) (Yano et al., 2003), a small TPR containing immunophilin constituent of arylhydrocarbon receptor-Hsp90 complexes (Bell and Poland, 2000). Furthermore, AIP was found to form complexes with preproteins and Hsp70, and it was hypothesised that Hsp70 and AIP existed as a chaperone complex with preprotein in the cytosol. Tom20 would then displace Hsp70

from AIP by competition of their C-terminal peptides and the preprotein would be transferred from AIP to Tom20.

The recently identified mammalian Tom34 has also been proposed to function in the import of mitochondrial preproteins (Chewawiwat et al., 1999). It contains two TPR-clamp domains and has been shown to interact with Hsp90 (Young et al., 1998). However, it is suggested that Tom34 is cytosolic and functions to maintain preproteins in an unfolded state suitable for mitochondrial import.

1.4.2.4. A link between protein folding and degradation

In order to maintain accurate protein quality control the cell has to maintain a tight balance between protein folding and degradation (see section 1.5 for what happens when this goes wrong). One protein central in determining the cellular fate of proteins processed by the Hsp70/Hsp90 chaperone machinery is Chip (C-terminus of Hsp70 interacting protein). Chip is a 35 kDa protein consisting of a TPR-clamp motif capable of interacting with both Hsp70 and Hsp90 and a U-box domain that functions to target proteins for ubiquitination and subsequent proteasome-dependent degradation (Murata et al., 2001). Chip has inhibitory effects on the chaperone cycles of both Hsp70 and Hsp90 leading to destabilisation of chaperone-substrate complexes and subsequent targeting for proteasomal degradation (McDonough and Patterson, 2003). This is thought to be executed by a Chip-Hsp70-client complex and Hsp90 clients must be transferred to Hsp70 first. The U-box domain of Chip has intrinsic E3 ubiquitin ligase activity and catalyses the conjugation of ubiquitin chains to target proteins (Xu et al., 2002). In addition, Chip has been proposed to escort proteins actively to the proteasome with evidence of a direct interaction between Chip and proteasomal subunits (Connell et al., 2001).

1.4.3. Mechanism of binding

TPR domains have evolved to interact with Hsp70 and/or Hsp90 via an interaction with the extreme C-terminal which, fascinatingly, is common between Hsp70 and Hsp90; both eukaryotic cytosolic proteins terminate in a flexible loop with an extreme EEVD-COOH motif. As a result of this converged mechanism of binding, many TPR co-chaperones are capable of interacting with both Hsp70 and Hsp90. Comparative sequence analysis of Hsp70/Hsp90 binding TPR motifs revealed conservation of a number of hydrophobic and charged amino acids on the concave surface of the domain. Mutation of some of these residues diminished or abrogated binding in the isolated TPR domain from PP5 (Russell et

al., 1999). Further, mutation of the EEVD motif also disrupted binding in several co-chaperone interactions (Liu et al., 1999; Russell et al., 1999).
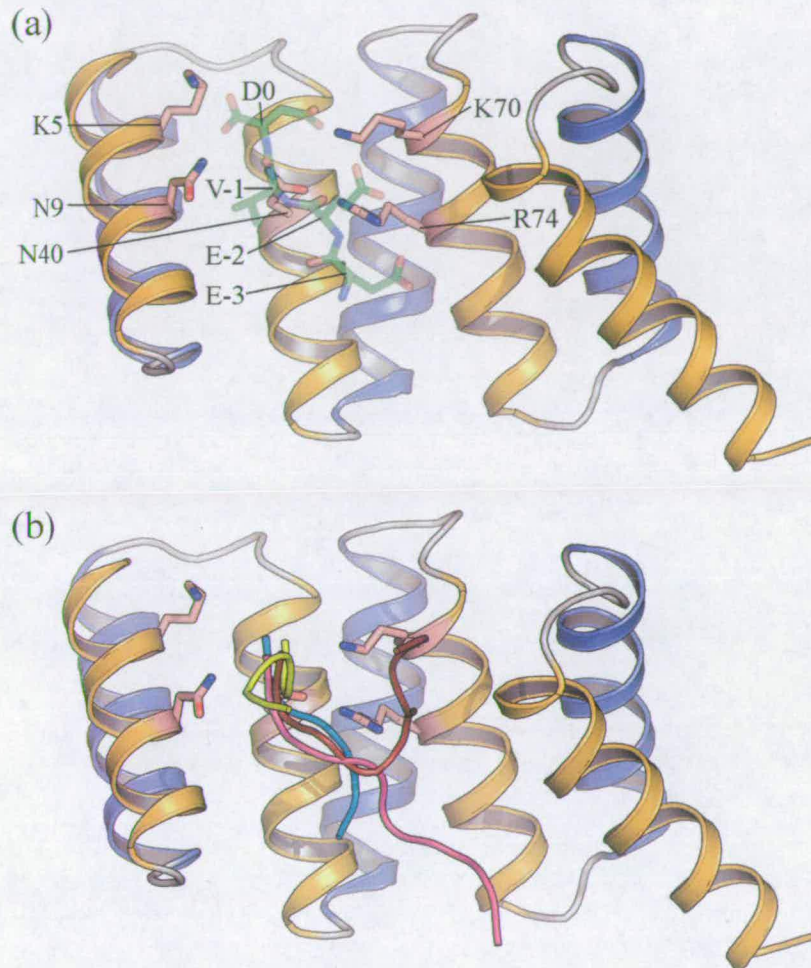
Figure 1-21 The carboxylate clamp binding mechanism. (a) Conserved positively charged residues on the concave surface of the TPR domain form a network of charge-charge interactions with the negatively charged substrate residues. (b) Although the carboxylate clamp provides a general anchor, peptide-TPR complexes from several different TPR domains show a range of binding orientations. hopTPR2A-MEEVD - cyan, hopTPR1-GPTIEEVD - purple, PP5-MEEVD - yellow and Chip-MEEVD - red.

Definitive evidence for this interaction came with the crystal structures of two TPR domains from Hop in complex with the C-terminal peptides from Hsp70 and Hsp90 (Scheufler et al., 2000); TPR1 in complex with the Hsp70 peptide GPTIEEVD and TPR2A in complex with the Hsp90 peptide MEEVD. These show a common network of electrostatic interactions between the conserved positively charged residues on the TPR surface and the EEVD peptide motifs in a manner termed the "two-carboxylate clamp" (Figure 1-21a). Peptide residues Asp^0 and Val^{-1} were shown to act as a general anchor with the highly conserved glutamates critical for Hsp90 binding but less so for the interaction with Hsp70 (Brinker et al., 2002). Hydrophobic sequences upstream of the EEVD and specific residues within the TPR groove were found to influence the discrimination between the TPR domains.

Several further structures of TPR-Hsp70/90 peptide complexes have since been solved confirming the general two-carboxylate binding mechanism (Cliff et al., 2006; Wu et al., 2004a; Zhang et al., 2005). These structures, however, also illustrate the diversity of evolved binding modes with a great deal of variation in the orientation of the peptides downstream from the VD anchor (Figure 1-21b).

Disease	Proteins	Cause
Alzheimer's disease	β-amyloid protein/tau	Aggregation
Parkinson's disease	α-Synuclein	Aggregation
Huntington's disease	Huntingtin	Aggregation
Creutzfeld Jakob disease	Prion protein (PrP)	Aggregation
Cystic fibrosis	CFTR protein	Trafficking
Cancer	p53	Trafficking
Sickle cell anemia	Hemoglobin	Aggregation
Gaucher's disease	β-glucosidase	Trafficking
Nephrogenic diabetes insipidus	V2 vasopressin receptor	Trafficking
Transthyretin amyloidoses	Transthyretin	Aggregation
Retinitis pigmentosa	Rhodopsin	Trafficking
$αB_{1B}$-Antitrypsin	$αB_{1B}$-Antitrypsin	Trafficking/aggregation
Fabry	α-Galactosidase	Trafficking

Table 1-4 Diseases and specific proteins associated with protein misfolding and aggregation.

1.5.   Protein folding and disease

The importance of the fidelity of the protein folding system is illustrated by the ever growing list of disorders resulting from aberrant folding reactions (Table 1-4) and protein folding has been implicated in up to half of human diseases (Bradbury, 2003). Not only can misfolded proteins lose their function, such as CFTR in cystic fibrosis which is prematurely targeted for proteasomal degradation, but may also form toxic species, including oligomers or larger

aggregates characteristic of amyloidosis conditions such as many neurodegenerative disorders. The involvement of protein chaperones in these disorders is becoming increasingly evident and activation or inhibition of chaperone pathways are proving effective targets for therapeutic intervention. In addition, protein misfolding is responsible for many cancers and Hsp90, required by many mutated and misfolded clients for aberrant function, is proving an effective target for treatment for multiple malignancies.

1.5.1. Neurodegenerative disorders

Many neurodegenerative disorders are due to protein misfolding and a common feature is the intra- or extracellular accumulation of misfolded, aggregated or ubiquitinated proteins (Berke and Paulson, 2003). Common conditions include Alzheimer's, Parkinson's and Huntington's diseases (for an overview see Chaudhuri and Paul, 2006).

Alzheimer's disease is an age-onset progressive degenerative disease of the brain causing memory loss and impaired behaviour. It is characterised by the extracellular deposition of β-amyloid protein (Aβ) and neurofibrillary tangles (NFT) in the brain and the intraneuronal accumulation of Aβ-42 and the microtubule associated protein tau. Hsp70 and Hsp90 have been implicated in maintaining Aβ-42 and tau solubility and suppressing aggregation (Ansar et al., 2007; Sahara et al., 2005).

Parkinson's disease is characterised by muscular rigidity, tremor and slowing of physical movement. The symptoms result from the loss of dopaminergic neurons in the substantia nigra due to the accumulation of intracellular inclusion bodies called Lewy bodies (Lansbury and Brice, 2002). These are commonly composed of aggregated α-synuclein leading to mitochondrial dysfunction, oxidative stress and caspase degradation (Chaudhuri and Paul, 2006). Genetic polymorphisms in Hsp70 have been associated with risk (Wu et al., 2004b) and up-regulation of both Hsp70 and Hsp90 has been associated with protection against Lewy body toxicity (Auluck et al., 2002; McLean et al., 2001).

Polyglutamine diseases are caused by CAG trinucleotide repeats that result in polyglutamine tracts and proteins likely to misfold. Huntington's disease is the most common polyglutamine disease, caused by a mutant version of the protein huntingtin rendering the protein aggregation-prone (Hague et al., 2005). Overexpression of Hsp70 in flies and mouse models has been shown to increase resistance to polyglutamine-induced toxicity (Dedeoglu et al., 2002; Shulman et al., 2003).

1.5.2. Cystic fibrosis

Cystic fibrosis is a hereditary disorder that affects about 0.05% of Caucasians. It is caused by mutations in the CFTR protein and characterised by thick mucous secretions in the lung and intestines (Welch, 2004). CFTR is a membrane protein possessing 12 transmembrane domains, two nucleotide-binding domains and a highly charged regulatory hydrophilic domain. CFTR is related to other adenine nucleotide-binding cassette (ABC) transporters and CFTR has been shown to function as a cAMP-regulated chloride channel in epithelial cells (Welsh et al., 1992). Although a large number of mutations have been found, a deletion of one phenylalanine (ΔF508) is attributed to 70% of cystic fibrosis cases (Riordan et al., 1989). The ΔF508 allele of CFTR has been confirmed as a trafficking mutation that blocks maturation of the protein in the ER and targets it for premature proteolysis (Kunzelmann and Nitschke, 2000).

Molecular chaperones Hsp70 and Hsp90 have both been found to interact with wild-type and mutant nascent CFTR. This interaction is found to be transient with the wild-type protein but more long-lived with the mutant and it has been proposed that this targets the protein for premature degradation (Skach, 2006). The correct processing of CFTR is temperature sensitive with protein targeted for degradation at 37 °C but a portion reaching the native state at <30 °C suggesting the extended interaction is due to misfolding (Denning et al., 1992). Further, sodium 4-phenylbutyrate (4PBA) has been shown to inhibit the interaction of ΔF508-CFTR with Hsp70, allowing mutant CFTR to escape targeting for degradation (Rubenstein and Zeitlin, 2000).

1.5.3. Hsp90 and cancer

Hsp90 is involved in maintaining the conformation, stability, activity and cellular localisation of multiple signal transduction proteins, many of which are oncoproteins or tumour suppressors e.g. v-Src, c-Erb2, Raf-1, Akt, Bcr-Abl, and p53 (Wegele et al., 2004). Mutations in some of these proteins result in constitutively active but conformationally unstable mutants that require Hsp90 to maintain their aberrant function. Small-molecule inhibitors of Hsp90, such as the unrelated geldanamycin and radicicol, disrupt Hsp90 chaperone complexes and have been shown to be highly effective in reducing cellular levels of oncogenic client proteins via degradation by the proteasomal pathway (Sharp and Workman, 2006). One geldanamycin derivative, 12-AAG, proved effective in phase I clinical trials demonstrating Hsp90 as a valid pharmacological target and showed clinical efficacy in patients with melanoma, breast and prostate cancers (Sharp and Workman, 2006).

1.6.   *C. elegans* as a model system

C. elegans is a free-living nematode, about 1 mm in length, which lives in temperate soil environments. Its use as a model organism in molecular and developmental biology research began in the 1960s, pioneered by Sydney Brenner (Brenner, 1974). *C. elegans* represents a genetically tractable multicellular eukaryotic organism simple enough to be studied in great detail. It is easy to culture and manipulate and it shares many organ systems with higher animals including a nervous system. Wild-type individuals contain a constant 959 cells and the complete cell lineage, depicting which cells are derived from which, was completed in the 1980s by John Sulston. The *C. elegans* genome, first published in 1998 (Consortium, 1998), represented the first sequenced genome of a multicellular organism. *C. elegans* homologues have been identified for 60-80% of human genes making it an attractive organism in the study of human disease. Significant biomedical discoveries in Alzheimer's disease, Type II diabetes and depression have been enabled by *C. elegans* research (for review of *C. elegans* as a model system see Kaletta and Hengartner, 2006). Facilitating this, *C. elegans* are especially amenable to disruption of the function of specific genes by RNA interference allowing the study of knock-down phenotypes (Tabara et al., 1998).

C. elegans has been used a model system of stress responses and diseases of protein misfolding. In particular, model systems have been established for the study of Alzheimer's, Parkinson's and Huntington's disease (Kaletta and Hengartner, 2006). *C. elegans* contains homologues for Hsp70 and Hsp90 in addition to many regulatory co-chaperones. The *C. elegans* Hsp70 multigene family consists of 11 isoforms and includes orthologues for the Hsp70 family members Hsp70, Hsc70, Grp78/BiP and Grp75/mtHsp70 (Figure 1-22) (Heschl and Baillie, 1989; Heschl and Baillie, 1990). *C. elegans* has one cytosolic Hsp90, Daf-21, in addition to homologues for mitochondrial located TRAP1/mtHsp90 and Grp94 of the endoplasmic reticulum (Figure 1-23). Daf-21 is abundantly expressed, representing 3% of cDNA clones of a library isolated from mixed stage worms (Birnby et al., 2000) and, as with other eukaryotes, is essential for survival.

```
                                        1        10        20        30        40
hsp-1              ...................................MSKHNAVGIDLGTTYSCVGVFMHGKVEIIANDQGNKTTPSY
hsp-70             ...................................MSTCKAIGIDLGTTYSCVGIYQNGHVEILANSEGNKTTPSY
F44E5.4            ...................................MSSCKAIGIDLGTTYSCVGIYQNGKVEILANSEGNKTTPSY
hsp-3(Grp78/BiP)   MKTLFLLGLIALSAVSVYCEEEEKTEKKETKYGTIIGIDLGTTYSCVGVFKNGRVEIIANDQGNRITPSY
hsp-4(Grp78/BiP)   .MKVFSLILIAFVAN.AYCDEGASTEK.EEKMGTIIGIDLGTTYSCVGVFKNGRVEIIANDQGNRITPSY
hsp-6(Grp75/mtHsp70) ..MLSARSFLSSARTIARSSLMSARSLSDKPKGHVIGIDLGTTNSCVSIMEGKTPKVIENAEGVRTTPST
F11F1.1            ..............................MSIGIDLGTTFSCVAYYQNGGVHILENENGCRITPSV

                              50        60        70        80        90       100
hsp-1              VAFTDT..ERLIGDAAKNQVAMNPHNTVFDAKRLIGRKFDDPAVQSDMKHWPFKVISAEGAKPKVQVEYK
hsp-70             VAFTDT..ERLVGDAAKDQAARNPENTVFDAKRLIGRRFDEETVQSDIKHWPFTVKGKQG.KPVVEVEVK
F44E5.4            VAFTDT..ERLVGDAAKDQAARNPENTVFDAKRLIGRRFDESTVQSDCKHWPFGVKGKQG.KPVIEVEVK
hsp-3(Grp78/BiP)   VAFSGDQGDRLIGDAAKNQLTINPENTIFDAKRLIGRDYNDKTVQADIKHWPFKVIDKSN.KPSVEVKVG
hsp-4(Grp78/BiP)   VAFSGDQGERLIGDAAKNQLTINPENTIFDAKRLIGRFYNEKTVQQDIKHWPFKIVDKSN.KPNVEVKVG
hsp-6(Grp75/mtHsp70) VAFTAD.GERLVGAPAKRQAVTNSANTLFATKRLIGRRYEDPEVQKDLKVVPYKIVKASN....GDAWVE
F11F1.1            LAVGED.GDLLIGQHAKDVIG.KATNSLFDVKRIIGRRYDDVLLQRDMPLWPFKVEKDDNGIPYLEIHKN

                       110       120       130       140       150       160       170
hsp-1              GENKIFTPEEISSMVLLKMKETAEAFLGTTVKDAVVTVPAYFNDSQRQATKDAGAIAGLNVLRIINEPTA
hsp-70             GEKREFNAEEISAMVLQKMKETAEAVLGHSVKDAVITVPAYFNDSQRQATKDAATIAGLNAIRIINEPTA
F44E5.4            GEKRQFNPEEISAMVLQKMKETAESYLGKEVKDAVITVPAYFNDSQRQATKDAATIAGLNAIRIINEPTA
hsp-3(Grp78/BiP)   SDNKQFTPEEVSAMVLTKMKQIAESYLGHEVKNAVVTVPAYFNDAQRQATKDAGTIAGLNVVRIINEPTA
hsp-4(Grp78/BiP)   SETKQFTPEEVSAMVLTKMKQIAESYLGHEVTVNAVVTVPAYFNDAQKQATKDAGSIAGLNVVRIINEPTA
hsp-6(Grp75/mtHsp70) AKGKVYSPSQVGAFVLMKMKETAESYLGTTVNAVVTVPAYFNATQRRATEEAAKIAGLNVLRIINEPTA
F11F1.1            EKPLKFSAVTVSSLILRCLKYNAERKLGLEVKSAVITVPAYFNATQRRATEEAAKIAGLKVLRILNEPTA

                       180       190       200       210       220       230       240
hsp-1              AAIAYGLDKKGHGERNVLIFDLGGGTFDVSILTIEDG.IFEVKSTAGDTHLGGEDFDNRMVNHFCAEFKR
hsp-70             AALAYGLDKGITDEKNILIFDLGGGTFDVSILSIAEGS.IFEVKSTAGDTHLGGEDFDSRVLQHFMTEFKR
F44E5.4            AALAYGLDKGITDEKNILIFDLGGGTFDVSILSIAEGS.IFEVKSTAGDTHLGGEDFDQRMLQHFMNEFKR
hsp-3(Grp78/BiP)   AAIAYGLDKK.DGERNILVFDLGGGTFDVSMLTIDNG.VFEVLATNGDTHLGGEDFDQRVMEYFIKLYKK
hsp-4(Grp78/BiP)   AAIAYGLDKK.DGERNILVFDLGGGTFDVSLLTIDSG.VFEVLATNGDTHLGGEDFDQRVMEYFIKLYKK
hsp-6(Grp75/mtHsp70) AALAYGLDKD.AGDKIIAVYDLGGGTFDVSILEIQKG.VFEVKSTNGDTFLGGEDFHALVHHLVGEFKK
F11F1.1            AAIAYGLKGQRLSRRNILIYDLGGGTFDVAAVNUDGP.RITVKAKGGDTHLGGEDIDNIIMIKMLEEFKN

                  250       260       270       280       290       300       310
hsp-1              KHKKDLASNPRALRRLRTACERAKRTLSSSSQASIEIDSLFEGID....FYTNITRARFEELCADLFRST
hsp-70             KTGKDISSNPRAIRRLRTACERAKRTLSSSTEATVEVDSLFDGTD....FYSKITRARFEELCADLFRKT
F44E5.4            KTGKDISPNPRAIRRLRTACERAKRTLSSSSEACVEVDSLFEGID....LCTKITRARFEELCADLFRKT
hsp-3(Grp78/BiP)   KSGKDLRKDKRAVQKLRREVEKAKRALSTQHQTKVEIESLFDGED....FSETLTRAKFEELNMDLFRAT
hsp-4(Grp78/BiP)   KSGKDLKDNRAVQKLRREVEKAKRALSTQHQTKIEIESLFDGED....FSETLTRAKFEELNMDLFRAT
hsp-6(Grp75/mtHsp70) EQGVDLTKDPQAMQRLREAAEKAKCELSSTTQTDINLPYITMDQSGPKHLNLKLTRAKFEQIVGDLIKRT
F11F1.1            RHGIDLKGNYFALKRIRKAAEVAKITLSASSVARIELECLHLGID....FIMRISRTDEESWIENLLMAT

                       320       330       340       350       360       370       380
hsp-1              MDPVEKSLRDAKMDKSQVHDIVLVGGSTRIPKVQKLLSDLFS.GKELNKSINPDEAVAYGAAVQAAILSG
hsp-70             LEPVEKALRDAKTDKGRIDEVVLVGGSTRIPKIQKLKEFFN.GKDLNCSINPDEAVAFGAAVQAAILSG
F44E5.4            LEPVEKALRDAKLDKSKIDEVVLVGGSTRVPKIQKLLKDFFN.GKELNCSINPDEAVAFGAAVQAAVLSG
hsp-3(Grp78/BiP)   LKPVQKVLEDSDLKNDDVHEIVLVGGSTRIPKVQQLIKEFFN.GKEPSRGINPDEAVAYGAAVQGGVISG
hsp-4(Grp78/BiP)   LKPVQKVLEDADMKKTDVHEIVLVGGSTRIPKVQQLIKDYFN.GKEPSRGINPDEAVAYGAAVQGAVISG
hsp-6(Grp75/mtHsp70) IEPCRKALHDAEKSSQIADVLLVGGMSRMPKVQATVQEIF..GKVPSKAVNPDEAVAMGAAIQGAVLAG
F11F1.1            VIHVERVIREANLKKSQINEIVLVGGSTRIPILKNIIKQSFESNTRICESIHPDEAVAYGAAIMAAVLSG

                       390       400       410       420       430       440       450
hsp-1              DKSEAVQDLLLLDVAPLSLGIETAGGVMTALIKRNTTIPTKTAQTETTYSDNQPGVLTQVYEGERAMTKD
hsp-70             VKDDTIKDVLLLDVAPLSHGIETAGGVMTNLLDRNTRIPAKASKTETTYADNQPGVSIQVYEGERAMTRD
F44E5.4            VKDETIKDVLLLDVAPLSLGIETAGGVMTNLLDRNTRIPTKACKTETTYADNQPGVSIQVYEGERAMTRD
hsp-3(Grp78/BiP)   E..EDTGEIVLLDVNPLTMGIETVGGVMTKLIGRNTVIPTKKSQVFSTAADNQPTVTIQVFEGERPMTKD
hsp-4(Grp78/BiP)   V..ENTGDIVLLDVNPLTLGIETVGGVMTKLIGRNTVIPTKKSQVFSTAADSQSAVSIVIYEGERPMVMD
hsp-6(Grp75/mtHsp70) D.....VTDVLLLDVTPLSLGIETLGGIMTKLITRNTTIPTKKSQVFSTAADGQTQVQIKVFQGEREMATS
F11F1.1            A..EEMQDMRLVDVIPMSIGVQCNRDYMSVLIKKGTVFPCTIKRKTFINSDDFQTSINVPVYEGERVLCSH

                       460       470       480       490       500       510       520
hsp-1              NNLLGKFELSGIPPAPRGVPQIEVTFDIDANGILNVSATDKSTGKQNKITITNDKGRLSKDDIERMVNEA
hsp-70             NHRLGTFELTGIPPAPRGVPQVDVTFDIDANGILNVSAADKSTGRSNRITIRNEKGRLSQADIDRMVNEA
F44E5.4            NHRLGTFELSGIPPAPRGVPQIEVTFDIDVNGILNVSAEDKSTGKSNRITIQNEKGRLTQSDIDRMVHEA
hsp-3(Grp78/BiP)   NHQLGKFDLTGLPPAPRGVPQIEVTFEIDVNGILHVTAEDKGTGNKNKITITNDQNRLSPEDIERMINDA
hsp-4(Grp78/BiP)   NHKLGNFDVTGIPPAPRGVPQIEVTFEIDVNGILHVSAEDRGTGNKNKLTITNDHNRLSPEDIERMINDA
hsp-6(Grp75/mtHsp70) NKLLGQFSLVGIPPAPRGVPQIEVTFDIDANGIVNVSARDRGTGKEQQIVIQSS.GGLSKDQIENMIKEA
F11F1.1            NRQLGEISLP.IQPKRRGESVVDJTLEVDYNGILQATALDLNTKKAIITTILYDCCTYTQPEIDEMISOS
```

```
              530        540        550        560        570        580        590
hsp-1           EKYKADDEAQKDRIGAKNGLPSYAFNLKQTIEDEK.LKDKISPEDKKKIEDKCDEILKWLDSNQTAEKEK
hsp-70          KQFEREDAAQRDRISSRNQLPAYAFQVKQALEEH...GSLLSAEDAKRAKDAVEDTLRWMERNTLADKDE
F44E5.4         KQFEKEDGEQRERVQARNQLPAYAFQINQALDEY...SSKLSTEDASRAKEAVDETLRWLDSNSLAEKDE
hsp-3(Grp78/BiP) EKFAEDDKKVKDKAEARNELPSYAYNLKNQIEDKEKLGGKLDEDDKKTIEEAVEEAISWLGSNAEASAEE
hsp-4(Grp78/BiP) DKFAADDQAQKERVESRNELPAYAYQMKTQIADKEKLGGKLTDEDKVSIESAVERAIEWLGSNQDASTEE
hsp-6(Grp75/mtHsp70) EKNAAEDAKRKELVEVINQASGIIHDTEAKMTEF...ADQLPKDECEALRTKIADTKKILDNKDNETPEA
F11F1.1         EEDRIFDEHFKMRYKQLQKSEDMAYGYKHRLEKVK...GSVEPGKFSNLMETVEREISWLSVFPKVSPNE
```

```
              600        610        620        630        640
hsp-1           FEHQQKDLEGLANPIISKLYQSAGGAPPG...AAPGGAGGAG....GPTIEEVD
hsp-70          IEAKDKELKSICQDILTKMHQQEAQSGSG..CGNPGSGGFHSSNYPGGPTVEEVD
F44E5.4         IEAKDKELKSVCQSILTKMHQSGQSSASAGCAGNPGTAGFNPNNYPQGPTVEEVD
hsp-3(Grp78/BiP) LKEQKKDLESKVQPIVSKLYKDAGAGGEE....APEEG.........SDDKDEL.
hsp-4(Grp78/BiP) NKEQKKELESVVQPIVSKLYG.AGGQGEQ....ASEEP.........SEDHDEL.
hsp-6(Grp75/mtHsp70) IKEACNTLQQQSLKLFEAAYKNMAAKNSG...GDAQEAKTAE.....EPKKEQN.
F11F1.1         YMQERRALF...LKVYPTLFINLSFY..........................
```

Figure 1-22 Alignment of *C. elegans* Hsp70 family members. *C. elegans* contains three cytoplasmic Hsp70s (hsp-1, hsp-70 and F44E5.4) which terminate in the common C-terminal GPTIEEVD motif implicated in co-chaperone binding. In addition, there are two ER localised isoforms (hsp-3 and hsp-4), one mitochondrial homologue (hsp-6) and an isoform lacking the C-terminal flexible loop (F11F1.1). NBD underlined yellow, SBD β-sandwich subdomain underlined blue and SBD helical subdomain underlined green.

1.7. Project outlines

1.7.1. Structural studies of the C-terminal domain of *C. elegans* Hsp70

Despite abundant structural information regarding the NBD and the β-sandwich subdomain (section 1.1.1), structures of the C-terminal lid subdomain are limited to *E. coli* homologue DnaK and rat homologue Hsc70. Despite structural conservation of the NBD and β-sandwich between *E. coli* and rat, the C-terminal subdomains were observed to adopt significantly different conformations; a three-helix bundle in *E. coli* and an anti-parallel coiled-coil dimer in rat. Limited by the available data, it is unclear whether these reflect a true divergence between prokaryotes and eukaryotes. Alternatively, the structures could represent different conformational states or the rat structure, because it was solved as an isolated C-terminal subdomain, could be a crystallographic artefact. To investigate this further, the crystal structure of the C-terminal subdomain from *C. elegans* was solved. Unexpectedly, this revealed structural conservation with the helical lid from *E. coli*. Further, comparison with the rat structure revealed a domain-swap dimerisation mechanism potentially providing insight into the folding pathway of the small three-helix bundle subdomain. Work for this project contributed to two publications –

Worrall, L., and Walkinshaw, M. D. (2006). Crystallization and X-ray data analysis of the 10 kDa C-terminal lid subdomain from C. elegans Hsp70. Acta Cryst. F *62*, 938-943.

Worrall, L., and Walkinshaw, M. D. (2007). Crystal structure of the C-terminal three-helix subdomain from *C. elegans* Hsp70. Biochemical and Biophysical Research Communications *357*, 105-110. Coordinates and structure factors deposited under the PDB-ID 2P32.

```
                                                                              1
daf-21(Hsp90)     ...........................................................MSEN
T05E11.3(Grp94)   ...........MRFLLVGFVALLAVSAFIPNVYAEDEIEDAPKETKEETREEDSIKLDGLSVSQIKELRSK
R151.7(mtHsp75)   MSLPVRMLTGRIMRTMRYVAPITASHFTPSV...........................PTQRRWMASE


                     10        20        30        40        50        60        70
daf-21(Hsp90)     AETFAFQAETAQLMSLIINTFYSNPEIYLRELISNASDALDKIRYQALTEPSELDIGKELFIKITPNKEE
T05E11.3(Grp94)   AERHEFQAEVNRMMKLIINSLYRNKEIFLRELISNASDALDKIRLLSQIDPEQLRETEEMSVRIKADREN
R151.7(mtHsp75)   PQFHEFQAETRNLMDIVAKSLYSHGEVFVRELISNASDALEKRRYAELK..GDVAEGPS.EIRITTNKDK


                         80        90        100       110            120       130
daf-21(Hsp90)     RTLTIMDTGIGMTKADLVNNLGTIAKSGTKAFMEALQAGADIS.....MIGQFGVGFYSALVADKVVVT
T05E11.3(Grp94)   RLIHITDTGVGMTRQDLINNLGTIAKSGTSEFLSKLMDTATSSDQQQDLIGQFGVGFYAAFLVADRVVVT
R151.7(mtHsp75)   RTITFEDTGIGMNREDLVKFLGTIAKSGSKDFIENNKENAEA......VIGQFGVGFYSAFTVADSVVVT


                   140       150       160       170       180       190       200
daf-21(Hsp90)     SK...NNDDDSYCWESSAGGSFVVRPFNDPEVTPGTKIVMHIK.EDQIDFLEERKIKEIVKQHSQFIGYP
T05E11.3(Grp94)   TK...NNDDDQYIWESDSASFTISKDPRGNTLKRGTQITLYLK.EEAADFLEPDTLKNLVHYYSQFINFD
R151.7(mtHsp75)   TKVGSSDADGLQWTWNGDNSYEIAETGG..LQTGTKIEIRLKAVGDSATYAEEDRIKEVINKYSYFMSAP


                   210       220       230       240       250       260       270
daf-21(Hsp90)     TKLVVEKEREKEVEDEEAVEAKDEEKKEGEVENVADDADKKKKTKKIKEKYFEDEELNKTKPIWTRNPDDI
T05E11.3(Grp94)   IFLW....QSKTEMVEEAVEEEPATTEDGAVE...EEKEEKKTKKVEKTTWDWEKVNNVKPIWMRKPNQV
R151.7(mtHsp75)   ILV.......................................NGSRVNNLNALWTMQAREV


                     280       290       300       310       320       330
daf-21(Hsp90)     SNEEYAEFYKSLSND......WEDHLAVKHFSVEGQLEFRALLFVPQRAPFDLFEN.KKSKNSIKLYVRR
T05E11.3(Grp94)   EEDEYKQFYKSITKD......SEEPLSHVHFSAEGEVSFRGILYVPKKSPNDMFQNYGKVIENIKLYVRR
R151.7(mtHsp75)   MKEMHETFFKQLVKTQGKQEMYTRPQYTIHFQTDTPVSLRGVIYIPQTQFNQLTFMAQQTMCGLSLYARR


                   340       350       360       370       380       390       400
daf-21(Hsp90)     VFIMENCEELHPEYLNFIKGVVDSEDLPLNISREMLQQSKILKVIRKNLVKKCMELID.EVAEDKDNFKK
T05E11.3(Grp94)   VFITDDFADMLPKYLRFIRGIVDSDDUPLNVSRENLQCHKILKVIKSKLVRKVLDMLK...MLDGAQFDD
R151.7(mtHsp75)   VLIKPDAQELIPNYLRFVIGVVDSEDIPLNLSREMLQNNPVLRKLRKILTDKILGSLQSEMKKDPVKYSE


                   410       420       430       440       450       460       470
daf-21(Hsp90)     FYEQFGKNLKLGIH..EDSTNRKKLSDFLRYSTSA..GDEPQLKEYVSRMKENQTQIYYITGESKDVVA
T05E11.3(Grp94)   FWSEFSTNIKLGVM...EDPSNRMRLAKLIRFQGSN.DADKITTLAAYVERMKEKQDAIYYMAGTSRKEVE
R151.7(mtHsp75)   FFKIYSLYFKEGVVTEQDQGVKEDVAKLLLFEGSSCKKAGELTSLGDYVKRMCEGSKEIYYMYANNRQLAE


                     480       490       500       510       520       530       540
daf-21(Hsp90)     ASAPVDRVKSRGFEVLYMCDPIDEYCVQQLKEYDGKKLVSV...TKEGLELPETEEEKKKFEEDKVAYEN
T05E11.3(Grp94)   TSPVFVERLIARGYEVLFLTEAVDEYCIQAMPEYESKKFQNV...AKSGVTIDDGEKAKEAHKGLEEEFKP
R151.7(mtHsp75)   SSPYYEVIKSQNKEVLFLYDPADEVVFLGLGQFGMKQLVPVEKWAQEEAEKTGTDDKKDTKDFRDNEKNE


                   550       560       570       580       590       600
daf-21(Hsp90)     ILCKVIKD.ILEKKVEKVGVNRLVSSECCIVTSEYGWSANMERIMKAQAL...RDSSTMGYMAAKKHLET
T05E11.3(Grp94)   LTDWLKETALKDLIEKAVVSQRLVKSPSALVASSYGWSGNMERIMKSQAYAKAKDPTQDFYATQKKTFEI
R151.7(mtHsp75)   LLDWMKETLGSVRVSETSGNHRPSEHVVMVTVLDMG...RARHFLRTGEI...KDMEHLVYF..KPHVHV


                   610       620       630       640       650       660       670
daf-21(Hsp90)     NPDHAIMKTLRDRVEMDKNDKTVRDLVVLLFETALLASGFSLEEPQSHASRIYFMIKLGLDIGDDEIEDS
T05E11.3(Grp94)   NPRHPVIKELLKRVTASEEDTTAASTAKLLFETATLRSGFSLQDQMGFADRIEAVLRQSLDVSQDAQVET
R151.7(mtHsp75)   NLTHPLVKAMYKMRKTDKE.TAAILAEQIYDNALITAGL.IKDTSRMVGRLMKLL............


                     680       690       700
daf-21(Hsp90)     AVPSSCTAEAKIEGAEEDASRMEEVD...
T05E11.3(Grp94)   ...EQHIEEAEPEPEAAEETTIEEEHSEL
R151.7(mtHsp75)   .........TSLAGNKGTSTILTP.....
```

Figure 1-23 Alignment of *C. elegans* Hsp90 proteins. *C. elegans* contains one cytoplasmic Hsp90 (daf-21), ER isoform Grp94 (T05E11.3) and mitochondrial located mtHsp75 (R151.7).

1.7.2. Biochemical and structural studies of two putative TPR domain containing co-chaperones

Work contributing to a previous doctoral degree identified two *C. elegans* TPR domain containing co-chaperones likely to interact with Hsp90 (Opamawutthikul, 2005). These proteins were found to be the *C. elegans* homologues for small glutamine-rich TPR containing protein (SGT) and Hsp70/Hsp90 organising protein (HOP). In this study, these *C. elegans* proteins have been cloned, expressed and purified to near homogeneity with the objective of pursuing biochemical investigations regarding the native states of the proteins and their interactions with Hsp70/Hsp90. In addition, structural studies were also conducted.

Biochemical studies of native SGT revealed it formed high-affinity dimers with an elongated shape. Further, it was demonstrated that SGT interacted with the C-terminal peptides from HSp70 and Hsp90 with equal affinities. Despite the crystallisation of full-length SGT and its isolated TPR domain, the crystals were not of sufficient quality for X-ray diffraction.

Studies on *C. elegans* HOP suggested it might exist as a dimer in solution. In addition, a tight binding interaction was demonstrated with human and *C. elegans* Hsp90 homologues.

1.7.3. Prediction of the complete repertoire of C. elegans TPR co-chaperones

The aim of the final study was to define the complete repertoire of *C. elegans* TPR domain containing proteins capable of interacting with Hsp70 or Hsp90. TPR domains demonstrated to interact with Hsp70/Hsp90 have a well defined domain architecture with strict conservation to the consensus for the residues defining the carboxylate-clamp motif. A profile hidden Markov model (HMM) method was employed to search for Hsp70/Hsp90 interacting TPR domains in the *C. elegans* proteome and genome. This highlighted a family of 12 proteins; nine of which are homologues of proteins known to interact with Hsp70 or Hsp90. The remaining three are uncharacterised putative proteins and represent targets for further study.

1.8. References

Ali, M. M., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature *440*, 1013-1017.

Ansar, S., Burlison, J. A., Hadden, M. K., Yu, X. M., Desino, K. E., Bean, J., Neckers, L., Audus, K. L., Michaelis, M. L., and Blagg, B. S. (2007). A non-toxic Hsp90 inhibitor protects neurons from Abeta-induced toxicity. Bioorg Med Chem Lett.

Auluck, P. K., Chan, H. Y., Trojanowski, J. Q., Lee, V. M., and Bonini, N. M. (2002). Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson's disease. Science *295*, 865-868.

Bell, D. R., and Poland, A. (2000). Binding of aryl hydrocarbon receptor (AhR) to AhR-interacting protein. The role of hsp90. J Biol Chem *275*, 36407-36414.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature *386*, 414-417.

Berke, S. J., and Paulson, H. L. (2003). Protein aggregation and the ubiquitin proteasome pathway: gaining the UPPer hand on neurodegeneration. Curr Opin Genet Dev *13*, 253-261.

Birnby, D. A., Link, E. M., Vowels, J. J., Tian, H., Colacurcio, P. L., and Thomas, J. H. (2000). A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of chemosensory behaviors in caenorhabditis elegans. Genetics *155*, 85-104.

Boorstein, W. R., Ziegelhoffer, T., and Craig, E. A. (1994). Molecular evolution of the HSP70 multigene family. J Mol Evol *38*, 1-17.

Bradbury, J. (2003). Chaperones: keeping a close eye on protein folding. Lancet *361*, 1194-1195.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71-94.

Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F. U. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. J Biol Chem *277*, 19265-19275.

Brychzy, A., Rein, T., Winklhofer, K. F., Hartl, F. U., Young, J. C., and Obermann, W. M. (2003). Cofactor Tpr2 combines two TPR domains and a J domain to regulate the Hsp70/Hsp90 chaperone system. EMBO J *22*, 3613-3623.

Bukau, B., Deuerling, E., Pfund, C., and Craig, E. A. (2000). Getting newly synthesized proteins into shape. Cell *101*, 119-122.

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

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

Chaudhuri, T. K., and Paul, S. (2006). Protein-misfolding diseases and chaperone-based therapeutic approaches. Febs J *273*, 1331-1349.

Chewawiwat, N., Yano, M., Terada, K., Hoogenraad, N. J., and Mori, M. (1999). Characterization of the novel mitochondrial protein import component, Tom34, in mammalian cells. J Biochem (Tokyo) *125*, 721-727.

Chou, C. C., Forouhar, F., Yeh, Y. H., Shr, H. L., Wang, C., and Hsiao, C. D. (2003). Crystal structure of the C-terminal 10-kDa subdomain of Hsc70. J Biol Chem *278*, 30311-30316.

Cliff, M. J., Harris, R., Barford, D., Ladbury, J. E., and Williams, M. A. (2006). Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. Structure *14*, 415-426.

Connell, P., Ballinger, C. A., Jiang, J., Wu, Y., Thompson, L. J., Hohfeld, J., and Patterson, C. (2001). The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. Nat Cell Biol *3*, 93-96.

Consortium, T. C. e. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.

Cupp-Vickery, J. R., Peterson, J. C., Ta, D. T., and Vickery, L. E. (2004). Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. J Mol Biol *342*, 1265-1278.

D'Andrea, L. D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem Sci *28*, 655-662.

Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J *17*, 1192-1199.

Dedeoglu, A., Ferrante, R. J., Andreassen, O. A., Dillmann, W. H., and Beal, M. F. (2002). Mice overexpressing 70-kDa heat shock protein show increased resistance to malonate and 3-nitropropionic acid. Exp Neurol *176*, 262-265.

Denning, G. M., Anderson, M. P., Amara, J. F., Marshall, J., Smith, A. E., and Welsh, M. J. (1992). Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. Nature *358*, 761-764.

Denny, W. B., Valentine, D. L., Reynolds, P. D., Smith, D. F., and Scammell, J. G. (2000). Squirrel monkey immunophilin FKBP51 is a potent inhibitor of glucocorticoid receptor binding. Endocrinology *141*, 4107-4113.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci *25*, 24-28.

Fernandez-Saiz, V., Moro, F., Arizmendi, J. M., Acebron, S. P., and Muga, A. (2006). Ionic contacts at DnaK substrate binding domain involved in the allosteric regulation of lid dynamics. J Biol Chem *281*, 7479-7488.

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

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.

Freeman, B. C., Toft, D. O., and Morimoto, R. I. (1996). Molecular chaperone machines: chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. Science *274*, 1718-1720.

Galat, A. (2003). Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity--targets--functions. Curr Top Med Chem *3*, 1315-1347.

Gruschus, J. M., Greene, L. E., Eisenberg, E., and Ferretti, J. A. (2004). Experimentally biased model structure of the Hsc70/auxilin complex: substrate transfer and interdomain structural change. Protein Sci *13*, 2029-2044.

Hague, S. M., Klaffke, S., and Bandmann, O. (2005). Neurodegenerative disorders: Parkinson's disease and Huntington's disease. J Neurol Neurosurg Psychiatry *76*, 1058-1063.

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

Harrison, C. J., Hayer-Hartl, M., Di Liberto, M., Hartl, F., 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.

Hennessy, F., Nicoll, W. S., Zimmermann, R., Cheetham, M. E., and Blatch, G. L. (2005). Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. Protein Sci *14*, 1697-1709.

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

Heschl, M. F., and Baillie, D. L. (1989). Characterization of the hsp70 multigene family of Caenorhabditis elegans. DNA *8*, 233-243.

Heschl, M. F., and Baillie, D. L. (1990). The HSP70 multigene family of Caenorhabditis elegans. Comp Biochem Physiol B *96*, 633-637.

Hickey, E., Brandon, S. E., Smale, G., Lloyd, D., and Weber, L. A. (1989). Sequence and regulation of a gene encoding a human 89-kilodalton heat shock protein. Mol Cell Biol *9*, 2615-2626.

Hirano, T., Kinoshita, N., Morikawa, K., and Yanagida, M. (1990). Snap helix with knob and hole: essential repeats in S. pombe nuclear protein nuc2+. Cell *60*, 319-328.

Jackson, S. E., Queitsch, C., and Toft, D. (2004). Hsp90: from structure to phenotype. Nat Struct Mol Biol *11*, 1152-1155.

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

Jiang, J., Prasad, K., Lafer, E. M., and Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. Mol Cell *20*, 513-524.

Jiang, R., Gao, B., Prasad, K., Greene, L. E., and Eisenberg, E. (2000). Hsc70 chaperones clathrin and primes it to interact with vesicle membranes. J Biol Chem *275*, 8439-8447.

Kaletta, T., and Hengartner, M. O. (2006). Finding function in novel targets: C. elegans as a model organism. Nat Rev Drug Discov *5*, 387-398.

Kunzelmann, K., and Nitschke, R. (2000). Defects in processing and trafficking of cystic fibrosis transmembrane conductance regulator. Exp Nephrol *8*, 332-342.

Lai, B. T., Chin, N. W., Stanek, A. E., Keh, W., and Lanks, K. W. (1984). Quantitation and intracellular localization of the 85K heat shock protein by using monoclonal and polyclonal antibodies. Mol Cell Biol *4*, 2802-2810.

Lansbury, P. T., Jr., and Brice, A. (2002). Genetics of Parkinson's disease and biochemical studies of implicated gene products. Curr Opin Genet Dev *12*, 299-306.

Laufen, T., Mayer, M. P., Beisel, C., Klostermeier, D., Mogk, A., Reinstein, J., and Bukau, B. (1999). Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones. Proc Natl Acad Sci U S A *96*, 5452-5457.

Lemmon, S. K. (2001). Clathrin uncoating: Auxilin comes to life. Curr Biol *11*, R49-52.

Little, E., Ramakrishnan, M., Roy, B., Gazit, G., and Lee, A. S. (1994). The glucose-regulated proteins (GRP78 and GRP94): functions, gene regulation, and applications. Crit Rev Eukaryot Gene Expr *4*, 1-18.

Liu, F. H., Wu, S. J., Hu, S. M., Hsiao, C. D., and Wang, C. (1999). Specific interaction of the 70-kDa heat shock cognate protein with the tetratricopeptide repeats. J Biol Chem *274*, 34425-34432.

Mayer, M. P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci *62*, 670-684.

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

McDonough, H., and Patterson, C. (2003). CHIP: a link between the chaperone and proteasome systems. Cell Stress Chaperones 8, 303-308.

McLean, P. J., Kawamata, H., and Hyman, B. T. (2001). Alpha-synuclein-enhanced green fluorescent protein fusion proteins form proteasome sensitive inclusions in primary neurons. Neuroscience 104, 901-912.

Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S. M., Panaretou, B., Piper, P. W., and Pearl, L. H. (2003). Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. Mol Cell 11, 647-658.

Meyer, P., Prodromou, C., Liao, C., Hu, B., Roe, S. M., Vaughan, C. K., Vlasic, I., Panaretou, B., Piper, P. W., and Pearl, L. H. (2004). Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. EMBO J 23, 1402-1410.

Minami, Y., Kimura, Y., Kawasaki, H., Suzuki, K., and Yahara, I. (1994). The carboxy-terminal region of mammalian HSP90 is required for its dimerization and function in vivo. Mol Cell Biol 14, 1459-1464.

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. J Mol Biol 289, 1387-1403.

Murata, S., Minami, Y., Minami, M., Chiba, T., and Tanaka, K. (2001). CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. EMBO Rep 2, 1133-1138.

Nathan, D. F., Vos, M. H., and Lindquist, S. (1997). In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. Proc Natl Acad Sci U S A 94, 12949-12956.

Nicolet, C. M., and Craig, E. A. (1989). Isolation and characterization of STI1, a stress-inducible gene from Saccharomyces cerevisiae. Mol Cell Biol 9, 3638-3646.

Obermann, W. M., Sondermann, H., Russo, A. A., Pavletich, N. P., and Hartl, F. U. (1998). In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. J Cell Biol 143, 901-910.

Opamawutthikul, M. (2005) PhD thesis - Biochemical and Biophysical Studies of FK506 Binding Proteins, University of Edinburgh, Edinburgh.

Owens-Grillo, J. K., Czar, M. J., Hutchison, K. A., Hoffmann, K., Perdew, G. H., and Pratt, W. B. (1996). A model of protein targeting mediated by immunophilins and other proteins that bind to hsp90 via tetratricopeptide repeat domains. J Biol Chem 271, 13468-13475.

Panaretou, B., Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1998). ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. EMBO J 17, 4829-4836.

Pearl, L. H. (2005). Hsp90 and Cdc37 -- a chaperone cancer conspiracy. Curr Opin Genet Dev 15, 55-61.

Pearl, L. H., and Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem 75, 271-294.

Pellecchia, M., Szyperski, T., Wall, D., Georgopoulos, C., and Wuthrich, K. (1996). NMR structure of the J-domain and the Gly/Phe-rich region of the Escherichia coli DnaJ chaperone. J Mol Biol *260*, 236-250.

Pishvaee, B., Costaguta, G., Yeung, B. G., Ryazantsev, S., Greener, T., Greene, L. E., Eisenberg, E., McCaffery, J. M., and Payne, G. S. (2000). A yeast DNA J protein required for uncoating of clathrin-coated vesicles in vivo. Nat Cell Biol *2*, 958-963.

Pratt, W. B., Galigniana, M. D., Harrell, J. M., and DeFranco, D. B. (2004). Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. Cell Signal *16*, 857-872.

Pratt, W. B., and Toft, D. O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. Exp Biol Med (Maywood) *228*, 111-133.

Prodromou, C., Panaretou, B., Chohan, S., Siligardi, G., O'Brien, R., Ladbury, J. E., Roe, S. M., Piper, P. W., and Pearl, L. H. (2000). The ATPase cycle of Hsp90 drives a molecular 'clamp' via transient dimerization of the N-terminal domains. EMBO J *19*, 4383-4392.

Prodromou, C., Roe, S. M., Piper, P. W., and Pearl, L. H. (1997). A molecular clamp in the crystal structure of the N-terminal domain of the yeast Hsp90 chaperone. Nat Struct Biol *4*, 477-482.

Prodromou, C., Siligardi, G., O'Brien, R., Woolfson, D. N., Regan, L., Panaretou, B., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1999). Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. EMBO J *18*, 754-762.

Reynolds, P. D., Ruan, Y., Smith, D. F., and Scammell, J. G. (1999). Glucocorticoid resistance in the squirrel monkey is associated with overexpression of the immunophilin FKBP51. J Clin Endocrinol Metab *84*, 663-669.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., and et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066-1073.

Roe, S. M., Ali, M. M., Meyer, P., Vaughan, C. K., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2004). The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). Cell *116*, 87-98.

Rubenstein, R. C., and Zeitlin, P. L. (2000). Sodium 4-phenylbutyrate downregulates Hsc70: implications for intracellular trafficking of DeltaF508-CFTR. Am J Physiol Cell Physiol *278*, C259-267.

Russell, L. C., Whitt, S. R., Chen, M. S., and Chinkers, M. (1999). Identification of conserved residues required for the binding of a tetratricopeptide repeat domain to heat shock protein 90. J Biol Chem *274*, 20060-20063.

Sahara, N., Murayama, M., Mizoroki, T., Urushitani, M., Imai, Y., Takahashi, R., Murata, S., Tanaka, K., and Takashima, A. (2005). In vivo evidence of CHIP up-regulation attenuating tau aggregation. J Neurochem *94*, 1254-1263.

Scammell, J. G., Denny, W. B., Valentine, D. L., and Smith, D. F. (2001). Overexpression of the FK506-binding immunophilin FKBP51 is the common cause of glucocorticoid resistance in three New World primates. Gen Comp Endocrinol *124*, 152-165.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199-210.

Sharp, S., and Workman, P. (2006). Inhibitors of the HSP90 molecular chaperone: current status. Adv Cancer Res *95*, 323-348.

Shiau, A. K., Harris, S. F., Southworth, D. R., and Agard, D. A. (2006). Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. Cell *127*, 329-340.

Shomura, Y., Dragovic, Z., Chang, H. C., Tzvetkov, N., Young, J. C., Brodsky, J. L., Guerriero, V., Hartl, F. U., and Bracher, A. (2005). Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. Mol Cell *17*, 367-379.

Shulman, J. M., Shulman, L. M., Weiner, W. J., and Feany, M. B. (2003). From fruit fly to bedside: translating lessons from Drosophila models of neurodegenerative disease. Curr Opin Neurol *16*, 443-449.

Sikorski, R. S., Boguski, M. S., Goebl, M., and Hieter, P. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell *60*, 307-317.

Siligardi, G., Panaretou, B., Meyer, P., Singh, S., Woolfson, D. N., Piper, P. W., Pearl, L. H., and Prodromou, C. (2002). Regulation of Hsp90 ATPase activity by the co-chaperone Cdc37p/p50cdc37. J Biol Chem *277*, 20151-20159.

Silverstein, A. M., Galigniana, M. D., Chen, M. S., Owens-Grillo, J. K., Chinkers, M., and Pratt, W. B. (1997). Protein phosphatase 5 is a major component of glucocorticoid receptor.hsp90 complexes with properties of an FK506-binding immunophilin. J Biol Chem *272*, 16224-16230.

Skach, W. R. (2006). CFTR: new members join the fold. Cell *127*, 673-675.

Slepenkov, S. V., and Witt, S. N. (2002). The unfolding story of the Escherichia coli Hsp70 DnaK: is DnaK a holdase or an unfoldase? Mol Microbiol *45*, 1197-1206.

Sondermann, H., Scheufler, C., Schneider, C., Hohfeld, J., Hartl, F. U., and Moarefi, I. (2001). Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. Science *291*, 1553-1557.

Song, H. Y., Dunbar, J. D., Zhang, Y. X., Guo, D., and Donner, D. B. (1995). Identification of a protein with homology to hsp90 that binds the type 1 tumor necrosis factor receptor. J Biol Chem *270*, 3574-3581.

Sousa, R., and Lafer, E. M. (2006). Keep the traffic moving: mechanism of the Hsp70 motor. Traffic *7*, 1596-1603.

Stebbins, C. E., Russo, A. A., Schneider, C., Rosen, N., Hartl, F. U., and Pavletich, N. P. (1997). Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. Cell *89*, 239-250.

Stevens, S. Y., Cai, S., Pellecchia, M., and Zuiderweg, E. R. (2003). The solution structure of the bacterial HSP70 chaperone protein domain DnaK(393-507) in complex with the peptide NRLLLTG. Protein Sci *12*, 2588-2596.

Tabara, H., Grishok, A., and Mello, C. C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science *282*, 430-431.

Takeda, S., and McKay, D. B. (1996). Kinetics of peptide binding to the bovine 70 kDa heat shock cognate protein, a molecular chaperone. Biochemistry *35*, 4636-4644.

Tavaria, M., Gabriele, T., Kola, I., and Anderson, R. L. (1996). A hitchhiker's guide to the human Hsp70 family. Cell Stress Chaperones *1*, 23-28.

Tsai, J., and Douglas, M. G. (1996). A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. J Biol Chem *271*, 9347-9354.

Ungewickell, E., Ungewickell, H., and Holstein, S. E. (1997). Functional interaction of the auxilin J domain with the nucleotide- and substrate-binding modules of Hsc70. J Biol Chem *272*, 19594-19600.

van der Laan, M., Rissler, M., and Rehling, P. (2006). Mitochondrial preprotein translocases as dynamic molecular machines. FEMS Yeast Res *6*, 849-861.

Vaughan, C. K., Gohlke, U., Sobott, F., Good, V. M., Ali, M. M., Prodromou, C., Robinson, C. V., Saibil, H. R., and Pearl, L. H. (2006). Structure of an Hsp90-Cdc37-Cdk4 complex. Mol Cell *23*, 697-707.

Vogel, M., Bukau, B., and Mayer, M. P. (2006a). Allosteric regulation of Hsp70 chaperones by a proline switch. Mol Cell *21*, 359-367.

Vogel, M., Mayer, M. P., and Bukau, B. (2006b). Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. J Biol Chem *281*, 38705-38711.

Wang, H., Kurochkin, A. V., Pang, Y., Hu, W., Flynn, G. C., and Zuiderweg, E. R. (1998). NMR solution structure of the 21 kDa chaperone protein DnaK substrate binding domain: a preview of chaperone-protein interaction. Biochemistry *37*, 7929-7940.

Wegele, H., Muller, L., and Buchner, J. (2004). Hsp70 and Hsp90--a relay team for protein folding. Rev Physiol Biochem Pharmacol *151*, 1-44.

Welch, W. J. (2004). Role of quality control pathways in human diseases involving protein misfolding. Semin Cell Dev Biol *15*, 31-38.

Welch, W. J., and Feramisco, J. R. (1982). Purification of the major mammalian heat shock proteins. J Biol Chem *257*, 14949-14959.

Welsh, M. J., Anderson, M. P., Rich, D. P., Berger, H. A., Denning, G. M., Ostedgaard, L. S., Sheppard, D. N., Cheng, S. H., Gregory, R. J., and Smith, A. E. (1992). Cystic fibrosis transmembrane conductance regulator: a chloride channel with novel regulation. Neuron *8*, 821-829.

Wilbanks, S. M., Chen, L., Tsuruta, H., Hodgson, K. O., and McKay, D. B. (1995). Solution small-angle X-ray scattering study of the molecular chaperone Hsc70 and its subfragments. Biochemistry *34*, 12095-12106.

Wu, B., Li, P., Liu, Y., Lou, Z., Ding, Y., Shu, C., Ye, S., Bartlam, M., Shen, B., and Rao, Z. (2004a). 3D structure of human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. Proc Natl Acad Sci U S A *101*, 8348-8353.

Wu, Y. R., Wang, C. K., Chen, C. M., Hsu, Y., Lin, S. J., Lin, Y. Y., Fung, H. C., Chang, K. H., and Lee-Chen, G. J. (2004b). Analysis of heat-shock protein 70 gene polymorphisms and the risk of Parkinson's disease. Hum Genet *114*, 236-241.

Xiao, J., Kim, L. S., and Graham, T. R. (2006). Dissection of Swa2p/auxilin domain requirements for cochaperoning Hsp70 clathrin-uncoating activity in vivo. Mol Biol Cell *17*, 3281-3290.

Xu, W., Marcu, M., Yuan, X., Mimnaugh, E., Patterson, C., and Neckers, L. (2002). Chaperone-dependent E3 ubiquitin ligase CHIP mediates a degradative pathway for c-ErbB2/Neu. Proc Natl Acad Sci U S A *99*, 12847-12852.

Yano, M., Terada, K., and Mori, M. (2003). AIP is a mitochondrial import mediator that binds to both import receptor Tom20 and preproteins. J Cell Biol *163*, 45-56.

Young, J. C., Agashe, V. R., Siegers, K., and Hartl, F. U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol *5*, 781-791.

Young, J. C., and Hartl, F. U. (2000). Polypeptide release by Hsp90 involves ATP hydrolysis and is enhanced by the co-chaperone p23. EMBO J *19*, 5930-5940.

Young, J. C., Hoogenraad, N. J., and Hartl, F. U. (2003). Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. Cell *112*, 41-50.

Young, J. C., Obermann, W. M., and Hartl, F. U. (1998). Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. J Biol Chem *273*, 18007-18010.

Zhang, M., Windheim, M., Roe, S. M., Peggie, M., Cohen, P., Prodromou, C., and Pearl, L. H. (2005). Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. Mol Cell *20*, 525-538.

Zhao, R., Davey, M., Hsu, Y. C., Kaplanek, P., Tong, A., Parsons, A. B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., *et al.* (2005). Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell *120*, 715-727.

Zhu, X., Zhao, X., Burkholder, W. F., 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.

2. Crystal structure of the C-terminal 10 kDa helical subdomain from *C. elegans* Hsp70

2.1. Introduction

As discussed in detail in chapter 1, Hsp70 is an essential molecular chaperone involved in numerous protein folding processes. It belongs to a family of ubiquitously expressed proteins that exist in virtually all living organisms (Wegele et al., 2004). Hsp70 consists of two domains; a 40 kDa N-terminal nucleotide binding domain (NBD), which both binds and hydrolyses ATP, and a 30 kDa C-terminal substrate-binding domain (SBD) (Chappell et al., 1987). The SBD can be further divided into an 18 kDa beta-sandwich subdomain, which forms a hydrophobic peptide binding groove capable of binding exposed hydrophobic patches on proteins, and a 10 kDa helical-bundle subdomain that forms a lid over the peptide binding groove (Zhu et al., 1996). The only eukaryotic structure solved for the 10 kDa C-terminal lid is from rat (Chou et al., 2003), which has an anti-parallel coiled-coil mediated dimer. This is in contrast to the monomeric three-helical bundle observed in the *E. coli* homologue DnaK (Zhu et al., 1996), which shares approximately 17% sequence identity. Since it is the lid domain that restricts access to the peptide binding groove, structural knowledge of this domain should increase understanding of the process of client binding and release.

The *C. elegans* Hsp70 multigene family consists of 11 paralogues and includes orthologues for the Hsp70 family members Hsp70, Hsc70, GRP78/BiP and GRP75/mtHsp70 (see Figure 1-22). Gene *hsp-1* (Wormbase ID F26D10.3) encodes Hsp70A, the *C. elegans* Hsp70 orthologue. Hsp70A is closely related to the *Drosophila* heat inducible Hsp70s and the *S. cerevisiae* SSA Hsp70 subfamily. The *hsp-1* gene is normally expressed throughout development and upon heat-shock the *hsp-1* mRNA is enhanced 2-6 fold. Down-regulation of *hsp-1* via RNA interference results in a small reduction in the life-span of an *age-1* mutant indicating that *hsp-1* may play some role in regulating longevity (www.wormbase.com).

This chapter details the cloning, expression, purification, crystallisation, X-ray data analysis, phasing and refinement of the 10 kDa lid subdomain from the heat-inducible *C. elegans* Hsp70 homologue Hsp70A. This work has contributed to one publication discussing the preliminary X-ray data analysis (Worrall and Walkinshaw, 2006) and one presenting the final crystal structure (Worrall and Walkinshaw, 2007).
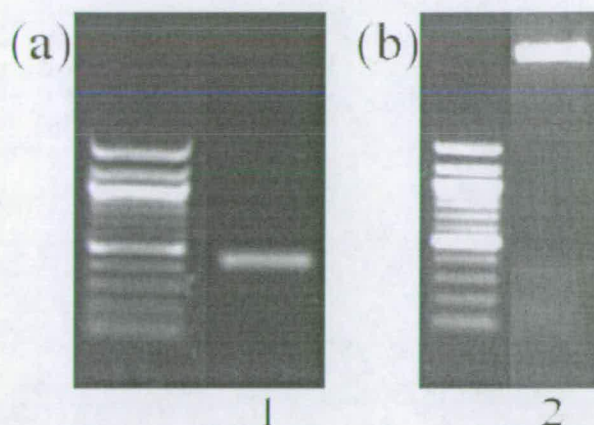
Figure 2-1 Cloning of ceHsp70-CT. (a) PCR of ceDNA corresponding to residues 542-640 of Hsp70A (lane 1). (b) Restriction digest of cloned PCR product. Double-digest showing band of appropriate size (faint; lane 2). 100 bp MW ladder used. in both gels.

2.2. Materials and methods

2.2.1. Cloning

The cDNA fragment corresponding to residues 542–640 of the *C. elegans* Hsp70 homologue Hsp70A was generated by polymerase chain reaction (PCR) using *C. elegans* mixed stage N2 cDNA as a template (Figure 2-1a). The sequence was amplified with the TaqPlus® precision PCR system (Stratagene) using forward (GCGGCATATGGGA CTCGAGTCATACGCCTTC) and reverse primers (GCGGGCGGCCGCTTAGT CGACCTCCTCGATC). The resulting PCR product was digested with NdeI and NotI (New England Biolabs) and ligated into a similarly digested pET-28a vector (Novagen), downstream of the 6XHis coding region (Figure 2-1b). The correct sequence was verified by sequencing and is now referred to as ceHsp70-CT.

2.2.2. Expression and purification

Recombinant ceHsp70-CT was expressed in Rosetta2(DE3) *E. coli* (Novagen) in LB liquid media containing kanamycin (25 μg/ml) and chloramphenicol (30 μg/ml). Cultures were grown with shaking at 37 °C until the A600 was ~ 0.6, and expression induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Expression was continued for 4 hours and cells were harvested by centrifugation (3000 x g for 15 minutes).
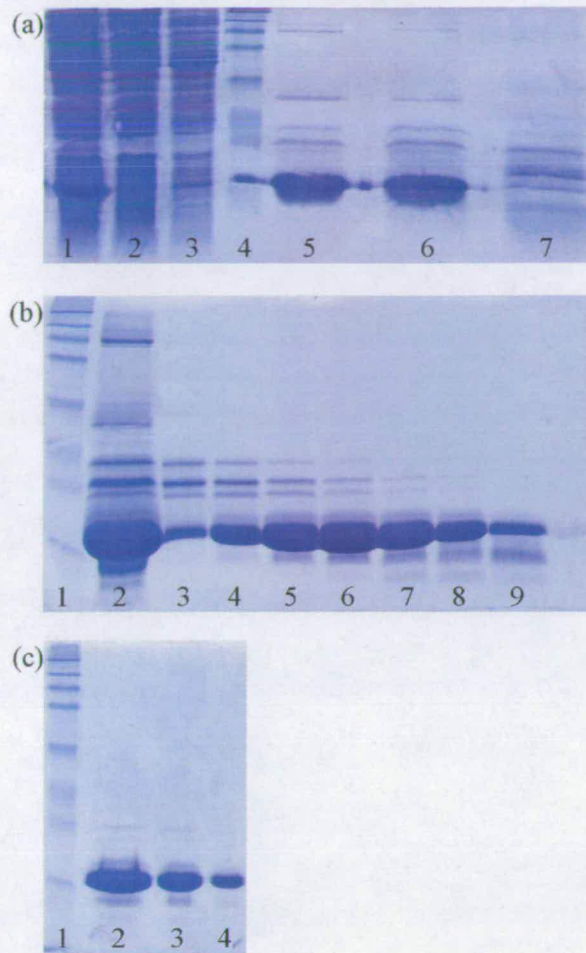
Figure 2-2 Purification of ceHsp70-CT. (a) Ni-NTA purification. 1. Whole-cell extract applied to column, 2. Flow-through, 3. Wash, 4. MW markers, 5. Elution from peak 1 (25 mM imidazole), 6. Elution from peak 2 (50 mM imidazole), 7. Elution from peak 3 (125 mM imidazole) Fractions 5 and 6 were pooled. (b) Gel-filtration purification with Superdex 75. 1. MW markers, 2. Sample loaded, 3. - Elution profile. Fractions 6-9 pooled. (c) Purified ceHsp70-CT. 1. MW markers, 2. 10 µg ceHsp70-CT, 3. 5 µg ceHsp70-CT, 4. 2 µg ceHsp70-CT.

Cell pellets were resuspended at 10% weight per volume in ice-cold lysis buffer (buffer A; 50mM sodium phosphate pH 8.0, 100 mM NaCl, 0.1mM benzamidine, 0.1mM PMSF) plus excess protease inhibitor cocktail (Roche) and sonicated on ice for 6 x 30 second bursts with 30 seconds cooling in between. The cell lysate was subjected to centrifugation at 30,000 g for 1 hour at 4 °C. The supernatant was filtered through a 0.2 µm filter and applied onto a 10ml Ni-NTA superflow column (Qiagen) pre-equilibrated in wash buffer (buffer B; 50mM sodium phosphate pH 8.0, 100 mM NaCl, 10 mM imidazole) (Figure 2-2a). Proteins were eluted with a stepped imidazole gradient (buffer C; 50mM sodium phosphate pH 8.0, 100 mM NaCl, 250 mM imidazole). Weakly binding protein was washed off with 10% buffer C and 6xHis tagged ceHsp70-CT eluted in two peaks with 20% (50 mM) and 50% (125 mM) buffer C respectively (Figure 2-2a). Fractions containing recombinant protein were pooled,

concentrated to ≤1 ml and loaded onto a Superdex 75 HR (Pharmacia) gel filtration column (V_t ~120 ml; 1.6 x 60 cm) equilibrated in buffer D (25 mM HEPES pH 7.5, 50 mM KCl, 1 mM DTT) (Figure 2-2b). Recombinant ceHsp70-CT eluted as a single peak and was more than 95% pure as judged by SDS-PAGE (Figure 2-2c). Protein was stored at 4 °C in buffer D.
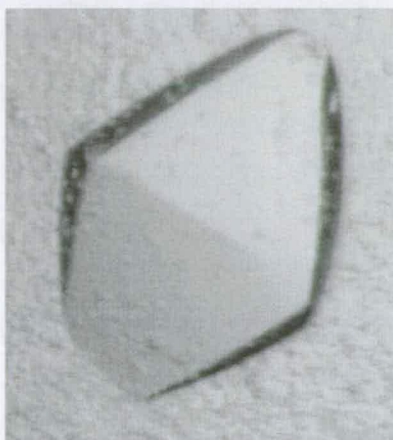
Figure 2-3 Example of ceHsp70-CT crystal grown in 62% saturated ammonium sulphate buffered by sodium citrate pH 6.0. Maximum dimension ~0.6 mm.

2.2.3. Crystallisation

2.2.3.1. Crystallisation of orthorhombic form I crystals and preparation of a heavy-metal derivative

Crystals of the 10 kDa subdomain, including the recombinant His tag, were grown by the hanging drop vapour diffusion method from a 12 mg ml^{-1} protein solution in buffer D at 18 °C. Initial conditions were identified with an ammonium sulphate grid screen. Best crystals were obtained using a well solution of 62% saturated ammonium sulphate buffered by 100 mM sodium citrate pH 6.0 with a 2 μl drop consisting of a 1:1 ratio of protein and well solution. Crystals appeared within 24 hours and grew to dimensions of 0.6 x 0.4 x 0.4 over 3 weeks (Figure 2-3). Conditions were comparable to those published for the rat homologue (Chou et al., 2003; Chou et al., 2001). A mercury derivative was obtained by soaking native crystals for 30 minutes in well solution containing 5 mM mercuric chloride followed by back-soaking for 30 seconds in well solution. All crystals were flash cooled in liquid nitrogen prior to data collection, either directly from mother liquor or with prior soaking in cryoprotectant containing 70% saturated ammonium sulphate and 10% glycerol.

2.2.3.2. Crystallisation of tetragonal form II crystals

An additive screen using Hampton Crystal Screen™ was employed to search for new crystallisation conditions (described in detail in section 2.3.2.1.). Four conditions produced good quality, single crystals with a similar octahedral habit seen for the orthorhombic crystal form (Figure 2-3). Condition 32, containing only 2 M ammonium sulphate, produced small single crystals. Conditions 16 (100 mM HEPES pH 7.5 and 1.5 M lithium sulphate) and 33 (4 M sodium formate) also produced good quality crystals. Finally, condition 39 (100 mM HEPES pH 7.5, 2 M ammonium sulphate and 2% PEG 400) resulted in multiple single crystals. Crystals from conditions 16, 33 and 39 were flash-cooled directly from the crystallisation drop.

2.2.4. Data collection and processing

All data were collected at either station 10.1, SRS, Daresbury, UK or station BM14, ESRF, Grenoble, France. All data were collected at 100 K. MAD data were collected from a single derivative crystal at 2 wavelengths, corresponding to the mercury L-III peak (1.005 Å) and the L-III edge inflection point (1.009 Å). All data were indexed and integrated using the MOSFLM (Leslie, 1992) package and scaled using SCALA (Collaborative Computational Project, 1994).

2.2.5. Data analysis, phasing and refinement

Specific methods are referred to in the Results section.
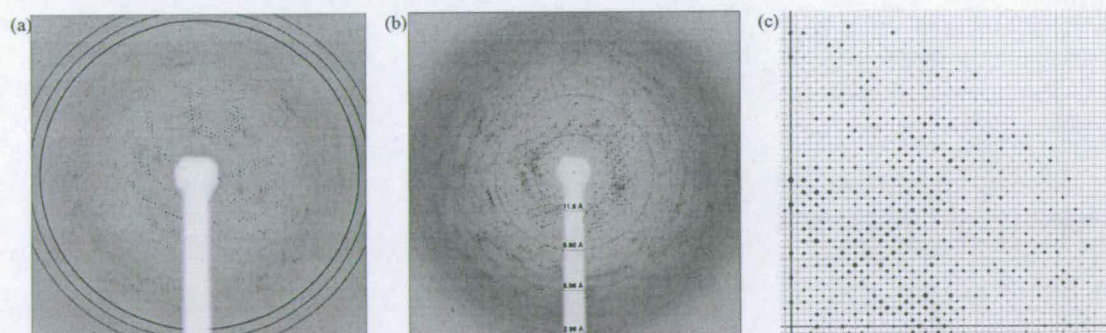
Figure 2-4 Diffraction patterns for ceHsp70-CT crystals. (a) Native crystal flash cooled directly from mother liquor (crystal native I). Hexagonal ice diffraction rings are present at ~3.9, ~3.62 and ~3.44 Å. (b) Native crystal flash cooled using 70% saturated ammonium sulphate, 10% glycerol as cryoprotectant (crystal native II). (c) Pseudo-precession image showing a section of the 0kl zone for native form I data processed in space-group $I2_12_12_1$. Reflections show (0k0) = 2n (x-axis) and (00l) = 4n (y-axis). Produced with XPREP (Bruker AXS, Madison, USA)

2.3. Results and discussion

2.3.1. Solving the structure of orthorhombic form I ceHsp70-CT crystals

Crystals of the 10 kDa C-terminal subdomain of *C. elegans* Hsp70 were initially produced that diffracted X-rays to ~3.5 Å for native and ~4.0 Å for derivative crystals.

2.3.1.1. X-ray data analysis

2.3.1.1.1. Space-group determination

ceHsp70-CT crystals flash cooled directly from mother liquor (62% saturated ammonium sulphate, 100 mM sodium citrate pH 6.0) diffracted X-rays to approximately 4 Å. Prominent hexagonal-ice diffraction rings at ~3.9, ~3.62 and ~3.45 Å were present but did not interfere with data processing (Figure 2-4a). Initial indexing identified the most likely Bravais lattice to be body-centred tetragonal with unit-cell dimensions $a = b = 196.9$ Å, $c = 200.6$ Å. Analysis of unmerged data with the program POINTLESS (Evans, 2006) suggested the cubic Laue group I m -3 m or the tetragonal Laue group I 4/m m m as possibilities. Inspection of the systematic absences revealed (h00) = 2n, (0k0) = 2n and (00l) = 4n (Figure 2-4c), consistent with tetragonal space-group $I4_122$. Data were successfully indexed and processed to an R_{sym} of ~10% for native and derivative crystals flash cooled without cryoprotectant. Unit-cell parameters and data reduction statistics are shown in table 2-1.

In an effort to improve the diffraction quality of the crystals, a series of cryoprotectants were screened. Of these, crystals vitrified in 70% ammonium sulphate, 10% glycerol and 100 mM sodium citrate pH 6.0 diffracted X-rays to ~3.4 Å and showed no ice rings (Figure 2-4b). However, whilst these crystals could be indexed in a tetragonal lattice, the merging statistics were poor (R_{sym} = 29.1% for data processed in $I4_122$). Reprocessing in the orthorhombic space-group I222 (or $I2_12_12_1$) resulted in an improved R_{sym} of 15.5%, albeit still high (Table 1, native I-II). $I4_122$ constitutes a minimal non-isomorphic supergroup of space-group $I2_12_12_1$ (Hahn, 2002), and data has been processed in both tetragonal and orthorhombic space-groups for comparison (Table 2-1). The possibility of twinning was investigated but no indication was present in the cumulative intensity distribution or in the plots of accentric and centric moments of E as output by the program TRUNCATE (French and Wilson, 1978).

Data set	Native I-I		Native I-II		Hg – L-III Peak		Hg – L-III inflection	
Cryoprotectant	Mother liquor		70% ammonium sulphate, 10% glycerol		Mother liquor		Mother liquor	
Beamline	SRS, 10.1		ESRF, BM14		ESRF, BM14		ESRF, BM14	
Φ-rotation	120° at 1° step		180° at 1.5° step		100° at 1° step		100° at 1° step	
Temperature	100K		100K		100K		100K	
Wavelength (Å)	1.005		0.978		1.005		1.009	
Space-group	I4₁22	I2₁2₁2₁	I4₁22	I2₁2₁2₁	I4₁22	I2₁2₁2₁	I4₁22	I2₁2₁2₁
Unit-cell parameters (Å)	a=196.9 b=196.9 c=200.6	a=196.8 b=196.9 c=200.6	a=194.7 b=194.7 c=200.8	a=194.6 b=195 c=200.8	a=195.1 b=195.1 c=202.8	a=195.1 b=195.2 c=202.8	a=195.5 b=195.5 c=203.5	a=195.5 b=195.6 c=203.4
Resolution range (Å)	45 - 3.95 (4.16 - 3.95)	45 - 3.95 (4.16 - 3.95)	35 - 3.4 (3.58 - 3.4)	35 – 3.4 (3.58 - 3.4)	42 – 3.95 (4.16 - 3.95)	42 – 3.95 (4.16 - 3.95)	42 – 4.1 (4.32 - 4.1)	42 – 4.1 (4.32 - 4.1)
No. observations	131543 (16631)	131659 (16730)	399719 (58219)	388320 (56298)	142589 (21135)	142701 (21069)	128025 (18741)	128245 (18752)
No. unique reflections	17565 (2516)	32857 (4723)	26797 (3841)	52659 (7591)	16504 (2420)	31088 (4612)	14849 (2151)	27844 (4089)
Completeness	99.8 (99.8)	95.3 (95.5)	99.9 (100)	99.9 (99.9)	95 (96.1)	91.5 (93.2)	94.8 (95.8)	90.9 (92.5)
Anomalous completeness					94.6 (95.6)	86.1 (87.5)	94.2 (95.2)	85 (86.3)
Multiplicity	7.8	4.0 (3.5)	14.9 (15.2)	7.4 (7.4)	8.6 (8.7)	4.6 (4.5)	8.6 (8.7)	4.6 (4.6)
Anomalous multiplicity					4.6 (4.5)	2.5 (2.4)	4.6 (4.5)	2.5 (2.5)
$R_{sym}{}^a$ (%)	12.8 (88.8)	11.3 (76.8)	29.1 (116.3)	15.5 (111.1)	11 (73.9)	10.1 (70.8)	10.4 (102.9)	12 (105.4)
$R_{p.i.m.}{}^b$ (%)	5.0 (35.5)	6.0 (44.3)	7.4 (25.4)	6.1 (43.5)	4.7 (27.3)	6.2 (37.4)	4.1 (37.5)	6.6 (63.5)
I/σ(I)	14.2 (3.0)	10.5 (2.4)	10.1 (1.6)	10.1 (1.6)	13.7 (2.5)	10 (1.9)	13.9 (2.1)	9.9 (1.4)

a $R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i |I_i(hkl)|$ b $R_{p.i.m.} = \sum_{hkl}[1/N-1]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i |I_i(hkl)|$

Table 2-1 Reflection data statistics for data processed in space-groups I4₁22 and I2₁2₁2₁. Values in parentheses are for the highest resolution bin.

K = 90°   K = 120°   K = 180°
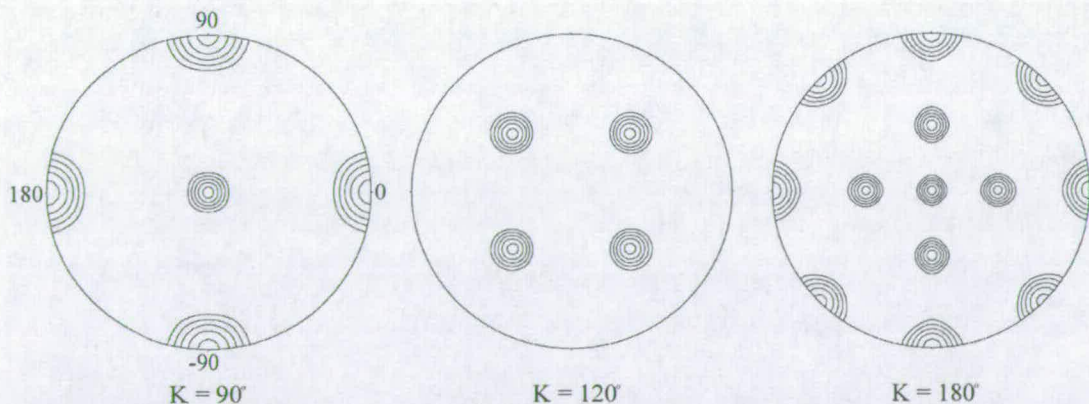
Figure 2-5 Stereographic projection plots of the κ = 90°, 120° and 180° sections of the self-rotation function of the native form II data set. Calculated from data in the resolution range 35–4.0 Å, integration radius 20 Å, showing peaks >30% origin, contour steps of 15%. The data were reduced in I2₁2₁2₁ but the plot shows 432 symmetry. ω rotation angle along the radial axis (ω = 0° in the middle, ω = 90° on the perimeter) and φ rotation angle around the perimeter. Figure prepared with the programs POLARRFN (Collaborative Computational Project 1994).

2.3.1.1.2. Content of the asymmetric unit

For a protein with molecular weight 13094 Da, the Matthew's equation (Matthews, 1968) indicates there to be between 9 ($V_m = 4.09$, solvent = 70%) and 21 monomers ($V_m = 1.75$, solvent = 30%) per asymmetric unit for packing in $I4_122$ and between 18 and 42 for space-group $I2_12_12_1$. Hsp70 has been shown to form dimers, trimers and higher-order oligomers in solution (Benaroudj et al., 1995; Benaroudj et al., 1997; Benaroudj et al., 1996; Chou et al., 2003; Fouchaq et al., 1999; Nemoto et al., 2006) and a related domain from rat crystallised with 4 monomers in the asymmetric unit, as two dimers in a cruciform like arrangement (Chou et al., 2003).

A self-rotation Patterson map calculated using data processed in space-group $I2_12_12_1$ reveals a high degree of rotational non-crystallographic symmetry (Figure 2-5). Three orthogonal fourfold axes are present parallel to the crystallographic twofold axes at $\kappa = 90°$. Four mutually orthogonal threefold axes are present aligned parallel to the cell body-diagonals at $\kappa = 120°$. Additionally, there are twofold peaks at $\kappa = 180°$ every 45° in the ab plane and every 90° parallel to the ac and bc plane face-diagonals. All peaks are approximately the height of the origin and show 432 point group symmetry suggesting a pseudo-cubic packing symmetry. Inspection of the native Patterson map also indicates the presence of translational non-crystallographic symmetry, with three large non-origin peaks approximately 20% the height of the origin observed for crystals of space-group $I2_12_12_1$ (Figure 2-6). This is consistent with the presence of a dimer of trimers or a trimer of dimers related by translational non-crystallographic symmetry at four positions, resulting in 24 monomers in the asymmetric unit and a solvent content of ~60% ($V_m = 3.03$).

2.3.1.2. Phasing

2.3.1.2.1. Molecular replacement

The C-terminal domain of *C. elegans* Hsp70 has 76% sequence identity to the previously published rat structure. Extensive molecular replacement trials were carried out using various programs [AMoRe (Navaza, 1994), Beast (Read, 2001), MolRep (Vagin and Teplyakov, 1997), Phaser (Storoni et al., 2004)] with multiple models using both the rat structure and the *E. coli* structure as a template, however, all failed to yield a satisfactory solution. Low signal-to-noise ratio due to multiple monomers in the asymmetric unit and translational non-crystallographic symmetry can be problematic in molecular replacement but this could also suggest a significantly different conformation.
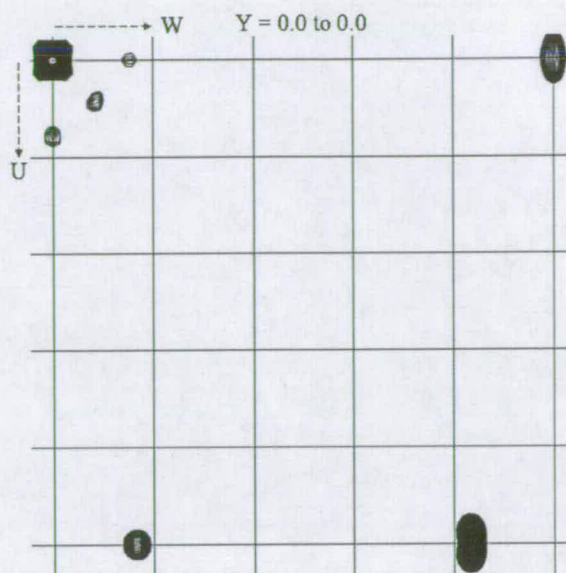
Figure 2-6 Native Patterson map (0 < u < 0.5, v = 0, 0 < w < 0.5) calculated from the native data processed in space-group I2₁2₁2₁ using reflections in the resolution range 40-4 Å with Fobs ≥ 3σ(Fobs). Three large non-origin peaks are observed at (0.4864, 0.0000, 0.4154), (0.5000, 0.0193, 0.0847) and (0.0142, 0.0200, 0.5000) approximately 20% the height of the origin. For data processed in I4₁22, peaks (0.4864, 0.0000, 0.4154) and (0.5000, 0.0193, 0.0847) are symmetry related. Figure prepared with the programs NPO and XPLOT84DRIVER (Collaborative Computational Project 1994).

2.3.1.2.2.   Multiwavelength anomalous dispersion

For these reasons a heavy-metal derivative was sought to enable structure determination either by isomorphous replacement or anomalous dispersion methods. A mercury derivative was produced that diffracted X-rays to ~ 4 Å, with slightly altered unit-cell dimensions along all three axes (Table 2-1) and MAD data were collected at two wavelengths from the same crystal corresponding to the mercury LIII peak (1.005 Å) and inflection point (1.009 Å).

2.1.1.1.1.1.   Analysis of MAD data

Analysis for significant anomalous signal and heavy-atom location was performed with SHELXC (Sheldrick, 2004) and SHELXD (Schneider and Sheldrick, 2002). An anomalous signal-to-noise ratio, based on the mean value of the ratio between the anomalous differences $|F^+ - F^-|$ and the estimated standard deviation of these differences, of greater than 1.2 was deemed significant (Sheldrick, 2004). The peak/inflection point datasets processed in space-groups I4₁22 and I2₁2₁2₁ were estimated to have anomalous signal to 5.4 Å/6.5 Å and 5.6Å/7Å respectively (Table 2-2). The maximum resolution to be included for heavy-atom location, based on a correlation coefficient between the signed anomalous differences ΔF of

greater than 30% (Sheldrick, 2004), was estimated to be 5.4 Å ($I4_122$) and 5.6 Å ($I2_12_12_1$) (Table 2-2). ceHsp70-CT contains one cysteine, thus 12 mercury atoms were predicted to be bound in the asymmetric unit for data processed in $I4_122$ and 24 for data processed in $I2_12_12_1$.

Statistics of the anomalous signal-to-noise ratio against resolution, $<|F^+ - F^-|/\sigma(F^+ - F^-)>$											
Resolution (Å)	∞-8.0	8.0-6.0	6.0-5.6	5.6-5.4	5.4-5.2	5.2-5.0	5-4.8	4.8-4.6	4.6-4.4	4.4-4.2	4.2-3.95
$I4_122$											
Peak	2.93	1.87	1.39	1.17	1.12	1.07	0.98	0.95	0.84	0.81	0.78
Inflection	1.82	1.06	0.90	0.91	0.83	0.75	0.78	0.74	0.80	0.83	0.79
$I2_12_12_1$											
Peak	2.26	1.52	1.17	1.01	1.02	0.94	0.89	0.85	0.80	0.74	0.74
Inflection	1.46	0.95	0.84	0.78	0.72	0.75	0.77	0.79	0.83	0.75	0.73
Correlation coefficient between the signed anomalous differences against resolution CC $(F^+ - F^-)i$, $(F^+ - F^-)j$											
$I4_122$											
Peak/inflection	84.5	55.8	39.0	30.7	16.0	23.9	18.2	3.5	-0.3	11.1	16.5
$I2_12_12_1$											
Peak/inflection	58.9	43.7	29.0	20.1	8.3	12.8	16.0	7.3	-0.4	5.8	10.6

Table 2-2 Anomalous signal statistics for the mercury derivative data processed in space-groups $I4_122$ and $I2_12_12_1$. A signal-to-noise ratio greater than 1.2 indicates a significant anomalous signal where 0.8 indicates noise. Data were truncated where the correlation coefficient between the signed anomalous differences was greater than ~30%.

2.1.1.1.1.2.   MAD phasing

For data processed in $I4_122$, SHELXD found 3 strong heavy-atom positions and an additional 10 weaker positions, with a sharp drop off in occupancy between the third and fourth sites (68% and 34%). The heavy-atom substructure was passed to SHARP (La Fortelle and Bricogne, 1997) for maximum-likelihood heavy-atom parameter refinement followed by density modification with SOLOMON (Abrahams and Leslie, 1996) using a solvent content of 60%, resulting in a final correlation coefficient on $|E^2|$ of 0.623. The resulting map, whilst noisy, had readily interpretable regions of α-helical secondary structure encompassing the top three heavy-atom positions, with the remaining predicted heavy-atom sites located in areas of disordered density. Despite the interpretable features in the map, the

solution was not in agreement with the analysis of the data. Only three monomers in the asymmetric unit would mean a Matthews coefficient of 12.28 and a predicted solvent content of 90%. In addition, whilst the solution was consistent with the self-rotation function (Figure 2-5), the large non-origin Patterson peaks were not (Figure 2-6). Regions of disordered density were observable in areas consistent with the native Patterson vectors (Figure 2-7a) but attempts to locate further heavy-atoms failed.
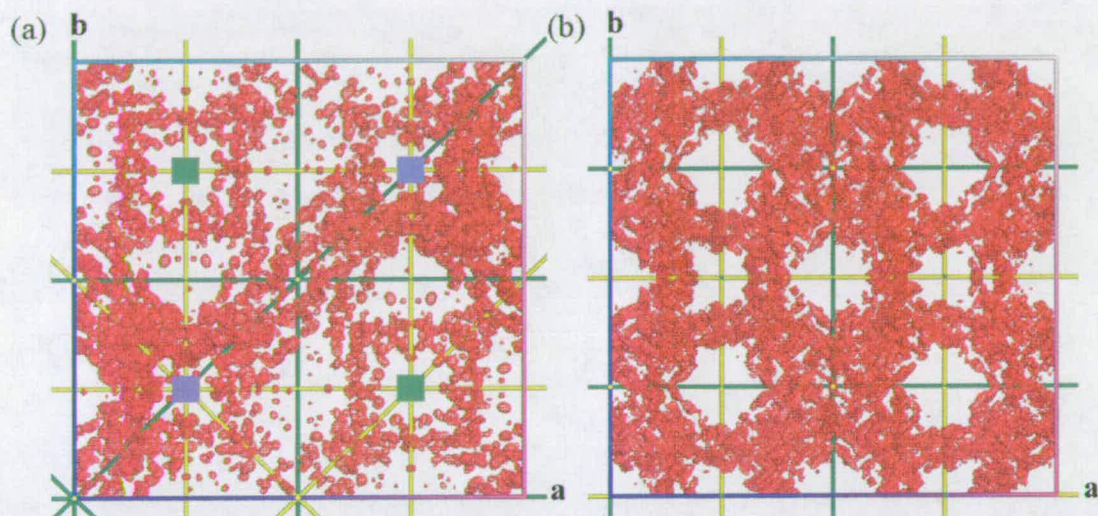
Figure 2-7 Experimental electron-density maps. (a) Electron-density for data processed in $I4_122$. Interpretable electron density only accounts for 10% of the unit-cell, with disordered density evident, related by translations consistent with the native Patterson analysis. (b) Electron-density for data processed in $I2_12_12_1$. Density accounts for all predicted 24 monomers in the asymmetric unit. Unit-cell translated one quarter along b axis compared to (a). 2-fold axes in yellow, 2_1-screw axes in green, 4_1-screw axes indicated by green panel, 4_3-screw axes indicated by blue panel.

Repeating the procedure with data processed in space-group $I2_12_12_1$, SHELXD located 27 heavy-atom positions with occupancies ranging from 100% to 13%, with 15 of the sites greater than 50% and no clear drop off in occupancy. Heavy-atom substructure refinement with SHARP resulted in 3 sites being discarded and density modification with SOLOMON lead to a map with a final correlation coefficient on $|E^2|$ of 0.684. The resulting map had a clear protein-solvent boundary with the disordered regions from the $I4_122$ solution now interpretable and all 24 heavy-atoms located in regions of protein density (Figure 2-7b). The crystal-lattice is made up of four identical sub-lattices related by the non-crystallographic translations defined by the native Patterson analysis (Figure 2-6).

2.3.1.3. Model building and refinement

The electron density maps resulting from MAD phasing were of sufficient quality to begin model building. It soon became apparent that the *C. elegans* C-terminal subdomain adopted an alternative conformation to the closely related rat homologue and resembled the more distantly related bacterial DnaK from *E. coli*. Furthermore, close comparison of the *C. elegans* electron density with the rat and *E. coli* structures revealed the rat structure had undergone a 3D domain-swap (discussed in chapter 3). The significance of this was that a composite monomer, consisting of helices αB and αC (Mse^{541}-Gln^{585}) from rat chain A and helices αD and αE (Glu^{588}-Ser^{613}) from rat chain B, could be generated that fit the electron density very well.

The "search for model in map" option in MOLREP was used to position all 24 rat composite monomers using the experimentally derived phases (Figure 2-8a). The resulting model had Cys^{574}, conserved with Cys^{575} in Hsp70A, positioned 3 Å from a mercury atom (Figure 2-8b). Rigid-body refinement in REFMAC, using the Hendrickson-Lattman coefficients as input, with each monomer as a separate rigid body followed by restrained refinement with tight NCS restraints, overall B-factors and inclusion of the 24 heavy-atom positions yielded a starting R_{cryst}/R_{free} of 49.5/50.1%. In addition, density was also evident for the unmodelled N-terminal residues and the loop connecting helices αC and αD missing from the rat structure. To take advantage of the better diffracting form I-II crystals (3.5 Å compared to 4 Å) the initial model was used as a molecular replacement search model using the form I-II data. Rigid-body and restrained refinement, as before, yielded starting R_{cryst}/R_{free} values of 46.2/47.2%.

Due to the problems in refining a structure with 24 monomers in the asymmetric unit at poor resolution, optimisation of crystallisation conditions was carried out in parallel and a higher symmetry better diffracting form was obtained. Consequently, refinement of form I crystals was put on hold until a better model was obtained (section 2.5.5).

(a)

(b)
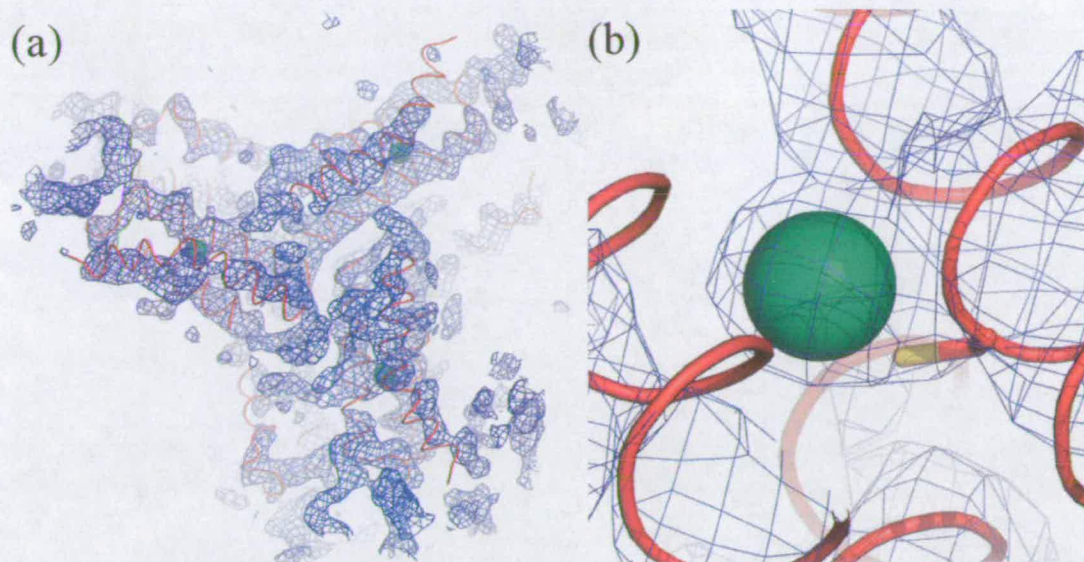
Figure 2-8 MAD phasing of ceHsp70-CT form I data. (a) Electron density map generated using experimentally derived phases (section 2.1.1.1.1.2. with rat Hsc70 composite monomer fitted with MOLREP. Hg positions marked in green. (b) Rat structure fitted with Cys-574 (conserved with Cys[575] in C. elegans) 3 Å from Hg position.

2.3.2. Solving the structure of tetragonal form II ceHsp70-CT crystals

Refinement of the orthorhombic crystal form was hampered by low resolution and multiple monomers in the asymmetric unit. Based on the pseudo-tetragonal and pseudo-cubic nature of the crystal lattice it was hypothesised that only a slight modification in the crystal packing would be required to adopt a higher symmetry space-group. This would reduce the number of monomers in the asymmetric-unit making refinement easier and hopefully improve the resolution of diffraction.

2.3.2.1. Optimisation of crystallisation conditions

Optimisation of crystallisation conditions in order to find higher symmetry better diffracting crystals was carried out using an additive screen described by Birtley and Curry (Birtley and Curry, 2005). The simple screen samples multiple new slightly transformed conditions by mixing 75% existing conditions with 25% Hampton Crystal Screen™ conditions. Crystals of ranging quality were observed in approximately 40% of the new conditions. A common additive that appeared to be beneficial for crystallisation was various molecular weight PEGs. In particular, four conditions produced nice, single crystals with a similar octahedral habit seen for the orthorhombic crystal form (Figure 2-3). Condition 32, containing only 2 M ammonium sulphate, produced small single crystals, unsurprising since the final contents are virtually identical to the original condition. Conditions 16 (100 mM HEPES pH 7.5 and 1.5

M lithium sulphate) and 33 (4 M sodium formate) also produced good quality crystals; however, crystals from condition 33 offered no improvement in resolution and symmetry (Figure 2-9a) and crystals from conditions 16 diffracted very poorly (Figure 2-9b). Finally, condition 39 (100 mM HEPES pH 7.5, 2 M ammonium sulphate and 2% PEG 400) resulted in multiple single crystals. These diffracted beyond 3.5 Å with some spots observed at around 3 Å (Figure 2-9c). Furthermore, preliminary indexing suggested a primitive tetragonal space-group.
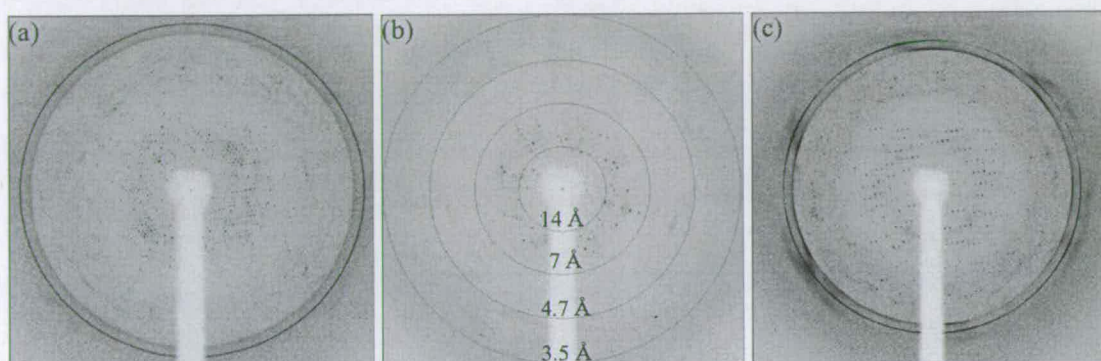
Figure 2-9 Diffraction images for new crystals. (a) Diffraction for crystal grown from Hampton Structure Screen condition 16. Same space-group as before. Ice-rings at 3.9 and 3.65 Å. (b) Diffraction for crystal grown from Hampton Structure Screen condition 33. (c) Diffraction for crystal grown from Hampton Structure Screen condition 39. Ice-rings at 3.9 and 3.65 Å. Data is tetragonal and extends to ~3 Å.

2.3.2.2. Analysis of X-ray data

Indexing in MOSFLM suggested a primitive tetragonal space-group with a clear gap in the solution penalty to the next best solution. Data were processed in space-group P4 with a unit-cell dimensions of a = b = 139 Å, c = 100.6 Å. Analysis of the unmerged intensities processed in P4 with POINTLESS (Evans, 2006) confirmed the Laue group P 4/m m m with a four-fold rotation axis parallel to the unit-cell c axis and two-fold axes down a, b and c. Inspection of systematic absences revealed l = 2n and h = 2n consistent with space-group $P4_22_12$. Data were processed in space-group $P4_22_12$ and scaled with an R_{sym} of 13%. Data collection and processing statistics can be found in table 2-3.

Data collection statistics	
Wavelength (Å)	0.978
Space-group	$P4_22_12$
Unit-cell parameters (Å)	$a = b = 138.9$, $c = 100.6$
Resolution range (Å)	36 – 3.2 (3.37 – 3.2)
No. observations	146865 (21737)
No. unique reflections	16809 (2399)
Completeness (%)	99.9 (100)
Redundancy	8.7(9.1)
$R_{sym}{}^a$ (%)	13.6 (93.6)
$R_{p.i.m.}{}^b$ (%)	5.1 (33.2)
$I/\sigma(I)$	12.9 (2.0)

$$a\ R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i |I_i(hkl)| \quad b\ R_{p.i.m.} = \sum_{hkl}[1/N-1]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i |I_i(hkl)|$$

Table 2-3 Reflection data statistics for data processed in space-groups $P4_22_12$. Values in parentheses are for the highest resolution bin.
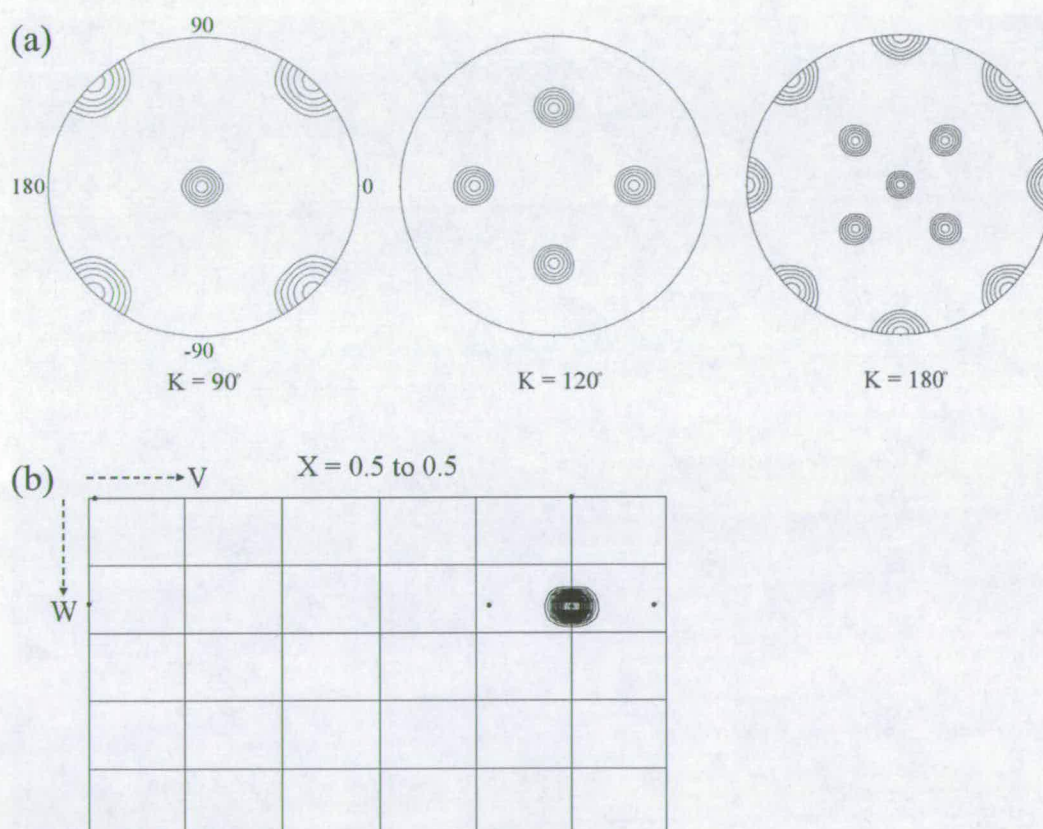
Figure 2-10 Self-rotation and native Patterson analysis of form II data processed in space-group $P4_22_12$. (a) Self-rotation map showing peaks at κ = 90, 120 and 180°. (b) Native Patterson map (0 < v < 0.6, 0 < w < 0.5) showing large non-origin peak at (0.5, 0.5, 0.16).

2.3.2.2.1. Content of the asymmetric unit

For a protein with molecular weight 13094 Da, the Matthew's equation (Matthews, 1968) indicates there to be between 4 ($V_m = 4.64$, solvent = 73%) and 10 monomers ($V_m = 1.86$, solvent = 34%) per asymmetric unit. Calculation of a self-rotation Patterson reveals the same pseudo-cubic general packing as the orthorhombic form (Figure 2-5) but with a 90° rotation of the unit-cell around the c-axis (Figure 2-10a). Non-crystallographic translational symmetry is also evident (Figure 2-10b) with a strong peak roughly half that of the origin present at (0.5, 0.5, 0.16).

The unit-cell volume (1.9×10^6 Å^3) is one-quarter that of the $I2_12_12_1$ form (7.7×10^6 Å^3) and was predicted to consist of one of the four hexamers witnessed in the form I asymmetric-unit. Consequently, some of the non-crystallographic translations between the hexamers are now predicted to have been transformed to crystallographic translations of the unit-cell.

2.3.2.3. Phasing

Based on the predictions of the content of the asymmetric-unit, molecular replacement was carried out using one hexamer from the $I2_12_12_1$ solution as a search model. Using PHASER, three clear solutions were found with log-likelihood gains in excess of 3000 and Z-scores over 60 indicating a very significant solution. As expected, the three solutions, generated by rotation of the hexamer about the three-fold NCS axes, revealed the same crystal packing as before.

2.3.2.4. Model building and refinement

Model building and refinement was carried out with REFMAC (Collaborative Computational Project, 1994) and COOT (Emsley and Cowtan, 2004). An initial round of rigid body refinement was carried out with the molecular replacement solution resulting in an R_{cryst}/R_{free} of 51.6/52.4%. Restrained refinement using tight main-chain and side-chain non-crystallographic restraints and an overall B-factor was continued until convergence at an R_{cryst}/R_{free} of 46.0/47.6%. At this stage the model still contained the rat sequence and was incomplete, with the ordered N-terminal tag residues and 2 residues linking helices αC and αD absent. Incorrect residues were mutated and positioned using the rotamer library from within COOT followed by refinement to an R_{cryst}/R_{free} of 39.9/41.3%. Loop 2 residues linking helices αC and αD were built and refined to R_{cryst}/R_{free} of 38.1/39.8. Finally, 8 residues belonging to the N-terminal affinity tag were added, 5 of which belonged to helix αB, and the complete model was refined to an R_{cryst}/R_{free} of 32.0/33.3%. TLS (translation-libration-screw) refinement was used in the final rounds of refinement with pronounced

results. Use of TLS parameters allows the modelling of anisotropic atomic displacement factors describing a rigid group and is especially suited to medium to low resolution refinement due to the low parameter to observation ratio (Painter and Merritt, 2006; Winn et al., 2001). The number of TLS groups to include in the refinement was assessed with the TLSMD server (Painter and Merritt, 2006) and one TLS group for each monomer was used, leading to an R_{cryst}/R_{free} of 27.6/29.0%. The progress of the R-factors throughout the refinement is charted in figure 2-11.
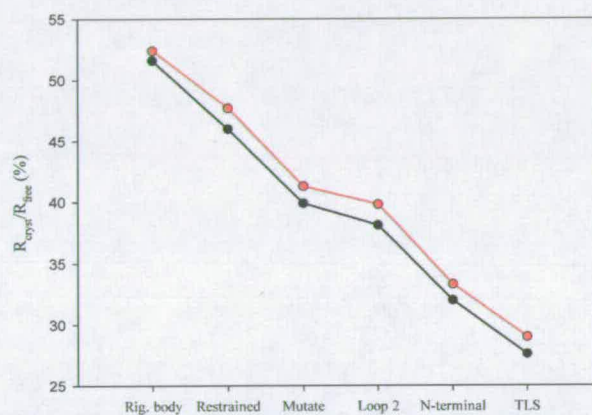
Figure 2-11 Refinement of ceHsp70-CT against form II data. R_{cryst} (black) and R_{free} (red) throughout refinement of model against form II data.

The final model comprises 6 protomers of 82 residues each (total 3966 atoms). No water molecules were included due to the resolution of the data. The model shows good geometry and is supported by R_{cryst} and R_{free} values of 27.6% and 29.0% respectively for all data measured between 36-3.2 Å. The final statistics for refinement and structural details can be found in table 2-4.

The quality of the final model was assessed with PROCHECK (Laskowski et al., 1993). The Ramachandran plot (Figure 2-12) shows that 75.7% of non-glycine and non-proline residues are in the most favoured regions, 16.2% in the additionally allowed regions and 8.1% in the generously allowed regions. No residues are in the disallowed regions. This is better than a typical structure of 3.2 Å resolution which is expected to have 61.7% ± 10% of residues in the most favourable region. Comparison with the secondary structure reveals that the most favoured regions correlate well with the helical areas whilst the allowed and generously allowed residues are largely confined to the connecting loops (Figure 2-12b and c).

Refinement against form I-II data (section 2.4.5) was continued using the final model. The four hexamers were positioned using SSM in COOT and refined with one TLS group per monomer to a final R_{cryst}/R_{free} of 28.7/32.0% (Table 2-4).
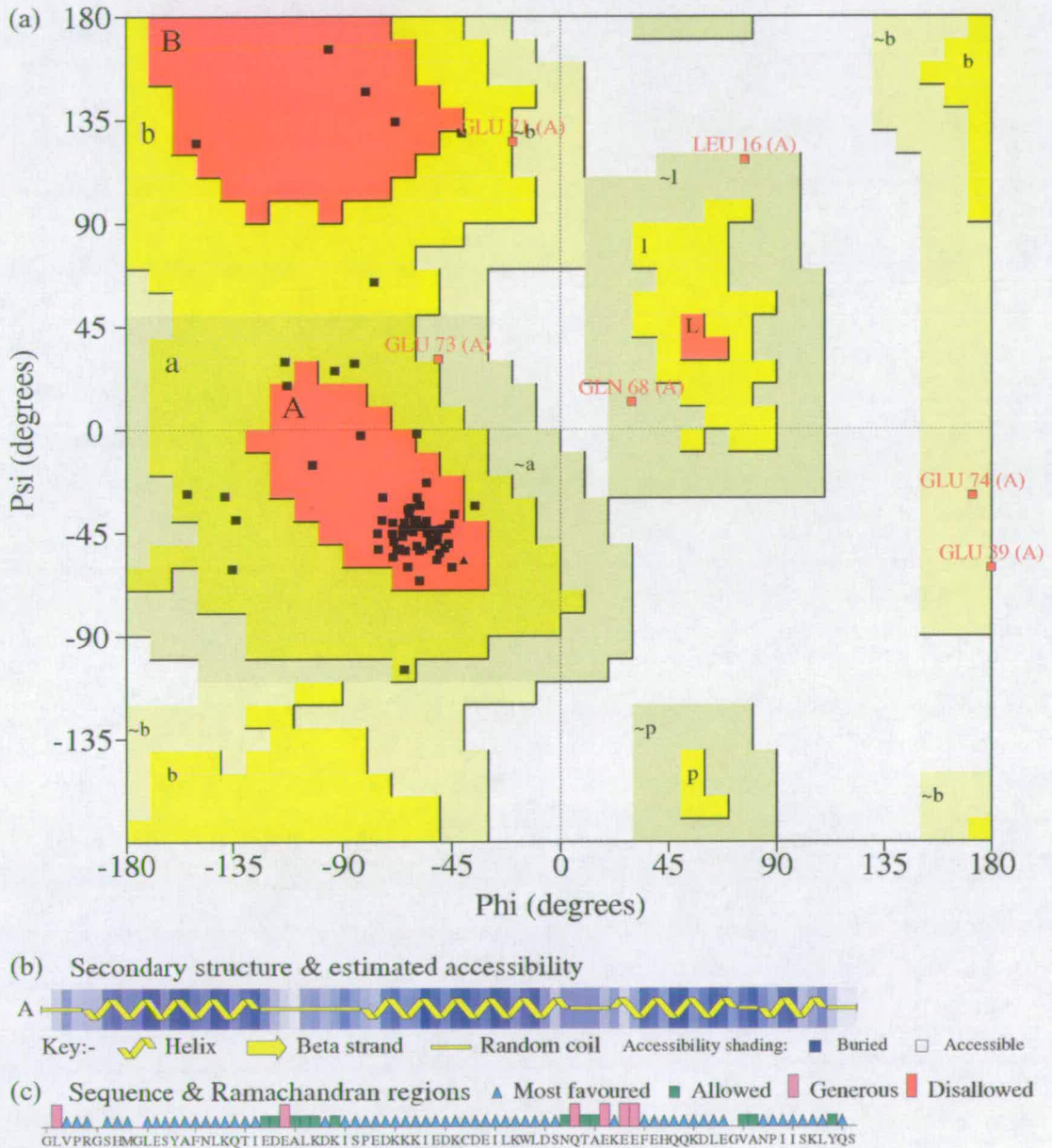
Figure 2-12 Analysis of sterochemical properties of ceHsp70-CT model with PROCHECK. (a) Ramachandran plot. 75.7% of residues in the most favoured region. (b) and (c) Secondary structure prediction and sequence Ramachandran regions. Allowed and generously allowed residues located in and around loop regions.

Crystal	Form I-II	Form II
Space-group	$I2_12_12_1$	$P4_22_12$
Unit-cell dimensions (Å)	a = 196, b = 196.1, c = 200	a = b = 139, c = 100.6
Resolution range (Å)	40 - 3.5	36 - 3.2
$R_{cryst}{}^1$/ $R_{free}{}^2$	28.7 / 32	27.6 / 29
Average B-factor (Å^2)	111	88
r.m.s.d. bonds (Å) /angles (°)	0.017 / 1.744	0.015 / 1.489
Ramachandran plot		
Most favoured (%)	75.3	72.1
Additionally allowed (%)	13.7	22.6
Generously allowed (%)	9.6	5.3
Disallowed	1.4	0

$$_1 \quad R_{cryst} = \sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right| \sum_{hkl} |F_{obs}|$$

^{2} R_{free} as R_{cryst} but summed over a 5% test set of reflections

Table 2-4 Refinement statistics for form I and form II ceHsp70-CT crystals.

2.3.3. Description of the ceHsp70-CT crystal structure

ceHsp70-CT was crystallised as a recombinant protein, incorporating a 2.3 kDa (21 residue) vector encoded N-terminal 6xHis tag, in two forms; an orthorhombic form belonging to space-group $I2_12_12_1$ (form I; section 2.3.1.) and a tetragonal form belonging to space-group $P4_22_12$ (form II; section 2.3.2.). Both the monomeric structure and crystal-lattice packing are virtually identical between the two crystal forms. The asymmetric-unit of form I crystals consists of four hexamers related by translational non-crystallographic symmetry whilst the higher symmetry form II crystals only have one hexamer per asymmetric-unit. Discussion of the monomeric and asymmetric-unit structure will concentrate on the form II structure but applies equally to both forms.

The final model consists of six protomers per asymmetric unit. Most residues are well modelled except the first 12 N-terminal residues, encompassing the 6xHis sequence, and the last 26 C-terminal residues. Tight NCS restraints were applied throughout refinement and all six ceHsp70-CT protomers are identical with RMSDs <0.05 Å. The six protomers form two back-to-back trimers related by 32 point group symmetry with an NCS three-fold axis through the centre of the trimer and three orthogonal evenly spaced two-fold axes relating the two trimers (Figure 2-13a).
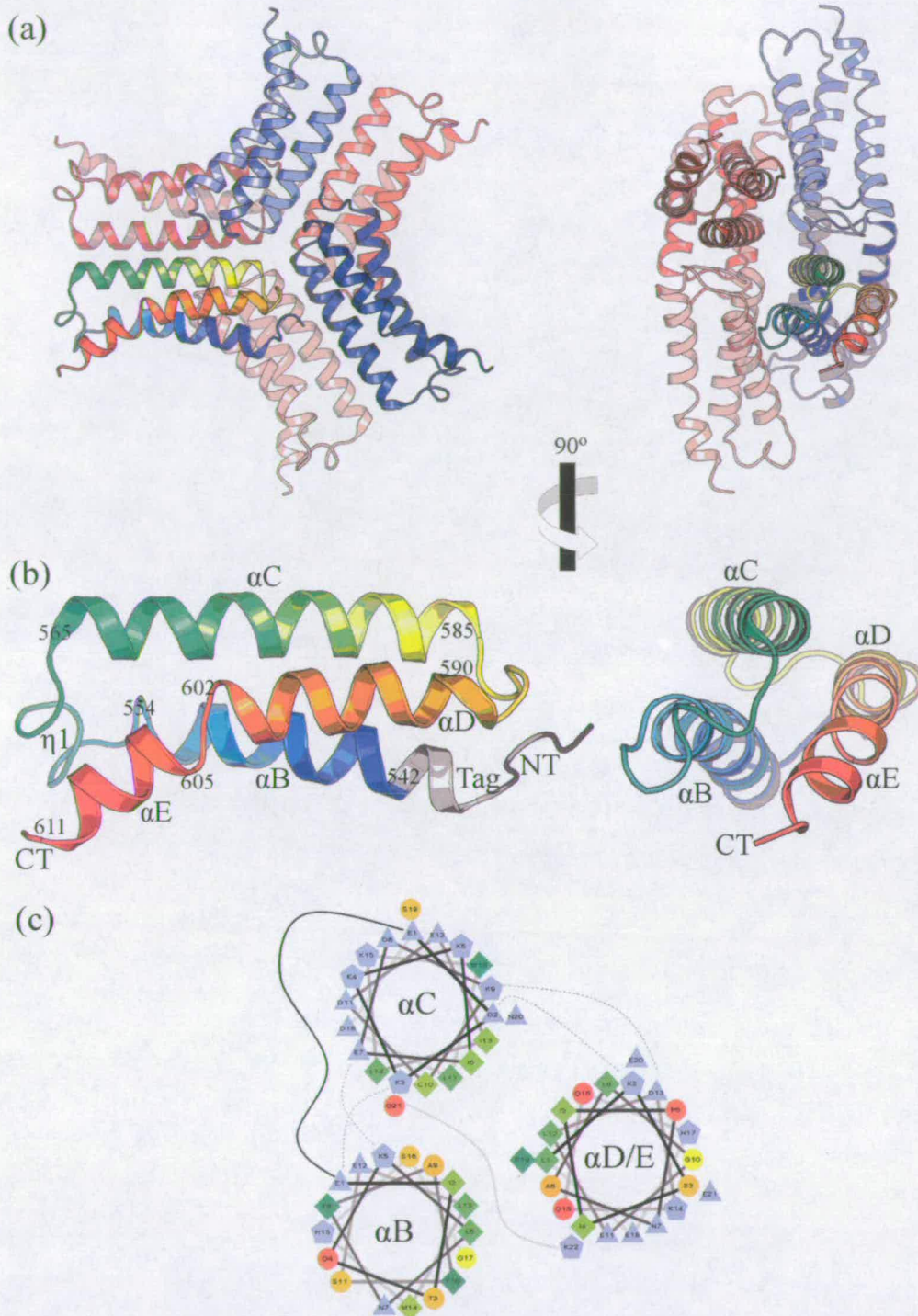
Figure 2-13 Asymmetric-unit and monomeric ceHsp70-CT structure. (a) Structure of the ceHsp70-CT asymmetric unit viewed down the three-fold NCS axis and the orthogonal two-fold NCS axis. Asymmetric unit consists of six protomers arranged as back-to-back trimers, coloured red and blue. One monomer coloured in a gradient from N-terminal (blue) to C-terminal (red). (b) Monomeric structure of ceHsp70-CT. Coloured in a gradient from N-terminal (Blue) to C-terminal (red). ceHsp70-Ct contains four helices, αB–αE (beginning/end position numbered), which form a three-helix bundle. The loop connecting helices αB and αC contains a short 3_{10}-helix (η1). The recombinant 6xHis tag contributes five residues to helix αB (coloured grey). (c) Helical-wheel diagram showing hydrophobic packing in the core of the structure and intra-chain electrostatic interactions (indicated with dotted lines). Green – hydrophobic, blue – charged, red – polar.

Within each monomer the secondary structure is all helical, comprising four α-helices, αB-αE (Table 2-5) and a helical content of 72% according to PROMOTIF (Hutchinson and Thornton, 1996). Successive interhelical angles of 153°, 164° and 32° produce an anti-parallel three-helix bundle, with helices αB-αD arranged in an anti-clockwise up-down-up topology. Helix αE is contiguous with helix αD but kinked 32° at Ala[604] and extends under the loop connecting helices αB and αC (Figure 2-13b). The helices have a classical amphipathic nature with a well defined hydrophobic core and are stabilised by intra- and inter-chain electrostatic interactions (Figure 2-13c). The primary structure of the three helices is similar to the heptad repeat motif found in coiled-coils, with predominantly hydrophobic residues located at the first and fourth positions.

Helix	Range (# res)	Length (Å)	Sequence
αB	(-5)542-554 (18)	27.62	(P)RGSHMGLESYAFNLKQTI(E)
αC	565-585 (21)	30.80	(S)PEDKKKIEDKCDEILKWLDSN(Q)
αD	590-602 (13)	19.92	(E)KEEFEHQQKDLEG(V)
αE	605-611 (7)	10.41	(A)NPIISKL(Y)

Table 2-5 Secondary structure content of ceHsp70-CT.

Residues	Sequence	Type	i to i+3 dist. (Å)
556-559	DEKL	IV	5.2
557-560	EKLK	IV	5
559-562	LKDK	I	5.7
560-563	KDKI	I	6.4
585-588	NQTA	IV	5.9

Table 2-6 ceHsp70-CT β-turns. First four turns form a short stretch of 3_{10}-helix in other C-terminal structures. Final turn is in loop connecting helices αC and αD.

Analysis with PROMOTIF also reports the presence of five β-turns (Table 2-6). The first four are consecutive turns positioned in the loop connecting helices αB and αC (Residues Asp[556]–Ile[563]). Comparison with the same regions from homologous structures from rat (Chou et al., 2003) and *E. coli* (Zhu et al., 1996) reveal that this region forms two turns of 3_{10}-helix (labelled η1 in figure 3-1a). The final β-turn is located immediately at the C-terminal of helix αC (residues Asn[585]-Ala[588]) and forms the hairpin turn allowing the structure to fold back on itself.

In accordance with solution studies of *E. coli* DnaK (Bertelsen et al., 1999) and the crystal structure of rat Hsc70 (Chou et al., 2003), the final 26 C-terminal residues were found to be

disordered. This highly mobile region is enriched in glycine and proline residues in many Hsp70 family members and contains the conserved co-chaperone binding GPTIEEVD motif at the extreme C-terminus.

Nine residues of the recombinant 6xHis tag are visible in the electron density. Interestingly, the five immediately preceding the start of the *C. elegans* sequence (residue 542) adopt a helical secondary structure and form the beginning of helix αB (Figure 2-12b, coloured grey) and contribute side-chains to the hydrophobic core of the three-helix bundle.

2.3.4. Comparison of the orthorhombic and tetragonal crystal lattice

Both orthorhombic form I crystals and tetragonal form II crystals have the same general packing. Form I crystals belong to space-group $I2_12_12_1$ with unit-cell dimensions $a = 194.6$, $b = 195.0$, $c = 200.8$ Å whilst form II crystals belong to space-group $P4_22_12$ with unit-cell dimensions $a = b = 138.9$, $c = 100.6$ Å.

2.3.4.1. Description of the orthorhombic form I crystal-lattice

Form I crystals have 24 monomers in the asymmetric-unit. These can be defined as four hexamers, as described in section 2.6.1, in the same orientation related by translation non-crystallographic symmetry (Figure 2-14a). Translational NCS was first suggested by the large non-origin peaks in the native Patterson map (section 2.4.3.2; Figure 2-14b) approximately one-quarter the height of the origin peak at (0.4864, 0.0000, 0.4154), (0.5000, 0.0193, 0.0847) and (0.0142, 0.0200, 0.5000).
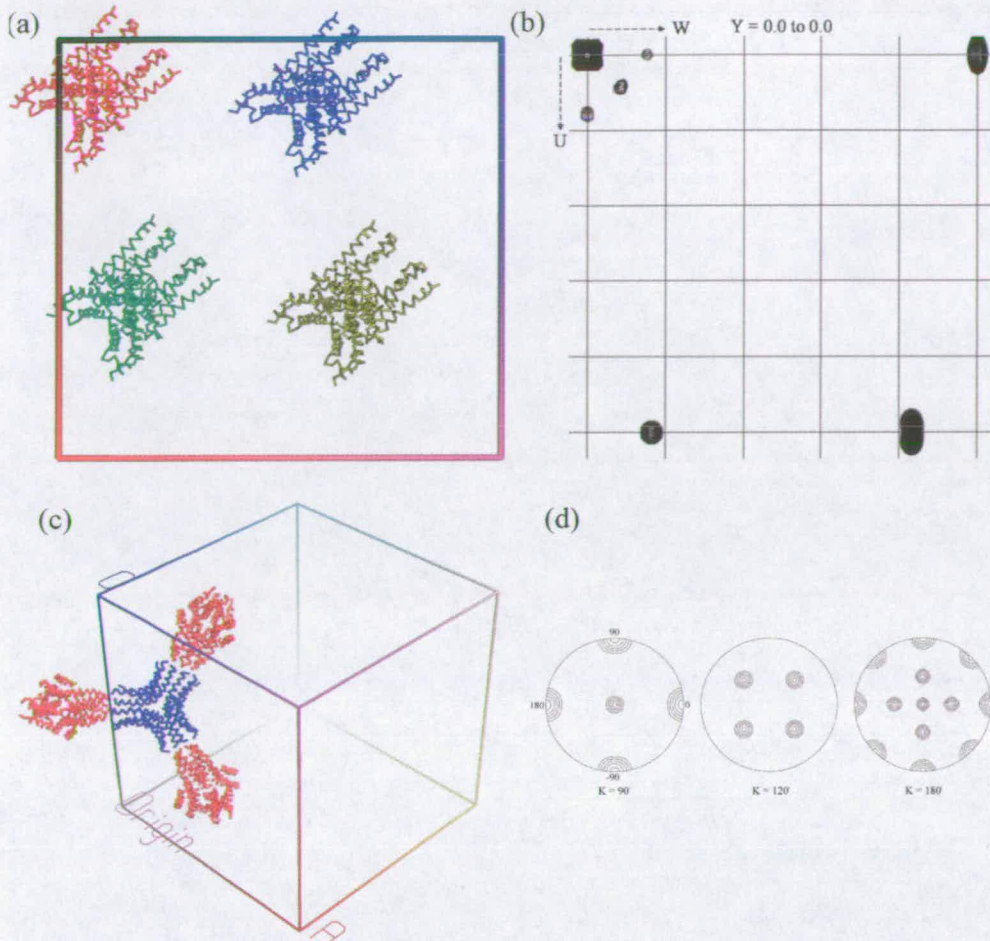
Figure 2-14 Translational and rotational NCS of the asymmetric-unit in space-group $I2_12_12_1$. (a) The asymmetric-unit can be defined as four hexamers related by the native Patterson vectors in (b). (c) Each hexamer is aligned with its local three-fold NCS rotation axis parallel to the unit-cell body diagonal. Each hexamer is related to three neighbouring hexamers by two-fold NCS axes parallel to the unit-cell edges or face diagonals. (d) The two- and three-fold NCS is illustrated in the self-rotation map with peaks at $\kappa = 120°$ and $90°$.

Each hexamer, a pair of back-to-back trimers, has 32 point-group symmetry and is orientated such that its local three-fold rotation axis is aligned parallel to a unit-cell body diagonal (Figure 2-14c). Each hexamer belongs to a distinct sub-lattice and is related to three neighbouring hexamers by a two-fold rotation axes parallel to either the unit-cell edges or the face diagonals (Figure 2-14c). Both these two-fold axes, and the three-fold axes concomitantly generated aligned parallel to each unit-cell body diagonal, are non-crystallographic in space-group $I2_12_12_1$, with the two-fold NCS axes parallel to the cell edges also parallel to crystallographic two-folds. This is nicely illustrated in the self-rotation Patterson map which shows 432 point-group symmetry with large peaks at $\kappa = 90°$, $120°$ and $180°$ (Figure 2-14d).
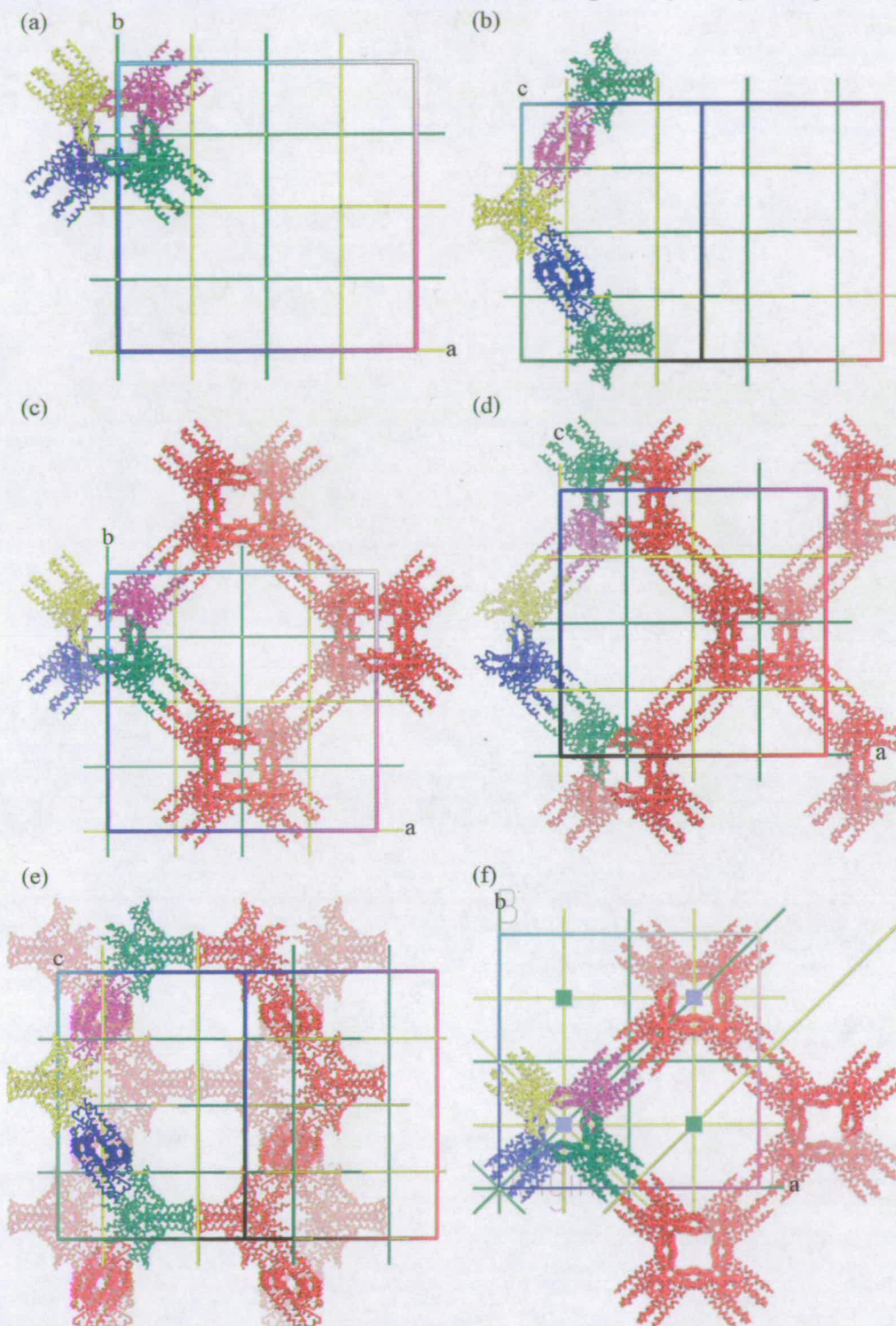
Figure 2-15 Crystal packing of one lattice in space-group I2₁2₁2₁. (a) and (b) The asymmetric-unit can be defined as four hexamers (coloured green, blue, yellow and purple) related by a 90° rotation and one-quarter unit-cell translation along the c-axis forming a single-stranded helical turn here viewed down the c-axis (a) and the b-axis (b). (c) Sub-lattice packing viewed down in the c-axis. (d) Sub-lattice packing viewed down in the b-axis. (e) Viewed parallel to ab face diagonal. (f) This sub-lattice also has I4₁22 symmetry with the 4₃-screw axis (blue box) aligned down the helices and the 4₁-screw axis aligned between in minor helices.

The asymmetric-unit can also be defined as four hexamers belonging to the same sub-lattice (Figure 2-15a and b). Successive hexamers - coloured green, blue, yellow and purple - are related by a unit-rise of ~50 Å along the c-axis. Coupled with a 90° rotation, this generates a single-stranded left-handed helix extending parallel to the c-axis with equivalent positions defining the unit-cell dimension (~200 Å) (Figure 2-15c). This can be thought of like a left-handed four-sided staircase. This packing is repeated along both the a- and b- axes (Figure 2-15d), however, the unit-rise in both cases is ~49 Å giving rise to both the a- and b- axes being approximately 4 Å shorter than the c-axis. This creates a honeycomb like sub-lattice packing (Figure 2-15e) and accounts for approximately 15% of the unit-cell.

Interestingly, this sub-lattice has $I4_122$ symmetry with the unit-cell shifted (0.25, 0, -0.125) (Figure 2-15f). The longitudinal helices along the c-axis are related by the 4_3-screw axis whilst the 4_1-screw axis describes the relationship between these helices. $I4_122$ is a maximal non-isomorphic super-group of $I2_12_12_1$ and there was an ambiguity during space-group determination with the data collected from form I crystals flash-cooled directly from mother liquor scaling equally well in $I2_12_12_1$ or $I4_122$ (section 2.4.3). Furthermore, MAD phasing using data processed in $I4_122$ gave a clear solution corresponding to this sub-lattice (section 2.4.4.2.2).

The sub-lattice is repeated four times in total, with hexamers in the same orientation in each sub-lattice related by the native Patterson vectors as described in figure 2-14b. Two of the sub-lattices are related by the NCS translation vector (0.0142, 0.02, 0.5). When viewed down the c-axis these two sub-lattices overlay (Figure 2-16b and c; red and blue sub-lattices) and, due to the one-half unit-cell translation along c, form a left-handed double-stranded helix running down the c-axis (Figure 2-16d). Due to the accompanying small translations along the a- and b- axes, these sub-lattices overlay imperfectly and are only related by a two-fold rotation axis. This imperfect packing breaks the four-fold screw axis and explains why the $I4_122$ symmetry does not hold for the complete crystal-lattice. The remaining two sub-lattices, related by the same packing, are generated by the NCS translation vector (0.5000, 0.0193, 0.0847) (Figure 2-16e and f; green and olive sub-lattices).
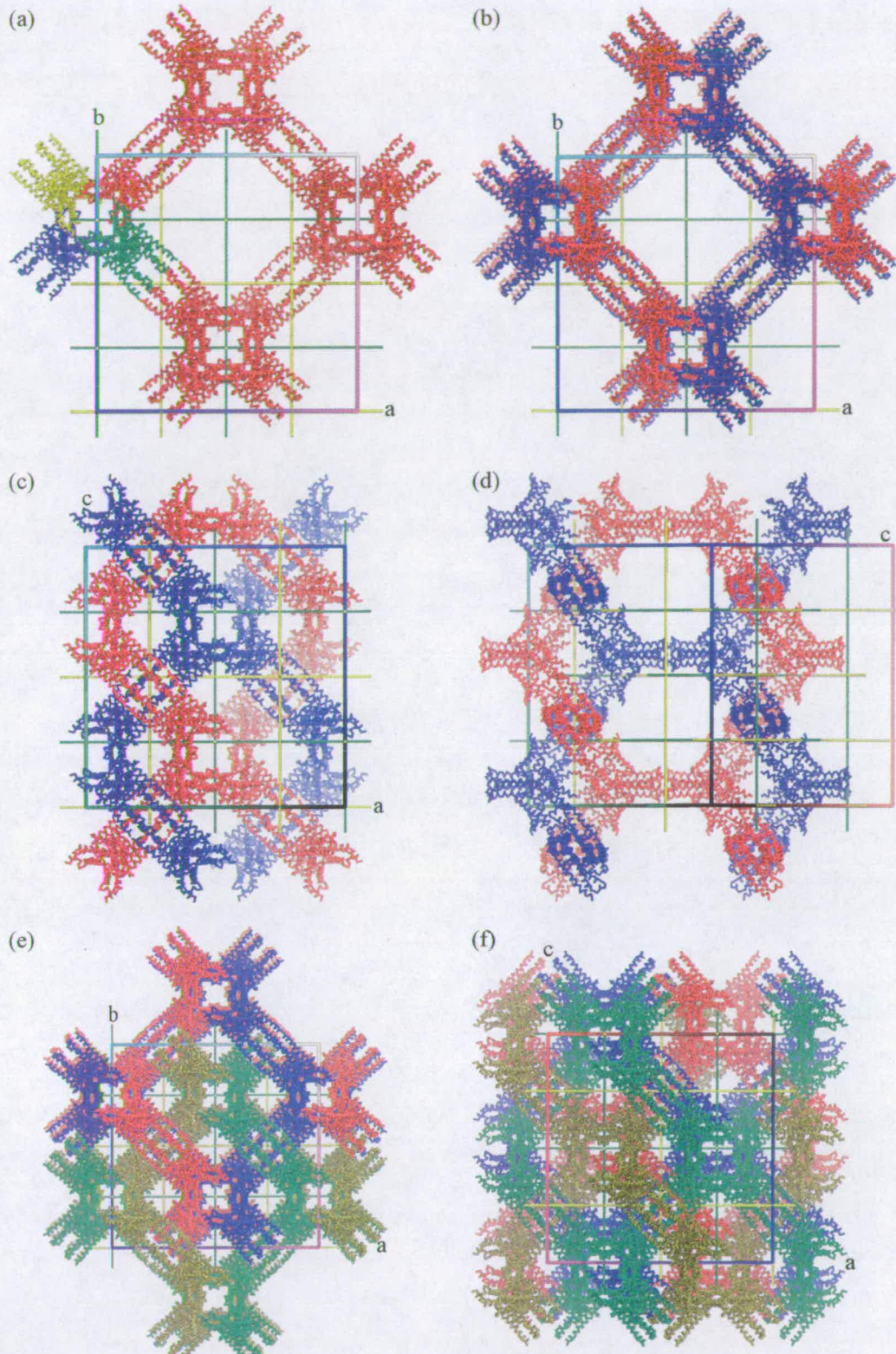
Figure 2-16 Packing of sub-lattices in space-group I2₁2₁2₁. (a) One sub-lattice viewed down unit-cell c-axis, hexamers of one helical turn coloured blue, yellow, red and green. (b) Two sub-lattices (red and blue) related by NCS translation vector (0.0142, 0.02, 0.5) viewed down c-axis. (c) Two sub-lattices viewed down b-axis. (d) Two sub-lattices viewed along ab face-diagonal showing the helical relationship along the c-axis. (e) Packing of all four sub-lattices (red, blue, green, olive) viewed down c-axis (f) Packing of all four sub-lattices viewed down b-axis

2.3.4.2. Description of the tetragonal form II crystal-lattice

Tetragonal crystals belonging to space-group $P4_22_12$ were grown from the same conditions by the addition of a small amount of PEG 400. These crystals have the same general packing as form I crystals, as initially evidenced by the related self-rotation Patterson map (Figure 2-5 and 2-10a), however the unit-cell dimensions are now $a = b = 138.9$, $c = 100.1$ Å. The unit-cell volume is one-quarter that of the orthorhombic form and the asymmetric-unit consists of only one hexamer.

Analysis of the crystal-lattice reveals the same packing, however, the non-crystallographic translation relating two sub-lattices along the unit-cell c-axis, described by the native Patterson vector (0.0142, 0.02, 0.5), is now a pure translation along the c-axis. Consequently, the unit-cell c-axis is now defined by trimers in equivalent positions but on the opposite strand of the helix and is half the length as in $I2_12_12_1$. As the sub-lattices overlay perfectly, they are now related by a 4_2-screw axis (Figure 2-17). This also allows the unit-cell to be described by a primitive lattice.
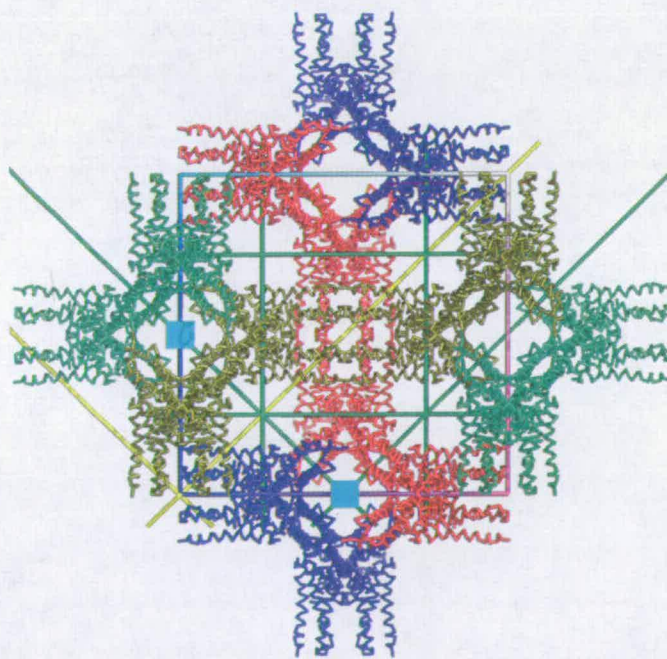
Figure 2-17 Crystal packing in space-group $P4_22_12$. (a) Unit-cell viewed down c-axis. Separate sub-lattices coloured red, blue, green and olive. 4_2-screw axis represented by blue box, 2_1-screw axis in green, 2-fold rotation axis in yellow.

2.3.4.3. Space-group relationships in the ceHsp70-CT crystal lattice

The packing between orthorhombic form I crystals and tetragonal form II crystals is very similar. A slight re-ordering of the intra-sub-lattice packing establishes a four-fold screw axis relating the double-stranded helices extending parallel to the c-axis which are only described by a two-fold rotational symmetry in form I crystals. Consequently, the unit-cell c axis is halved to ~100 Å and coupled with the transformation to a primitive cell means the unit-cell is one-quarter the volume of $I2_12_12_1$.

The $P4_22_12$ unit-cell is related to the $I2_12_12_1$ unit-cell by a 45° rotation about the c axis, first suggested by the self-rotation Patterson analysis, and a translation of (0.25, 0, -0.25). This is clearly illustrated by inspection of the space-group symmetry (Figure 2-18). Rotation of a $P4_22_12$ cell about the c axis orientates the 4_2-screw axes with two-fold rotation axes of space-group $I2_12_12_1$ (2-17b and d), made possible by the re-ordering of the intra-sub-lattice packing. Further, the two-fold rotation axes at the unit-cell corners in $P4_22_12$ are now aligned with 2_1-axes in the $I2_12_12_1$ cell and a translation of -0.25 along the c-axis (not shown) is also required to align the symmetry axes parallel the a and b axes.

Data collected from crystals vitrified directly from mother-liquor were initially thought to belong to space-group $I4_122$ and mercury derivative data processed as such even gave a clear solution corresponding to one of the sub-lattices (section 2.4.3.) The space-group assignment proved to be incorrect but comparison of the space-group symmetry shows the close relationship between the two space-groups (Figure 2-17a and c). An $I4_122$ unit-cell is related to an $I2_12_12_1$ cell of the same size by a translation of the origin by (-0.25,0 ,-0.125). This aligns the tetragonal 4_1- and 4_3- axes with the two-fold rotational axes of the orthorhombic cell. Each sub-lattice can be defined by a body-centred tetragonal cell; however, as discussed, the relative orientation of the two sub-lattices that intertwine down the c axis transforms the four-fold screw axes to two-fold rotation axes.

2.4.   Conclusions

In summary, the structure of the C-terminal subdomain of *C. elegans* Hsp70 has been solved. Orthorhombic $I2_12_12_1$ and tetragonal $P4_22_12$ crystal forms were produced which exhibited very similar crystal packing. The final model, refined to R_{cryst} and R_{free} values of 27.6% and 29.0% respectively, shows a compact three-helix bundle dramatically different in conformation to the only eukaryotic structure from rat Hsc70. Further analysis of the structure is presented in Chapter 3.

(a)

Orthorhombic I2$_1$2$_1$2$_1$
a = 194.6, b = 195, c = 200.8 Å

Tetragonal P4$_2$2$_1$2
a = b = 139, b = 100.1 Å

Tetragonal I4$_1$22
a = b = 195, c = 200.8 Å

(b)

P4$_2$2$_1$2 origin
(0.25,0,-0.25)

(c)

I4$_1$22 origin
(-0.25,0,-0.125)

(d)
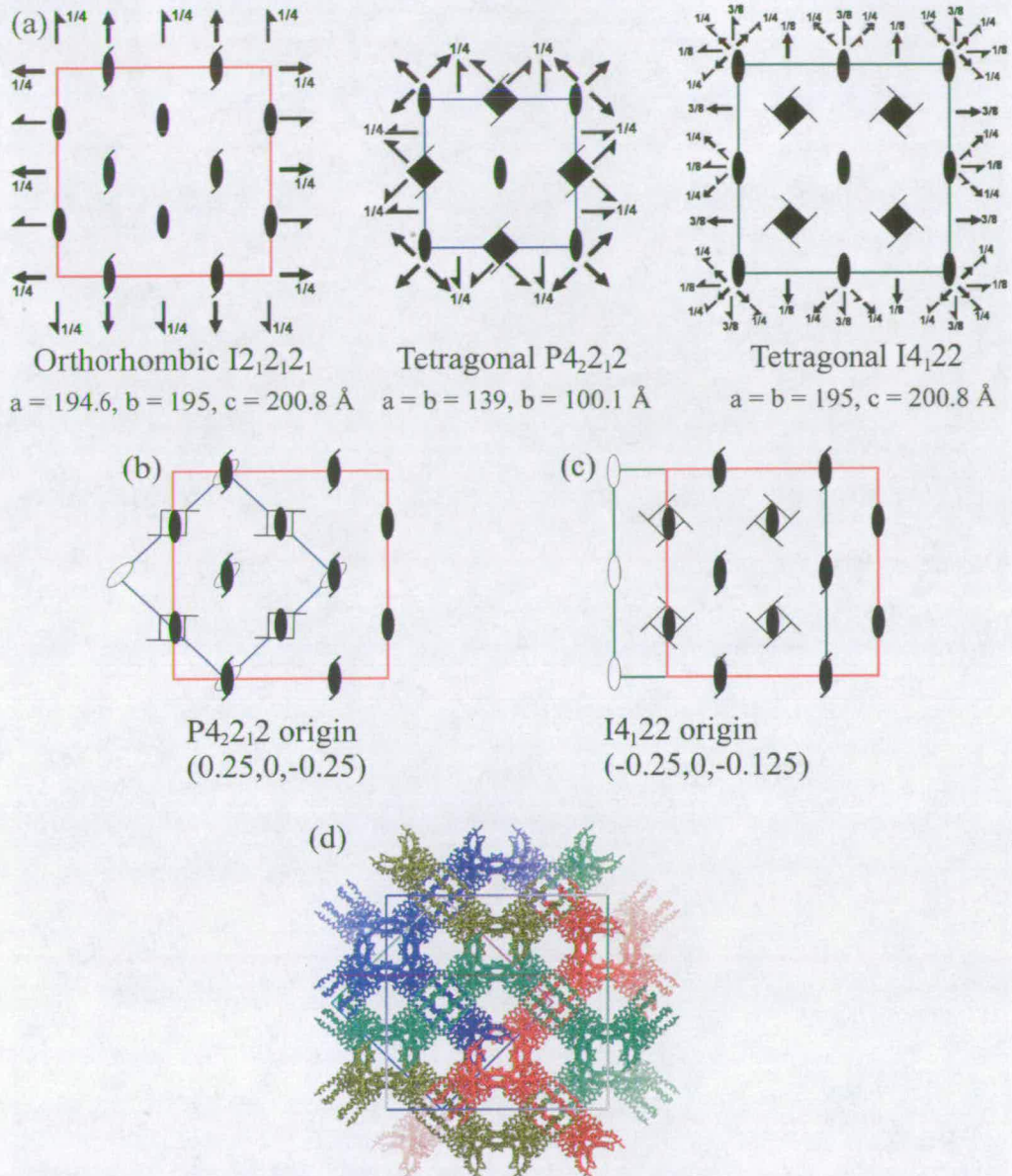
Figure 2-18 Comparison of space-groups I2$_1$2$_1$2$_1$, P4$_2$2$_1$2 and I2$_1$2$_1$2$_1$. (a) Symmetry operators for space-groups I2$_1$2$_1$2$_1$, P4$_2$2$_1$2, I4$_1$22. (b) Overlay of I2$_1$2$_1$2$_1$ (red) and P4$_2$2$_1$2 (blue) with P4$_2$2$_1$2 origin at (0.25, 0, -0.25). 42-screw axes overlay with two-fold rotation axes. (c) Overlay of I2$_1$2$_1$2$_1$ (red) and I4$_1$22 (green) with I4$_1$22 origin at (-0.25, 0, -0.125). 4$_1$- and 4$_3$- screw axes overlay with two-fold axes. (d) Unit-cells for ceHsp70-CT packing in I2$_1$2$_1$2$_1$ and P4$_2$2$_1$2 as in (b) Sub-lattices coloured red, blue, green and olive.

2.5.   References

Abrahams, J. P., and Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Benaroudj, N., Batelier, G., Triniolles, F., and Ladjimi, M. M. (1995). Self-association of the molecular chaperone HSC70. Biochemistry *34*, 15282-15290.

Benaroudj, N., Fouchaq, B., and Ladjimi, M. M. (1997). The COOH-terminal peptide binding domain is essential for self-association of the molecular chaperone HSC70. J Biol Chem *272*, 8744-8751.

Benaroudj, N., Triniolles, F., and Ladjimi, M. M. (1996). Effect of nucleotides, peptides, and unfolded proteins on the self-association of the molecular chaperone HSC70. J Biol Chem *271*, 18471-18476.

Bertelsen, E. B., Zhou, H., Lowry, D. F., Flynn, G. C., and Dahlquist, F. W. (1999). Topology and dynamics of the 10 kDa C-terminal domain of DnaK in solution. Protein Sci *8*, 343-354.

Birtley, J. R., and Curry, S. (2005). Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. Acta Crystallogr D Biol Crystallogr *61*, 646-650.

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

Chou, C. C., Forouhar, F., Yeh, Y. H., Shr, H. L., Wang, C., and Hsiao, C. D. (2003). Crystal structure of the C-terminal 10-kDa subdomain of Hsc70. J Biol Chem *278*, 30311-30316.

Chou, C. C., Wang, C., Sun, Y. J., Shr, H. L., and Hsiao, C. D. (2001). Crystallization and preliminary X-ray diffraction analysis of the 10 kDa C-terminal subdomain of 70 kDa heat-shock cognate protein. Acta Crystallogr D Biol Crystallogr *57*, 1928-1930.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.

Fouchaq, B., Benaroudj, N., Ebel, C., and Ladjimi, M. M. (1999). Oligomerization of the 17-kDa peptide-binding domain of the molecular chaperone HSC70. Eur J Biochem *259*, 379-384.

French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. Acta Crystallogr A *34*, 517-525.

Hahn, T. (2002). International Tables for Crystallography, Vol. A, 5th ed., Dordrecht: Kluwer Academic Publishers.).

Hutchinson, E. G., and Thornton, J. M. (1996). PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci *5*, 212-220.

La Fortelle, E. d., and Bricogne, G. (1997). Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. Methods Enzymol *276*, 472-494.

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

Leslie, A. G. W. (1992). Mosflm. Jnt CCP4/ESF-EACBM Newsl Protein Crystallogr *26*.

Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement

doi:10.1107/S0108767393007597. Acta Crystallographica Section A *50*, 157-163.

Nemoto, T. K., Fukuma, Y., Itoh, H., Takagi, T., and Ono, T. (2006). A Disulfide Bridge Mediated by Cysteine 574 Is Formed in the Dimer of the 70-kDa Heat Shock Protein. J Biochem (Tokyo) *139*, 677-687.

Painter, J., and Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr *62*, 439-450.

Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr D Biol Crystallogr *57*, 1373-1382.

Schneider, T. R., and Sheldrick, G. M. (2002). Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr *58*, 1772-1779.

Sheldrick, G. M. (2004). High-Throughput Phasing with SHELXC/D/E. http://shelxuni-acgwdgde/SHELX/.

Storoni, L. C., McCoy, A. J., and Read, R. J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr *60*, 432-438.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement

doi:10.1107/S0021889897006766. J Appl Crystallogr *30*, 1022-1025.

Wegele, H., Muller, L., and Buchner, J. (2004). Hsp70 and Hsp90--a relay team for protein folding. Rev Physiol Biochem Pharmacol *151*, 1-44.

Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr *57*, 122-133.

Worrall, L., and Walkinshaw, M. D. (2006). Crystallization and X-ray data analysis of the 10 kDa C-terminal lid subdomain from Caenorhabditis elegans Hsp70. Acta Crystallograph Sect F Struct Biol Cryst Commun *62*, 938-943.

Worrall, L., and Walkinshaw, M. D. (2007). Crystal structure of the C-terminal three-helix bundle domain from *C. elegans* Hsp70. Biochem. Biophys. Res. Commun. *Accepted*.

Zhu, X., Zhao, X., Burkholder, W. F., 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.

3. Analysis of the *C. elegans* Hsp70 C-terminal 10 kDa subdomain structure

3.1. Introduction

The Hsp70 C-terminal helical-bundle subdomain is implicated in regulation of client binding, self-association and co-chaperone binding. The only eukaryotic structure solved for the 10 kDa C-terminal lid domain is from rat (Chou et al., 2003), which has an anti-parallel coiled-coil mediated dimer. This is in contrast to the monomeric three-helical bundle observed in the *E. coli* homologue DnaK (Zhu et al., 1996), which shares approximately 17% sequence identity with rat Hsc70.

Chapter 3 presents an analysis of the C-terminal subdomain from *C. elegans* Hsp70. Comparison of the structure with *E. coli* and rat homologues shows structural conservation with the distantly related bacterial proteins and also reveals a domain-swapped dimerisation mechanism for self-association of the C-terminal subdomain.

3.2. Materials and methods

3.2.1. Structural analysis

Evolutionary conservation analysis was carried out with ConSurf (Glaser et al., 2003) using the empirical Bayesian method. Sequence alignments generated with MUSCLE (Edgar, 2004) using a non-redundant dataset with sequences corresponding to the 10 kDa C-terminal lid subdomain of all eukaryotic cytoplasmic Hsp70 proteins found in the UniProt database (for alignment see appendix A.2). Residues coloured according to conservation ranging from 1 (variable) to 9 (conserved).

Electrostatic-potential maps were calculated with APBS (Baker et al., 2001) using a PyMol plug-in (http://www-personal.umich.edu/~mlerner/PyMOL/). Charges were assigned using PDB2PQR (Dolinsky et al., 2004) and an AMBER forcefield (Case et al., 2005).

Analysis of the biological relevance of the oligomeric complexes within the crystal structure was carried out with the server PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html).

A homology model of residues 1–533 of *C. elegans* Hsp70A was produced with SWISS-MODEL using bovine Hsc70 as a template (unwound helix αB residues trimmed from template). The helical subdomain solved in this study was subsequently positioned based on

the SBD structures of *E. coli* DnaK and HscA, with the final eight unmodelled helical residues (Lys^{534}-Asn^{541}) connecting the model and structure filled in manually.

All graphical figures were produced with PyMol (http://www.pymol.org).

3.2.2. Determination of solution oligomeric state of ceHsp70-CT

The oligomeric state of ceHsp70-CT was investigated using gel filtration and glutaraldehyde cross-linking. Gel-filtration was carried out on an AKTA explorer FPLC using a Superdex 75 HR 30/10 column (Amersham Bioscience) at 4 $^{\circ}$C. 200 μL *ce*Hsp70-CT (2 μM, 5μM and 80 μM) in storage buffer (25 mM HEPES pH 7.5, 50 mM KCl, 1 mM DTT) was applied to the column equilibrated in the same buffer and run at 0.5 ml min^{-1}. The column was calibrated with protein standards with sizes ranging from 16.4 Å (13.7 kDa) to 85 Å (669 kDa).

Cross-linking was carried out using the homobifunctional amine reactive cross-linker glutaraldehyde. 5 μg total protein in 15 μl 25 mM HEPES pH 7.5 was cross-linked with addition of a 1/10th volume of 10 × glutaraldehyde stock, 0.1% and 0.2% final glutaraldehyde concentrations were used. The reaction was quenched at various time points by addition of a 1/10th volume of 1M tris pH 7.5 and subjected to gel-filtration and/or SDS-PAGE analysis.

3.2.3. Thermal-denaturation studies

3.2.3.1. Far-UV CD spectroscopy

Far-UV CD spectra were recorded using 10 μM protein in 25 mM citrate buffer pH 6.5 or pH 4.5. Spectra were obtained with a Jasco J-810 spectrometer equipped with a peltier temperature controller. Individual CD spectra were collected at 20, 30, 40, 50, 60, 70 and 80 $^{\circ}$C in the range of 200–250 nm with a 0.1 cm pathlength cuvette. A resolution of 0.5 nm and a scanning speed of 20 nm min^{-1} were used. In a separate experiment, CD signals at 222 nm were monitored as a function of temperature. The protein samples were heated from 20 to 80 $^{\circ}$C with a heating rate of 30 $^{\circ}$C h^{-1} with the measurements recorded every 0.5 $^{\circ}$C.

3.2.3.2. Trp fluorescence spectroscopy

Fluorescence spectroscopy experiments were carried out on a FluoroMax-3 (HORIBA Jobin Yvon) luminescence spectrometer equipped with a circulating water bath to control the temperature. Protein spectra were recorded using 10 μM protein (in 25 mM citrate buffer pH 6.5 or pH 4.5) and a 1 cm path length cuvette. The protein sample was excited at 295 nm and

the resultant Trp emission spectra were collected over the range of 305–425 nm. An integration time of 1 s and a resolution of 1 nm were used. Thermal analysis was performed over the range of 20–85 °C with emission spectra recorded every 3 °C, allowing 5 min equilibration at each temperature.

3.3.   Results and discussion

3.3.1.   Analysis of evolutionary conservation and electrostatic properties

ConSurf, a web-server for analysis of evolutionary conservation, was used to predict functionally and structurally important residues. The analysis was carried out using all eukaryotic cytoplasmic Hsp70 family members in the UniProt database (for alignment see appendix A.2). Amino acids with above average conservation scores could be grouped into two categories; those involved in the defining the overall structure and those predicted to participate in the "latch" interactions with the β-sandwich subdomain (Figure 3-1a and b).

Helices αB-αE have a classical amphipathic nature with most of the hydrophobic residues in the core of the fold subject to above average conservation, in particular residues Leu^{543}, Leu^{600} and Ile^{607} cluster with the most conserved residues (Figure 3-1b). Residues important for correct folding include Asn^{585}, Ala^{588} and Glu^{592}, which are important in defining the tight hairpin loop connecting helices αC and αD; Pro^{606}, which occurs at the N-cap +1 position of helix αE immediately proceeding the kink separating helices αD and αE; and Lys^{562} of loop 1 which interacts with the backbone carbonyl of the penultimate ordered amino acid Gln^{613}.

Excluding variable Gly^{542}, all residues belonging to helix αB are subject to above average conservation. In particular, Glu^{544} and Ser^{545} are very well conserved. Glu^{544} is conserved with Asp^{540} in *E. coli* DnaK which, along with other C-terminal helix αB residues, was shown to interact with residues of the outer-loops of the β-sandwich (see section 3.3.2.) (Zhu et al., 1996). These interactions, termed the "latch", have has since been shown to be important for correct Hsp70 function (Fernandez-Saiz et al., 2006).

Interestingly, Tyr^{612} at the C-terminus of helix αE is conserved not only across eukaryotic Hsp70s but also across many prokaryotic family members suggesting an important function, the nature of which is unclear. The solvent exposed part of helix αC stands out as the least conserved region with all residues clustering as highly variable.

ceHsp70-CT contains many charged residues and carries a net negative charge of -6 eV. The electrostatic surface shows two electronegative patches. The first is located around the latch residues of helix αB, which covers the substrate-binding groove, perhaps involved in the

preference of Hsp70 for substrate peptides flanked by positively charged residues. The second is located at the surface generated by the C-terminal surface of helix αC, loop 2 and the N-terminal surface of helix αD (Figure 3-1c). In the predicted full-length model (section 3.3.2.) this region is close to the NBD-SBD domain interface and also the predicted binding site of the J domain of Hsp40 proteins.
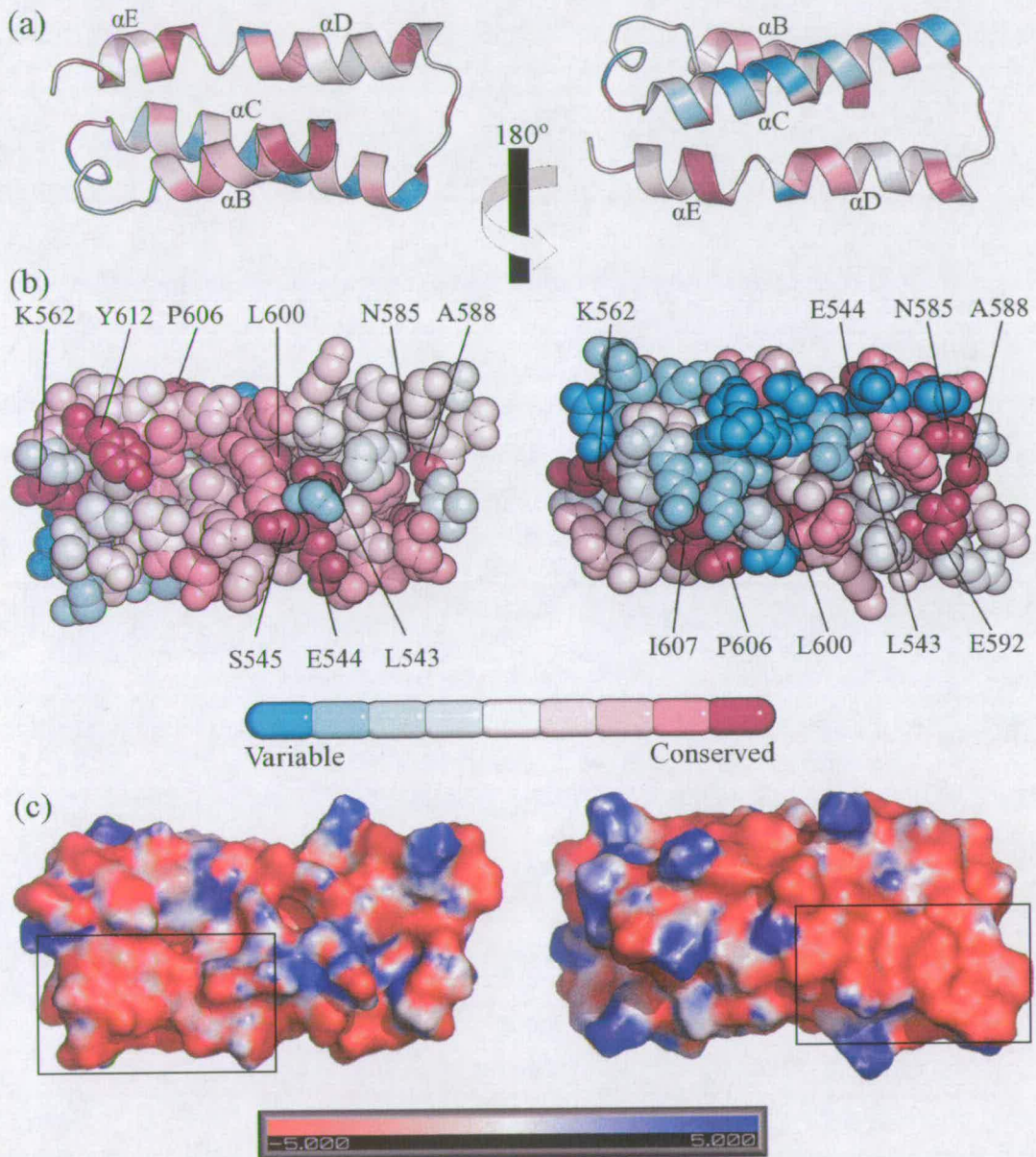
Figure 3-1 Evolutionary conservation and electrostatic properties of ceHsp70-CT. (a) Cartoon representation coloured according to ConSurf analysis at 0 and 180°. (b) Sphere representation in same orientation as (a). (c) Electrostatic surface calculated with the APBS plug-in for Pymol in same orientation as (a). Electronegative patches boxed. Left – region of helix αB covering substrate-binding groove, right – N-terminal region of helix αD near inter-domain interface and J domain binding site.

3.3.2. Model of the complete *C. elegans* Hsp70 structure

The structure of a full-length Hsp70 protein, either in the open or closed conformation, remains elusive with the most complete structure published to date being bovine Hsc70 (residues 1 - 554; PDB 1YUW) (Jiang et al., 2005) but lacking most of the C-terminal helical subdomain presented here. Although nucleotide-free, the structure is also thought to resemble the ADP-bound "closed" high-affinity state. As with some other SBD structures, the C-terminal region of helix αB is locally unwound and bound in the peptide-binding groove. The only structures of complete SBDs encompassing the β-sandwich subdomain and the helical subdomain are from *E. coli* DnaK and HscA, both peptide bound and in the high-affinity closed conformation. These show the relative position of the helical lid and β-sandwich subdomains.

A model of residues 1–533 of *C. elegans* Hsp70A was produced with SWISS-MODEL using bovine Hsc70 as a template (unwound helix αB residues trimmed from template). The helical subdomain solved in this study was subsequently positioned based on the structures of *E. coli* DnaK and HscA, with the final 8 unmodelled residues (Lys^{534}-Asn^{541}) connecting the model and structure filled in (Figure 3-2a). The model, representing the ADP-bound high-affinity conformation, illustrates the positioning of the helical lid covering the peptide-binding groove and the latch-like contact of helix αB and outer-loops of the β-sandwich. Analysis of evolutionary conservation with ConSurf highlights the important residues involved in this interaction (Figure 3-2b).

Based on the relative orientation of the lid and β-sandwich subdomains, small side-chain conformational changes upon lid closure would allow the direct interaction of highly conserved residues Glu^{544} on helix αB and Arg^{470} on loop $L_{5,6}$ seen in the DnaK structure (residues Arg^{467} and Asp^{540} in DnaK), with highly conserved Ser^{545} potentially hydrogen bonding with the guanidinium group of Arg^{470}. Additional conserved solvent exposed helix αB residues Phe^{548} and Gln^{552} are also likely to contribute to the stabilisation of the closed state.
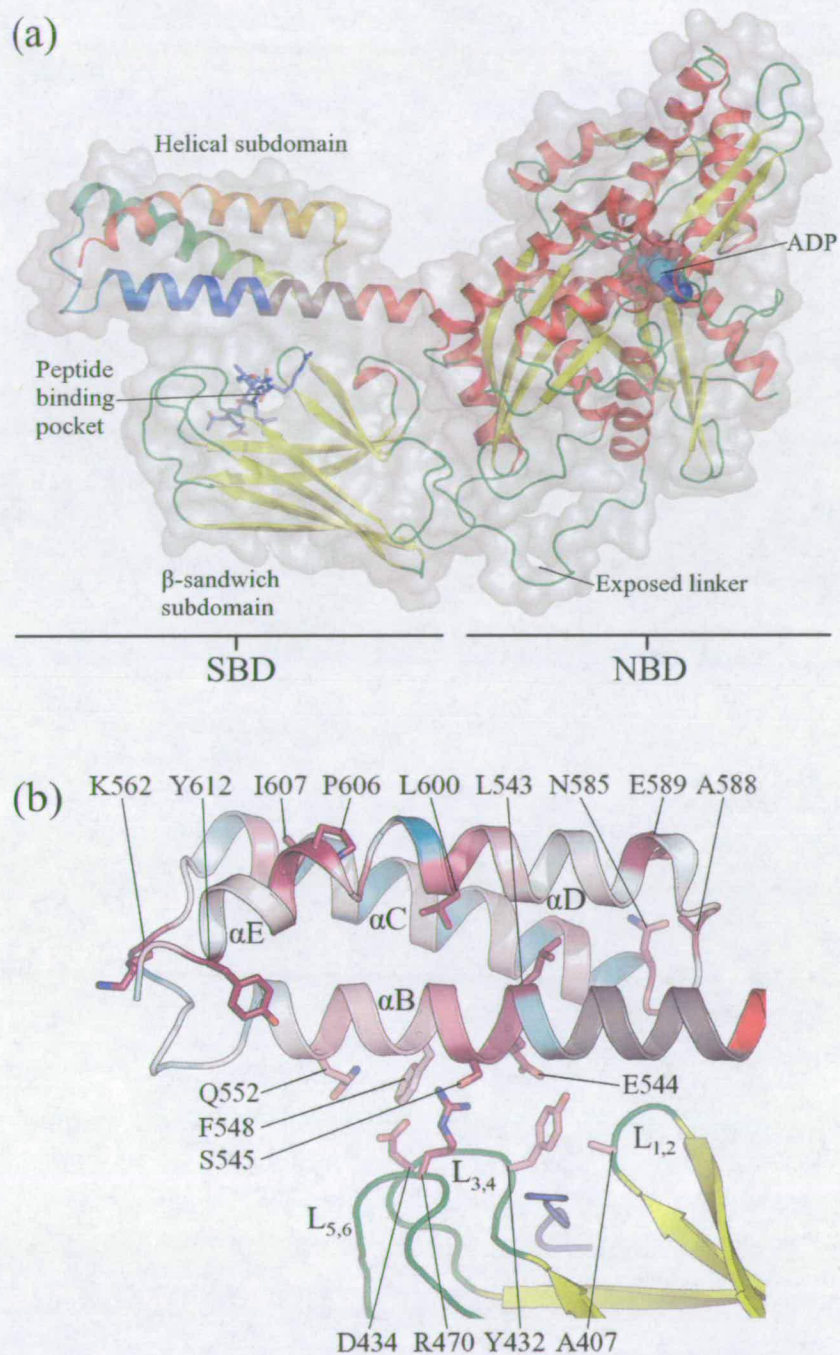
Figure 3-2 Model of the complete ceHsp70 structure in the closed high-affinity conformation. (a) Full-length model. Helical sub-domain presented in this study coloured in gradient. Manually built helix coloured grey. (b) Interactions between the lid subdomain and the outer loops of the β-sandwich sub-domain. Lid and all residues coloured according to ConSurf results with conserved residues coloured dark red and variable residues coloured cyan. Conserved residues on underside of helix αB form latch-like interactions with residues on loop $L_{3,4}$ and $L_{5,6}$.

3.3.2.1. Analysis of the oligomerisation state of ceHsp70-CT

Hsp70 family members - including Hsp70, Hsc70 and GRP75/BiP - predominantly exist as monomers, but have also been reported to both dimerise and further oligomerise in a concentration dependent manner (Benaroudj et al., 1995; Benaroudj et al., 1997; Benaroudj et al., 1996; Chou et al., 2003; Fouchaq et al., 1999). Hsp70 proteins have a tendency to aggregate and successful crystallisation of the full-length protein has only been achieved using a construct lacking most of the C-terminal subdomain (Jiang et al., 2006; Jiang et al., 2005). It has been suggested that the SBD is both necessary and sufficient for self-association; however, there are conflicting views on the exact mechanisms and both the 18 kDa β-sandwich subdomain (Benaroudj et al., 1997; Fouchaq et al., 1999) and the 10 kDa helical lid subdomain (Chou et al., 2003) have been proposed to mediate oligomerisation. A dimer of the C-terminal domain from rat Hsc70 was observed in the crystal state with this domain both necessary and sufficient for oligomerisation in solution (Chou et al., 2003). Conversely, the 18 kDa peptide-binding subdomain of bovine Hsc70 was shown to oligomerise in a peptide-sensitive manner comparable to the whole protein and also that oligomerisation of a 60 kDa fragment, lacking the 10 kDa C-terminal subdomain, was both peptide and ATP sensitive (Benaroudj et al., 1996). Finally, a recent study has implicated regions of both domains, with the β-sandwich subdomain and N-terminal regions of the helical subdomain found to be necessary for dimerisation of human Hsp70 (Nemoto et al., 2006).

ceHsp70-CT crystallised as a hexameric complex as two back-to-back trimers (Figure 2-13) with putative dimeric, trimeric and hexameric assemblies. Protein crystals are inherently composed of multiple protein-protein interfaces and it is not always easy to distinguish between a biologically relevant contacts and crystal lattice contacts. The web-server PISA (Protein Interfaces, Structures and Assemblies; http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) attempts to asses the biological significance of quaternary structures within crystal packing based on the structural and chemical properties of the interface.

There are three main interfaces that define the crystal packing (Figure 3-3); two are involved in the packing of the hexamer whilst the other is involved in the interaction between neighbouring hexamers. According to PISA, none of the crystal interfaces are biologically relevant. The two interfaces which define the hexamer have buried surface areas (BSA) of ~465 Å^2 (Figure 3-3a) and ~336 Å^2 (Figure 3-3b) whilst the dimeric interface relating to

hexamers has a BSA of ~310 Å^2 (Figure 3-3c). All interfaces have PISA complexation significance scores of 0.0 implying the interfaces only play a role in crystal packing.

(a)

(b)

(c)
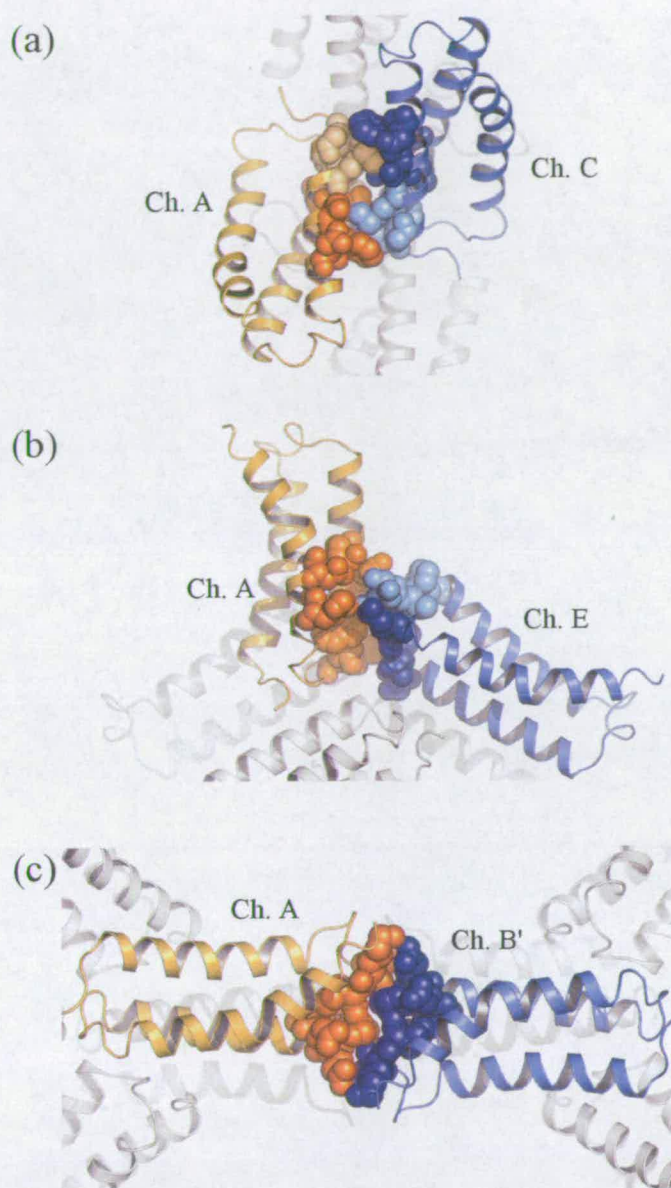
ID	Structure 1		Structure 2		BSA (Å^2)	# HB	# SB	CSS
	Chain	# Res (tag)	Chain	# Res (tag)				
(a)	A	11 (4)	C	11 (4)	465	4	8	0.0
(b)	A	11 (5)	E	11	336	4	1	0.0
(c)	A	8	B'	8	310	4	4	0.0

Figure 3-3 Structural interfaces in the ceHsp70-CT crystal packing. Residues involved in each interface are represented as spheres. Respective chains are coloured orange and blue with light-orange/blue indicating residues belonging to the recombinant N-terminal tag which participate in interfaces (a) and (b).

Gel-filtration was used to investigate the oligomerisation properties of ceHsp70-CT in solution. Whereas the C-terminal domain from rat Hsc70 was shown to exist in various oligomeric states in solution, ceHsp70-CT eluted as a single peak regardless of concentration (500 nM-10 µM) with a predicted Stokes radius consistent with the dimensions of the monomeric crystal structure (max. dimension ~45 Å) (Figure 3-4). To confirm the single peak represented the monomeric species, glutaraldehyde cross-linking was carried out prior to gel-filtration. SDS-PAGE analysis of the eluate from the three resolved peaks confirmed the smallest species corresponded to the monomeric protein (Figure 3-4).

Taken together, the analysis of the crystal interfaces and solution studies support the conclusion that the C-terminal subdomain exists exclusively as a monomer in solution and that regions outside the subdomain are required for oligomerisation. These results are consistent with a study on human Hsp70 which demonstrated that, although sequences within the N-terminal portion of the helical domain were necessary for dimerisation, the lid domain alone could not form dimers (Nemoto et al., 2006).
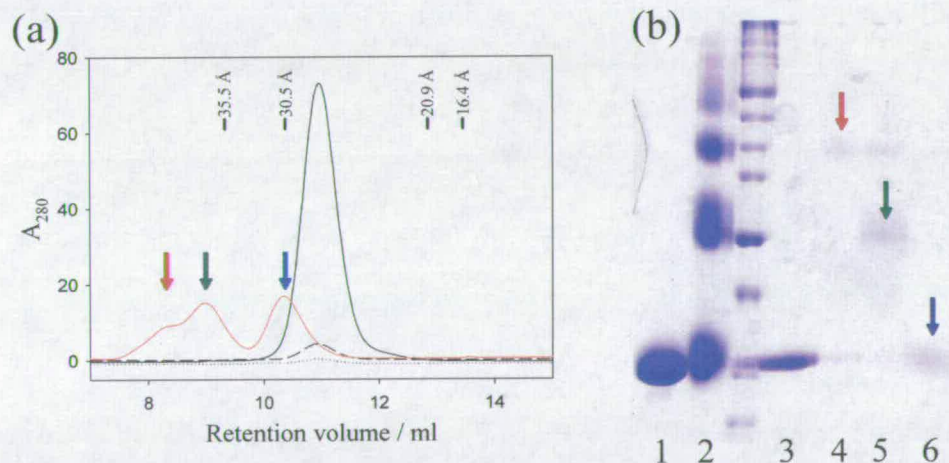
Figure 3-4 Gel-filtration analysis of ceHsp70-CT. (a) 80 µM (solid), 5µM (dashed) or 2µM (dotted) protein was resolved on a Superdex-75 HR column. Retention volumes of standards with known Stokes radius indicated. ceHsp70-CT elutes as a single peak at all concentrations with an estimated Stokes radius consistent with the dimensions of the monomeric crystal structure. Elution profile of glutaraldehyde cross-linked ceHsp70-CT indicated in red. (b) Glutaraldehyde cross-linking. Lanes 1 and 2 show native and cross-linked ceHsp70-CT respectively. Lanes 3-6 after gel-filtration. Lane 3 is from native peak, lanes 4-6 correspond to cross-linked peaks marked with coloured arrows in (a) and (b). Results show that ceHsp70-CT has a estimated Stokes radius consistent with the dimensions of the monomeric crystal structure. This is supported by cross-linking prior to gel-filtration.

3.3.3. Comparison with C-terminal structures from *E. coli* and rat

Structures of the NBD (cow, human *and E. coli*) and the SBD β-sandwich subdomain (cow, rat and *E. coli*) from several species reveal structural conservation from bacteria to humans. Structures of the C-terminal 10 kDa helical subdomain are, however, limited to one prokaryotic homologue; *E. coli* DnaK solved as part of the complete SBD (Zhu et al., 1996), and one eukaryotic homologue; rat Hsc70 solved as an isolated helical subdomain (Chou et al., 2003). In contrast to the NBD and SBD β-sandwich subdomain, the helical subdomains of DnaK and rat Hsc70 are significantly diverged with DnaK adopting a monomeric three-helix bundle and rat Hsc70 forming a helix-loop-helix that dimerises via an anti-parallel coiled-coil like interaction. In addition, the SBD from the distantly related *E. coli* paralogue HscA, a specialised bacterial Hsp70-class molecular chaperone, was shown to adopt a near identical conformation to DnaK (Cupp-Vickery et al., 2004).

Across the C-terminal 10 kDa subdomain *C. elegans* Hsp70 shares 69% sequence identity with rat Hsc70, 16% sequence identity with DnaK and only 5% sequence identity with HscA (Figure 3-5a). It was surprising therefore that the *C. elegans* structure presented here adopts the same three-helix bundle conformation as the bacterial homologues DnaK and HscA (Figure 3-5b). ceHsp70-CT and DnaK superimpose with an RMSD of 2.3 Å with the overall topology well conserved. There are two small insertions of two and three residues respectively in loops 1 and 2. The kink between helices αD and αE is positioned in the same place, at residue Ala[614] in the *C. elegans* sequence; however it is bent at an angle of ~70° in the *E. coli* structure compared to 32° in *C. elegans*. Accordingly, ceHsp70-CT and the more distantly related HscA superimpose with an RMSD of 2.5 Å (Figure 3-5).

In contrast, ceHsp70-CT adopts a structurally diverged conformation to the closely related rat homologue. Anti-parallel helices αC and αD in the *C. elegans* structure form one extended helix in rat Hsc70 that serves as a dimerisation interface. Hsiao and colleagues observed that helix αB from rat Hsc70 and DnaK superimposed relatively well but helices αC-αE did not superimpose at all (Chou et al., 2003). Careful superimposition of all the helical subdomains, however, reveals both helix αB and αC from all structures superimpose well (Figure 3-5). In addition, it is evident that dimerisation of rat Hsc70 is mediated via a domain-swap mechanism.
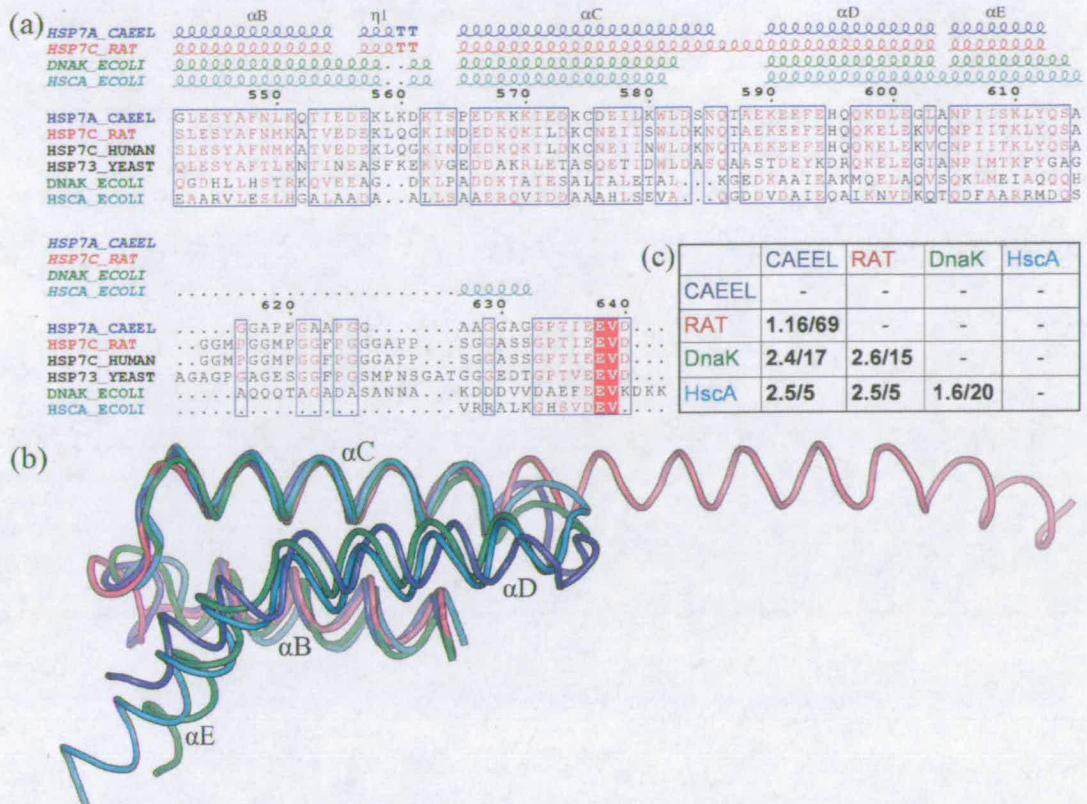
Figure 3-5 (a) Multiple sequence alignment of the C-terminal helical lid domain. Secondary structure of homologues with known structure is indicated. Sequences are labelled with SWISS-PROT IDs, HSP7A_CAEEL is *C. elegans* homologue used in this study. αB - αE: α-helices, η1: 3_{10}-helix, TT: type I β-turn. (b) Structural alignment of the C-terminal domains from *ce*Hsp70-CT, rat Hsc70, *E. coli* DnaK and *E. coli* HscA. Coloured according to sequence alignment. (c) Table of RMSDs/percent sequence identity of Hsp70 proteins over range in alignment.

3.3.4. A 3D domain-swap relates the *C. elegans* monomer and rat dimer

Superimposition of the *C. elegans* and rat structures reveals that the rat C-terminal structure seen in the crystal is a 3D domain-swapped dimeric form of the *C. elegans* monomer. Domain-swapping is a process in which one protein molecule exchanges an identical structural element ("domain") with an identical partner leading to oligomerisation.

Corresponding residues of the C-terminal subdomains from *C. elegans* and rat crystal structures superimpose with a backbone RMSD of 1.16 Å. Helices αB and αC (Leu^{543}-Asn^{585}) from *C. elegans* superimpose with the corresponding region from rat Hsc70 chain A whilst helices αD and αE (Lys^{590}-Ser^{614}) superimpose with the same residues from rat chain B (Figure 3-6a). Aside from the conformation of the ordered N-terminal affinity tag residues, the only significant area of difference is loop 2 (Gln^{586}-Glu^{589}), the hairpin loop connecting helices αC and αD in ceHsp70-CT. This region, the hinge region for the domain swap, forms

one helical turn in the rat structure resulting in the elongated αC/D/E helix. This loop - helix transition leads to dimerisation via the exchange of helices αD and αE such that helices αB and αC of monomer A interact with helices αD' and αE' of monomer B and vice versa.
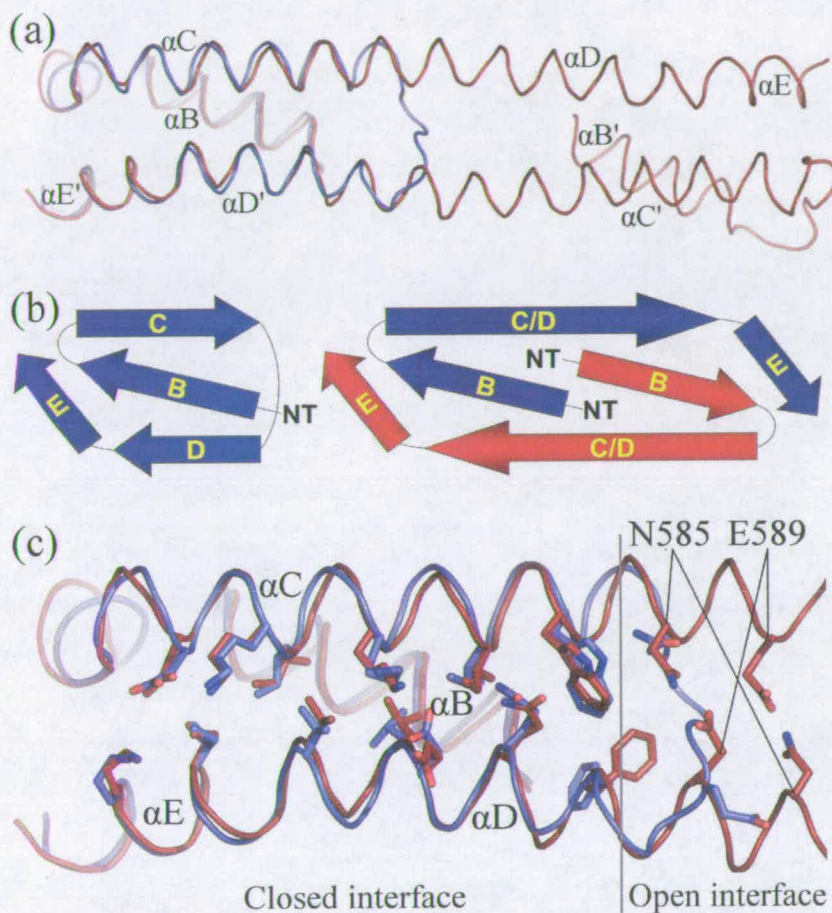
Figure 3-6 (a) Superimposition of ceHsp70-CT monomer (blue) and rat domain-swapped dimer (red). Structures superimpose with a backbone RMSD of 1.16 Å. (b) Topological representation showing packing of helices in the monomeric three-helix bundle and the domain-swapped dimer. (c) Superimposition of the ceHsp70-CT and rat structures illustrating the conserved hydrophobic and electrostatic packing of the closed interface and the newly formed interactions of the open interface.

The closed interface - the interface found in both the monomer and oligomer - is well conserved between the *C. elegans* and rat structures with analogous hydrophobic packing in the core of the structure and conserved intra-chain electrostatic interactions (Figure 3-6c). In addition, domain-swapping results in the formation of a new open interface - interactions absent in the monomer - with two symmetrical inter-chain hydrogen bonded interactions between hinge residues Asn[585] and Glu[589] from opposite chains (Figure 3-6c). Domain swapping results in an extended interface between the dimer subunits calculated at 1447 Å^2.

Whether this represents a biologically relevant means of dimerisation in Hsp70 proteins remains unclear. The monomeric and domain-swapped dimeric structures are isolated examples from different homologues and must, as such, be considered an example of a quasi-domain swap. The C-terminal subdomain of rat Hsc70 was shown to exist in monomeric and dimeric forms in solution by gel-filtration. In contrast, ceHsp70-CT eluted as a single species regardless of concentration with a retention volume consistent with the dimensions of the monomer seen in the crystal (Figure 3-4).
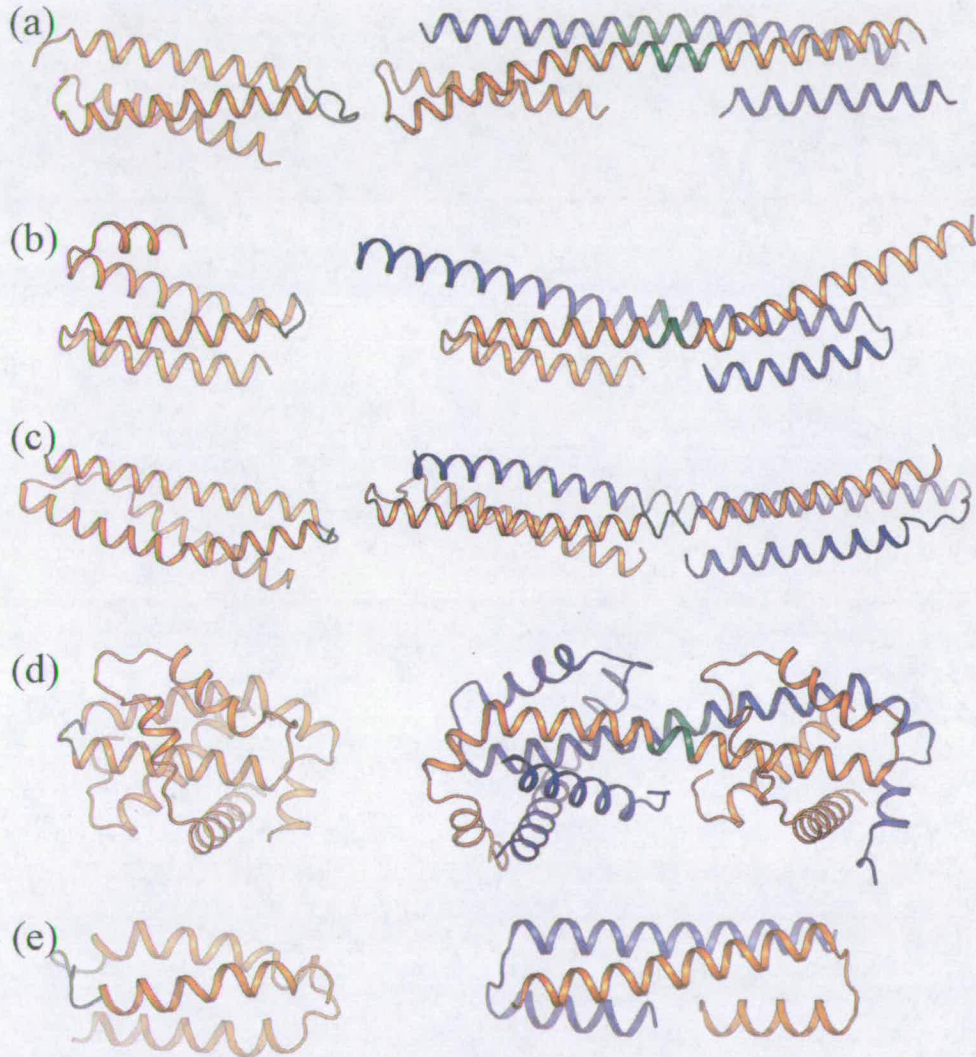
Figure 3-7 Examples of domain-swapping in helical-bundles. Hinge region for domain-swap coloured green in each case. (a) Monomeric (PDB-ID 2C5I) and domain-swapped (2C5J) forms of yeast TLG-1; involved in endosome-golgi trafficking. (b) Monomeric (1IHG) and domain-swapped (1IIP) forms of co-chaperone cyclophilin-40. (c) Monomeric (1CUN) and domain-swapped (2SPC) forms of cytoskeletal protein α-spectrin. (d) Monomeric (1R2D) and domain-swapped (2B48) forms of apoptotic factor BCL-XL. (e) Monomeric engineered three-helix bundle coil-ser (2A3D) and related domain-swapped dimer (1G6U). Domain-swap was triggered by deletion of loop in coil-ser (coloured green).

3.3.4.1. Domain-swapping and insight into ceHsp70-CT folding

Biological or not, domain-swapped structures can provide valuable information about folding pathways and protein flexibility. There are several examples of helical-bundle mediated (quasi)domain-swap dimerisation (Figure 3-7).

A common feature amongst the helical-bundle domain-swapped structures is that the hinge loop forms an α-helix generating an extended helical dimerisation interface (Figure 3-7; coloured green). Significantly, folding pathways of small three-helix bundles have been proposed to be populated by "open" two-helix intermediates suitable for domain-swapped dimer formation (Mayor et al., 2003; Zhou and Karplus, 1999). There is a high activation energy required for transition from a folded "closed" monomer to an unfolded "open" intermediate (Bennett et al., 1995). However, this can be reduced under certain conditions such as low pH and high ionic-concentration leading to accumulation of long-lived unfolded intermediates (Oliveberg and Fersht, 1996). Further, high protein concentrations can favour inter-chain over intra-chain interactions leading to self-assembly in the form of domain-swapped dimers. Thus, domain-swapped structures artificially triggered by non-physiological conditions may provide observable snapshots of protein folding intermediates.
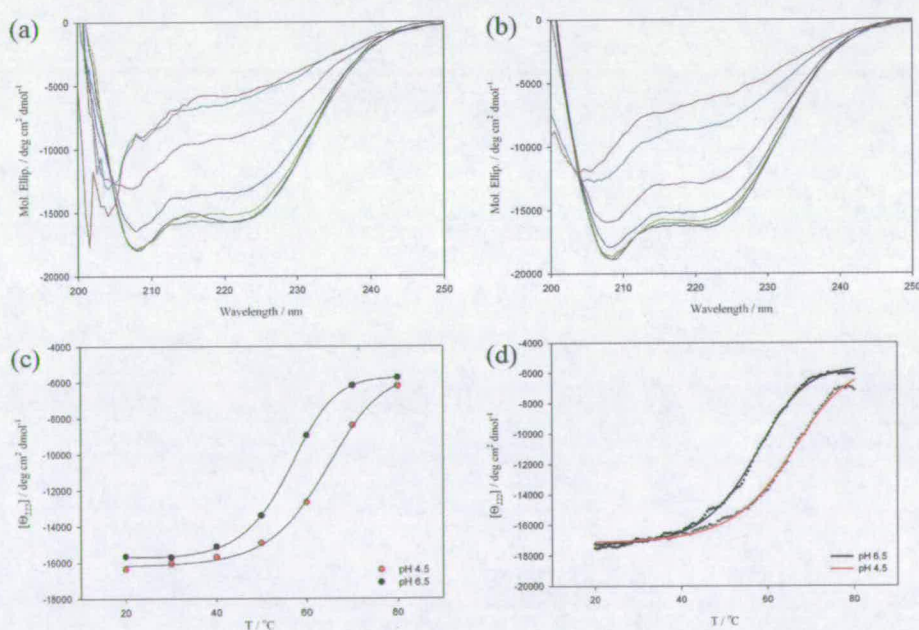
Figure 3-8 Far-UV CD thermal denaturation of ceHsp70-CT and pH 6.5 and 4.5. (a) Far-UV CD spectra recorded every 10 °C from 20 to 80 °C at pH 6.5. (b) Same as (a) at pH 4.5. (c) Molar ellipticity at 222 nM with increasing temperature from (a) and (b). (d) Molar ellipticity at 222 nM measured every 0.5 °C. (c) and (d) show a pH dependent transition of 55.9 °C at pH 6.5 and 64 °C at pH 4.5.

3.3.4.1.1.   Thermal denaturation of ceHsp70-CT

To explore the hypothesis that the extended helix-loop-helix conformation adopted by rat Hsc70-CT could represent a common folding intermediate of the Hsp70 C-terminal helical bundle, the thermal stability of ceHsp70-CT was investigated. Circular dichroism (CD) spectroscopy in the far-UV range was used to probe secondary structure content whilst the tertiary structure was analysed using the intrinsic tryptophan fluorescence of the single tryptophan (Trp[556]), located at the C-terminus of helix αC immediately preceding the hinge loop for the domain swap.

Figures 3-8a and 3-8b show the far-UV CD spectra of ceHsp70-CT at pH 6.5 and pH 4.5 respectively, recorded in 10 °C steps from 20 to 80 °C. At 20 °C, the spectra measured at pH 6.5 and pH 4.5 are virtually identical (Figure 3-8c). Deconvolution of the CD spectra with CONTIN (Provencher and Glockner, 1981) predicts a secondary structure content of 51% helix, 11% turn and 38% unordered in very good agreement with the crystal structure. At both pHs there is a temperature dependent change in the far-UV CD spectra indicative of a loss of secondary structure, although even at 80 °C the CD spectra reveal a protein with approximately 16% helix, 13% sheet, 8% turn and 63% unordered. Comparison of the molar ellipticity at 222 nm reveals a significant difference in thermal sensitivity of ceHsp70-CT at pH 6.5 and 4.5 (Figure 3-8c and d). Using a two-state model, the melting curves show transition temperatures (T_m) of 55.9 °C at pH 6.5 and 64 °C at pH 4.5

At pH 6.5 the fluorescence emission spectra presents at 348 nm (Figure 3-9a), higher than would be expected for a residue buried in a hydrophobic environment. As the temperature increases the fluorescence emission is shifted to 357 nm reflecting the exposure of the tryptophan to solvent with an associated T_m of 51 °C (Figure 3-9c). The peak emission intensity is quenched to approximately 50% of the native value although there is an enhancement between ~42 °C and ~58 °C (Figure 3-9d). At pH 4.5, the emission peak presents a 2 nm red shift compared to pH 6.5 indicating a marginally less hydrophobic environment (Figure 3-9b). Heating causes a shift in peak emission to 356.5 nm (Figure 3-9c) with a T_m of 65°C although there is only a linear quench in intensity (Figure 3-9d).

For comparison of CD and fluorescence monitored thermodynamic stability, measurements were converted to an apparent unfolded fraction (F_{app}) using the following equation:

$$F_{app} = (Y - Y_N)/(Y_D - Y_N)$$

Where Y is the observed signal, and Y_N and Y_D are the corresponding signals for the native and unfolded proteins respectively. Figure 3-10a shows the unfolded fraction as recorded by both CD and tryptophan fluorescence at pH 6.5 and 4.5.
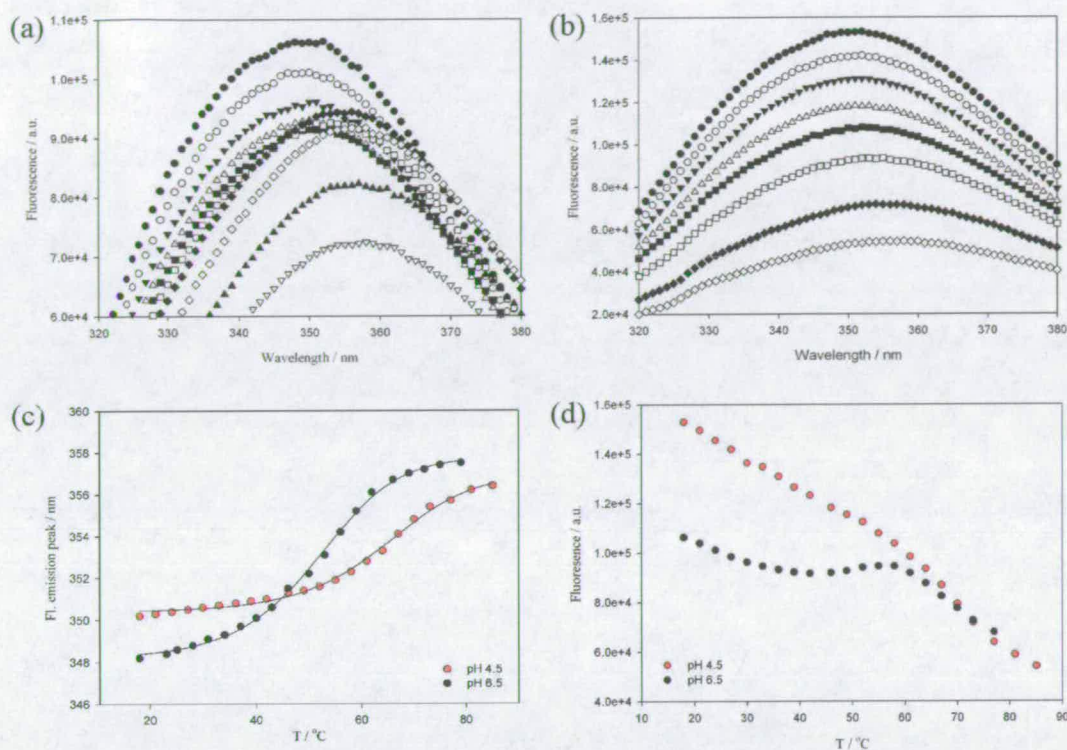
Figure 3-9 Intrinsic tryptophan fluorescence thermal denaturation of ceHsp70-CT. (a) Fluorescence spectra measured every ~4 °C at pH 6.5. (b) As (a) at pH 4.5. (c) Peak emission wavelength with increasing temperature. There is a pH dependent transition of ~51 °C at pH 6.5 and ~64 °C at pH 4.5. (d) Peak emission intensity with increasing temperature.

Two things are immediately apparent when comparing the CD and fluorescence monitored thermodynamic stability. Firstly there is a large pH dependence on protein stability with transitions measured by both methods significantly higher at pH 4.5. This inverse relationship between pH and protein stability was further confirmed with fluorescence monitored thermal denaturation studies over a range of pH values (Figure 3-10b). Secondly, whilst thermal denaturation appears to follow a simple two-state model at pH 4.5, at pH 6.5 unfolding appears to proceed via at least one intermediate (Figure 3-10a). The thermal transitions monitored by CD and fluorescence differ by approximately 5 °C at pH 6.5 indicating the accumulation of a species with local unfolding of the three-helix bundle prior to the loss of secondary structure. Comparison of the temperature dependent fluorescence emission intensity at pH 6.5 (Figure 3-10; indicated by crosses) with the CD and

fluorescence melting curves suggests the enhancement witnessed between ~42 °C and ~58 °C occurs as the loss of secondary structure catches up with the loss of tertiary structure. This pattern was only observed at pH 6.5, with mostly linear quenches in intensity observed at lower pHs. At pH 4.5, the unfolding appears to be only two-state with good agreement of the T_m values calculated from both methods.
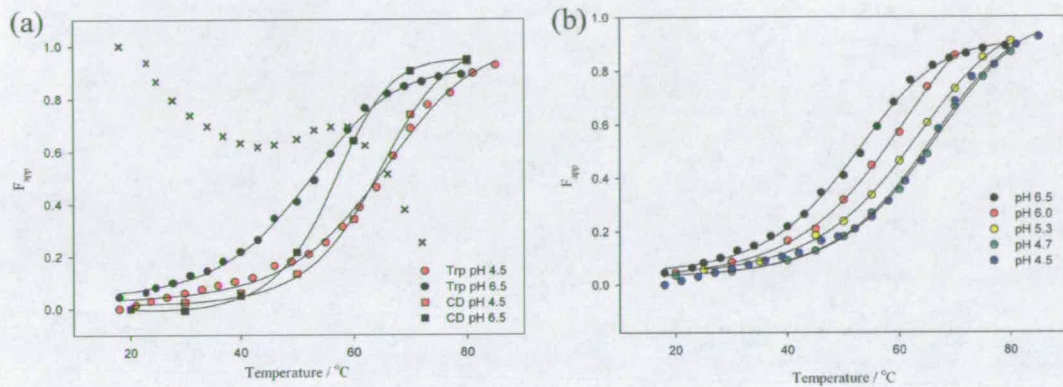
Figure 3-10 Apparent unfolded fraction (F_{app}). (a) F_{app} calculated from CD and trp-fluorescence data at pH 6.5 and 4.5. Peak emission intensity at pH 6.5 overlaid (x). (b) F_{app} at decreasing pHs. There is an inverse relationship between thermal sensitivity and pH.

These results agree in part with studies on human Hsp70 (Fuertes et al., 2004) and Hsc70 (Fan et al., 2006). Both the isolated SBD from human Hsp70 and full-length human Hsc70 were shown to unfold via several intermediates with local unfolding of the C-terminal subdomain preceding loss of secondary structure. However, whereas ceHsp70-CT is more stable at pH 4.5, pHs deviating from physiological had destabilising effects with the more complete proteins.

Domain-swapping is enthalpically favourable with additional backbone hydrogen bonds as a result of the higher helical content coupled with the new interactions of the open interface. The increased stability of ceHsp70-CT at pH 4.5 and more solvent exposed environment of the single tryptophan raise the possibility that the low pH triggers the domain-swap. Gel-filtration at pH 4.5, however, failed to support this with no change in retention volume compared to pH 6.5 (data not shown). An additional explanation of the increased stability could be due to charge distribution of the three-helix bundle. ceHsp70-CT contains many charged residues and carries a net negative charge of -6 eV at pH 7.0. The pKa of glutamate and aspartate residues is around pH 4.5 so reducing the pH would have quite a significant

overall effect and, if repulsive charge-charge interactions affected stability, result in a more stable structure.

At present, further experiments are required to characterise the thermal denaturation characteristics of ceHsp70-CT in more detail.
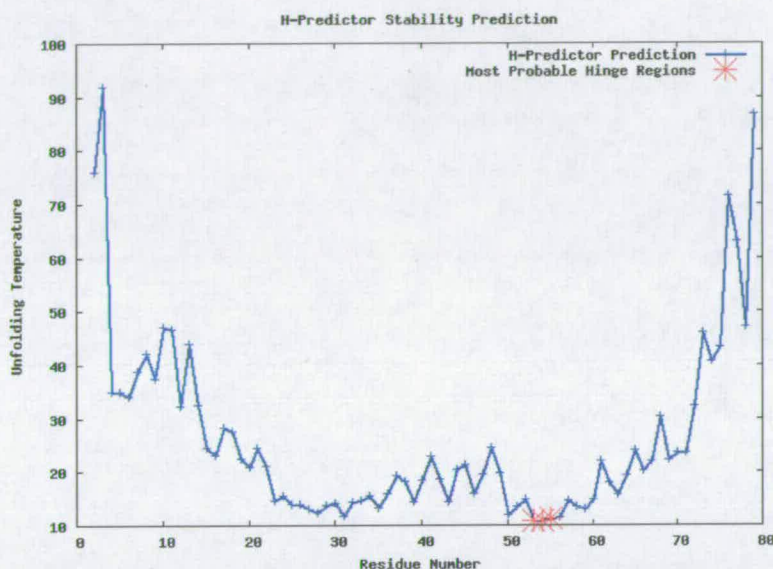
Figure 3-11 H-predictor analysis of ceHsp70-CT. Hinge residues Gln[586]-Ala[588] (53-55 in the above analysis) are predicted to be most probable hinge region.

3.3.5. Proposed folding pathway of ceHsp70-CT and other three-helix bundles

The observation that the three-helix bundle unfolds prior to loss in secondary structure at pH 6.5 supports the hypothesis that the rat domain-swap dimer represents a snapshot of a folding intermediate. Local unfolding of the C-terminal domain around loop 2, exposing Trp[556] but not losing secondary structure, would disrupt the packing of the three-helix bundle, exposing the hydrophobic core. Formation of a more stable domain-swapped dimer would re-establish the closed interface with added enthalpic contributions of the open interface.

Further support comes from analysis of the ceHsp70-CT structure with H-predictor, a domain-swap hinge region predictor (Ding et al., 2006). H-predictor computes for each residue the effective temperature to populate an intermediate state, where the protein unfolds around this residue into two subdomains each of which maintains their native-like structure. In the case where a protein features folding intermediates, it can also provide hints regarding the weakest regions that unfold prior to compete unfolding. Analysis with H-predictor

identifies residues Gln^{586}-Ala^{588} to have a structural propensity to constitute a hinge region, in agreement with the crystallographic evidence (Figure 3-11).

The domain-swapped rat dimer thus possibly represents a stabilised intermediate formed in response to the non-physiological environment of the crystallisation conditions. This destabilisation-compensation mechanism of domain-swap formation, similar to that demonstrated for the domain-swap mediated trimerisation of barnase at pH 4.5 (Oliveberg and Fersht, 1996; Zegers et al., 1999), is likely to be a common mechanism for the formation of helical-bundle domain-swapped oligomers in protein crystallisation where non-physiological pHs, ionic strengths and protein concentrations are common (Figure 3-12).
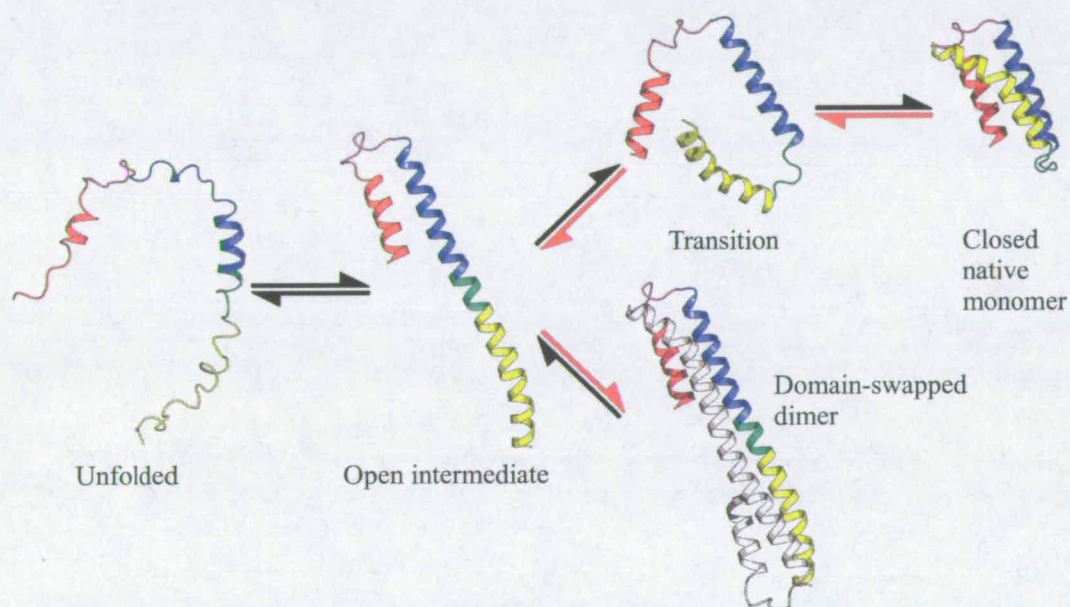
Figure 3-12 Proposed folding pathway for native and domain-swapped helical bundles. Colours refer to secondary structure elements in native structure; helix A (red), loop AB (magenta), helix B (blue), loop BC (green) and helix C (yellow). Under denaturing conditions protein exists as an unfolded random-coil. Folding proceeds via an open two-helix intermediate with loop BC (green) forming one helical turn. There is a large energy barrier between closed and open monomers although non-physiological conditions may lower this promoting the population of long-lived open intermediates and favouring domain-swapped dimerisation (pathway indicated by red arrows).

A key unresolved question is why, considering the similar sequence and crystallisation conditions, did the C-terminal 10 kDa subdomain of rat Hsc70 form a domain-swapped dimer in the crystal structure when the *C. elegans* subdomain did not. A possible explanation is in the start positions of the rat and *C. elegans* clones. The *C. elegans* C-terminal construct begins at residue Gly^{542} whereas the rat construct begins at Leu^{543} (*C. elegans* numbering).

The *C. elegans* sequence was based on the rat construct but with the slight modification due to the observation by Hendrickson and colleagues that residues 538-607 (*E. coli* Gln[538] aligns with *C. elegans* Gly[542]) from the complete DnaK structure formed a relatively stable functional unit with a well-defined hydrophobic core (Zhu et al., 1996). However, in both the *C. elegans* and rat structures, residues belonging to the recombinant affinity tag form the beginning of helix αB and contributed side-chains to the hydrophobic core of the three-helix bundle. Indeed, reanalysis of the DnaK structure reveals that an additional seven residues (from Glu[531]) should have been defined with the three-helix bundle, in agreement with NMR solution studies on the isolated DnaK subdomain which predicted residues 531-608 to form a compact well-ordered structure (Bertelsen et al., 1999). The alternate recombinant residues presented to the hydrophobic core in the *C. elegans* and rat polypeptides could thus define the stability of compact bundle and the open domain-swapped dimer. Significantly, comparison of the *C. elegans* and *E. coli* structures reveals that *C. elegans* tag residue Val[-7] occupies the same position and makes similar contacts as DnaK Leu[532], which forms the first layer of hydrophobic interactions in the three-helix bundle (Figure 3-13). Due to the one residue shift in start positions, this contact is absent in the rat structure and perhaps contributes to the destabilisation of the three-helix bundle, favouring the formation of the domain-swapped dimer.

This highlights the importance, when working with subdomains in isolation, of the careful selection of an appropriate and biologically relevant region and also the caution which must be exercised in the interpretation of any crystal structure, especially when confronted with novel and interesting conformations such as domain-swaps.

3.4. Conclusions

In summary, the C-terminal helical subdomain from *C. elegans* Hsp70 has been shown to adopt a three-helix bundle conserved with distantly related bacterial homologues and significantly distinct from the closely related rat Hsc70 structure. Comparison with the rat structure, however, reveals the rat dimer is a domain-swapped form of the *C. elegans* monomer. In contrast to the rat C-terminal subdomain, which was shown form dimers in solution, the isolated *C. elegans* subdomain was shown to only exist in the monomeric state supporting the theory that regions outwith this subdomain are required for Hsp70 self-association. Although the structure presented here casts doubt on the physiological relevance of the rat dimer, comparison of the monomeric and domain-swapped form may provide useful information regarding the folding pathway of the three-helix bundle. Thermal-denaturation studies on ceHsp70-CT at pH 6.5 suggested unfolding proceeds via the

accumulation of an unfolded intermediate. The open monomer observed in the rat domain-swapped structure may provide a snapshot of such a folding intermediate.
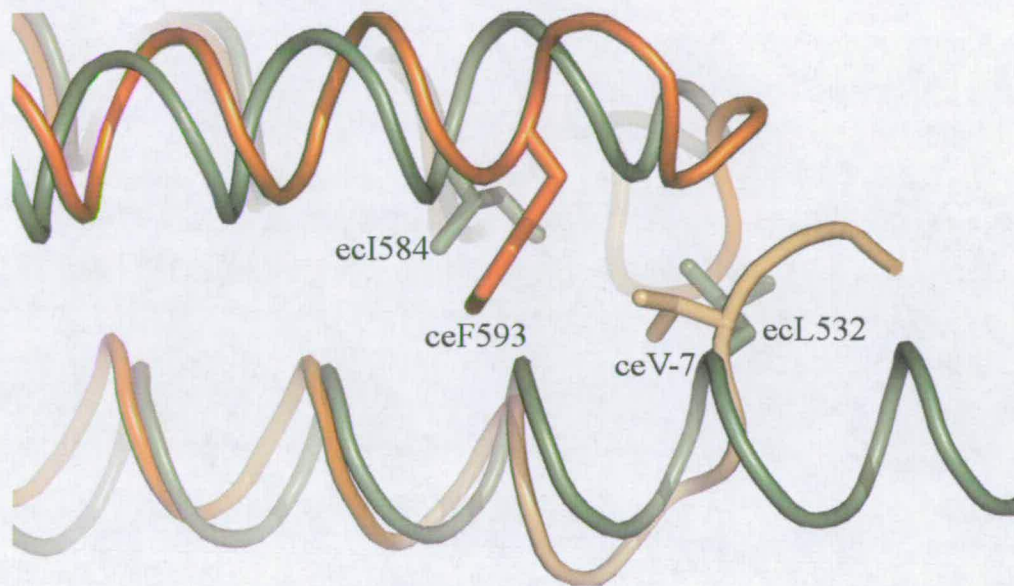
Figure 3-13 Contacts between helix αB and αD. Tag residue Val^{-7} (ceV-7) and helix αD residue Phe^{593} (ceF593) in the *C. elegans* structure (orange; tag residues in light orange) make a similar hydrophobic contact as *E. coli* (green) residues Leu^{532} (ecL532) and Ile^{584} (ecI584) and may contribute to the stabilisation of the three-helix bundle.

3.5.   References

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.

Benaroudj, N., Batelier, G., Triniolles, F., and Ladjimi, M. M. (1995). Self-association of the molecular chaperone HSC70. Biochemistry *34*, 15282-15290.

Benaroudj, N., Fouchaq, B., and Ladjimi, M. M. (1997). The COOH-terminal peptide binding domain is essential for self-association of the molecular chaperone HSC70. J Biol Chem *272*, 8744-8751.

Benaroudj, N., Triniolles, F., and Ladjimi, M. M. (1996). Effect of nucleotides, peptides, and unfolded proteins on the self-association of the molecular chaperone HSC70. J Biol Chem *271*, 18471-18476.

Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. Protein Sci *4*, 2455-2468.

Bertelsen, E. B., Zhou, H., Lowry, D. F., Flynn, G. C., and Dahlquist, F. W. (1999). Topology and dynamics of the 10 kDa C-terminal domain of DnaK in solution. Protein Sci *8*, 343-354.

Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005). The Amber biomolecular simulation programs. J Comput Chem *26*, 1668-1688.

Chou, C. C., Forouhar, F., Yeh, Y. H., Shr, H. L., Wang, C., and Hsiao, C. D. (2003). Crystal structure of the C-terminal 10-kDa subdomain of Hsc70. J Biol Chem *278*, 30311-30316.

Cupp-Vickery, J. R., Peterson, J. C., Ta, D. T., and Vickery, L. E. (2004). Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. J Mol Biol *342*, 1265-1278.

Ding, F., Prutzman, K. C., Campbell, S. L., and Dokholyan, N. V. (2006). Topological determinants of protein domain swapping. Structure *14*, 5-14.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res *32*, W665-667.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Fan, H., Kashi, R. S., and Middaugh, C. R. (2006). Conformational lability of two molecular chaperones Hsc70 and gp96: effects of pH and temperature. Arch Biochem Biophys *447*, 34-45.

Fernandez-Saiz, V., Moro, F., Arizmendi, J. M., Acebron, S. P., and Muga, A. (2006). Ionic contacts at DnaK substrate binding domain involved in the allosteric regulation of lid dynamics. J Biol Chem *281*, 7479-7488.

Fouchaq, B., Benaroudj, N., Ebel, C., and Ladjimi, M. M. (1999). Oligomerization of the 17-kDa peptide-binding domain of the molecular chaperone HSC70. Eur J Biochem *259*, 379-384.

Fuertes, M. A., Perez, J. M., Soto, M., Menendez, M., and Alonso, C. (2004). Thermodynamic stability of the C-terminal domain of the human inducible heat shock protein 70. Biochim Biophys Acta *1699*, 45-56.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163-164.

Jiang, J., Lafer, E. M., and Sousa, R. (2006). Crystallization of a functionally intact Hsc70 chaperone. Acta Crystallograph Sect F Struct Biol Cryst Commun *62*, 39-43.

Jiang, J., Prasad, K., Lafer, E. M., and Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. Mol Cell *20*, 513-524.

Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V., and Fersht, A. R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. Nature *421*, 863-867.

Nemoto, T. K., Fukuma, Y., Itoh, H., Takagi, T., and Ono, T. (2006). A Disulfide Bridge Mediated by Cysteine 574 Is Formed in the Dimer of the 70-kDa Heat Shock Protein. J Biochem (Tokyo) *139*, 677-687.

Oliveberg, M., and Fersht, A. R. (1996). Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. Biochemistry *35*, 2738-2749.

Provencher, S. W., and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. Biochemistry *20*, 33-37.

Zegers, I., Deswarte, J., and Wyns, L. (1999). Trimeric domain-swapped barnase. Proc Natl Acad Sci U S A *96*, 818-822.

Zhou, Y., and Karplus, M. (1999). Interpreting the folding kinetics of helical proteins. Nature *401*, 400-403.

Zhu, X., Zhao, X., Burkholder, W. F., 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.

4. Biochemical characterisation of *C. elegans* SGT

4.1. Introduction

The tetratricopeptide repeat (TPR) domain is a versatile all-helical structural motif, found in proteins from almost all organisms, which mediates protein-protein interactions in numerous different cellular processes (D'Andrea and Regan, 2003). TPR domain associated molecular recognition is central to the interaction between the chaperone proteins Hsp70 and Hsp90 and the numerous co-chaperone binding partners including the TPR containing immunophilins cylophilin 40, FKBP-51, FKBP-52 (Owens-Grillo et al., 1996; Owens-Grillo et al., 1995; Ratajczak and Carrello, 1996; Young et al., 1998).

C. elegans contains a single large immunophilin (FKB-6) which, like the human homologues, contains 3 TPR motifs. Interestingly however, *fkb-6* expression is predominantly restricted to neuronal cells (Anthony Page, University of Glasgow, unpublished). This prompted a search of the published *C. elegans* genome for other, sequence related TPR-domain containing proteins that may have a wider cellular distribution. This highlighted two hypothetical proteins, the first was found to be the *C. elegans* orthologue for small glutamine-rich tetratricopeptide repeat-containing protein (SGT) and will be discussed in this chapter; and the second was found to be related to the Hsp70/Hsp90 organising protein (Hop) which will be discussed in chapter 5.

SGT was initially identified as a human protein that interacted with two components of the human immunodeficiency type-1 virus (HIV-1) genome; viral protein U (Vpu) and the structural precursor polyprotein Group specific Antigen (GAG) (Callahan et al., 1998). Rat SGT was independently identified to interact with the non-structural protein NS-1 from autonomous parovirus H-1 (Cziepluch et al., 1998). Subsequently, SGT has been implicated in a myriad of functions ranging from maintenance of a normal synapse (Bai et al., 2007; Natochin et al., 2005; Swayne et al., 2006; Tobaben et al., 2001) to roles in cell cycle progression (Wang et al., 2005; Winnefeld et al., 2004) and apoptosis (Wang et al., 2005; Winnefeld et al., 2006; Yin et al., 2006). The common underlying theme is the action of SGT as a co-chaperone, directly interacting with and modifying the behaviour of the Hsp70/Hsp90 chaperone machinery. The alpha isoform of human SGT has been shown to negatively regulate the chaperone activity of Hsp70 via interaction with the C-terminal region (Angeletti et al., 2002) and also directly bind to the C-terminus of Hsp90 (Angeletti et al., 2002; Liou and Wang, 2005).
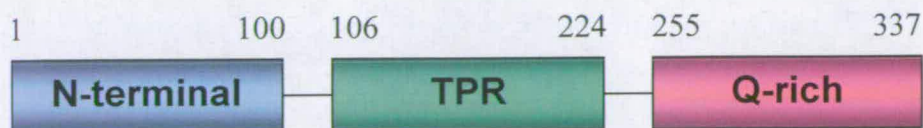
Figure 4-1 SGT Domain architecture. Full-length ceSGT is a 337 residue modular protein with three functionally distinct domains - the N-terminal domain implicated in self-association, the TPR domain and the Q-rich C-terminal domain

Human SGT consists of three functionally distinct domains (Figure 4-1) (Liou and Wang, 2005). The central domain is comprised of three TPR repeats shown to bind human Hsp70 and Hsp90 (Angeletti et al., 2002; Liou and Wang, 2005; Tobaben et al., 2003). The N- and C-terminal regions of the molecule lack significant homology to any known domain, however the N-terminus has been shown to be both necessary and sufficient for self-association (Liou and Wang, 2005; Tobaben et al., 2003). The C-terminal region shows slight enrichment in the amino acid glutamine in addition to containing several asparagine/proline repeats. *C. elegans* SGT (ceSGT) is a 36.5 kDa protein consisting of 337 amino acids; it has 34% sequence identity with the human homologue, with the highest degree of conservation across the TPR domain (Figure 4-2).

The aims of this project were to clone, express and purify *C. elegans* SGT; to characterise the solution state of ceSGT and its interaction with Hsp90/Hsp70; and to use X-ray crystallography to solve the molecular structure of the full-length protein or smaller domain constructs.

```
ceSGT        ooooooooooooooooooooooooooooooooooooooooooo
                        10        20        30        40        50        60
ceSGT       MSEEIKPSAPDAPTASPIVTRDEQNLVVSFLQFIRQKVSQNQATAEQAEALEVAIQCLEHSFGLDD
Drosophila  ...............MPDAVQQSFVRSFIDYLKK..QGDVMSPDQTESIEVAIQCLQAAFDLGD
Chicken     ...............MADQKRLAYSIIQFLHDQLQNGGLSPDAQESLEVAIQCLETAFGVSL
Yeast       ..............MSASKEEIAALIVNYFSSIVEKKEISEDGADSLNVAMDCISEAFGFER
Human_A     ...............MDNKKRLAYAIIQFLHDQLRHGGLSSDAQESLEVAIQCLETAFGVTV
Human_B     ............MSSIKHLVYAVIRFLREQSQMDTYISDEQESLEVAIQCLETVFKISP
                                                        ***********

ceSGT              o......ooooooooo                              oo...
                        70        80        90        100
ceSGT       ..........ASYAFQPSRP......TLELFKSAEGLPEGESALPTPSD..........SDI...
Drosophila  DVEAAPAAAGEEQATTQSSSTASAPDDDAVASGSGAGIGAAAAAVPNNIDMFELFQSLYTERNPESL
Chicken     ..........EDQGLAVSRT.....LPEIFEAAAG..KEPEHIRANSEPVTP.....SEDDI...
Yeast       ..........EAVSGILGKSEFKGQHLADILNSASRVPESNKKDDAENVEINI.....PEDDAETK
Human_A     ..........EDSDLALPQT.....LPEIFEAAATGKEMPQDLRSPARTPP......SEEDS...
Human_B     ..........EDTHLAVSQP.....LTEMFTSSF..........CKNDVLPL.....SNSVPEDV
                                                                       ***

ceSGT        oooooooo       o   oooooooooooooooo      oooooooooooooooooooooooo
                   110       120       130       140       150       160
ceSGT       SQANKLKEEGNDLMKASQFEAAVQKYNAAIKLN.RDPVYFCNRAAAYCRLEQYDLAIQDCRTALAL
Drosophila  ALAESIKNEGNRLMKENKYNEALLQYNRAIAFDPKNPIFYCNRAAAHIRLGENERAVTDCKSALVY
Chicken     AEAERLKTEGNEQMKAENFEAAVSFYGKAIELNPSNAVYFCNRAAAYSKLGNYAGAVRDCERAIGI
Yeast       AKAEDLKMQGNKAMANKDYELAINKYTEAIKVLPTNAIYYANRAAAHSSLKEYDQAVKDAESAISI
Human_A     AEAERLKTEGNEQMKEENYAAAVDCYTQAIELDENNAVYYCNRAAAQSKLGHYTDAIKDCERAIAI
Human_B     GKADQLKDEGNNHMKEENYAAAVDCYTQAIELDENNAVYYCNRAAAQSKLGHYTDAIKDCEKAIAI
            *******************************  ****************************

ceSGT        oooooooooo     ooooooooo                  oo   ooooooooo  α1
                   170       180       190       200        210       220
ceSGT       DPSYSKAWGRMGLAYSCQNRYEHAAEAYKKALELEPNQES......YKNNLKIAEDKLKELESSRP
Drosophila  NNNYSKAYCRLGVAYSNMGNFEKAEQAYAKAIELEPDNEV......YKSNLEAARNARNQ....PP
Chicken     DPNYSKAYGRMGLALSSLNKHTEAVVYYKKALELDPDNDT......YKSNLKIAEQKMKE....TP
Yeast       DPAYSKAYGRMGLALSSLNKHVEAVAYYKKVLDIEGDNATEAMKRDYESAKKKVEQSLNLEKTVPE
Human_A     DPSYFRGYSRLGFAKYAQGKPEEALEAYKKALELDPDNET......YKSNLKIAELKLRE....AP
Human_B     DSKYSKAYGRMGLALTALNKFEEAVTSYQKALDLDPENDS......YKSNLKIAEQKLRE....VS
            ***********************************  ****************

ceSGT         oooooooo     oo        oooooooo
                   230       240       250       260       270       280
ceSGT       APGANPLAGLLGAMGGGPGGMPAMPGMPNMQNLMNEPGLMQAASQMMSDP..ALSDMFNNMMS...
Drosophila  Q.............TGRLRE.......DLINMLSQP...MVRNLFNNAEIDVEQLLSMLQNPMI
Chicken     SP............TGGPGGF.......DLAGLLNNPGFMSMASTLMEQST.STTSSLSGMISGGH
Yeast       QSRDADVDASQGASAGGLPDLGSLLG.GGLGGLMNNPQLMQAAQKMMSNPG.AMQNIQKMMQDPSI
Human_A     SP............TGGVGSF.......DIAGLLNNPGFMSMASNLMNNP..QIQQLMSGMISGGN
Human_B     SP............TGTGLSF.......DMASLINNPAFISMAASLMQNP..QVQQLMSGMMT...

ceSGT         ooooooooooooooooo    ooooooooo
                   290       300       310       320       330
ceSGT       ............GNGSIADLMAAGQOMAARMQETNPELIENLRRQFGPG...ADGGAPPPNPPQ
Drosophila  MNTIRQQFGGGNAPTLPNDM......VQMIYNMTSQLTGFPNSNANPGNEQQPGNQPPPST..
Chicken     NPMGAAGTS.PSTNDLASLIQAGQQFAQQMQQQNPELIEQLRSQIRSR...TPSASNEDQQE.
Yeast       RQMAEGFASGGGTPNLSDLM...........NNPALRNMAGNLFGGAGAQSTDETPDNENKQ
Human_A     NPLGTPGTS.PSQNDLASLIQAGQQFAQQMQQQNPELIEQLRSQIRSR...TPSASNDDQQE.
Human_B     NAIGGPAAGVGGLTDLSSLIQAGQQFAQQIQQQNPELIEQLRNHIRSR...SFSSSAEEHS..
```

Figure 4-2 Protein sequence alignment of SGT homologues. PHD secondary structure prediction of ceSGT indicated. Predicted N-terminal coiled-coil domain highlighted with blue stars and TPR domain with green stars.

4.2. Materials and methods

4.2.1. Cloning

cDNA corresponding to full-length ceSGT (residues 1-337) and the SGT TPR domain (residues 101-226) were generated by PCR using *C. elegans* mixed stage N2 cDNA as a template. Sequences were amplified with the TaqPlus® precision PCR system (Stratagene) using the forward and reverse primers found in Table 4-1. The resulting PCR products were cloned into a pCR®2.1 TOPO vector (Invitrogen), verified by sequencing and digested with *NdeI* and *XhoI* (New England Biolabs). The digested inserts were ligated into a similarly digested pET-30a vector (Novagen) and verified by DNA sequencing (Figure 4-3).

Clone (Sequence #)		Primer sequence, restriction site in italics	Restriction enzyme
ceSGT (1-337)	Forward	GCGG*CATATG*TCCGAGGAGATCAAGCCTTCTG	NdeI
	Reverse	GGCG*CTCGAG*CTATCGCGAGCTTTCCAGCTCCTT	XhoI
ceSGT-TPR (101-226)	Forward	GCGG*CATATG*AGTGATATTTCTCAAGCTAACAAG	NdeI
	Reverse	GGCG*CTCGAG*CTATCGCGAGCTTTCCAGCTCCTT	XhoI

Table 4-1 ceSGT and ceSGT-TPR cloning information.
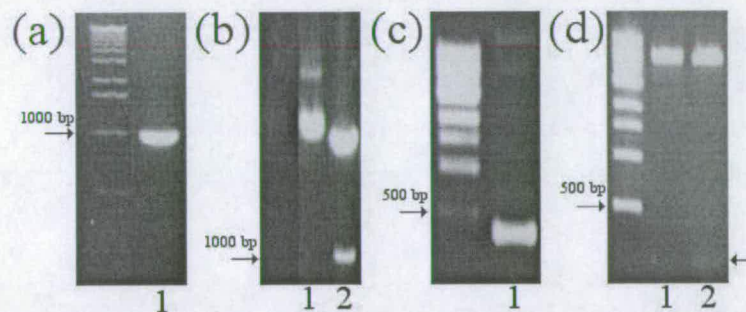
Figure 4-3 Cloning of ceSGT and ceSGT-TPR. (a) PCR of full-length ceSGT cDNA (1011 bp), (b) Restriction digest of cloned full-length ceSGT plasmid. 1 - Uncut plasmid, 2 - Restriction digest showing insert of correct size. (c) PCR of ceSGT-TPR cDNA (375 bp). (d) Restriction digest of cloned plasmid. 1 - Uncut plasmid, 2 - Restriction digest showing insert of correct size.

4.2.2. Expression and purification

Recombinant proteins were expressed in BL21(DE3)-Rosetta 2 *E. coli* (Novagen) in LB liquid media containing kanamycin (25 μg/ml) and chloramphenicol (30 μg/ml). Cultures were grown with shaking at 37 ˚C until the A_{600} was ~0.6, over-expression induced by addition of IPTG to 1 mM and growth continued for a further 4 hours at 37 ˚C. Cells were

harvested by centrifugation (3000 xg for 15 min), resuspended at 10% weight per volume in ice-cold lysis buffer (50mM tris pH 7.5, 5mM EDTA, 1mM DTT, 0.1mM benzamidine, 0.1mM PMSF) plus excess protease inhibitor cocktail (Roche), and sonicated on ice for 6 x 30 second bursts, with 30 seconds cooling in between. The cell lysate was subjected to centrifugation at 30,000 xg for 1 hr at 4 °C.

Figure 4-4 Purification of full-length ceSGT. (a) ceSGT enriched from cell-extract using anion exchange with source-30Q resin at pH 5.5. ceSGT elutes around 100 mM KCl. (b) Purification steps. Lane 1 - cell-extract, lane 2 - anion exchange round 1, lane 3 - repeat anion exchange with finer gradient, 4 - gel-filtration, purified to about 95% purity. 26 kDa degradation product (marked with black arrow) difficult to completely separate.

4.2.2.1. Purification of ceSGT

Untagged ceSGT was purified by a two-step strategy consisting of anion exchange and gel filtration. The calculated isoelectric point (http://www.embl-heidelberg.de/cgi/pi-wrapper.pl) for ceSGT was 4.5 and a bis-tris pH 5.5 buffer was selected for anion exchange. Clarified cell lysate was dialysed overnight against buffer A (50mM bis-tris pH 5.8, 1mM EDTA, 1mM DTT, 0.1mM PMSF, 1mM azide), filtered through a 0.2 μm filter and applied to a Source-Q 30 μm (Pharmacia) column (V_t ~ 10 ml; 2 x 5 cm) pre-equilibrated in buffer A. ceSGT was eluted with a 0-500 mM KCl gradient in buffer A over 100 mls and analysed by SDS-PAGE (Figure 4-4). Fractions containing ceSGT, eluting between 100 and 200 mM KCl, were pooled and concentrated. Protein was then applied to Superdex 200 HR 30/10 column (Amersham Bioscience) pre-equilibrated in buffer B (25mM HEPES pH 7.5, 100mM NaCl and 1mM DTT) and analysed by SDS-PAGE. Fractions containing ceSGT

were pooled and stored on ice at 4°C in buffer B. ceSGT was >95% pure as judged by SDS-PAGE (Figure 4-4).

4.2.2.2. Purification of ceSGT-TPR

Untagged ceSGT-TPR was purified by the same strategy as above. The calculated P_i is 6.9 and a tris pH 9.0 buffer was selected for anion exchange. Clarified cell lysate was dialysed overnight against buffer D (50mM tris pH 9.0, 1mM EDTA, 1mM DTT, 0.1mM PMSF, 1mM azide), filtered through a 0.2 μm filter and applied to the source 30Q column pre-equilibrated in buffer D. ceSGT-TPR failed to bind the resin but was sufficiently enriched in the flow-through (Figure 4-5a). ceSGT-TPR was further purified using a Superdex 75 HR 30/10 column (Amersham Bioscience) equilibrated in buffer C (Figure 4-5b). ceSGT-TPR was >95% pure as judged by SDS-PAGE, and stored on ice at 4 °C in buffer C (Figure 4-5c).
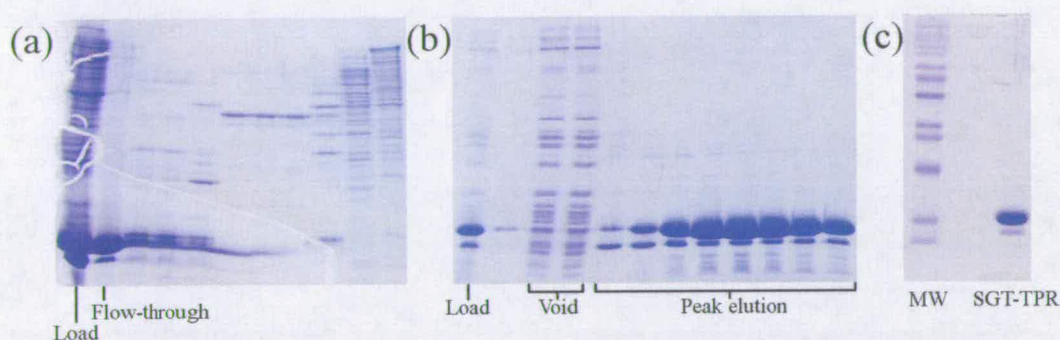
Figure 4-5 Purification of ceSGT-TPR. (a) Anion exchange (Source-30Q) at pH 9 used as first step. ceSGT-TPR highly enriched in flow-through. (b) Further cleaned with gel-filtration using a Superdex 75 column. (c) Purified protein, estimated >95% purity. Difficult to separate from smaller protein carried over from expression.

4.2.3. Mass Spectrometry

The mass spectrometry analyses were carried out on a Voyager DE-STR MALDI-TOF (Applied Biosystems) instrument using α-cyano-4-hydroxycinnamic acid (CHCA) matrix for peptides and sinapinic acid matrix for proteins. Proteins within gel pieces were first reduced, carboxyamidomethylated, and then digested with trypsin (Promega) prior to peptide mass fingerprinting.

4.2.4. Protein Cross-linking

5 μg total protein in 15 μl buffer B was cross-linked with addition of a 1/10th volume of 10 × glutaraldehyde stock made up in buffer B, 0.1% and 0.2% final glutaraldehyde concentrations were used. The reaction was quenched at various time points by addition of a 1/10th volume of 1M tris pH 7.5 and subjected to SDS-PAGE analysis.

4.2.5. Analytical Gel Filtration

Gel filtration studies were carried out on an AKTA explorer FPLC using either a Superdex 200 HR 30/10 column or a Superdex 75 HR 30/10 column at 4 °C. The column was equilibrated with buffer B and calibrated using the following molecular weight standards - ribonuclease A (15.6 kDa), chymotrypsinogen A (20.4 kDa), ovalbumin (49.1 kDa), albumin (67.4 kDa), aldolase (176 kDa), catalase (219 kDa), ferritin (416 kDa) and thyroglobulin (699 kDa). 200 μl protein at concentrations ranging from 10-500 μM was applied to the column and run at 0.5 ml min^{-1}. Calibration curves for both molecular weight and Stokes radius were generated using SigmaPlot 9.0. For molecular weight a plot of K_{av} against log molecular weight was used where $K_{av} = (V_e - V_o)/(V_t - V_o)$; V_e = retention volume, V_o = void volume, V_t = column bed volume. The Laurent and Killander solution (Laurent and Killander, 1964) for the calculation of the Stokes radius was used where $\sqrt{(-\log K_{av})}$ is plotted against the Stokes radius.

4.2.6. Circular Dichroism Spectroscopy

All CD spectra were recorded with a nitrogen-flushed Jasco J-810 spectropolarimeter. For far-UV CD measurements, proteins were dialysed against buffer containing 25 mM NaPO$_4$ and 50 mM KF prior to analysis. 10-50 μM protein was used with a path length of 0.02 cm. Data were recorded from 250 to 185 nm and accumulated over 2 runs using a 2 s time constant, 10 nm min^{-1} scan speed and a spectral bandwidth of 1 nm. Spectra were corrected for buffer and the secondary structure content of the protein was estimated by deconvolution with CDSSTR using the web server DICHROWEB (Whitmore and Wallace, 2004).

For near-UV CD measurements, 100 μM protein in buffer B was used with a path length of 1 cm. Data were recorded from 340 to 250 nm and accumulated over 3 runs using a 4 s time constant, 5 nm min^{-1} scan speed and a spectral bandwidth of 1 nm. For ligand binding experiments, titrations were conducted in a stepwise manner from a stock peptide solution of 4 mM with the volume of added peptide (in buffer B) not exceeding 15% of the starting volume. Peptides were titrated directly into the cuvette containing an initial volume of 1500

μl, mixed by pipetting and allowed to equilibrate for 10 mins. Titrated spectra were corrected for both dilution and intrinsic CD of the buffer/peptide alone.

To determine binding affinities, the plot of titration CD data signal at a single wavelength against the molar concentration of the ligand was analysed by non-linear regression analysis using the software package SigmaPlot 9.0. Data were fit to the equation

$$F = F_{free} + (F_{saturated} - F_{free})((K_d + [A] + [B]) - \sqrt{((K_d + [A] + [B]^2 - 4[A][B]))/2[A]}$$

where F = CD signal at a specified wavelength, K_d = dissociation constant, [A] = protein concentration and [B] = peptide concentration.

4.2.7. Isothermal titration calorimetry

Binding of Hsp90/Hsp70 C-terminal peptides MEEVD and GPTIEEVD to the ceSGT TPR domain was measured by isothermal titration calorimetry (ITC) using a MicroCal VP-ITC titration calorimeter (MicroCal Inc., Northhampton, USA). 50 to 100 aliquots of 2-5 μl peptide solution (4-8 mM; dissolved in buffer B and pH adjusted to match the protein solution) were titrated at 25 °C by injection into ~1.3 ml SGT-TPR solution (30-100 μM) in the chamber. Peptide only controls were conducted to allow for determination of heats of ligand dilution. After subtraction of dilution heats, calorimetric data were analyzed using the evaluation software provided by the manufacturer.

4.2.8. Protein crystallisation and crystal screening

Crystallisation trials were conducted using the hanging drop vapour diffusion method from a 10 mg ml^{-1} protein solution in buffer B at 4 and 18 °C. Hampton Crystal Screen™ and Crystal Screen™ II were used for screening with a 2 μl drop consisting of a 1:1 ratio of protein and well solution. Trials were carried out in the absence and presence of the Hsp90/Hsp70 peptides MEEVD and GPTIEEVD. Peptides were reconstituted in distilled H$_2$0 and mixed with protein at a ratio of 1:1.3. Optimisation of initial hits was conducted using a grid screen around promising conditions. Putative protein crystals were screened at station 10.1, SRS, Daresbury, UK or station BM14, ESRF, Grenoble.

4.2.9. Structure analysis

A homology model of the *C. elegans* SGT TPR domain (residues 100-225) was generated with SWISS-MODEL using several structures of the TPR domain from PP5 as a template (PDB-IDs 2BUG, 1WAO and 1A17). Evolutionary conservation analysis carried out with ConSurf (Glaser et al., 2003) using the empirical Bayesian method. Homologous sequences

for TPR domains from ceSGT, Hop (TPR1 and TPR2A), PP5 and Chip were extracted from the UniProt database using BLAST (Altschul et al., 1997) with a cut-off E-value of $1e^{-20}$. Alignments were generated with 3D-COFFEE (Armougom et al., 2006) (Appendix A.3.1.-A.3.3.) prior to analysis with ConSurf. Electrostatic-potential maps were calculated with APBS (Baker et al., 2001) using a PyMol plug-in (http://www-personal.umich.edu/~mlerner/PyMOL/). Charges were assigned using PDB2PQR (Dolinsky et al., 2004) and an AMBER forcefield (Case et al., 2005).

4.3. Results and discussion

4.3.1. Purification

ceSGT and ceSGT-TPR were successfully expressed and purified to over 90% purity based on estimation of relative band densities from coomassie stained SDS-PAGE gels (Figures 4-4c and 4-5c). During the purification of ceSGT, a 26 kDa protein with similar elution characteristics was difficult to completely separate. MALDI-TOF analysis of a trypsin digest of this 26 kDa protein revealed it to be a degradation product lacking the C-terminal region, with cleavage after residue 212, now designated ceSGTΔC. The presence of this contaminant with similar physiochemical properties to the full-length protein hampered the purification and resulted in a low final yield of approximately 0.5 mgs per litre culture.

4.3.2. Biophysical and biochemical characterisation of ceSGT and ceSGT-TPR

4.3.2.1. Far-UV CD spectroscopy analysis of ceSGT and ceSGT-TPR

Far-UV CD experiments sensitive to protein secondary structure were used to study the folded state of ceSGT and ceSGT-TPR. Both spectra are characterised by positive maxima at 193 nm and negative minima at 208 and 222 nm typical of helical proteins (Figure 4-6). Deconvolution of the CD spectra with CDSSTR predicts secondary structure content of 58% helix, 10% strand, 15% turn and 19% unordered for ceSGT; and 81% helix, 10% turn and 9% unordered for ceSGT-TPR. These are in good agreement with secondary structure prediction algorithm PHD (Rost, 1996), which predicts a full-length protein with 60-70% helix and 30 - 40% loop (Figure 4-2). The highest concentration of helix is across the TPR domain which was shown to be virtually all helical indicating that the protein was correctly folded in comparison to related TPR domains of known structure (Das et al., 1998; Granzin et al., 2006; Scheufler et al., 2000; Taylor et al., 2001; Wu and Sha, 2006; Zhang et al., 2005).
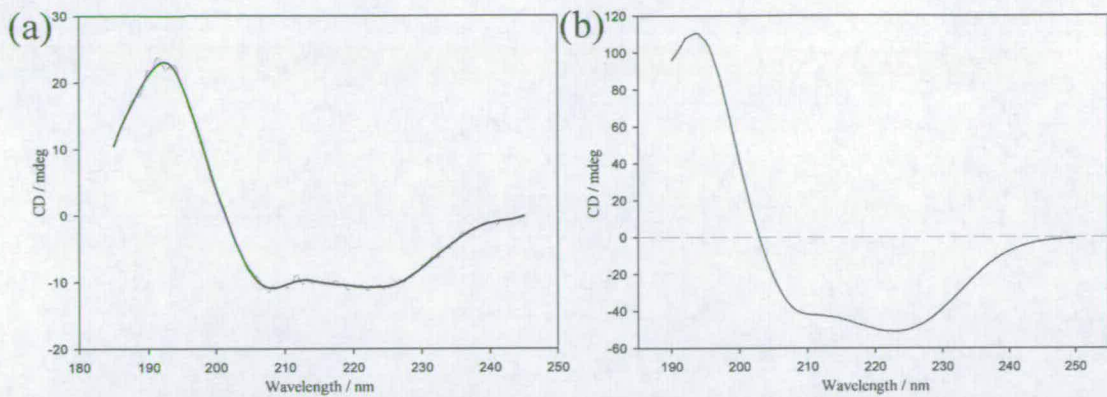
Figure 4-6 Circular dichroism spectroscopy analysis of ceSGT and ceSGT-TPR. (a) Spectra for ceSGT. Deconvolution with CDSSTR predicts a protein with 58% helix, 10% strand, 15% turn and 19% unordered. (b) Spectra for ceSGT-TPR. Deconvolution with CDSSTR predicts a protein with 85% helix, 10% turn and 5% unordered.

4.3.2.2. Glutaraldehyde cross-linking of ceSGT

Human SGT has previously been shown to oligomerise (Liou and Wang, 2005; Tobaben et al., 2003). Based on cross-linking studies, Liou and Wang (Liou and Wang, 2005) suggested the formation of dimers although the result, as analysed by SDS-PAGE, was more consistent with the molecular weight of a trimer. To investigate the self-association properties of *C. elegans* SGT, and also to clarify the oligomeric state, glutaraldehyde cross-linking was carried out. Glutaraldehyde is a small homobifunctional amine reactive cross-linker. Visualisation of cross-linked complexes on coomassie stained SDS-PAGE gels clearly shows the accumulation, with respect to both time and glutaraldehyde concentration, of a high molecular weight species mirrored by a concomitant decrease in concentration of monomeric protein (Figure 4-7). In addition, there is also evidence for the accumulation of a smaller molecular weight complex likely corresponding to the oligomerisation of the full-length protein and ceSGTΔC. From the gel, the larger more abundant complex is estimated to have a mass in the region of 110-120 kDa and the smaller complex 90-100 kDa, consistent with the molecular weight of a trimeric complex.
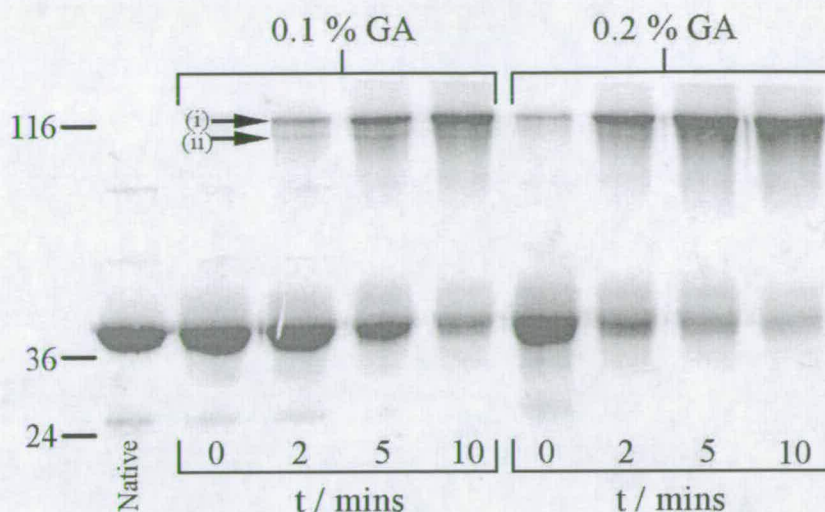
Figure 4-7 Glutaraldehyde cross-linking of ceSGT. Purified ceSGT was cross-linked with glutaraldehyde (GA) for indicated time periods. The reaction was quenched with 1M tris pH 7.5 and analysed by SDS-PAGE. Accumulation of a complex in the region of 110-120 kDa in size, and a less abundant complex 90-100 kDa in size, is evident with respect to both time and GA concentration, labelled (i) and (ii) respectively. Molecular weight markers are indicated to the left of the gel.

4.3.2.3. MALDI-TOF mass-spectroscopy of ceSGT oligomers

MALDI-TOF mass spectrometry was used to accurately determine the oligomeric state. Native ceSGT yielded a major monomeric peak of 36440 Da (Figure 4-8a). A peak approximately one third the magnitude of the ceSGT peak was observed at 26499 Da, corresponding to degradation product ceSGTΔC. The ceSGT oligomeric complex was also captured in the analysis of the native protein with peaks corresponding to full-length homo-dimer (MW = 73552 Da) and ceSGT/ceSGTΔC hetero-dimer (MW = 63264 Da) detected. Analysis was repeated with glutaraldehyde cross-linked proteins, resulting in increased intensity spectra for the oligomeric species coupled with an increase in mass according to the incorporation of glutaraldehyde (MW homo-dimer = 75069 Da, MW hetero-dimer = 64463; MW glutaraldehyde = 100.1). In addition, homo-dimerisation of the ceSGTΔC was also evident (MW = 54825) (Figure 4-8b). This provides direct evidence that ceSGT is capable of forming dimers and also that the ability to self-associate is maintained in the absence of the C-terminal domain.
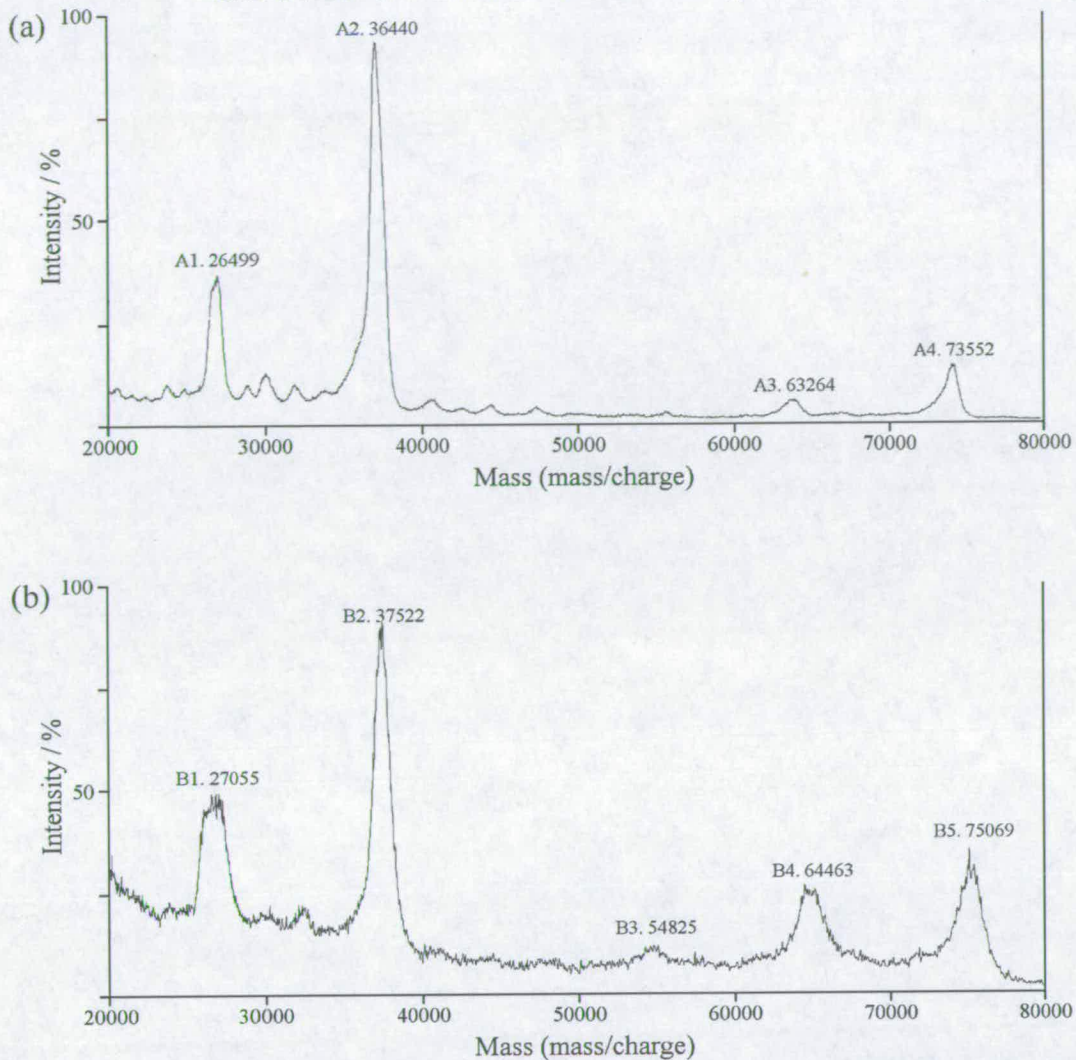
Figure 4-8 MALDI-TOF mass spectrometry analysis of native and cross-linked ceSGT. The mass of the ceSGT oligomer was investigated using MALDI-TOF mass spectrometry. a) Spectra for native ceSGT reveals a major peak corresponding to singly charged monomeric ceSGT with a mass of 36440 Da (A2). Degradation product ceSGTΔC has a mass of 26499 Da (A1). Peaks corresponding to full length homo-dimer (A4. 73552 Da) and ceSGT/ceSGTΔC hetero-dimer (A3. 63264 Da) are also recorded. b) Spectra for glutaraldehyde cross-linked ceSGT. Respective dimer peaks increase in intensity after cross-linking (B4 and B5) and the presence of ceSGTΔC homo-dimer is also recorded (B3). There is an increase in mass for all species indicating incorporation of glutaraldehyde (MW = 100.1 Da) molecules into protein.

The observation of the dimeric complex in both the native and cross-linked samples was surprising. MALDI-TOF mass-spectrometry using traditional low pH conditions is generally believed to only be able to detect covalently bound species. This suggests that ceSGT forms high affinity dimers. No evidence was found for the existence of any higher molecular weight oligomers suggesting the exclusive formation of dimers. The significantly larger molecular weight estimated from SDS-PAGE analysis of cross-linked SGT is likely to be an artefact imposed by the cross-linking process, with either lower amounts of SDS being

incorporated into the denatured protein and/or conformational constraints imposed by cross-linking affecting gel migration.

4.3.2.4. Gel filtration analysis of ceSGT and ceSGT-TPR

Semi-analytical gel filtration was used to investigate the hydrodynamic properties of ceSGT and ceSGT-TPR. ceSGT was resolved on a Superdex 200 HR 30/10 column (Figure 4-9a). At all concentrations tested, ceSGT eluted predominantly as single peak suggesting a mono-disperse population and an obligate dimer with a very tight association. The elution profile was, however, consistent with a protein of significantly higher molecular weight with a mean retention volume of 11.65 ± 0.1 ml (mean ± SEM; n=3). Using the calculated calibration curves, ceSGT is estimated to have an apparent molecular weight of 216.8 kDa (actual MW of dimer = 73 kDa). More accurately, gel filtration separates particles based on their hydrodynamic properties and ceSGT elutes with a predicted Stokes radius of 51.2 Å. A globular protein of 73 kDa would have a Stokes radius of approximately 33 Å. The discrepancy between the actual molecular weight and hydrodynamic properties is indicative of a non-globular conformation.

Protein	MW (Da)	Predicted MW (Da)	Stokes radius (Å)
ceSGT	36467	216800	51.2
ceSGT-TPR	14509	26500	19.3
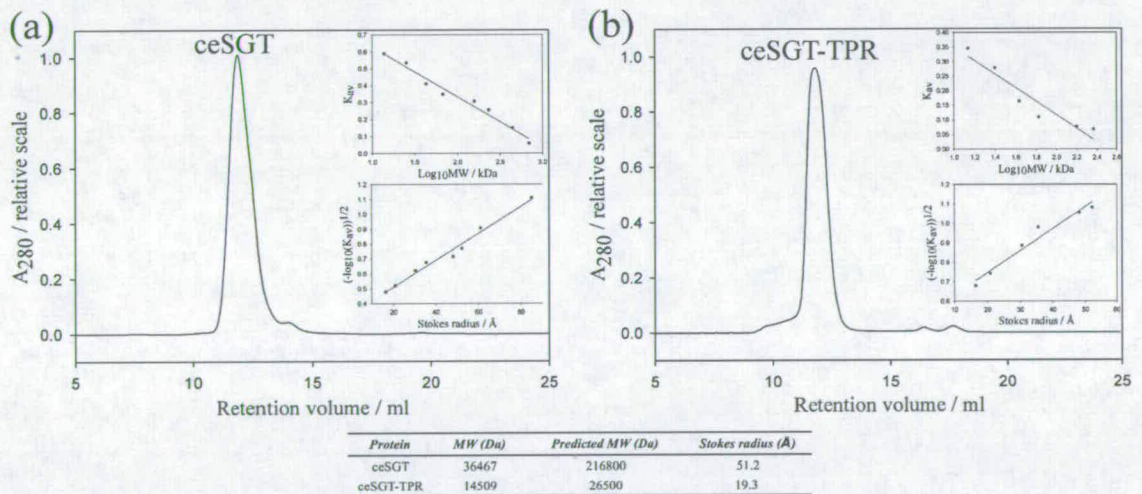
Figure 4-9 Analytical gel-filtration analysis of ceSGT and ceSGT-TPR. (a) Hydrodynamic properties of ceSGT were analysed using a Superdex 200 HR 10/30 gel-filtration column. ceSGT elutes with a retention volume of 11.65 ± 0.1 ml (mean ± SEM; n=3) and has an apparent molecular weight of 217.1 kDa and a Stokes radius of 52 Å. (b) Hydrodynamic properties of ceSGT-TPR analysed using a Superdex 75 HR 10/30 gel-filtration column. ceSGT-TPR elutes with a retention volume of 11.84 ± 0.12 ml (mean ± SEM; n=3) and has an apparent molecular weight of 26.5 kDa and Stokes radius of 19.3 Å. (c) Table of actual MW, predicted MW and Stokes radii.

Due to its smaller size ceSGT-TPR was resolved on a Superdex 75 HR 30/10 column, eluting with a retention volume of 11.84 ± 0.12 ml (mean $\pm$ SEM; n=3) and a predicted molecular weight and Stokes radius of 26.5 kDa (actual molecular weight 14.5 kDa) and 19.3 Å respectively (Figure 4-9b). This suggests a more compact tertiary fold and is consistent with the monomeric dimensions of related TPR domains of known structure.

Taken together, with a mass of 73 kDa and a Stokes radius of 5.2 nm, the hydrodynamic dimensions of the ceSGT dimer in relation to its molecular weight would suggest a protein with a low level of compactness and an extended conformation. The dimerisation of ceSGTΔC and the apparent monomeric state of ceSGT-TPR provide indirect evidence that the self-association properties of ceSGT are mapped to the N-terminal domain, consistent with studies on human SGT. Primary structure analysis using the program COILS (Lupas et al., 1991) highlights the significant probability of the formation of a small two-stranded coiled-coil in the N-terminus which is a likely candidate for mediating the self-association (Figure 4-2). A structural model is proposed whereby dimerisation of ceSGT mediated by a parallel coiled-coil interaction in the N-terminal domain results in an extended V-shaped dimer (Figure 4-10).
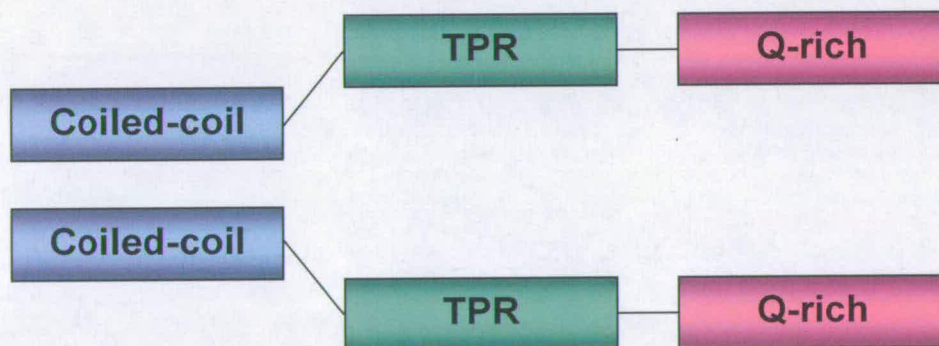
Figure 4-10 Predicted quaternary structure of ceSGT. SGT dimerises via the N-terminal coiled-coil forming an elongated V-shaped structure.

4.3.3. Characterisation of the interaction between ceSGT and Hsp90/Hsp70

4.3.3.1. Interaction between ceSGT and human Hsp90α-631

The interaction of ceSGT and a C-terminal construct of human Hsp90α was investigated using semi-analytical gel filtration. 30 μM ceSGT and 100 μM Hsp90α-631 were resolved on a Superdex 200 HR 30/10 column with retention volumes of 11.65 and 12.92 ml respectively (Figure 4-11a). 30 μM ceSGT and 100 μM hHsp90α-631 were then incubated for 1 hour at room temperature prior to application to the column. The incubated protein eluted with two peaks with retention volumes of 11.19 and 12.92 ml. The shift in elution position of the higher molecular weight peak from 11.65 ml for ceSGT alone to 11.19 ml for the two proteins incubated together suggested the co-migration of a ceSGT-hHsp90α-631 complex. To verify this, the eluate from fractions collected every 0.5 ml was subjected to SDS-PAGE analysis (Figure 4-11b). When run together, the elution profile of ceSGT is altered with the peak position shifted by about 0.5 ml and ceSGT detected in greater abundance earlier in the elution. Furthermore, although the majority of hHsp90α-631 elutes in the same position, a small fraction is detected with a significant shift in retention volume of >1 ml suggesting the interaction of ceSGT and hHsp90α-631.

Additional attempts to demonstrate an interaction using a larger *C. elegans* Hsp90 construct or the full-length protein were inconclusive. This may reflect a weak and transient nature of the interaction.

4.3.3.2. Interaction of the ceSGT TPR domain with the C-terminal Hsp90/70 peptides

Studies of the interaction between the TPR domains TPR1 and TPR2A of Hop with Hsp70 and Hsp90 respectively have shown that the isolated TPR domains interact with the isolated C-terminal peptides of Hsp90 or Hsp70 with comparable affinities to the full-length proteins (Scheufler et al., 2000). For this reason, the interaction of the isolated ceSGT TPR domain with the extreme C-terminal Hsp90 peptide MEEVD and the C-terminal Hsp70 peptide GPTIEEVD was also investigated using two methods – isothermal titration calorimetry (ITC) and circular dichroism (CD) spectroscopy.
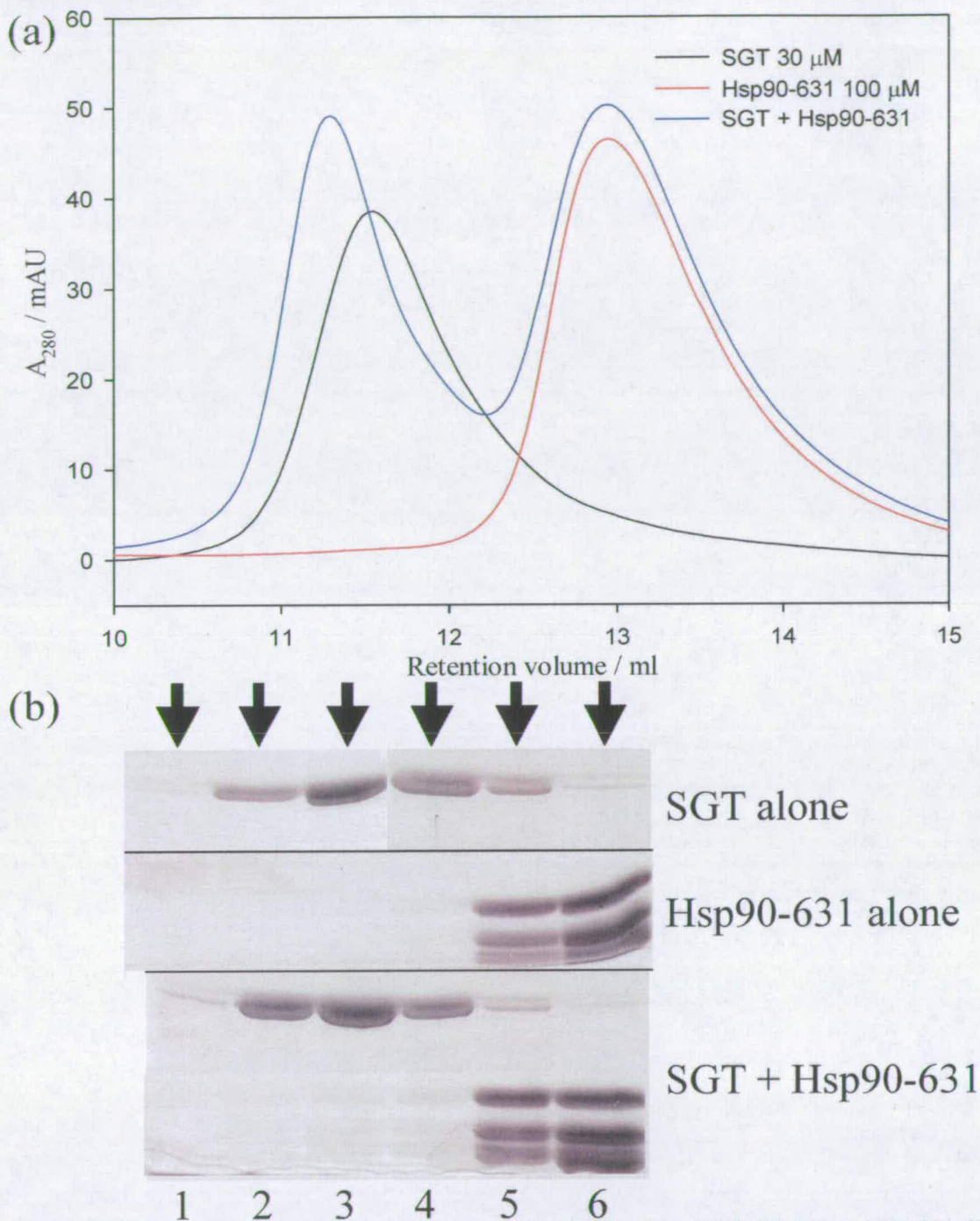
Figure 4-11 Interaction of ceSGT with human Hsp90-631. (a) ceSGT (30μM) and Hsp90α-631 (100 μM), alone and incubated together, were resolved on a superdex 200 gel-filtration column. A shift in retention volume of the high molecular weight peak when incubated together (blue) compared to alone (black and red) suggests complex formation. (b) SDS-PAGE analysis of eluate from gel-filtration (arrows indicate elution position of eluate) When run alone ceSGT peaks between lanes 3 and 4 and Hsp90-631 can be detected in lane 4. However, when run together the peak of ceSGT is shifted slightly to lane 3 and Hsp90α-631 can be detected in lane 2 corresponding to a shift of ~1 ml in retention volume
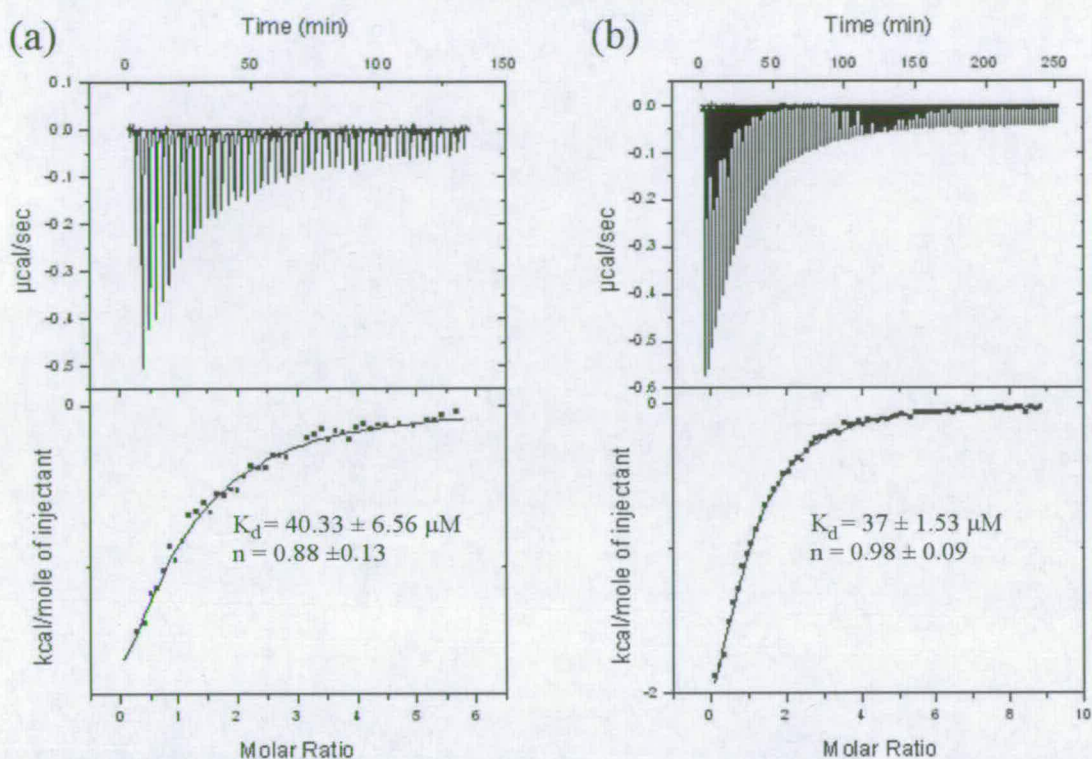
Figure 4-12 ITC analysis of the interaction between ceSGT-TPR and C-terminal peptides from Hsp90 and Hsp70. Top, raw data on heat change obtained after injection of peptides into ceSGT-TPR. Bottom, the heat changes as a function of the molar ratio of peptide and protein. (a) ceSGT-TPR plus Hsp90 peptide MEEVD. Results from representative run shown. K_d calculated from 3 separate runs, quoted as mean ± SEM. (b) ceSGT-TPR plus Hsp70 peptide GPTIEEVD. Results from representative run shown. K_d calculated from 3 separate runs, quoted as mean ± SEM.

ITC is a method of choice for measuring protein-ligand interactions. ITC experiments directly measure the heat absorbed or released upon interaction and allow the calculation of binding enthalpy (ΔH), the equilibrium dissociation constant (K_d) and the stoichiometry (n). The binding of Hsp70 peptide GPTIEEVD and Hsp90 peptide MEEVD to the TPR domain of ceSGT was investigated using ITC. Figure 4-12 shows the results of a representative experiment in which the TPR domain was titrated with the peptide. The affinity values measured for both peptides were very similar with GPTIEEVD and MEEVD having K_d values of 37 ± 1.53 μM and 40.33 ± 6.56 μM respectively. As expected, both peptides bound with a 1:1 stoichiometry. The changes in enthalpy on binding in both cases were negative indicating an exothermic reaction; however, this was greater for the binding of the longer Hsp70 heptapeptide (-3412.33 ± 183.45 cal/mol (mean ± SEM; n = 3)) compared with the Hsp90 pentapeptide (-2530.67 ± 86.46 cal/mol (mean ± SEM; n = 3)).
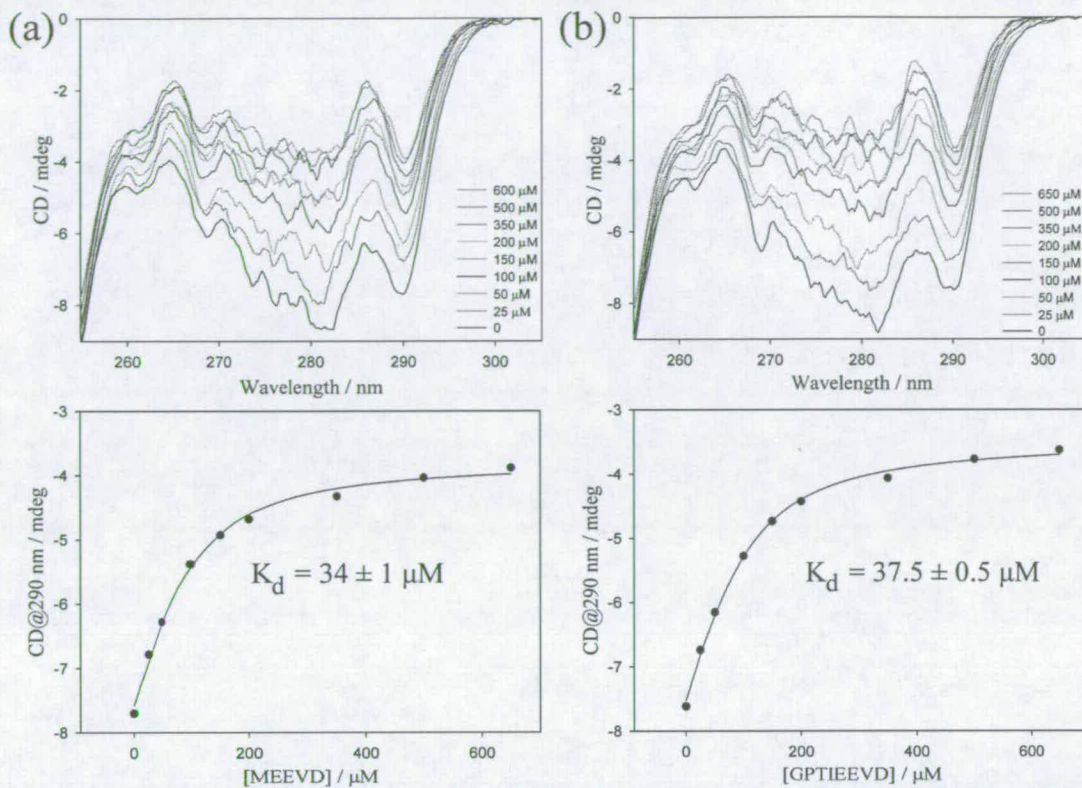
Figure 4-13 Near-UV CD analysis of the interaction between ceSGT-TPR and C-terminal peptides from Hsp90 and Hsp70. Near-UV CD spectra for ceSGT-TPR with increasing concentrations of Hsp90 peptide MEEVD (a) and Hsp70 peptide GPTIEEVD (b). The plot of CD at 290 nm verses concentration was fit to a tight-binding equation using non-linear regression to calculate the binding affinity. Spectra from representative experiment shown, repeated twice for each peptide. K_d quoted as mean ± SEM.

The peptide-TPR domain interaction was also assessed with CD spectroscopy. The CD spectrum of a protein in the near-UV spectral region (250-350 nm) can be sensitive to certain aspects of tertiary structure and therefore can be used to study macromolecular interactions. The chromophores are the aromatic residues phenylalanine, tyrosine and tryptophan, and also disulphide bonds. Each peptide alone lacked any near-UV CD; however, titration of either peptide into ceSGT-TPR caused saturatable dose-dependent perturbations in the spectra indicative of changes in the environment of aromatic residues as a result of molecular interaction (Figure 4-13). By assuming that the change in CD intensity at a given wavelength was proportional to the extent of peptide binding, the K_d for the interaction was calculated by fitting a tight-binding equation (see section 5.2.6) to a plot of CD signal at 290 nm to the molar concentration of peptide. Resulting K_d values of 37.5 ± 0.5 µM and 34 ± 1 µM (mean ± SEM; n = 2) for GPTIEEVD and MEEVD respectively were in good agreement with those measured by ITC.

Based on experiments with the TPR domain from protein phosphatase PP5 it has been proposed that a coupled folding and binding mechanism may be a common feature of TPR domain recognition (Cliff et al., 2005). To investigate whether any change in secondary structure accompanied binding of either peptide to ceSGT-TPR, far-UV CD spectra were recorded in the absence and presence of saturating concentrations of peptide. No difference in far-UV CD spectra was observed in the presence of either peptide (Figure 4-14); indeed, the isolated domain was judged to be almost completely folded with a predicted secondary structure content of 81% helical, 10% turn and 9% unordered.
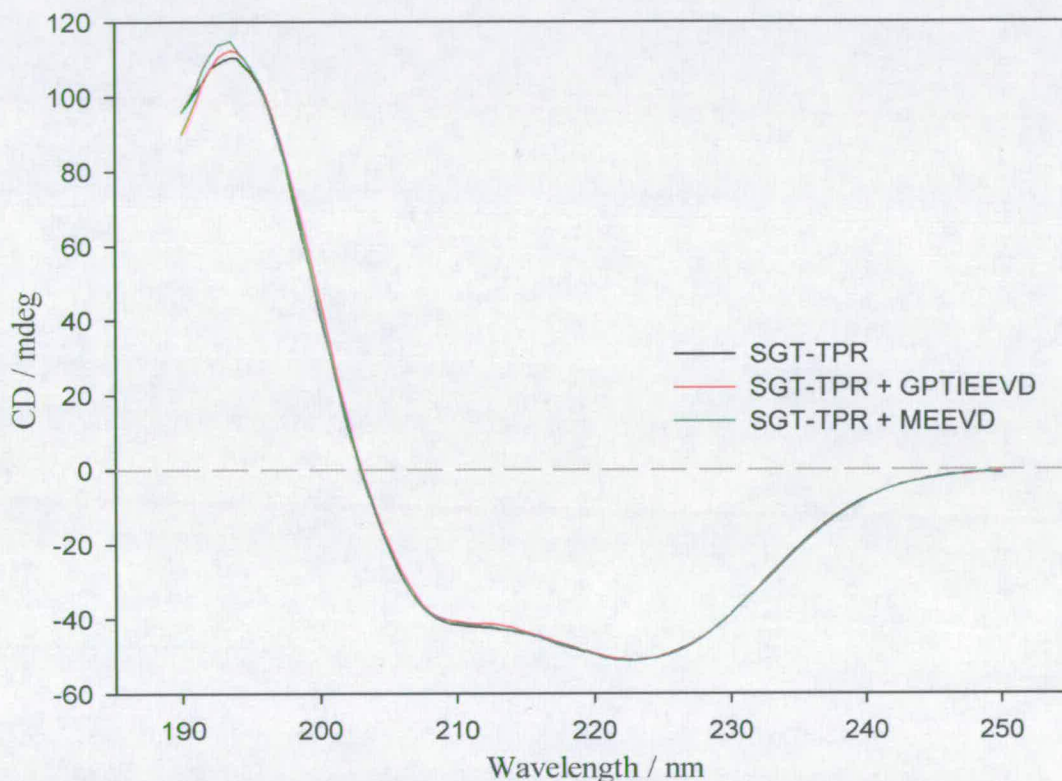
Figure 4-14 Far-UV CD analysis of ceSGT-TPR in complex with Hsp70/90 C-terminal peptides. Peptide-free and peptide-bound spectra are superimposable indicating no change in secondary structure upon peptide binding.

SGT has been shown to interact with both Hsp70 and Hsp90 both *in vitro* and *in vivo* so it was unsurprising that the TPR domain was found to interact with both peptides (Angeletti et al., 2002; Liou and Wang, 2005; Tobaben et al., 2003; Yin et al., 2006); however, studies have shown a preference for Hsp70 over Hsp90 (Angeletti et al., 2002) so it was anticipated that there would be a higher affinity for the Hsp70 peptide. It is possible that for the

interaction with SGT, sequences upstream of the short C-terminal peptides may provide additional specificity. Indeed the arginine residue at peptide position -5, preceding the methionine, in Hsp90 has been shown to reduce affinity (Brinker et al., 2002; Scheufler et al., 2000). The affinity of a 24-mer Hsp70 peptide for the TPR domain of human SGT, the only published SGT interaction kinetic data, is 225 µM (Cortajarena and Regan, 2006). This is almost one order of magnitude weaker than results here, although other K_d values from the same study were also proportionally higher than reported elsewhere.

The promiscuous interaction of TPR domains with both Hsp70 and Hsp90 is not uncommon and is exhibited in numerous co-chaperones including Chip, cyclophilin-40 and Tom70. In fact, with the exception of the TPR1 domain from Hop, TPR domains interacting with Hsp70 also have been shown to interact with Hsp90. The structures of 10 TPR domains known to interact with Hsp70 or Hsp90 have been solved. All, excluding a domain-swapped form of bovine cyclophilin-40 (Taylor et al., 2001), are formed by three TPR repeats which form a concave peptide interaction surface and all, with the exception of Arabidopsis FKBP42 (Granzin et al., 2006), are capped with a C-terminal helix with important functional significance.

Five of the TPR domains of known structure have been solved in complex with an Hsp70/90 C-terminal peptide (Cliff et al., 2006; Scheufler et al., 2000; Wu et al., 2004; Zhang et al., 2005). In all cases the interaction involves a similar set of TPR residues although the peptide orientation and precise ensemble of interacting residues differ between the structures. The canonical anchoring mechanism shared in all structures was first termed the "two-carboxylate clamp" (Scheufler et al., 2000) and consists of two lysine residues (Lys[4] and Lys[70]) that interact with the peptide Asp[0] side-chain and C-terminal carboxylates. Additional conserved clamp residues (Asn[9], Asn[40] and Arg[74]) form direct interactions with the peptide backbone and residues Glu[-2] and Glu[-3] (Figure 4-15). A hydrophobic pocket made of strictly conserved residues Ala[20] and Tyr[24] and a more variable hydrophobic residue as position 12 of the first TPR domain accommodates the Val[-1] residue.

Figure 4-15 Hsp70/90 interacting TPR domains. (a) Alignment of Hsp70/90 interacting TPR domains. Two-carboxylate clamp residues marked with red stars, additional important residues with yellow stars. (b) General structure of TPR domain-EEVD complex with highlighted residues.

Figure 4-16 shows the alternate conformations the C-terminal peptides adopt in complex with TPR domains from Hop, PP5 and Chip. All show conserved anchoring of the VD motif by the two-carboxylate clamp; however, all exhibit remarkably different upstream peptide binding orientations. The crystal structure of the domains of TPR1 and TPR2A from Hop, solved with Hsp70 and Hsp90 peptides respectively, show that both bind in an extended conformation with differences in the N-terminal hydrophobic peptide residues important for determining the specificity of the interaction. In contrast, the NMR structure of the PP5-Hsp90 peptide complex shows the peptide kinked away from the concave face at residue Val^{-1} with the upstream residues not forming any measurable interactions. Further, upstream of Glu^{-3} the Hsp90 peptide in complex with Chip departs significantly from the Hop bound conformation. A large hydrophobic pocket formed by the third TPR repeat and the C-terminal capping helix accommodates Met^{-4} and orientates the peptide out away from the TPR channel.
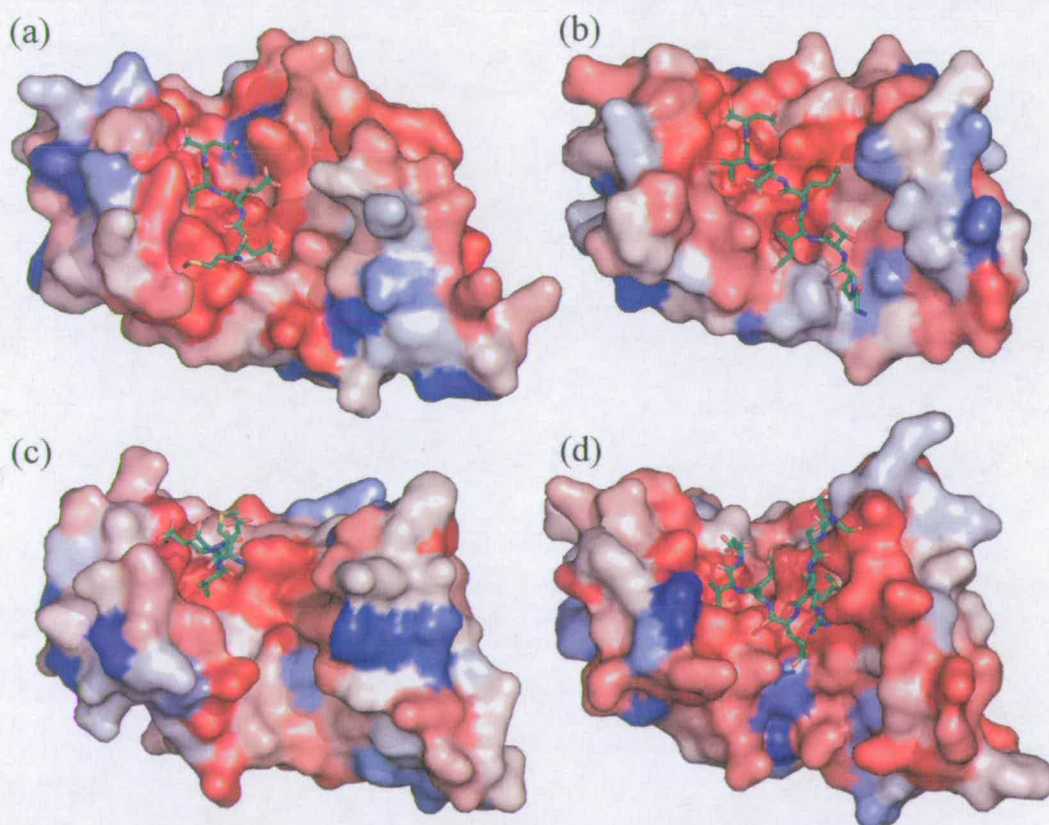
Figure 4-16 Evolutionary conservation of TPR domain-peptide complexes. All structures are coloured according to evolutionary conservation from red (highly conserved) to blue (variable) (a) hopTPR2A-MEEVD complex (1ELR). (b) TPR1-GPTIEEVD complex (1ELW). (c) PP5-TPR-MEEVD complex (2BUG). (d) CHIP-DTSRMEEVD complex (2C2L). All structures exhibit a pattern of evolutionary conservation reflective of peptide binding mechanism.

Interestingly, the pattern of surface residue conservation for each protein correlates well with the observed peptide binding orientation in all four examples (Figure 4-16). Evolutionary analysis of a model of the ceSGT TPR domain in conjunction with overlaying the TPR bound peptide structures indicates that ceSGT-TPR could accommodate peptides in an analogous manner to TPR1 of Hop, PP5 or Chip but not TPR2A of Hop. Moreover, NMR chemical shift analysis of the interaction between human SGT and an Hsp70 24-mer peptide has shown residues the length of the channel to be affected by peptide binding (Cortajarena and Regan, 2006). The pattern of binding site conservation and chemical shift analysis intimate a similar extended binding orientation for the Hsp70 peptide as witnessed in the Hop TPR1 peptide structure (Figure 4-17c, coloured yellow). The Hsp90 peptide could interact in the same fashion although the pattern of conservation and surface characteristics could also suggest similar mechanisms as for PP5 or Chip. In support of alternate peptide binding orientations for TPR domains which bind both Hsp70 and Hsp90, a mutational analysis of key TPR residues in cyclophilin-40 found different residues had different affects on Hsp70 or Hsp90 binding (Carrello et al., 2004).

The biological relevance of either the homo-dimerisation of SGT or the dual recognition of Hsp70 and Hsp90 is unclear. SGT has been shown to affect a myriad of cellular functions and important roles are emerging in neuronal synaptic transmission (Bai et al., 2007; Natochin et al., 2005; Swayne et al., 2006; Tobaben et al., 2001), the cell cycle (Winnefeld et al., 2004), apoptosis (Wang et al., 2005; Winnefeld et al., 2006; Yin et al., 2006), and viral replication (Callahan et al., 1998; Cziepluch et al., 2000; Handley et al., 2001). Many of these functions are linked to the ability of SGT to interact with Hsp70 or Hsp90. Interestingly, SGT has been shown to be a chaperone in its own right (Tobaben et al., 2001) and the ability to interact with Hsp70 and Hsp90 may allow the transfer of proteins along different folding pathways. Such chaperone communication has been documented for Hsp40, which binds newly synthesised polypeptides and passes them to Hsp70; and Hop, which binds in tandem to both Hsp70 and Hsp90, mediating the transfer of a family of Hsp90 client proteins. The existence of SGT as a dimer raises the possibility that it could interact simultaneously with both Hsp70 and Hsp90 in a manner similar to Hop although further work is required to investigate this.

(a)

(b)

Conserved | Variable

−5.000 | 5.000

(c)

(d)
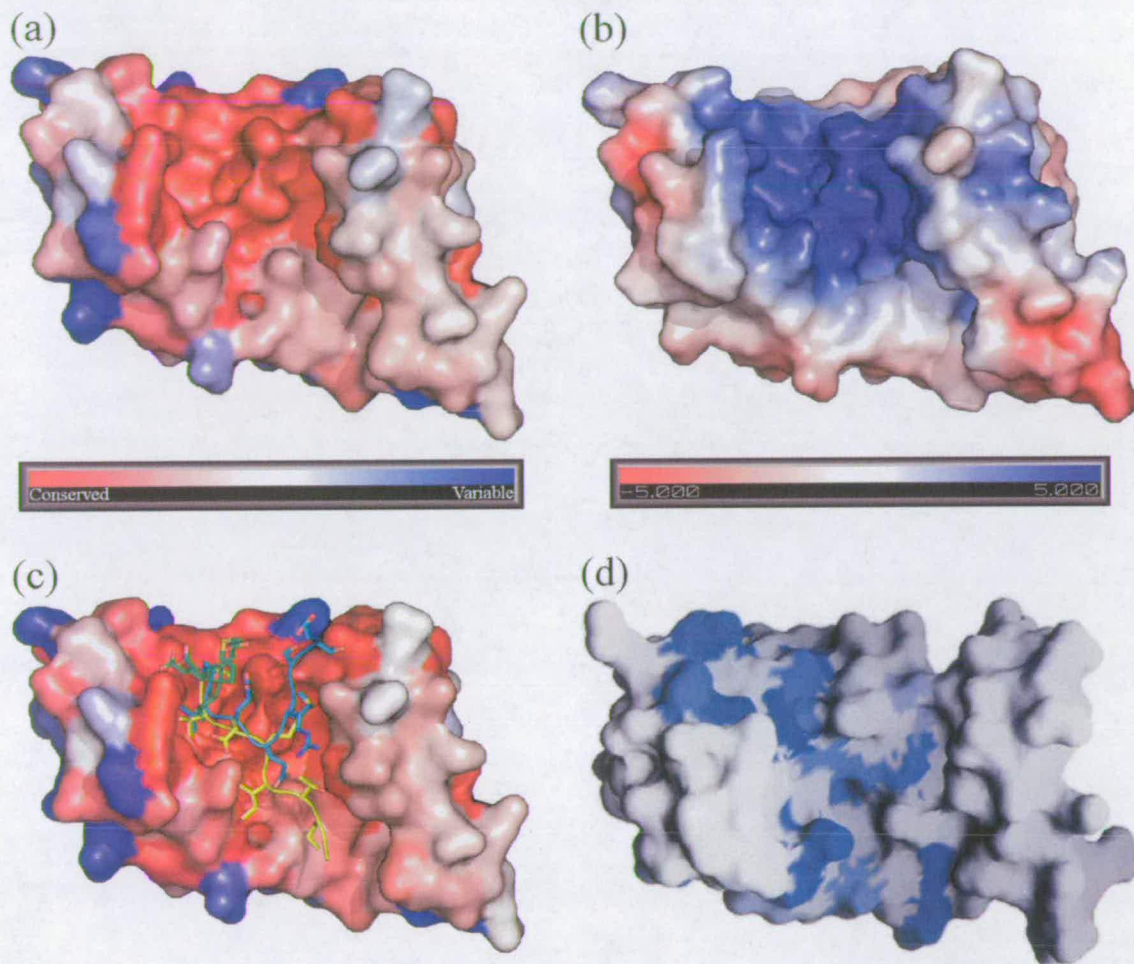
Figure 4-17 Surface properties of modelled ceSGT-TPR. (a) ConSurf analysis, coloured from red (conserved) to blue (variable). (b) Electrostatic surface coloured from red (-5 eV) to blue (5 eV). (c) ConSurf coloured surface with hopTPR1, PP5 and CHIP peptides docked. (d) NMR chemical shift analysis with blue indicating residues involved in human SGT-TPR:Hsp70 peptide interaction (Figure taken from Cortajarena and Regan, 2006).

Of interest, in a transgenic *C. elegans* model for Alzheimer disease, both ceSGT and *C. elegans* Hsp70 homologue Hsp70A were found to interact, directly or indirectly, with human β amyloid (Aβ) (Fonte et al., 2002). Hsp70A contains the C-terminal motif GPTIEEVD shown here to interact with the TPR domain from ceSGT, suggesting the formation of an Hsp70A-ceSGT complex. Intriguingly, in the same study, RNAi targeted against the ceSGT transcript was shown to reduce Aβ expression induced toxicity. Human SGT has been shown to negatively regulate the activity of Hsp70 (Angeletti et al., 2002) implicating a protective function of Hsp70 in modulating Aβ toxicity. Thus, the interaction between the SGT TPR domain and Hsp70 may represent a potential target for small-molecule peptidomimetics for the treatment of Alzheimer's disease.

4.3.4. Crystallisation trials of ceSGT and ceSGT-TPR

Hampton sparse matrix screens Crystal Screen™ and PEG/Ion Screen™ were used for initial crystallisation trials. For ceSGT, small plates with an hexagonal habit were grown from condition 14 of the PEG/Ion Screen™ (200 mM potassium thiocyanate, 20% PEG 3350, pH 7.0) at 4 °C. Optimisation around these conditions at 4 °C failed to reproduce the crystals; however, at 18 °C very fine needles were grown by lowering the PEG 3350 concentration to 5% (Figure 4-18a). Both the small plates and fine needles failed to diffract. ceSGT-TPR crystallisation trials produced several promising hits. Small spherulite-like particles were observed in Crystal Screen™ condition 17 (200 mM lithium sulphate, 100 mM tris pH 8.5 and 30% PEG 4000) (Figure 4-18b). Furthermore, small rod shaped crystals grew from Crystal Screen™ condition 35 (100 mM HEPES pH 7.5, 800 mM sodium phosphate, 800 mM potassium phosphate) (Figure 4-18c). Using synchrotron radiation and under cryo-conditions these diffracted to about 5 Å (Figure 4-18d). Further optimisation is ongoing for both full-length ceSGT and ceSGT-TPR.
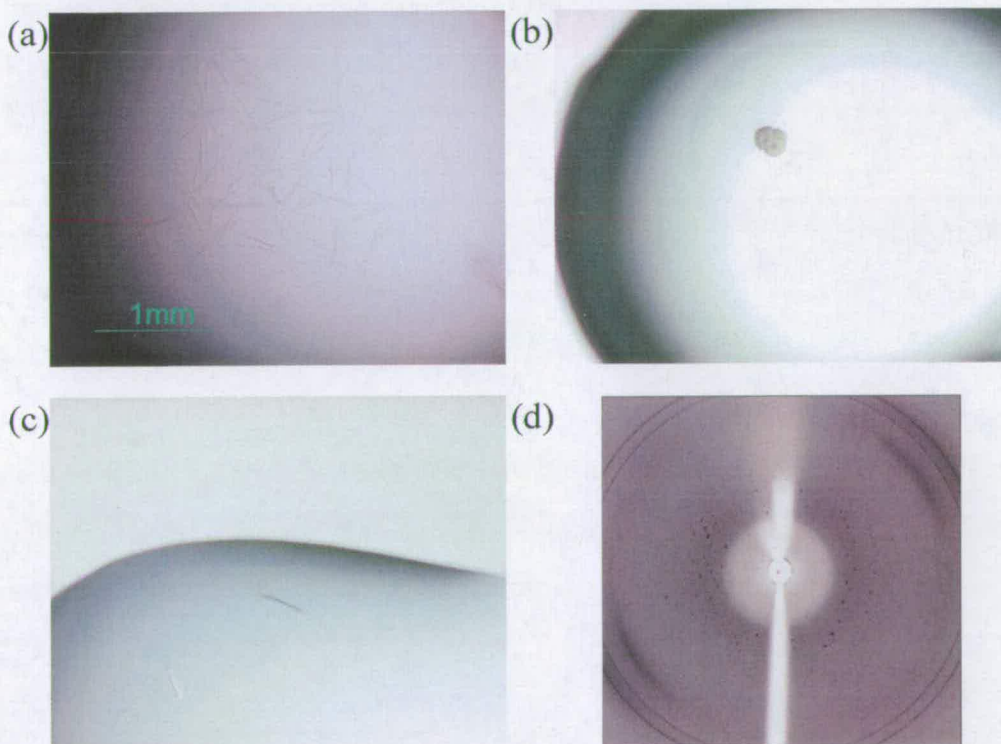
Figure 4-18 Crystallisation of ceSGT and ceSGT-TPR. (a) Fine needles of ceSGT; 200 mM KSCN, 5% PEG 3350 (b) SGT-TPR spherulite; 200 mM lithium sulphate, 100 mM tris pH 8.5, 30% PEG 4000 (c) Small SGT-TPR crystals; 100 mM HEPES pH 7.5, 800 mM sodium phosphate, 800 mM potassium phosphate. (d) Diffraction image of ceSGT-TPR crystals from (c). Innermost ice-ring is at 3.9 Å resolution, diffraction data extends to about 5 Å

4.4. Conclusions

The putative protein product of WormBase gene R05F9.10 was identified as a TPR domain containing co-chaperone likely to interact with *C. elegans* Hsp90. Comparative sequence analysis revealed the protein was the *C. elegans* homologue for small glutamine-rich tetratricopeptide repeat-containing protein. ceSGT has been successfully cloned, expressed and purified. Biochemical and biophysical characterisation has shown that SGT exists as a homo-dimer with a non-compact extended structure. The TPR domain of ceSGT was shown to bind the C-terminal peptides of Hsp90 and Hsp70 with a similar affinity of approximately 35 μM. Further work is required to discover if this result translates to the full-length proteins. Similarly, further optimisation of crystallisation conditions for both full-length ceSGT and ceSGT-TPR is required in order to obtain diffraction quality crystals.

As a final note, before this project *C. elegans* gene R05F9.10 referred only to a hypothetical protein. As a result of work carried out herein, the gene name *sgt-1* has been assigned to the WormBase entry.

4.5. References

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

Angeletti, P. C., Walker, D., and Panganiban, A. T. (2002). Small glutamine-rich protein/viral protein U-binding protein is a novel cochaperone that affects heat shock protein 70 activity. Cell Stress Chaperones *7*, 258-268.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res *34*, W604-608.

Bai, L., Swayne, L. A., and Braun, J. E. (2007). The CSPalpha/G protein complex in PC12 cells. Biochem Biophys Res Commun *352*, 123-129.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.

Brinker, A., Scheufler, C., Von Der Mulbe, F., Fleckenstein, B., Herrmann, C., Jung, G., Moarefi, I., and Hartl, F. U. (2002). Ligand discrimination by TPR domains. Relevance and selectivity of EEVD-recognition in Hsp70 x Hop x Hsp90 complexes. J Biol Chem *277*, 19265-19275.

Callahan, M. A., Handley, M. A., Lee, Y. H., Talbot, K. J., Harper, J. W., and Panganiban, A. T. (1998). Functional interaction of human immunodeficiency virus type 1 Vpu and Gag with a novel member of the tetratricopeptide repeat protein family. J Virol *72*, 8461.

Carrello, A., Allan, R. K., Morgan, S. L., Owen, B. A., Mok, D., Ward, B. K., Minchin, R. F., Toft, D. O., and Ratajczak, T. (2004). Interaction of the Hsp90 cochaperone cyclophilin 40 with Hsc70. Cell Stress Chaperones *9*, 167-181.

Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005). The Amber biomolecular simulation programs. J Comput Chem *26*, 1668-1688.

Cliff, M. J., Harris, R., Barford, D., Ladbury, J. E., and Williams, M. A. (2006). Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. Structure *14*, 415-426.

Cliff, M. J., Williams, M. A., Brooke-Smith, J., Barford, D., and Ladbury, J. E. (2005). Molecular recognition via coupled folding and binding in a TPR domain. J Mol Biol *346*, 717-732.

Cortajarena, A. L., and Regan, L. (2006). Ligand binding by TPR domains. Protein Sci *15*, 1193-1198.

Cziepluch, C., Kordes, E., Poirey, R., Grewenig, A., Rommelaere, J., and Jauniaux, J. C. (1998). Identification of a novel cellular TPR-containing protein, SGT, that interacts with the nonstructural protein NS1 of parvovirus H-1. J Virol *72*, 4149-4156.

Cziepluch, C., Lampel, S., Grewenig, A., Grund, C., Lichter, P., and Rommelaere, J. (2000). H-1 parvovirus-associated replication bodies: a distinct virus-induced nuclear structure. J Virol *74*, 4807-4815.

D'Andrea, L. D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem Sci *28*, 655-662.

Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. Embo J *17*, 1192-1199.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res *32*, W665-667.

Fonte, V., Kapulkin, V., Taft, A., Fluet, A., Friedman, D., and Link, C. D. (2002). Interaction of intracellular beta amyloid peptide with chaperone proteins. Proc Natl Acad Sci U S A *99*, 9439-9444.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163-164.

Granzin, J., Eckhoff, A., and Weiergraber, O. H. (2006). Crystal structure of a multi-domain immunophilin from Arabidopsis thaliana: a paradigm for regulation of plant ABC transporters. J Mol Biol *364*, 799-809.

Handley, M. A., Paddock, S., Dall, A., and Panganiban, A. T. (2001). Association of Vpu-binding protein with microtubules and Vpu-dependent redistribution of HIV-1 Gag protein. Virology *291*, 198-207.

Laurent, T. C., and Killander, J. (1964). A theory of gel filtration and its experimental verification. Journal of Chromatography A *14*, 317-330.

Liou, S. T., and Wang, C. (2005). Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. Arch Biochem Biophys *435*, 253-263.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Natochin, M., Campbell, T. N., Barren, B., Miller, L. C., Hameed, S., Artemyev, N. O., and Braun, J. E. (2005). Characterization of the G alpha(s) regulator cysteine string protein. J Biol Chem *280*, 30236-30241.

Owens-Grillo, J. K., Czar, M. J., Hutchison, K. A., Hoffmann, K., Perdew, G. H., and Pratt, W. B. (1996). A model of protein targeting mediated by immunophilins and other proteins that bind to hsp90 via tetratricopeptide repeat domains. J Biol Chem *271*, 13468-13475.

Owens-Grillo, J. K., Hoffmann, K., Hutchison, K. A., Yem, A. W., Deibel, M. R., Jr., Handschumacher, R. E., and Pratt, W. B. (1995). The cyclosporin A-binding immunophilin CyP-40 and the FK506-binding immunophilin hsp56 bind to a common site on hsp90 and exist in independent cytosolic heterocomplexes with the untransformed glucocorticoid receptor. J Biol Chem *270*, 20479-20484.

Ratajczak, T., and Carrello, A. (1996). Cyclophilin 40 (CyP-40), mapping of its hsp90 binding domain and evidence that FKBP52 competes with CyP-40 for hsp90 binding. J Biol Chem *271*, 2961-2965.

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol *266*, 525-539.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199-210.

Swayne, L. A., Beck, K. E., and Braun, J. E. (2006). The cysteine string protein multimeric complex. Biochem Biophys Res Commun *348*, 83-91.

Taylor, P., Dornan, J., Carrello, A., Minchin, R. F., Ratajczak, T., and Walkinshaw, M. D. (2001). Two structures of cyclophilin 40: folding and fidelity in the TPR domains. Structure *9*, 431-438.

Tobaben, S., Thakur, P., Fernandez-Chacon, R., Sudhof, T. C., Rettig, J., and Stahl, B. (2001). A trimeric protein complex functions as a synaptic chaperone machine. Neuron *31*, 987-999.

Tobaben, S., Varoqueaux, F., Brose, N., Stahl, B., and Meyer, G. (2003). A brain-specific isoform of small glutamine-rich tetratricopeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. J Biol Chem *278*, 38376-38383.

Wang, H., Shen, H., Wang, Y., Li, Z., Yin, H., Zong, H., Jiang, J., and Gu, J. (2005). Overexpression of small glutamine-rich TPR-containing protein promotes apoptosis in 7721 cells. FEBS Lett *579*, 1279-1284.

Whitmore, L., and Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res *32*, W668-673.

Winnefeld, M., Grewenig, A., Schnolzer, M., Spring, H., Knoch, T. A., Gan, E. C., Rommelaere, J., and Cziepluch, C. (2006). Human SGT interacts with Bag-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. Exp Cell Res *312*, 2500-2514.

Winnefeld, M., Rommelaere, J., and Cziepluch, C. (2004). The human small glutamine-rich TPR-containing protein is required for progress through cell division. Exp Cell Res *293*, 43-57.

Wu, B., Li, P., Liu, Y., Lou, Z., Ding, Y., Shu, C., Ye, S., Bartlam, M., Shen, B., and Rao, Z. (2004). 3D structure of human FK506-binding protein 52: implications for the assembly of

the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. Proc Natl Acad Sci U S A *101*, 8348-8353.

Wu, Y., and Sha, B. (2006). Crystal structure of yeast mitochondrial outer membrane translocon member Tom70p. Nat Struct Mol Biol *13*, 589-593.

Yin, H., Wang, H., Zong, H., Chen, X., Wang, Y., Yun, X., Wu, Y., Wang, J., and Gu, J. (2006). SGT, a Hsp90beta binding partner, is accumulated in the nucleus during cell apoptosis. Biochem Biophys Res Commun *343*, 1153-1158.

Young, J. C., Obermann, W. M., and Hartl, F. U. (1998). Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. J Biol Chem *273*, 18007-18010.

Zhang, M., Windheim, M., Roe, S. M., Peggie, M., Cohen, P., Prodromou, C., and Pearl, L. H. (2005). Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. Mol Cell *20*, 525-538.

5. Biochemical characterisation of *C. elegans* Hop

5.1.   Introduction

As introduced in chapter 4, two putative TPR domain containing proteins predicted to interact with the Hsp70/Hsp90 chaperone machinery were identified. The *C. elegans* orthologue for small glutamine-rich tetratricopeptide repeat-containing protein (SGT) was discussed in chapter 4. The second protein was the product of gene R09E12.3 and will be discussed in this chapter.

Gene R09E12.3 is a four exon gene located at the beginning of chromosome V (Figure 5-1). It encodes a 320 residue protein with a calculated molecular weight of 36.9 kDa. Comparative sequence analysis using BLAST highlights extensive sequence similarity to the C-terminal half of Hsp70/Hsp90 organising protein (Hop). Alignment with human Hop shows 56% sequence identity to the C-terminal 318 residues and the complete absence of the N-terminal 220 residues (Figure 5-2). Human Hop is composed of 9 TPR repeats segregated into three distinct domains named TPR1, TPR2A and TPR2B. In addition there are two regions with aspartate-proline (DP) repeats located between TPR1 and TPR2A, and at the C-terminal. The putative *C. elegans* Hop homologue lacks domain TPR1 and the first DP repeat region (Figure 5-1). The possibility of an incorrect gene prediction was investigated but no evidence of any upstream exons coding for the missing sequence was evident.
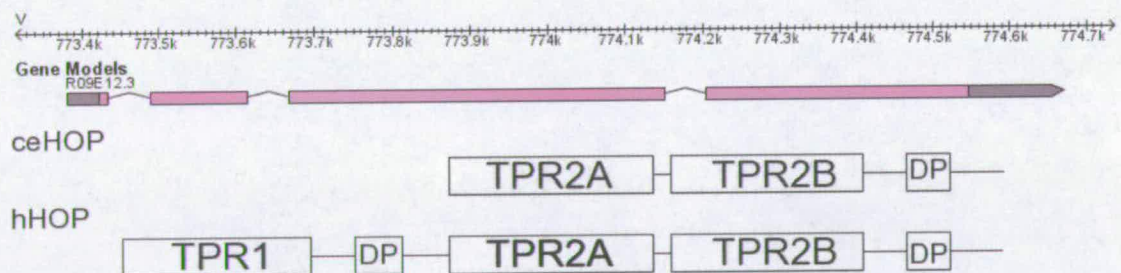
Figure 5-1 ceHop gene and protein architecture. WormBase gene R09E12.3 is a four exon gene located on the forward strand of *C. elegans* chromosome V. It encodes a 320 residue protein with a high degree of similarity to the C-terminal half of human Hop but lacking domains TPR1 and the first DP repeats.
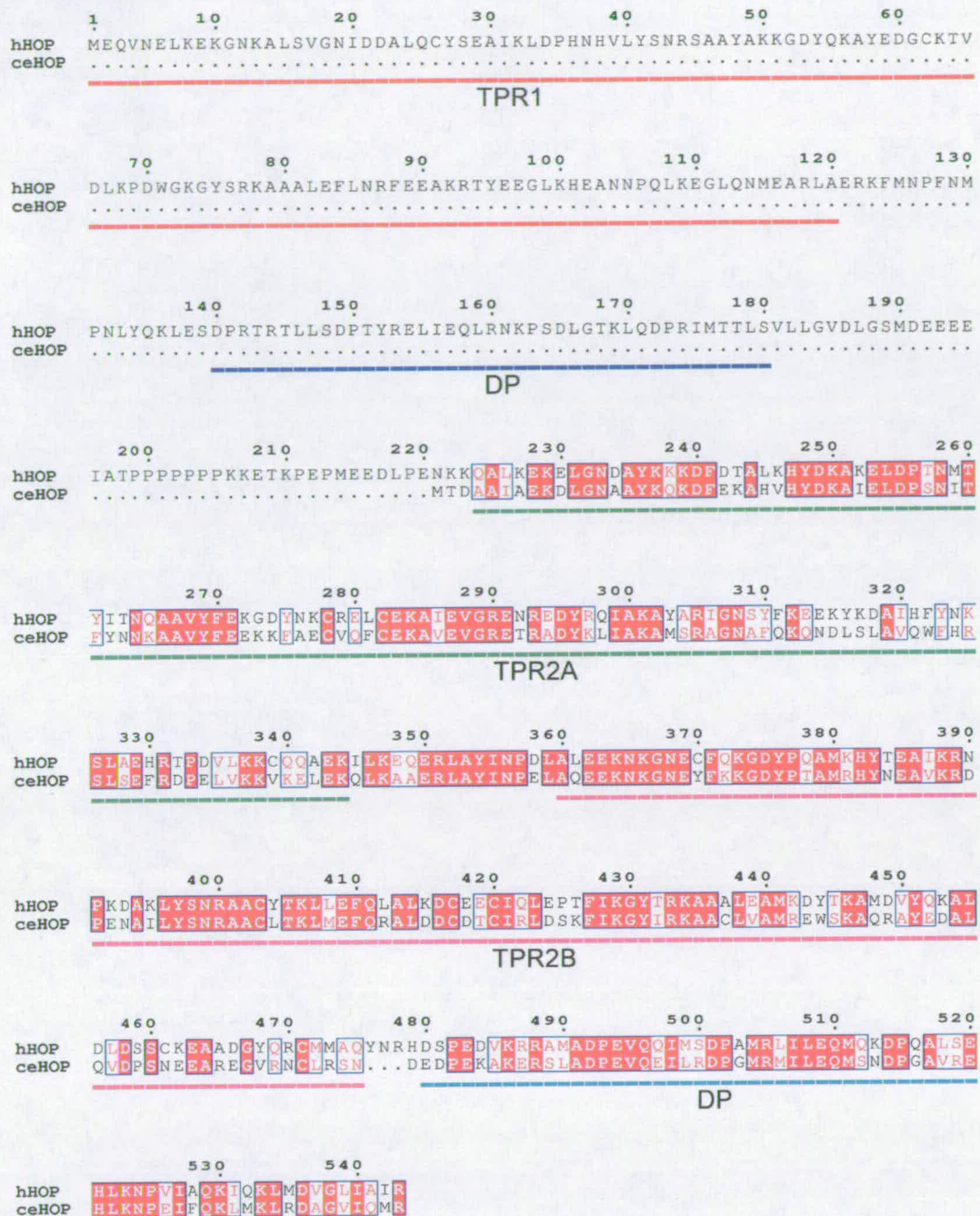
Figure 5-2 Alignment of Human and *C. elegans* Hop homologues. ceHop, which lacks the TPR1 domain and the N-terminal DP repeat, shares 56% sequence identity with the C-terminal 318 residues of human Hop.

Hop was first identified in yeast and named Sti1 for stress inducible protein 1 (also see section 1.4.2.1.) (Nicolet and Craig, 1989). Homologues have since been identified in a wide range of species including higher mammals, plants, insects and parasites (Odunuga et al., 2004). The predominant function of Hop appears to be as an adaptor protein linking the

Hsp70 and Hsp90 chaperone pathways (Odunuga et al., 2004). Hop is able to interact with both Hsp70 and Hsp90 via an interaction between the extreme C-terminal Hsp peptides and distinct Hop TPR domains; domain TPR1 is responsible for the Hsp70 interaction whilst domain TPR2A interacts with Hsp90 (Figure 5-3). Crystal structures of the isolated TPR1 and TPR2A domains in complex with Hsp70 and Hsp90 peptides respectively defined the two-carboxylate clamp mechanism (Scheufler et al., 2000). This showed the common Hsp70/90 C-terminal EEVD motif coordinated by a cluster of polar residues on the concave surface of the TPR domain with residues directly upstream of EEVD providing binding specificity and selectivity (see Figures 4-15 and 4-16). The binding partner for domain TPR2B remains unclear. The carboxylate-clamp motif required for Hsp70/90 binding is conserved in TPR2B but the domain in isolation was shown to bind Hsp70 or Hsp90 very poorly (Scheufler et al., 2000). Conversely, mutations in the TPR2B domain have been shown to impact interactions with both Hsp70 and Hsp90 suggesting an overlapping function with the other TPR domains (Carrigan et al., 2004; Chen et al., 1998).
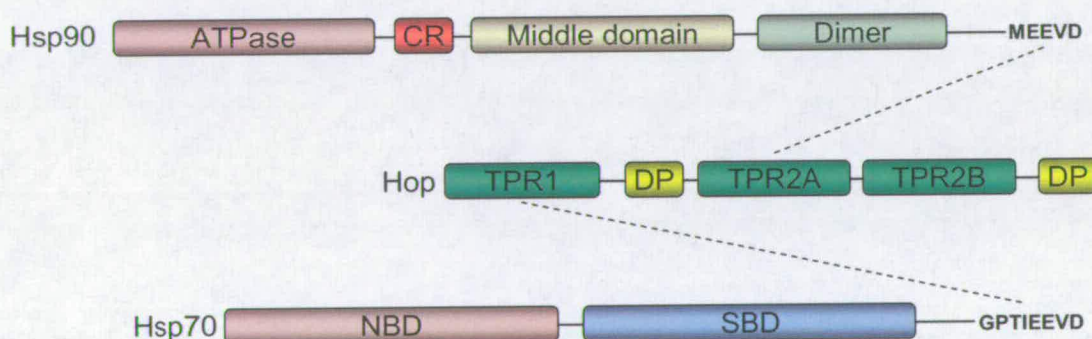
Figure 5-3 Hop can interact with both Hsp70 and Hsp90 via distinct TPR domains. TPR1 interacts with Hsp70 and TPR2A interacts with Hsp90. The ligand for TPR2B is unknown although it has been implicated in the interaction with Hsp70 and Hsp90.

Hsp90 serves as a chaperone for a sub-set of client proteins with the assembly of several Hsp90 based multi-protein complexes requiring communication with the Hsp70 chaperone machinery (see section 1.3.). In these cases, client proteins bind first to Hsp40 targeting them to Hsp70. Transfer to Hsp90 is then facilitated by the simultaneous interaction of the Hsp70-client complex and Hsp90 with Hop. Much of the current understanding of the Hop mediated interplay between the Hsp70/90 chaperone machinery has come from the study of the maturation of steroid-hormone receptors (see section 1.3).

The aims of this project were to initially clone, express and purify *C. elegans* Hop (ceHop). Hop has been proposed to exist as a dimer (Prodromou et al., 1999; van der Spuy et al., 2001) so the solution state of ceHop was to be investigated. Further, the expected interaction with Hsp90 via the TPR2A domain was to be investigated as was the ability of ceHop to interact with Hsp70 in view of the absent TPR1 domain. Finally, crystallisation experiments were carried out with the aim of obtaining diffraction quality crystals to allow the solving of the three-dimensional atomic structure.

5.2.　Materials and methods

5.2.1.　Cloning

cDNA corresponding to full-length ceHop (residues 1-320) were generated by PCR using *C. elegans* mixed stage N2 cDNA as a template. Sequences were amplified with the TaqPlus® precision PCR system (Stratagene) using the forward (GCGG *CATATG* ACG GAC GCC GCG ATT GCT G) and reverse primers (GGCG *GCGGCCGC* TTA GCG CAT CTG AAT GAC TCC). The resulting PCR products were cloned into a pCR®2.1 TOPO vector (Invitrogen), verified by sequencing and digested with *NdeI* and *NotI* (New England Biolabs). The digested inserts were ligated into a similarly digested pET-30a vector (Novagen) and verified by DNA sequencing (Figure 5-4).
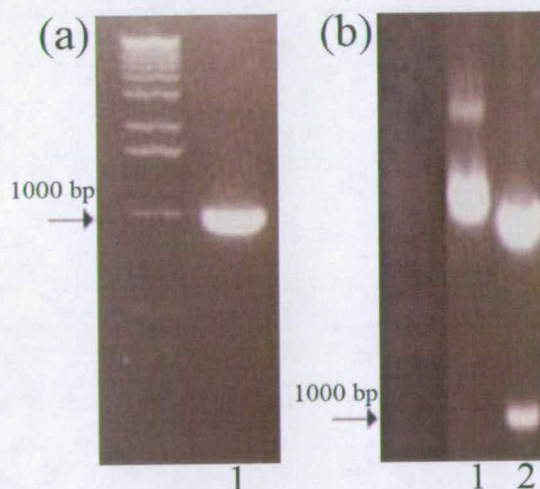
Figure 5-4 Cloning of ceHop. (a) PCR of ceHop from mixed stage N2 *C. elegans* cDNA. (b) Double-digest of recombinant pET-30a plasmid showing presence of ceHop sequence. 1 - pET-30a-ceHop plasmid. 2 - Double-digest with NdeI and NotI, ~1000 bp band shows ceHop sequence.

5.2.2. Expression and purification

ceHop was expressed in BL21(DE3)-Rosetta 2 *E. coli* (Novagen) in LB liquid media containing kanamycin (25 μg/ml) and chloramphenicol (30 μg/ml). Cultures were grown with shaking at 37°C until the A_{600} was ~ 0.6, over-expression induced by addition of IPTG to 1 mM and growth continued for a further 4 hours at 37°C. Cells were harvested by centrifugation (3000 xg for 15 min), resuspended at 10% weight per volume in ice-cold lysis buffer (50mM tris pH 7.5, 5mM EDTA, 1mM DTT, 0.1mM benzamidine, 0.1mM PMSF) plus excess protease inhibitor cocktail (Roche), and sonicated on ice for 6 x 30 second bursts, with 30 seconds cooling in between. The cell lysate was subjected to centrifugation at 30,000 xg for 1 hr at 4°C.

Figure 5-5 Two-step Purification of ceHop. (a) Anion-exchange chromatogram. ceHop was eluted with a KCl gradient (solid line). ceHop eluted around 50-100 mM KCl. (b) Gel-filtration chromatogram. ceHop was >90% pure based on SDS-PAGE analysis.

Untagged ceHop was purified by a two-step strategy consisting of anion exchange and gel filtration. The calculated isoelectric point (http://www.embl-heidelberg.de/cgi/pi-wrapper.pl) was 6.58 and a tris pH 8.7 buffer was selected for anion exchange. Clarified cell lysate was dialysed overnight against buffer A (50mM tris, pH 8.7, 1mM EDTA, 1mM DTT, 0.1mM PMSF, 1mM azide), filtered through a 0.2 μm filter and applied to a Source-Q 30 μm (Pharmacia) column (V_t ~ 10 ml; 2 x 5 cm) pre-equilibrated in buffer C. ceHop was eluted with a 0-500 mM KCl gradient in buffer A over 100 mls and analysed by SDS-PAGE (Figure 5-5). Fractions containing ceHop, eluting between 50 and 100 mM KCl, were pooled and concentrated. Protein was then applied to Superdex 200 HR 30/10 column (Amersham Bioscience) pre-equilibrated in buffer B (25mM HEPES pH 7.5, 100mM NaCl and 1mM DTT) and analysed by SDS-PAGE. Fractions containing ceHop were pooled and stored on ice at 4°C in buffer B. ceHop was > 90% pure as judged by SDS-PAGE (Figure 5-5).

Human Hsp90α-631 and *C. elegans* DAF21-492 used in the interaction studies were kindly provided by Dr. Amir Rabu (Rabu, 2006).

5.2.3. Protein Cross-linking

5 μg total protein in 15 μl buffer B was cross-linked with addition of a 1/10th volume of 10 × glutaraldehyde stock made up in buffer B, 0.1% and 0.2% final glutaraldehyde concentrations were used. The reaction was quenched at various time points by addition of a 1/10th volume of 1M tris pH 7.5 and subjected to SDS-PAGE analysis.

5.2.4. Analytical Gel Filtration

Gel filtration studies were carried out on an AKTA explorer FPLC using either a Superdex 200 HR 30/10 column at 4°C. The column was equilibrated with buffer B and calibrated as before (see section 4.2.5). 200 μl protein was applied to the column and run at 0.5 ml min^{-1}. For interaction studies, proteins were incubated for 1 hour at room temperature prior to application. To determine binding affinities, the plot of retention volume of the complex against the molar concentration was analysed by non-linear regression analysis using the software package SigmaPlot 9.0. Data were fit to the equation

$$F = F_{free} + (F_{saturated} - F_{free})((K_d + [A] + [B]) - \sqrt{((K_d + [A] + [B]^2 - 4[A][B]))/2[A]}$$

where F = retention volume of DAF21-492:ceHop complex, K_d = dissociation constant, [A] = ceHop concentration and [B] = DAF21-492 concentration.

5.2.5. Protein crystallisation

Crystallisation trials were conducted using the hanging drop vapour diffusion method from a 10 mg ml^{-1} protein solution in buffer B at 4 and 18 °C. Hampton Structure Screen^{TM} and Structure Screen II were used for screening with a 2 μl drop consisting of a 1:1 ratio of protein and well solution. Trials were carried out in the absence and presence of the Hsp90/Hsp70 peptides MEEVD and GPTIEEVD. Peptides were reconstituted in distilled H$_2$O and mixed with protein at a ratio of 1:1.3. Optimisation of initial hits was conducted using a grid screen around promising conditions.

Crystallisation trials and optimisation failed to produce any crystals.

5.3. Results and discussion

ceHop was successfully expressed and purified to greater than 90% purity based on estimation of relative band densities from SDS-PAGE analysis (Figure 5-5).

5.3.1. ceHop appears to exist as a dimer

Both yeast (Prodromou et al., 1999) and mouse (van der Spuy et al., 2001) Hop homologues
have been reported to dimerise. Using gel-filtration, ceHop elutes as a single species with a
retention volume of 14.2 ml consistent with an apparent molecular weight of ~70 kDa and a
Stokes radius of 34.5 Å (Figure 5-6). With a monomeric molecular weight of 36.9 kDa this
suggests ceHop exists as a dimer or in a non-globular conformation.
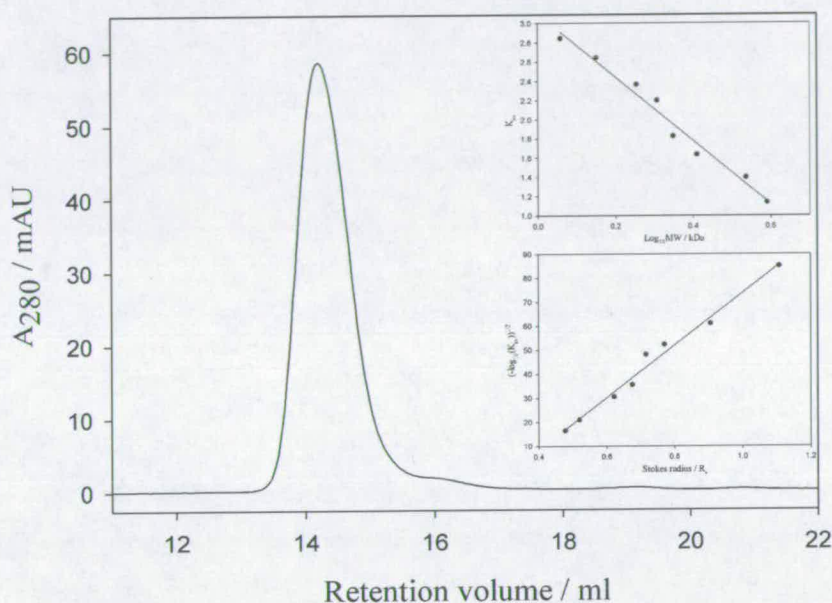
Figure 5-6 Gel-filtration analysis of ceHop. ceHop was resolved on a Superdex 200 HR 30/10
column. ceHop eluted as a single peak with a retention volume of 14.2 ml. This corresponds to a
molecular weight of ~70 kDa and a Stokes radius of about 34.5 Å. This is consistent with a dimer
(actual MW = 36.9 kDa) or a non-globular conformation.

The oligomerisation state of ceHop was further investigated with glutaraldehyde cross-
linking. Cross-linking results in the formation of a more compact monomeric species and the
accumulation of various higher-order oligomeric species including a dimeric form (Figure 5-
7). The formation of a lower molecular weight monomeric form in the presence of
glutaraldehyde is consistent with intra-molecular cross-linking locking the protein in a more
compact conformation which migrates faster in SDS-PAGE analysis. This result is in
agreement with cross-linking studies on mouse Hop and suggests a globular tertiary structure
(van der Spuy et al., 2001). The presence of higher-order oligomeric species alludes to dimer
formation and further oligomerisation but the ladder like distribution of oligomers perhaps
highlights the promiscuous nature of glutaraldehyde as a cross-linker.
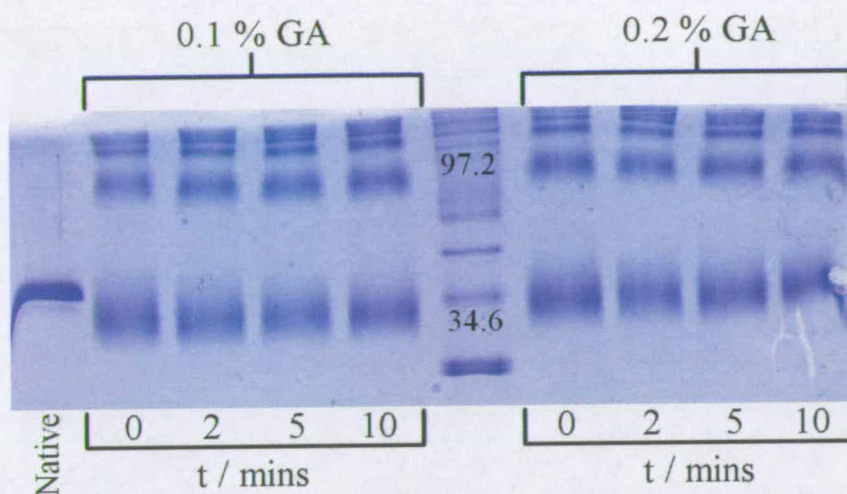
Figure 5-7 Glutaraldehyde cross-linking of ceHop. ceHop was incubated in the presence of 0.1% or 0.2% glutaraldehyde for 0, 2, 5 or 10 minutes and then analysed by SDS-PAGE. Cross-linking results in a more compact monomeric form and also the presence of higher molecular weight oligomers.

The structural and functional significance of Hop dimerisation remain unclear. Hsp90 has been shown to exist as a functional dimer and yeast Hop was shown to interact with Hsp90 in a 1:1 stoichiometry suggesting the interaction of dimeric forms (Prodromou et al., 1999). A recent study has shown that the TPR2A domain was necessary and sufficient for dimerisation of the yeast homologue (Flom et al., 2007), which is consistent with the self-association properties of the *C. elegans* protein.

5.3.2. ceHop interacts with both human and *C. elegans* Hsp90 homologues

The interaction of ceHop and human Hsp90α was investigated using semi-analytical gel filtration. C-terminal human Hsp90α construct hHsp90α-631 was used. This small 12 kDa C-terminal Hsp90 construct consists of the minimal helix-loop-helix Hsp90 dimerisation interface and the flexible C-terminal tail containing the MEEVD TPR binding motif.

50 µM ceHop and 100 µM Hsp90α-631 were resolved on a Superdex 200 HR 30/10 column with retention volumes of 14.2 and 13.94 ml respectively. 50 µM ceHop and 100 µM hHsp90α-631 were then incubated for 1 hour at room temperature prior to application to the column (Figure 5-8). The incubated protein eluted as a major peak with a retention volume of 13.05 ml, a shift of over 1 ml compared to ceHop alone. The peak contained a small shoulder on the trailing edge consistent with some free hHsp90α-631 and ceHop. The

significant shift in peak elution position when both proteins were incubated suggested complex formation.
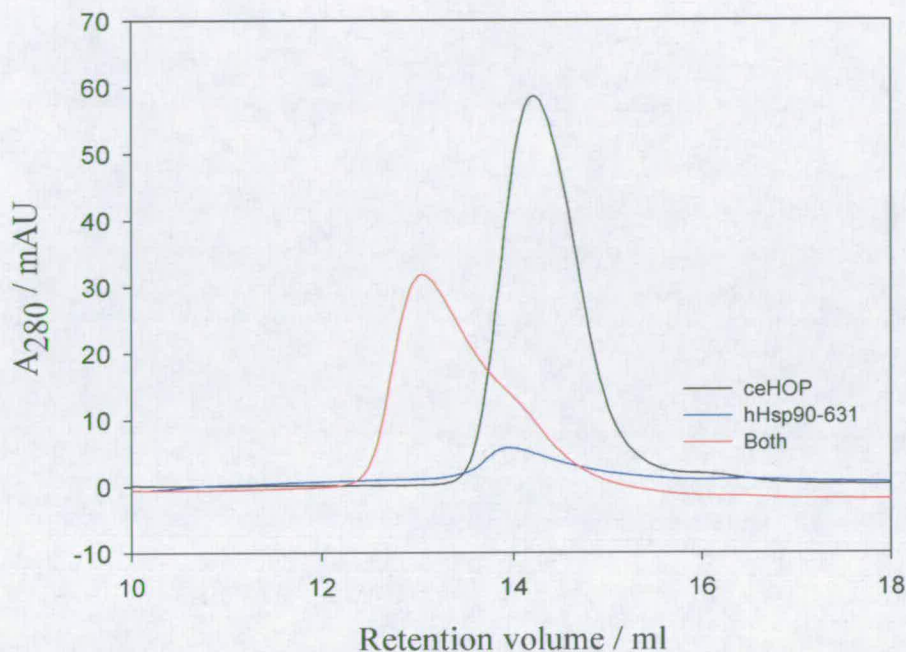
Figure 5-8 Gel-filtration analysis of the ceHop - hHsp90α-631 interaction. 50 μM ceHop and 100μM hHsp90α-631 were resolved separately and after a 1 hour incubation on a Superdex 200 HR 30/10 column. A clear shift in retention volume is apparent after incubation (red line) indicating an interaction.

The interaction of Hop and Hsp90 is primarily mediated by the extreme C-terminal pentapeptide MEEVD, a common feature to both *C. elegans* and human Hsp90 homologues. In addition to demonstrating an interaction between *C. elegans* Hop and human Hsp90, the interaction of ceHop with the *C. elegans* Hsp90 homologue Daf21 was also investigated. In this case a larger 24 kDa C-terminal construct corresponding to the complete C-terminal domain was selected (Daf21-492). The interaction was again analysed with gel-filtration; however, binding affinity was also explored. 10 μM ceHop was incubated with increasing concentrations of Daf21-492. Complex formation was suggested by the elution of a higher molecular weight peak (Figure 5-9a; retention volume 12.02 ml) and was verified by SDS-PAGE analysis of the eluate clearly demonstrating a significant shift in elution position and co-migration of both proteins (Figure 5-9b).

Incubation of ceHop with increasing concentrations of Daf21-492 revealed two interesting features. Firstly, instead of two high molecular weight peaks corresponding to the complex and free DAF21-492, only one peak was observed and secondly, the elution position of this peak changed with increasing concentrations of Daf21-492 reflecting an increased size of the migrating particle (Figure 5-9a). Elution of a single peak was interpreted to reflect rapid association/dissociation kinetics of complex formation, faster than the temporal resolution of the column, and the elution position of the peak was interpreted to reflect the equilibrium between bound and free ceHop and Daf21-492.

As peak position was predicted to reflect the binding equilibrium, a plot of retention volume verses Daf21 concentration was used to calculated the equilibrium dissociation constant (K_d). Fitting of a tight binding equation assuming 1:1 stoichiometry using non-linear regression resulted in a calculated K_d of 0.8 µM ($r^2 = 0.99$). This is in agreement with other studies which calculate the affinity to be in the range 0.09-0.33 µM for the protein-protein interaction (Hernandez et al., 2002; Prodromou et al., 1999; Siligardi et al., 2004) and 6-11 µM for the interaction of the isolated TPR2A domain and the C-terminal Hsp90 domain or the C-terminal MEEVD peptide (Scheufler et al., 2000).

Finally, the elution position of free ceHop was concomitantly shown to increase with increasing concentrations of Daf21. There are two possible explanations for this. The first is that an increased proportion of ceHop complexed with Daf21 would reduce the apparent free concentration of ceHop, as is reflected by the decreasing peak size for free ceHop. If ceHop existed in a monomer-dimer equilibrium with a dissociation constant in the low micromolar region then this decrease in ceHop concentration could shift in the equilibrium toward the monomeric form with a subsequent shift in retention volume. Additionally, it is possible that the transient nature of the interaction with Daf21 may influence the elution position of the free ceHop.
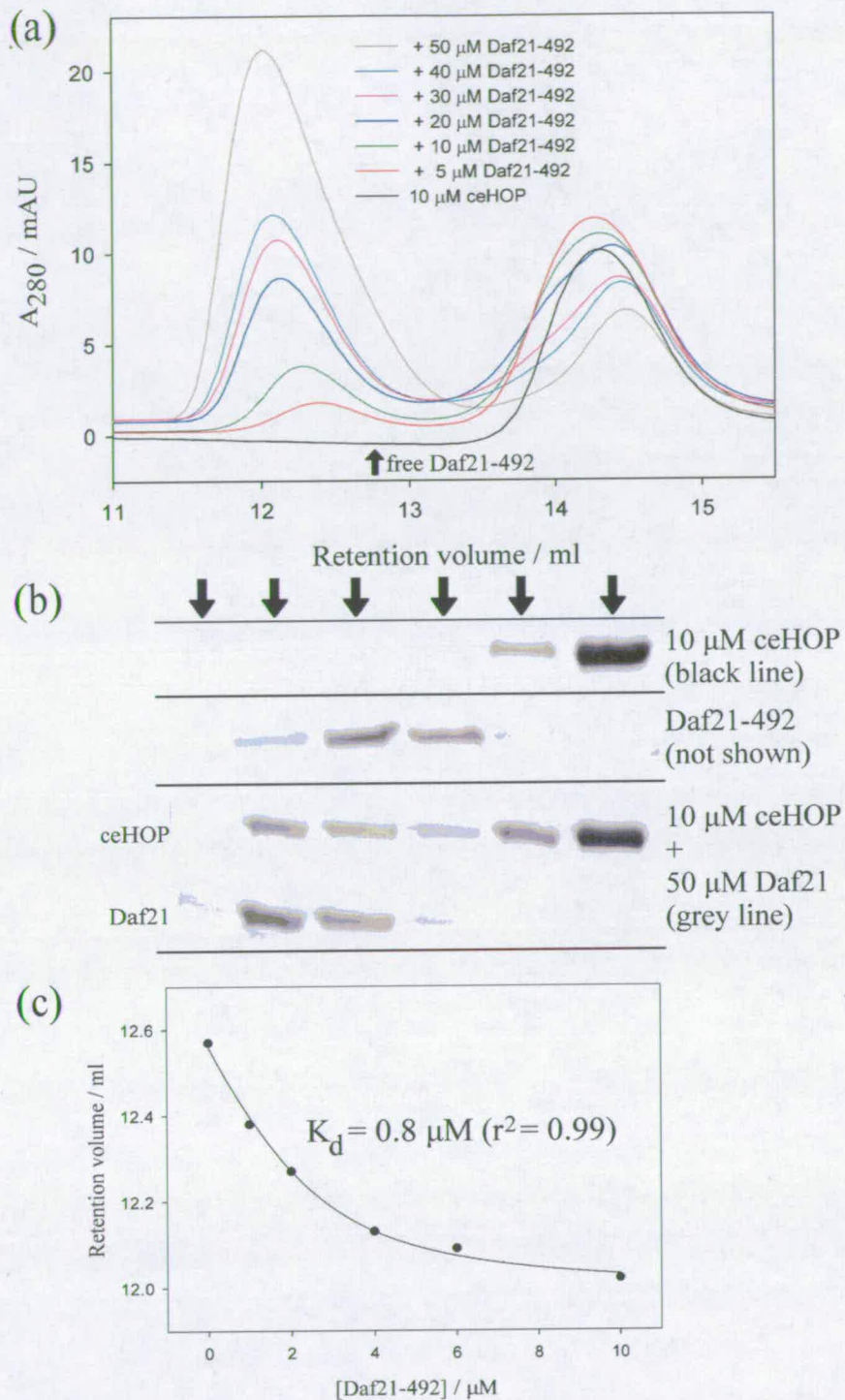
Figure 5-9 Gel-filtration analysis of the interaction between ceHop and a C-terminal construct of the *C. elegans* Hsp90 homologue Daf21. (a) 10 µM ceHop was resolved alone and in the presence of increasing concentrations of Daf21-492. The elution position of free Daf21-492 is indicated with an arrow. (b) Eluate (fraction indicated by arrows) from ceHop alone (black line in (a)), Daf21-492 alone and ceHop + Daf21-492 (grey line in (a)) analysed by SDS-PAGE. A clear shift in elution profiles of both proteins is evident. (c) A plot of retention volume verses molar concentration of Daf21-492 was analysed by non-linear regression using a tight binding equation giving a calculated K_d for the interaction of 0.8 µM.

5.3.3. Analysis of the putative interaction with Hsp70

The interaction of ceHop with Hsp70 was not experimentally investigated. The absence of the first TPR domain, TPR1 which has been isolated as the major interaction site for Hsp70 (Scheufler et al., 2000), and the first DP repeat might suggest a lack of Hsp70 interaction. However, several studies have demonstrated that domain TPR2B plays a significant role in the interaction of Hop and Hsp70 (Carrigan et al., 2004; Flom et al., 2007). Indeed, comparison of the primary structure of human Hop reveals a striking similarity between the TPR1-DP1 and TPR2B-DP2 regions (Figure 5-10) suggesting one of the regions has arisen by duplication. Analysis of the Ensembl (www.ensembl.org) Hop/Sti1 protein family, which contains 144 sequences, reveals seven sequences lacking the TPR1 and DP1 domains including homologues from *Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Anopheles gambiae*, *Aedes aegypti*, *Schistosoma japonicum*, *Candida albicans*, *Cryptosporidium hominis*. The contribution of TPR2B to the Hsp70 interaction and lack of TPR1 in some species suggests an overlapping redundant function. In support of this, a recent study has shown that a yeast recombinant Sti1 lacking the TPR1 and DP1 domains can rescue a lethal double knockout lacking wild-type Sti1 and HdJ1, an Hsp40-like Hsp70 co-chaperone (Flom et al., 2007). Furthermore, in the same study it was demonstrated that both TPR1 and TPR2B contributed to the Hsp70 interaction with mutations in both required to abrogate binding. *C. elegans* like mutants were, however, found to be dysfunctional in some specialised Hop functions such as activation of the glucocorticoid receptor. Overall, these results suggest that Hop homologues lacking the first TPR and DP domains are likely able to support some of the vital functions of the full-length protein and that the *C. elegans* protein is likely to interact with Hsp70, although experimental validation is required.

5.4. Conclusions and future work

In conclusion, the *C. elegans* Hop homologue has been successfully cloned, expressed and purified. Biochemical analysis suggests ceHop is capable of dimerising. This is in agreement with studies on yeast and mouse Hop and indicates that the region responsible for self-association maps to the C-terminal portion of the protein encompassing TPR2A, TPR2B and the C-terminal DP region. The observation of the increasing retention volume in response to decreasing free ceHop concentrations when looking at the interaction between ceHop and the *C. elegans* Hsp90 homologue DAF-21 provides the most convincing evidence of dimer formation and suggests a dissociation constant in the low micromolar range. ceHop was shown to interact with Hsp90 homologues from both human and *C. elegans* and in the case of DAF21 with a tight sub-micromolar binding affinity. Further work must be carried out to

explore the interaction with Hsp70 although recent evidence suggests an overlapping function of the TPR1 and TPR2B domains of yeast Hop indicating that *C. elegans* protein should be able to interact with Hsp70 in addition to Hsp90.
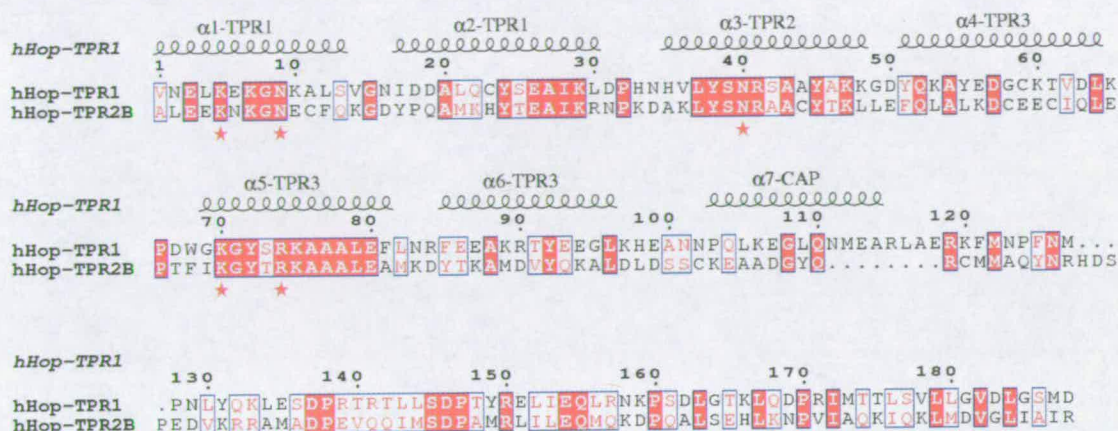
Figure 5-10 Alignment of the TPR1-DP1 and TPR2B-DP2 regions of human Hop. The regions share extensive similarity (~30% identity, ~50% similarity) suggesting TPR1-DP1 has arisen from a duplication event. Secondary structure from Hop TPR1 crystal structure indicated and carboxylate-clamp residues highlighted with red stars.

5.5.   References

Carrigan, P. E., Nelson, G. M., Roberts, P. J., Stoffer, J., Riggs, D. L., and Smith, D. F. (2004). Multiple domains of the co-chaperone Hop are important for Hsp70 binding. J Biol Chem *279*, 16185-16193.

Chen, S., Sullivan, W. P., Toft, D. O., and Smith, D. F. (1998). Differential interactions of p23 and the TPR-containing proteins Hop, Cyp40, FKBP52 and FKBP51 with Hsp90 mutants. Cell Stress Chaperones *3*, 118-129.

Flom, G., Behal, R. H., Rosen, L., Cole, D. G., and Johnson, J. L. (2007). Definition of the minimal fragments of Sti1 required for dimerization, interaction with Hsp70 and Hsp90 and in vivo functions. Biochem J.

Hernandez, M. P., Sullivan, W. P., and Toft, D. O. (2002). The assembly and intermolecular properties of the hsp70-Hop-hsp90 molecular chaperone complex. J Biol Chem *277*, 38294-38304.

Nicolet, C. M., and Craig, E. A. (1989). Isolation and characterization of STI1, a stress-inducible gene from Saccharomyces cerevisiae. Mol Cell Biol *9*, 3638-3646.

Odunuga, O. O., Longshaw, V. M., and Blatch, G. L. (2004). Hop: more than an Hsp70/Hsp90 adaptor protein. Bioessays *26*, 1058-1068.

Prodromou, C., Siligardi, G., O'Brien, R., Woolfson, D. N., Regan, L., Panaretou, B., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1999). Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. Embo J *18*, 754-762.

Rabu, A. (2006) PhD thesis - Biochemical and Biophysical Characterisation of Heat Shock Protein 90 and Its Domain Interactions, University of Edinburgh, Edinburgh.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199-210.

Siligardi, G., Hu, B., Panaretou, B., Piper, P. W., Pearl, L. H., and Prodromou, C. (2004). Co-chaperone regulation of conformational switching in the Hsp90 ATPase cycle. J Biol Chem *279*, 51989-51998.

van der Spuy, J., Cheetham, M. E., Dirr, H. W., and Blatch, G. L. (2001). The cochaperone murine stress-inducible protein 1: overexpression, purification, and characterization. Protein Expr Purif *21*, 462-469.

6. Prediction of the complete repertoire of *C. elegans* TPR co-chaperones

6.1. Introduction

ceSGT and ceHop, TPR domain containing co-chaperones discussed in chapters 4 and 5 respectively, were identified in a previous search of the *C. elegans* genome as putative Hsp90 interacting proteins (Opamawutthikul, 2005). The previous approach utilised the SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de) database; a web-based tool for the identification and annotation of protein domains. A text query of the SMART database for *C. elegans* TPR domains highlighted 41 proteins amongst which ceSGT and ceHop were identified based on conservation of the carboxylate-clamp motif; the key residues involved in the interaction with the C-terminal of Hsp70/Hsp90. However, interrogation of WormBase for proteins belonging to InterPro motif "Tetratricopeptide-like helical" (IPR011990) identifies 114 candidate TPR domain containing proteins. To investigate the existence of further unidentified *C. elegans* TPR domain-containing co-chaperones, a thorough analysis of the complete genome was repeated. Analysis of proteins using secondary databases, such as SMART, PROSITE and Pfam, relies on their presence in sequence databases and can vary from method to method. In an attempt to avoid these problems, a direct search of the published complete *C. elegans* proteome and genome was conducted using a profile hidden Markov model (HMM) specific to Hsp70/Hsp90 binding TPR domains.

An HMM is a probabilistic model suited for the analysis of time series or linear sequence models. HMMs have been most widely used in speech recognition and have since been found to be applicable for use in the computational analysis of biological sequences (Eddy, 1998). Introduced by Anders Krogh and colleagues in 1994 (Krogh et al., 1994), profile HMMs are statistical models of sequence alignments. They capture position-specific information regarding how conserved each column of an alignment is and provide a way of representing the consensus sequence of proteins belonging to the same family. Profile HMMs can be thought of as graphical models of protein families that emit a sequence of letters corresponding to amino acids. Each position in the model represents a column in the multiple alignment and has a match state (m), insert state (i) and delete state (d) (Figure 6-1). m and i states emit letters whilst d states are silent and represent gaps in the alignment. State emissions and transitions have associated probabilities, calculated from an input alignment, thus generating a full probabilistic profile of a protein family.
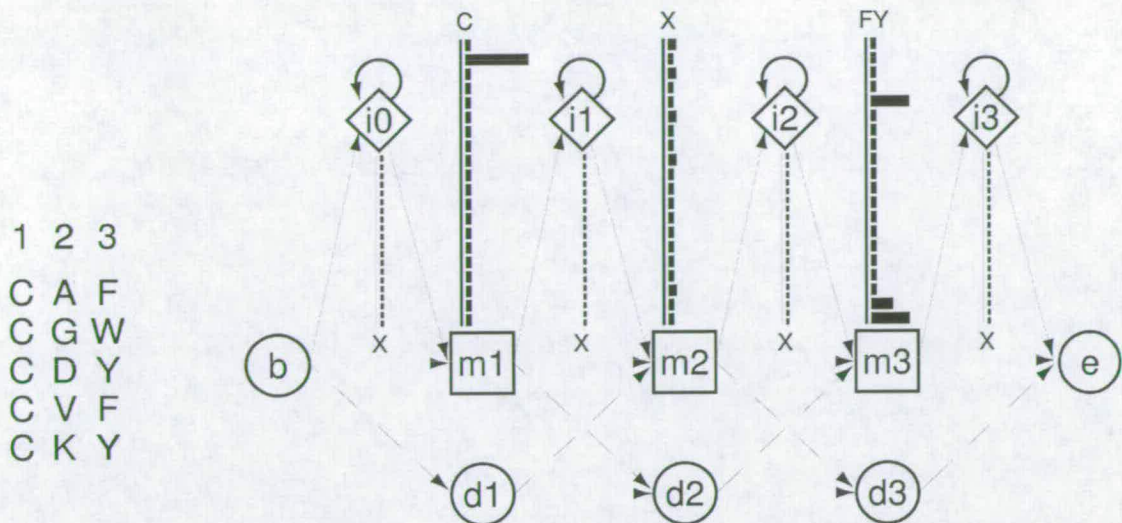
Figure 6-1 Model architecture of a simple profile HMM. Graphical representation of a profile HMM built from a simple multiple sequence alignment (left). States are beginning (b), match (m), insert (i), deletion (d) and end (e). Each match state represents a column in the sequence alignment with an associated emission probability for each residue. Transitions from one match state to another and to insert/deletion states also have associated probabilities. This allows the posterior calculation of the probability of the sequence being generated by the HMM. Figure adapted from Eddy, 1998.

One strength of profile HMMs is in the automatic annotation of the domain structure of proteins and profile HMMs form the basis of several automatic domain annotation databases including Pfam and SMART. An extension to this is the use of profile HMMs to search for members of a particular protein family. Instead of screening a profile HMM database such as Pfam against a new protein sequence, a sequence database is screened against a single profile HMM representing a single family or domain.

The aim of this study was to use a profile HMM approach to detect all *C. elegans* TPR domain containing Hsp70/Hsp90 co-chaperones. Hsp70/90 interacting TPR domains share the same characteristic domain architecture with three TPR repeats with a C-terminal capping helix in the majority of cases (Figure 6-2). In addition to the conserved TPR consensus sequence, a common five residue carboxylate-clamp motif has been described which is conserved in all TPR domains shown to interact with Hsp70 or Hsp90 (Scheufler et al., 2000). A profile HMM generated from an alignment of these domains provides a powerful stochastic model describing the overall TPR domain structure with the added sensitivity to distinguish carboxylate-clamp containing proteins.
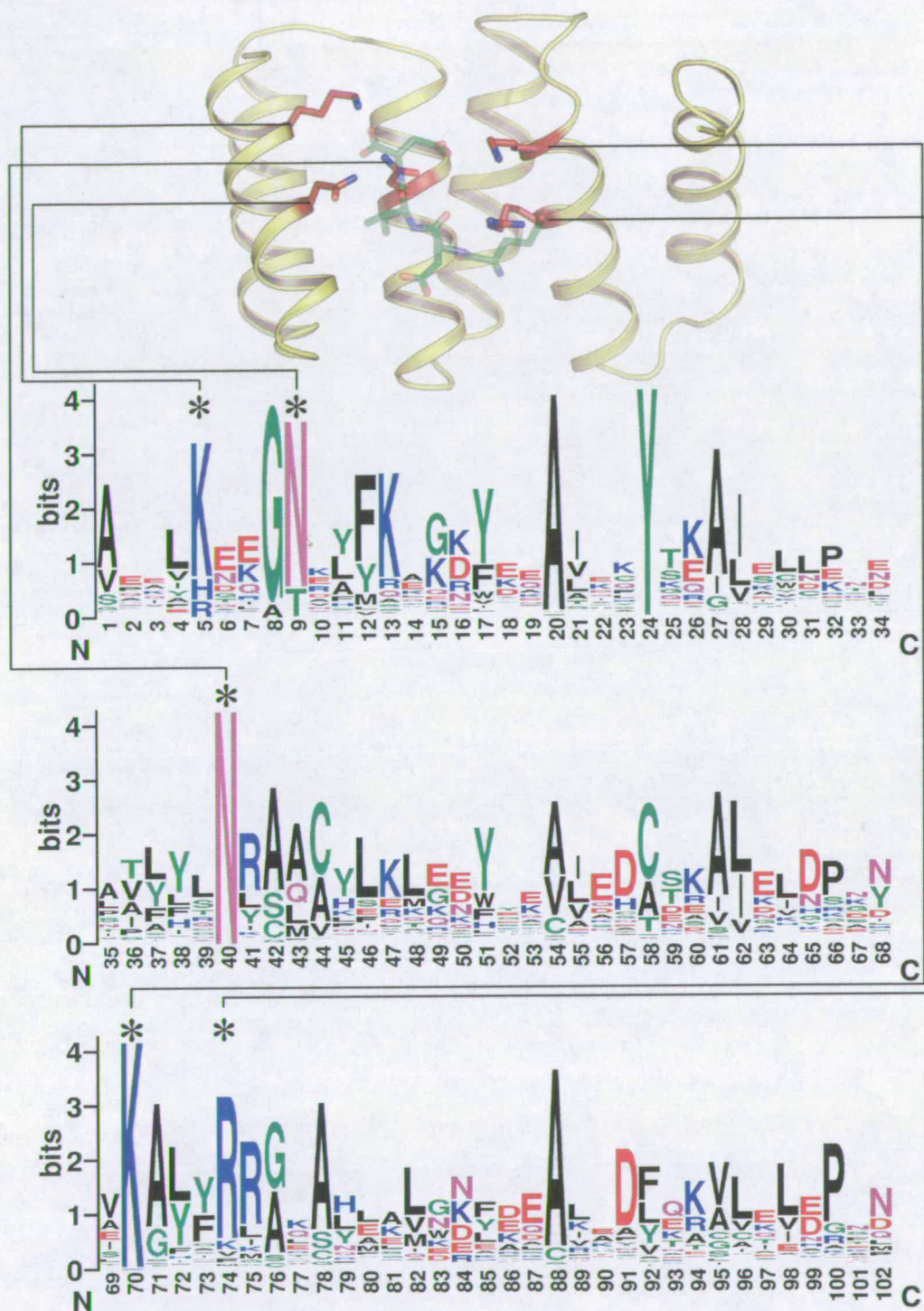
Figure 6-2 Consensus Hsp70/90 binding TPR domain sequence. Amino-acid frequency for all TPR domains shown to interact with Hsp70 or Hsp90. The five residue carboxylate-clamp motif key in the interaction are highlighted and shown in the TPR structure.

The complete sequence of the *C. elegans* genome was first published in 1998 (Consortium, 1998). The latest assembly (WS160, July 2006) has approximately 100 million base pairs and 20,060 genes. Two methods were used to search for *C. elegans* members of the Hsp70/90 interacting TPR domain containing protein family. Firstly, the complete *C. elegans* protein database, consisting of annotated and predicted proteins, was searched with the Hsp70/90 profile HMM. The *C. elegans* genome is well annotated and the complete protein repertoire should be represented in this database. The second approach was employed in case of omissions from the protein database. This method searched the complete genomic sequence using the program GeneWise (Birney et al., 2004). GeneWise is a sophisticated algorithm for the analysis of a DNA sequence at the level of its protein translation. It calculates the translation of a DNA sequence in all three frames, allowing frameshifts, introns and sequencing errors; and scores these against the input protein model including profile HMMs.

The analysis revealed 12 TPR domain containing co-chaperones predicated to interact with Hsp70 or Hsp90. This included six unannotated *C. elegans* proteins, three of which have no functionally annotated homologues in other species. The most interesting of these links the Hsp70/90 chaperone machinery to an undocumented role in fat metabolism.

6.2. Materials and methods

The complete set of *C. elegans* protein sequences and the complete *C. elegans* genome were obtained from the Ensembl database (release WS160, July 2006; www.ensembl.org).

The HMMer2.2 software (http://hmmer.janelia.org/) was used for HMM analysis. The package consists of programs for building profile HMMs (hmmbuild), searching protein databases (hmmsearch) and generating multiple alignments (hmmalign). Multiple sequence alignments were generated for all TPR domain containing protein families shown to interact with Hsp70 or Hsp90 including SGT, Hop (domains TPR1 and TPR2A), Cyp40, Chip, PP5, AIP, the FKBP family, Tom70, CNS-1 and UNC-45. Full-length sequences were obtained from the UniProt database using BLAST, aligned using MUSCLE and hand edited to select only the three 34 amino acid TPR repeats. Alignments from all families were then manually concatenated to ensure correct arrangement of the 34 residues TPR repeats. Loops between the TPR repeats were not included in the model, nor was the variable C-terminal capping helix. The final alignment, consisting of 635 sequences (see appendix A.3 for example alignments), was then used to build the profile HMM, figure 6-2 shows a logo representation of this model. This model was used to search both the *C. elegans* protein database (26,439

sequences) using HMMer and the complete *C. elegans* genome (100,281,235 base-pairs) using GeneWise. To avoid excessive demand of computational resources each *C. elegans* chromosome was divided into ~1.2 million bases with a 12,000 base-pair overlap prior to GeneWise analysis. All searches were run on a Dell 5150e with an Intel Pentium 4 HT processor and 512 MB RAM. Positive hits were manually inspected for conservation of the carboxylate-clamp residues.

Prediction (annotated name)	Gene location chr:start:stop:strand	Clamp motif	Description
R05F9.10 (sgt-1)	II:4902284:4903783:-1	K-N-N-K-R	SGT
R09E12.3 1	V:773389:774670:1	K-N-N-K-R	HOP TPR2A
R09E12.3 2	V:773389:774670:1	K-N-N-K-R	HOP TPR2B
Y39B6A.2 (pph-5)	V:19190339:19198715:1	K-N-N-K-R	PP5
F30H5.1 (unc-45)	III:491547:502061:1	R-N-N-K-R	UNC-45
T09B4.10 (chn-1)	I:6181318:6183882:-1	N-K-N-K-F	CHIP
C33H5.8	IV:7778004:7778946:-1	K-N-N-K-R	Unknown function
F31D4.3 (fkb-6)	V:20841374:20842783:1	K-T-N-K-R	FKBP
C17G10.2	II:5594706:5596327:1	K-N-N-K-R	CNS-1
C34B2.5	I:10675171:10676798:-1	K-N-N-K-R	Unknown function
ZK370.8	III:8752004:8754654:-1	K-N-N-K-R	TOM-70
C56C10.10	II:6592449:6594197:-1	R-N-N-K-R	AIP
T12D8.8.1	III:13614578:13622111:-1	R-Q-K-Q-F	HIP
Y22D7AL.9	III:1606834:1614022:-1	H-S-N-K-R	Unknown function
C55B6.2 (dnj-7)[1]	X:7194513:7198228:1	L-S-R-G-Q	J-domain co-chaperone
Y73E7A.9	I:1610391:1619944:1	R-S-N-K-R	Unknown function

[1]GeneWise prediction failed

Table 6-1 Predicted TPR domain containing proteins likely to interact with Hsp70 or Hsp90. T12D8.8.1 (HIP), C55B6.2 and Y22D7AL.9 were excluded from further analysis due to lack of conservation of the carboxylate-clamp residues or the presence of insertion/deletions which would interrupt the clamp residue positions.

6.3.   Results and discussion

A profile HMM search for *C. elegans* Hsp70/90 interacting TPR domains was conducted against the published *C. elegans* protein database and genomic sequence. Both produced very similar results identifying 12 TPR domain containing proteins predicted to interact with Hsp70 or Hsp90 (Table 6-1, Figure 6-3). These could be divided into three categories - annotated *C. elegans* co-chaperone homologues, unannotated *C. elegans* homologues of known co-chaperones, and TPR domain containing proteins of unknown function. The

HMMer output for the search against the *C. elegans* protein database is included in the appendices (Appendix A.4).
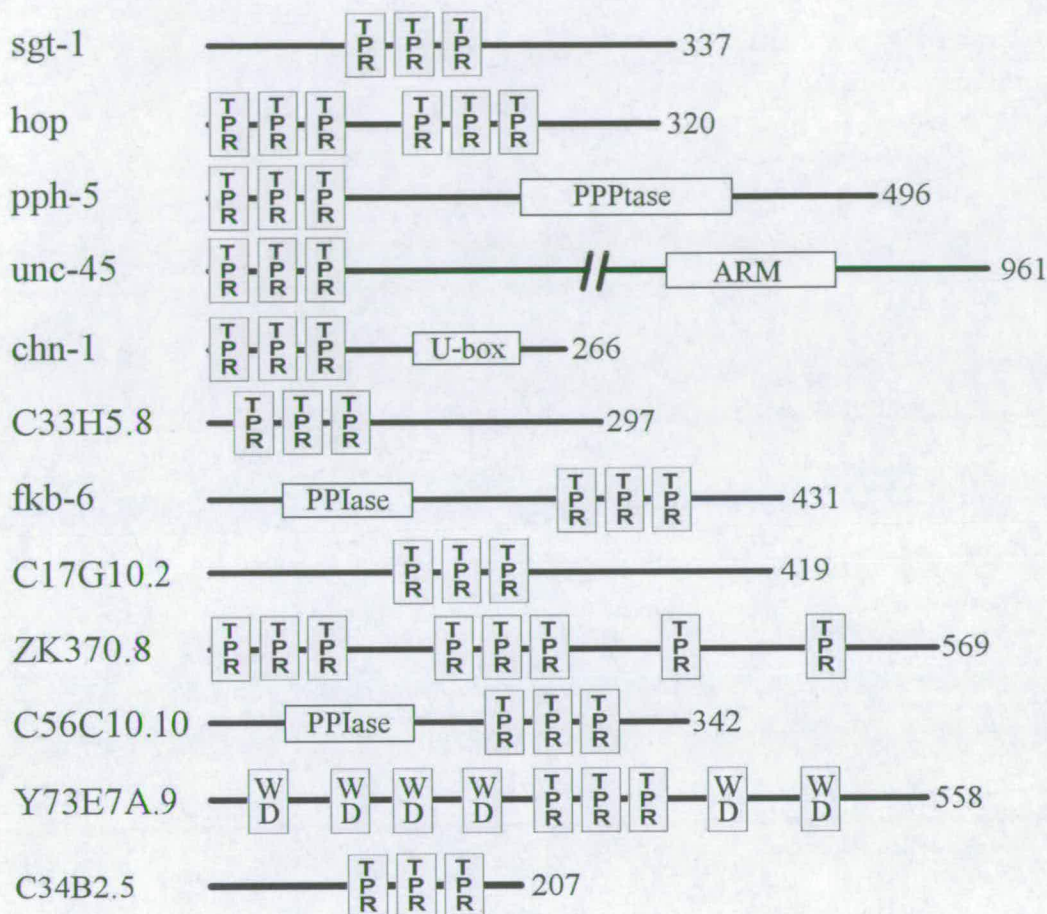
Figure 6-3 Domain architecture of predicted Hsp70/90 interacting TPR co-chaperones. PPPtase - protein phosphatase, ARM - armadillo repeat, U-box - E3 ubquitin ligase, PPIase - peptidyl-prolyl isomerase, WD - WD-40 repeat.

6.3.1. Identification of annotated *C. elegans* TPR domain containing proteins

In addition to ceSGT and ceHop (TPR domains TPR2A and TPR2B), discussed in chapters 4 and 5 respectively, annotated *C. elegans* co-chaperones identified included pph-5, the *C. elegans* homologue for protein phosphatase PP5; unc-45, a myosin co-chaperone important in muscle function (Barral et al., 2002); chn-1, the *C. elegans* homologue for the E3 ubiquitin-ligase Chip which links protein folding and degradation (Connell et al., 2001); and fkb-6, a member of the FK506-binding protein family and related to the large TPR domain containing immunophilins FKBP51 and FKBP52 (Opamawutthikul, 2005). All of these co-chaperones have been documented to be part of the Hsp70/90 chaperone machinery.

```
                1         10        20        30        40        50        60
C17G10.2    .MDIPQKKKFTESERAALAKKLDDDLDQFMEEMAARKSDKKEERKPFDFDDWCKEIDQHPAF
CNS1_YEAST  MSSVNANGGYTKPQKYVPGPG.DPELPPQLSEFKDKTS.........DEILKEMNRMPFF
CNS1-Human  ...|........MEQPGQDPTSDDVMDSFLEKFQSQPY....RGGFHEDQWEKEFEKVPLF
CNS1-Dros   MKQMAAQKAWTDEERLELAAQLDAELDAFIDGLEKKRY....EEGWPEDRWQEPMDKHPFF

                70        80        90        100       110       120
C17G10.2    MIEM.PT..DGKYQDTIEALQSMKYDKEDDEDKQMNAEHHKEEGNKHFRFRKYRWATDCYSN
CNS1_YEAST  MTKLDETDGAGGENVELEALKALAYEGEPHEI....AENFRKQGNELYKAKRFKDARELYSK
CNS1-Human  MSRA.PSEIDPRENPDLACLQSIIFDEERSPEEQ..AKTYRDEGNDYFREKDYKKAVISYTE
CNS1-Dros   MKRA.PQPGDD.VHPMFEGLQKLKYDPEENTRDEL.ALNYKEDGNFYMKHKKFRMAIYSFTE

                130       140       150       160       170       180
C17G10.2    GIKENSPDRKLNAVLYFNRAAAQKHLGNLRSAIKDCSMGRKFDPTHLKGVIRGAECLLELEY
CNS1_YEAST  GLAVECEDKSINESLYANRAACELELKNYRRCIEDCSKALTINPKNVKCYYRTSKAFFQLNK
CNS1-Human  GLKKKCADPDLNAVLYTNRAAQYYLGNFRSALNDVTAARKLKPCHLKAIIRGALCHLELKH
CNS1-Dros   GIKTKTDNPDVLAVLYNNRSAAHFFIKNYRSSLSDAQRALFYKPDYTKARWRSAQCAYELER

                190       200       210       220       230       240
C17G10.2    AKDALNWIPSSKKIFAFTKETSDTPDLTDDEKKFIDQLETLRVKSVELSLKEERDKRKSRAE
CNS1_YEAST  LEEAKSAATFANQR...........IDPENKSILNMLSVIDRKEQELKAKEEKQQREAQER
CNS1-Human  FAEAVNWCDEGLQI...........DAKEKKLLEM...RAKADKLKRIEQRDVRKANLK
CNS1-Dros   FDLCTQMCEELLEV...........DVDNEVAIALLHKNKMKKLEI...ERNQRKEAAE

                250       260       270       280       290       300
C17G10.2    ERKETESKKQLLDALKERNLNLCPRVPFDRPELMDMA..RLTVSLPLMHSHEC.VKFDDDLN
CNS1_YEAST  ENKKI....MPESAMTLRNIT.NIKTHSPVELLNEG......KIRLEDP....MDFESQ..
CNS1-Human  EKKERNQNEALLQAIKARNIRLSEAACEDEDSASEGLGELFLDGLSTENPHGARLSLDGQGR
CNS1-Dros   AKRRLTRFHRLRDAIEQRAIKFDDQKVGKKDVLSEEL..LYPKFLPLEDHP...VHLDEDGS

                310       320       330       340       350       360
C17G10.2    .LVWPILLQYPEAGQTDVLTETNELTTVGELLKEVLN...SPAQWDPEHKFNFENVRFFVSD
CNS1_YEAST  .LIYPALIMYPTQDEFDFVGEVSELTTVQELVDLVLEGPQERFKKEGKENFTPKKVLVFMET
CNS1-Human  .LSWPVLFLYPEYAQSDFISAFHEDSRFIDHLMVMFG...ETPSWDLEQKYCPDNLEVYFED
CNS1-Dros   TLIWPAAFSYPEFLYSDFYQQLPETTMRDCLATLLT...EKLPYDKAHNYRLGNVHVYYEN

                370       380       390       400       410
C17G10.2    EYDEYLMEVYEWNDFKSVLSM..PGYQIKQGLPVIMIMTKEKAASSLEALAEDENRYVAK.
CNS1_YEAST  K.AGGLIKAGKKLTFHDILKKESPDVPLFDNALKIYIVPKVESEGWISKWDKQKALERRSV
CNS1-Human  EDRAELYRVPAKSILLQVLQH..QRYFVKALTPAFLVCVGSSPFC..KNFLRGRKVYQIR.
CNS1-Dros   RKVGCVHKVDEEKQLAEIIAE..KGFFVSGGALLFYVVHKDSRVE..QEFINERRRPMVYS
```

Figure 6-4 Sequence alignment of *C. elegans* CNS-1 homologue (C17G10.2) with sequences from yeast, human and drosophila. TPR domain marked with blue bar and carboxylate-clamp motif residues highlighted with red stars. C17G10.2 shares ~30% sequence identity (~45% similarity) with human CNS-1.

6.3.2. Unannotated *C. elegans* TPR domain containing proteins with annotated homologues

6.3.2.1. C17G10.2 encodes the *C. elegans* CNS-1 homologue

Gene product C17G10.2 was identified to be the *C. elegans* homologue of co-chaperone CNS-1. The *C. elegans* protein aligns well with homologues from yeast, drosophila, xenopus and human, sharing ~30% sequence identity (~45% similarity) with the human sequence (Figure 6-4) with good conservation of the consensus carboxylate-clamp binding residues. CNS-1 was identified as an essential co-chaperone in yeast (Dolinski et al., 1998; Marsh et al., 1998); although little is known about its function it has been shown to bind and influence the chaperone activity of both Hsp70 (Hainzl et al., 2004) and Hsp90 (Lee et al., 2004; Tesic et al., 2003).

6.3.2.2. ZK370.8 encodes the *C. elegans* Tom70 homologue

The protein product of gene ZK370.8 is the *C. elegans* homologue for the mitochondrial import protein Tom70. The TOM (translocase of the mitochondrial outer membrane) complex contains receptors that mediate the targeting of proteins to the mitochondrial membrane and a general import pore complex through which proteins are translocated (Bains and Lithgow, 1999). The majority of mitochondrial proteins are nuclear encoded and synthesised in the cytoplasm. Prior to import, mitochondrial targeted proteins are bound to chaperones Hsc70/Hsp70 or Hsp90; and the Tom70 TPR domain provides a specific docking site for the chaperone-client complex. Although the *C. elegans* protein is not well conserved, sharing ~20% sequence identity (30% similarity) with the human homologue, it does align across the whole length of the protein and shares the same domain architecture with an N-terminal transmembrane domain, the Hsp70/90 binding TPR domain and a C-terminal array of TPR repeats (Figure 6-5). The carboxylate-clamp residues are strictly conserved across all species.
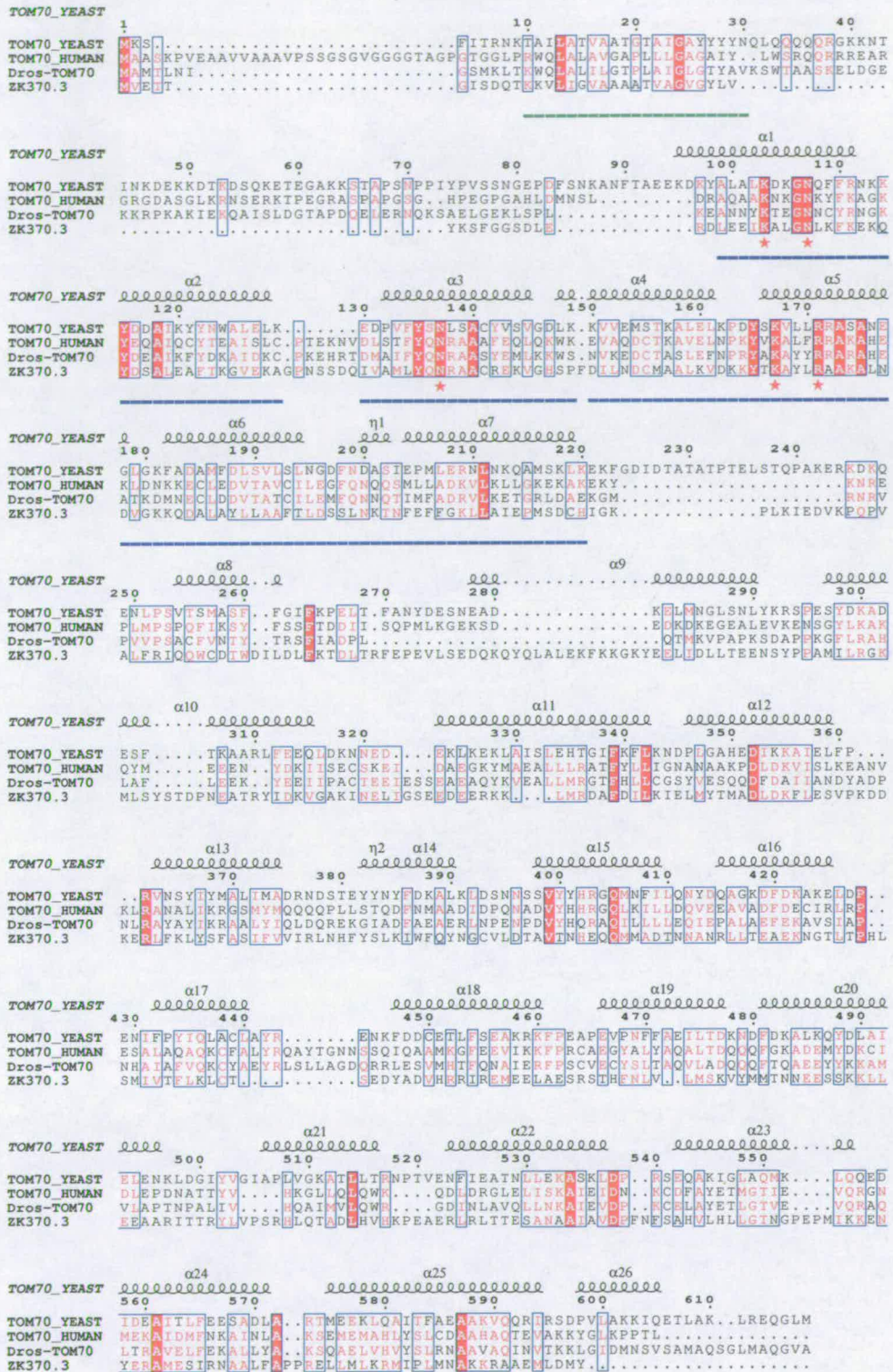
Figure 6-5 Sequence alignment of *C. elegans* Tom70 homologue (ZK370.3) with sequences from yeast, human and drosophila. TPR domain marked with blue bar and carboxylate-clamp motif residues highlighted with red stars; N-terminal transmembrane domain marked with green bar. Secondary structure from yeast crystal structure marked. ZK370.3 shares ~20% sequence identity (~30% similarity) with human CNS-1.

6.3.2.3. C56C10.10 encodes the *C. elegans* AIP homologue

Finally, gene product C56C10.10 belongs to the family of aromatic hydrocarbon receptor (AhR) interacting proteins (AIPs). AIP was first identified in mouse as a constituent of a cytosolic heterotrimeric complex consisting of AhR, AIP and Hsp90 (Ma and Whitlock, 1997). The AhR is a basic helix-loop-helix ligand inducible transcription factor which induces expression of enzymes involved in xenobiotic metabolism including the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (Pocar et al., 2005). AIP shares similarity with the immunophilin FKBP52, containing an N-terminal peptidylprolyl cis-trans isomerase domain and a C-terminal TPR domain; however, unlike FKBP52, AIP is unable to bind the immunosuppressant macrolide FK506. FKBP52, in addition to other imunophilins, exists as part of multi-protein steroid-hormone receptor complexes with Hsp90, and AIP is thought to be the analogous immunophilin component of the AhR receptor signalling complex (Riggs et al., 2004).

In addition, human AIP was also shown to participate in mitochondrial protein import (Yano et al., 2003). AIP was shown to interact with Tom20, a member of the TOM complex which binds an amino-terminal targeting sequence on mitochondrial preproteins via a TPR domain. AIP was shown to bind Hsp70 and Tom20 in the same manner, via a carboxylate-clamp mediated interaction with the extreme C-terminal peptides (EEVD in Hsp70 and DDVE in Tom20). In addition, AIP was shown to interact with the preprotein sequence via its PPIase and TPR domains. A model was thus proposed where Hsp70, AIP and a mitochondrial preprotein form a large complex in the cytosol. This complex is targeted to the TOM via an interaction of AIP with Tom20. The preprotein is then transferred to Tom20 for subsequent import into the mitochondria.

C. elegans AIP shares 34% sequence identity (55% sequence similarity) with human AIP. Positions 2-5 of the carboxylate-clamp motif are strictly conserved across all species; however, higher animals are unique in that they have a histidine at position 1, perhaps reflecting the ability of this TPR domain to recognise multiple C-terminal sequences (Figure 6-6). The TPR domain was shown to be the interacting site for Hsp90 with mutations in carboxylate-clamp motif positions 4 (lysine) and 5 (arginine) affecting the Hsp90 interaction (Bell and Poland, 2000).

Figure 6-6 Sequence alignment of *C. elegans* AIP homologue (C56C10.10) with sequences from human, drosophila and xenopus. TPR domain marked with blue bar and carboxylate-clamp motif residues highlighted with red stars; N-terminal PPIase domain marked with green bar. C56C10.10 shares ~34% sequence identity (~53% similarity) with human CNS-1.

6.3.3. Putative *C. elegans* Hsp70/Hsp90 TPR domain containing co-chaperones

The database search also highlighted several proteins with no functionally annotated homologues including gene products C33H5.8, C34B2.5, and Y73E7A.9. These putative proteins were subjected to further investigation.

6.3.3.1. Gene C33H5.8 encodes a protein with no known function

C33H5.8 is a two exon gene located on the reverse strand of chromosome IV (Figure 6-7), which encodes a 297 residue 34.4 kDa protein with an estimated isoelectric point of 5.26.

The protein contains similarities to InterPro and Pfam tetratricorepeat domains (IPR013026, IPR001440 and PF00515) and InterPro Protein phosphatase 5 domain (IPR011236).

Except across the TPR domain the protein has low similarity to any sequence in the UniProt database. It does, however, cluster with Ensembl family ENSF00000003666 (Mitochondrial import receptor subunit Tom34 translocase of outer membrane 34 kDa subunit) perhaps alluding to a function in the mitochondrial protein import machinery.

Figure 6-7 Predicted gene structure for C33H5.8. C33H5.8 is a two exon gene located on the reverse strand of chromosome IV. Its predicted protein product matches TPR and PP5 domains from InterPro and Pfam.

Figure 6-8 Predicted gene structure for C34B2.5. C34B2.5 is a four exon gene located on the reverse strand of chromosome I. Its predicted protein product matches TPR and PP5 domains from InterPro.

6.3.3.2. Gene C34B2.5 encodes an orthologue of tertratricorepeat protein 1

C34B2.5 is a four exon gene located on the reverse strand of chromosome I (Figure 6-8), which encodes a 207 residue 23.3 kDa protein with an estimated isoelectric point of 4.52. It is expressed in the larvae and adult in the intestine, renal glands and nervous system, and exhibits a late larval arrest RNAi phenotype. The protein contains a predicted Hsp70/90 binding TPR domain with a capping C-terminal helix (residues 18-141), which has similarities to InterPro Tetratricopeptide region (IPR013026), Protein phosphatase 5 (IPR011236) and Tetratricopeptide-like helical (IPR011990) domains.

```
C34B2.5              ........................................................
Q7Q9B6_ANOGA    1    ...QFEDAQSKEDF....................NERTVDEVIEKVMGLSVRRPPDSDAEE..F
Q9VEV3_DROME    1    ...MAGNANSDDEF....................HDAIGEEITEK..EATSTRQSEKDVDE..I
TTC1_MOUSE      1    MEEKSEDCKVPEDLFNGLKVADPQEGESASPMVSDPKGQHCQSKLPKAAEAHPQDDHVEEECF
TTC1_HUMAN      1    MGEKSENCGVPEDLLNGLKVTDTQEAECAGPPVPDPKNQHSQSKLLRDDEAHLQEDQGEEECF


C34B2.5         1    ......................MAIEEIF............EDEEDQITKVDSLKKE
Q7Q9B6_ANOGA   40    VDCVNGPEQENESCRQVAGTSEEDSPPE.DDLIDEDSQRDYECGLSEEELEANKTKADELKQQ
Q9VEV3_DROME   38    VEKQNQLALDDEAEQGAAGGDSIATPTTVDSELTIEELRREKDLSPEQLTANKEKADKLKVE
TTC1_MOUSE     64    HDCSASFEEEQPGAHVAGSKASDDS....SSELDEEYLIELEKNMPEEEKQKRREESAKLKEE
TTC1_HUMAN     64    HDCSASFEEE.PGADKVENKSNEDVN...SSELDEEYLIELEKNMSDEEKQKRREESTRLKEE
                                                                              ★

C34B2.5        25    GNNFFANGEFEKANEKYQEAIASCPPTSTEVQSILLSNSAAALIKLRKWESAVEAASKSIEIG
Q7Q9B6_ANOGA  102    GNELFKQGEHSRSLDLYTQALRLCPLDRKEARAILYANRAAAKAKLDRKQSALEDCTKALEYN
Q9VEV3_DROME  101    GNELFKNDDAEGAAKTYTEALDICPSASSKERAVLYGNRAAAKIKLEANKAAIDDCTKAIELW
TTC1_MOUSE    123    GNERFKRGDYMEAESSYSQALQMCPACFQKDRSVLFSNRAAARMKQDKKETAITDCSKAIQLN
TTC1_HUMAN    123    GNEQFKKGDYIEAESSYSRALEMCPSCFQKERSILFSNRAAARMKQDKKEMAINDCSKAIQLN
                        ★     ★                             ★

C34B2.5        88    ATNEKALERRAFAYSNMSEKYENSIEDYKQLQESLPKRRVELDRKIAEINEKINNRNEAMRAD
Q7Q9B6_ANOGA  165    FHYLKALLRRANLYEE.TDKLDESLEDYRKVLELEPGNGEARSAQV.RLPPKIAERNERLKEE
Q9VEV3_DROME  164    PEYVRVLLRRAKLYEQ.EDKPDEALEDYKKVTEIDPGQQEAREAQI.RLPPIINERNEKLKNE
TTC1_MOUSE    186    PTYIRAILRRAELYEK.TDKLDEALEDYKSVLEKDPSVHQAREACM.RLPKQIEERNERLKEE
TTC1_HUMAN    186    ESYIRAILRRAELYEK.TDKLDEALEDYKSILEKDPSIHQAREACM.RLPKQIEERNERLKEE
                       ★    ★

C34B2.5       151    VMEKLKGFGNFCLSPFGLSTDSFEMVPNGN.GGFSVQMKGATNKQEGASQEKKEEENA
Q7Q9B6_ANOGA  226    MMGKLKDLGNMILRPFGLSTQNFEMKQDPQTGSYSINFNQGGKQ.............
Q9VEV3_DROME  225    MMSSLKDLGNMILKPFGLSTQNFQMQQDPNTGSYSINFK.................
TTC1_MOUSE    247    MLGKLKDLGNLVLRPFGLSTENFQIKQDSSTGSYSINFVQNPNNNR...........
TTC1_HUMAN    247    MLGKLKDLGNLVLRPFGLSTENFQIKQDSSTGSYSINFVQNPNNNR...........
```

Figure 6-9 Sequence alignment of *C. elegans* TTC1 homologue (C34B2.5) with sequences from mosquito, drosophila, mouse and human. TPR domain marked with blue bar and carboxylate-clamp motif residues highlighted with red stars. C34B2.5 lacks the first ~100 residues and shares ~35% sequence identity (~57% similarity) with human over aligned residues.

C34B2.5 shares the same InParanoid cluster (Remm et al., 2001), an orthologue clustering database, with a group of proteins named tertratricorepeat protein 1 (TTC1). Sequence alignment shows that C34B2.5 lacks the first ~100 residues compared to TTC1_HUMAN, sharing 35% sequence identity (57% similarity) over the aligned sequence comprising the

TPR domain and C-terminal (Figure 6-9). The carboxylate-clamp motif adheres to the consensus and is conserved across all species.

There is no literature regarding the function of the TTC1 family of proteins. In a genome-wide prediction of *C. elegans* genetic interactions (Zhong and Sternberg, 2006), the most significant predicted interaction partner of C34B2.5 was NADH-ubiquinone oxidoreductase flavoprotein 2 (NDUFV2; gene F53F4.10). This is a subunit of mitochondrial complex I (NADH-ubiquinone oxidoreductase); a 43 protein complex that catalyses the first step in the mitochondrial electron transport chain. NDUFV2 is a nuclear encoded protein requiring import into the mitochondria perhaps implicating C34B2.5 in the TOM or TIM (translocase of the mitochondrial inner membrane) complexes. Additionally, the *Drosophila* TTC1 homologue was shown to interact with a basic helix-loop-helix protein (Giot et al., 2003). Although *C. elegans* lacks an orthologue for this bHLH protein, this might suggest a role in the regulation of a cytoplasmic ligand-inducible transcription factor similar to the AhR or steroid-hormone receptors.
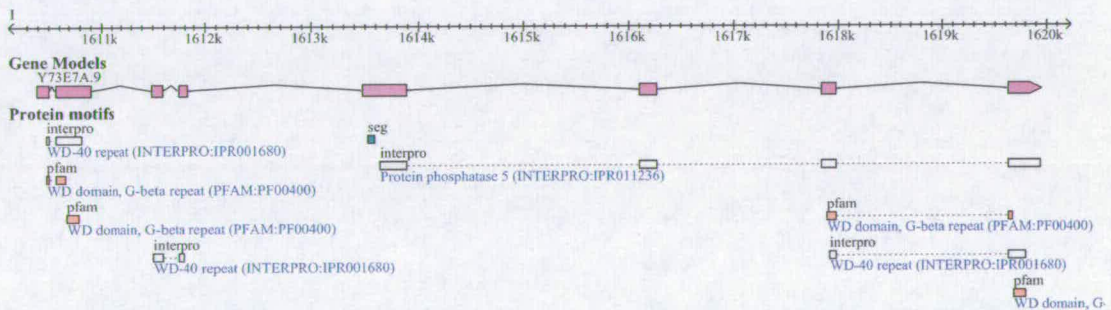
Figure 6-10 Predicted gene structure for Y73E7A.9. Y73E7A.9 is an 8 exon gene located on the forward strand of chromosome I. Its predicted protein product matches TPR and WD-40 domains from InterPro and Pfam.

6.3.3.3. Gene Y73E7A.9 encodes a conserved WD-40/TPR repeat protein implicated in fat metabolism

Y73E7A.9 is an eight exon gene located on the forward strand of chromosome I (Figure 6-10), which encodes a 558 residue 63.4 kDa protein with an estimated isoelectric point of 5.64. The protein contains similarities to the Interpro protein phosphatase 5 domain (IPR011236) with the TPR domain predicted from residue 285-414. This is flanked on the N- and C-termini by a series of WD-40 repeats matching both Interpro and PFAM domains. WD-40 repeats are short ~40 residue motifs existing in arrays of 4-16 repeats which form a circularised beta-propeller structure (Smith et al., 1999). Their primary function is as a

scaffold domain, coordinating the assembly of multi-protein complexes and they have been implicated in signal transduction, transcription, cell cycle control and apoptosis (Smith et al., 1999).
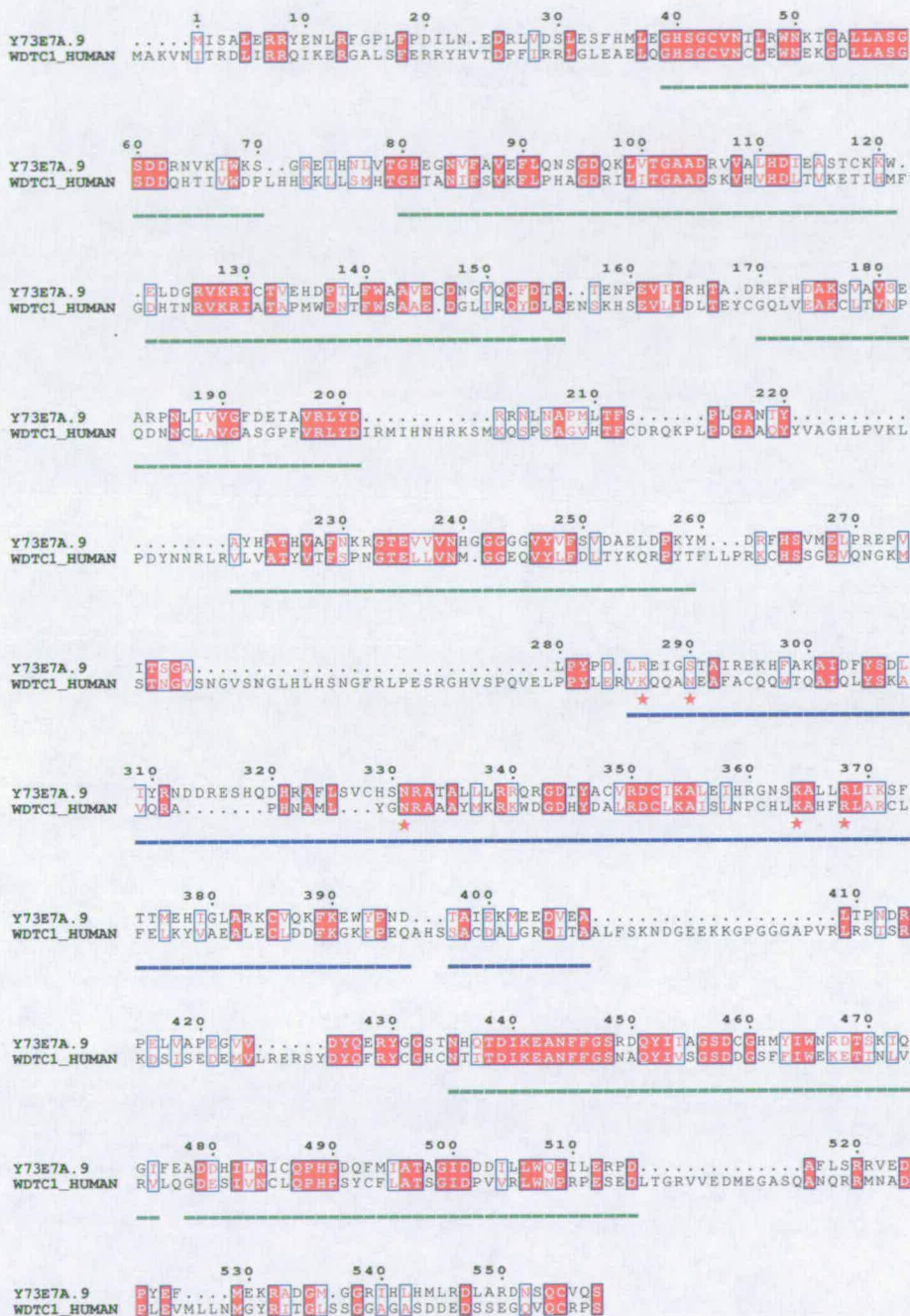
Figure 6-11 Sequence alignment of *C. elegans* adp homologue (Y73E7A.9) with human WDTC1_HUMAN. TPR domain marked with blue bar and carboxylate-clamp motif residues highlighted with red stars; WD-40 repeats marked with green bars. Y73E7A.9 shares ~35% sequence identity (~57% similarity) with human over aligned residues.

The carboxylate-clamp motif is different from the consensus with an arginine at position one and a serine at position two. Arginine is found at position one albeit at a lower frequency than lysine. Serine, however, is not found at position two in any of the known Hsp interacting TPR domains. *C. elegans* FKB-6 does, however, have a threonine substitution with similar physiochemical properties to serine.

The Y73E7A.9 amino acid sequence shares similarity with a family of WD-40/TPR repeat containing proteins that is well conserved from worms to humans (Figure 6-11). The protein possesses six predicted WD-40 repeats with a TPR domain between the fourth and fifth repeats. The carboxylate-clamp motif is well conserved across species with the non-canonical serine at position two of the *C. elegans* sequence a consensus lysine in most other species, in agreement with an interaction with Hsp70 or Hsp90.

Interestingly, the *D. melanogaster* homologue was identified to be mutated in *adp*[60] flies. These flies develop an obese phenotype if food supply permits with an increased accumulation of triglycerides, the lipid stores in the fly (Hader et al., 2003). Conversely, over-expression of the *adp* gene in the fat body, analogous to the vertebrate liver, caused a decrease in fat content but had no affect on cellular viability in the nervous system indicating a role in fat body cells. Consequently, this pathway has been identified as a potential target for pharmaceutical intervention in the treatment of obesity (Dohrmann, 2004). A model of the *D. melanogaster* homologue was produced showing a 7-stranded β-propeller with an associated TPR domain (Figure 6-12).

6.4. Conclusions

12 TPR domain containing proteins have been identified that are predicted to interact with Hsp70 or Hsp90 based on the conservation of the carboxylate-clamp motif. These proteins likely represent the complete repertoire of Hsp70/90 TPR co-chaperones in *C. elegans*. In addition to SGT and HOP, identified in a previous search, homologues of PP5, UNC-45, Chip and an FK506 binding immunophilin were identified that have previously been annotated. Six additional proteins were identified that were unannotated in WormBase. Three of these were found to be the *C. elegans* homologues of CNS-1, TOM70 and AIP. The remaining three are of unknown function although one, the product of gene Y73E7A.9, was found to be the homologue of a Drosophila protein involved in fat metabolism.

β-propeller   TPR motifs
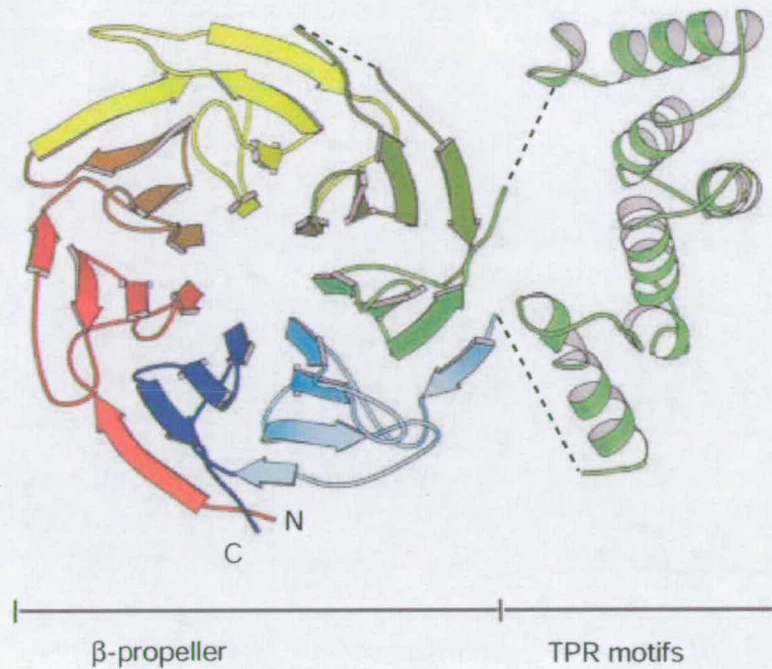
Figure 6-12 Modelled structure of the human adipose protein. Model building suggests that the adipose (adp) protein can form a seven-bladed beta-propeller with an associated TPR domain. Figure taken from Hader *et al.*, 2003.

6.5. References

Bains, G., and Lithgow, T. (1999). The Tom channel in the mitochondrial outer membrane: alive and kicking. Bioessays *21*, 1-4.

Barral, J. M., Hutagalung, A. H., Brinker, A., Hartl, F. U., and Epstein, H. F. (2002). Role of the myosin assembly protein UNC-45 as a molecular chaperone for myosin. Science *295*, 669-671.

Bell, D. R., and Poland, A. (2000). Binding of aryl hydrocarbon receptor (AhR) to AhR-interacting protein. The role of hsp90. J Biol Chem *275*, 36407-36414.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988-995.

Connell, P., Ballinger, C. A., Jiang, J., Wu, Y., Thompson, L. J., Hohfeld, J., and Patterson, C. (2001). The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. Nat Cell Biol *3*, 93-96.

Consortium, T. C. e. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.

Dohrmann, C. E. (2004). Target discovery in metabolic disease. Drug Discov Today *9*, 785-794.

Dolinski, K. J., Cardenas, M. E., and Heitman, J. (1998). CNS1 encodes an essential p60/Sti1 homolog in Saccharomyces cerevisiae that suppresses cyclophilin 40 mutations and interacts with Hsp90. Mol Cell Biol *18*, 7344-7352.

Eddy, S. R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755-763.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., *et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Hader, T., Muller, S., Aguilera, M., Eulenberg, K. G., Steuernagel, A., Ciossek, T., Kuhnlein, R. P., Lemaire, L., Fritsch, R., Dohrmann, C., *et al.* (2003). Control of triglyceride storage by a WD40/TPR-domain protein. EMBO Rep *4*, 511-516.

Hainzl, O., Wegele, H., Richter, K., and Buchner, J. (2004). Cns1 is an activator of the Ssa1 ATPase activity. J Biol Chem *279*, 23267-23273.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. J Mol Biol *235*, 1501-1531.

Lee, P., Shabbir, A., Cardozo, C., and Caplan, A. J. (2004). Sti1 and Cdc37 can stabilize Hsp90 in chaperone complexes with a protein kinase. Mol Biol Cell *15*, 1785-1792.

Ma, Q., and Whitlock, J. P., Jr. (1997). A novel cytoplasmic protein that interacts with the Ah receptor, contains tetratricopeptide repeat motifs, and augments the transcriptional response to 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Biol Chem *272*, 8878-8884.

Marsh, J. A., Kalton, H. M., and Gaber, R. F. (1998). Cns1 is an essential protein associated with the hsp90 chaperone complex in Saccharomyces cerevisiae that can restore cyclophilin 40-dependent functions in cpr7Delta cells. Mol Cell Biol *18*, 7353-7359.

Opamawutthikul, M. (2005) PhD thesis - Biochemical and Biophysical Studies of FK506 Binding Proteins, University of Edinburgh, Edinburgh.

Pocar, P., Fischer, B., Klonisch, T., and Hombach-Klonisch, S. (2005). Molecular interactions of the aryl hydrocarbon receptor and its biological and toxicological relevance for reproduction. Reproduction *129*, 379-389.

Remm, M., Storm, C. E., and Sonnhammer, E. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol *314*, 1041-1052.

Riggs, D. L., Cox, M. B., Cheung-Flynn, J., Prapapanich, V., Carrigan, P. E., and Smith, D. F. (2004). Functional specificity of co-chaperone interactions with Hsp90 client proteins. Crit Rev Biochem Mol Biol *39*, 279-295.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199-210.

Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999). The WD repeat: a common architecture for diverse functions. Trends Biochem Sci *24*, 181-185.

Tesic, M., Marsh, J. A., Cullinan, S. B., and Gaber, R. F. (2003). Functional interactions between Hsp90 and the co-chaperones Cns1 and Cpr7 in Saccharomyces cerevisiae. J Biol Chem *278*, 32692-32701.

Yano, M., Terada, K., and Mori, M. (2003). AIP is a mitochondrial import mediator that binds to both import receptor Tom20 and preproteins. J Cell Biol *163*, 45-56.

Zhong, W., and Sternberg, P. W. (2006). Genome-wide prediction of C. elegans genetic interactions. Science *311*, 1481-1484.

7. Summary and future work

This thesis has described the structural and biochemical studies of proteins belonging to the Hsp70/Hsp90 chaperone system in the nematode worm *Caenorhabditis elegans*. The work was broadly divided into three main projects:

1. The structural studies of the C-terminal 10 kDa subdomain of *C. elegans* Hsp70 (Chapters 2 and 3).

2. Biochemical characterisation of two putative TPR domain containing co-chaperones (Chapters 4 and 5).

3. Prediction of the complete repertoire of *C. elegans* TPR domain containing co-chaperones (Chapter 6).

This chapter will provide a brief overview of the major findings and conclusions of the projects and also outline areas requiring further work.

7.1. Structural studies of the C-terminal domain of *C. elegans* Hsp70

There is abundant structural information regarding the NBD and the β-sandwich subdomain of Hsp70. Structures of the C-terminal lid subdomain are, however, limited to *E. coli* homologues DnaK and HscA, and rat homologue Hsc70. Despite structural conservation of the NBD and β-sandwich between *E. coli* and rat, the C-terminal subdomains were observed to adopt significantly different conformations; a three-helix bundle in *E. coli* and an anti-parallel coiled-coil dimer in rat. Limited by the available data, it is unclear whether these reflect a true divergence between prokaryotes and eukaryotes. Alternatively, the structures could represent different conformational states or, indeed, the rat structure could be a crystallographic artefact.

7.1.1. Project aims

To provide further insight into the structure and properties of the Hsp70 C-terminal subdomain the aims of the project were:

- To solve the crystal structure of the *C. elegans* subdomain.

- To characterise the oligomeric state in solution.

7.1.2. Major findings and conclusions

- A recombinant C-terminal construct of *C. elegans* Hsp70 (ceHsp70-CT) was successfully generated (residues Gly^{542}-Asp^{640}), expressed and purified.

- ceHsp70-CT, including the N-terminal 6xHis tag was crystallised in two forms, an orthorhombic form belonging to space group $I2_12_12_1$ and a tetragonal form belonging to space group $P4_22_12$.

  - The orthorhombic crystal form was initially solved using MAD with data collected from a mercury derivative crystal; however, diffraction to only 4 Å and 24 monomers in the asymmetric unit hampered refinement (section 2.3.1.).

  - The tetragonal form, which diffracted X-rays to ~3 Å and contained six monomers in the asymmetric unit, was solved by molecular replacement using a hexameric search model constructed from the orthorhombic data (section 2.3.2.).

- The final model, refined to final R_{cryst}/R_{free} of 27.6%/29.0%, consists of six protomers arranged as a pair of back-to-back trimers with 32 point group symmetry. The monomeric structure consists of four α-helices folded into a compact three helix bundle (see section 2.3.3.).

- Comparison with structures from *E. coli* and rat surprisingly revealed structural conservation with the more distantly related bacterial homologues.

  - The structural conservation of the C-terminal domain has been proposed across all Hsp70 family members although, in light of the distinct rat C-terminal structure, the *C. elegans* structure represents the first direct evidence of this in eukaryotic Hsp70s (section 3.3.3.).

- Analysis of the *C. elegans* and rat structures revealed that the dimerisation of the rat domain in the crystal structure was mediated by a domain-swap mechanism (section 3.3.4.).

  - The alternate rat conformation was postulated to represent the substrate free SBD conformation. ATP binding, which allosterically triggers opening of the SBD and substrate release, has been shown to induce dissociation of Hsp70 oligomers to the monomeric form. This would be incompatible with

the rat elongated Hsc70 C-terminal structure which is shown to form domain-swapped dimers.

- The *C. elegans* subdomain only exists as a monomer in solution and, although it crystallised as a hexameric complex, none of the interfaces are predicted to be of biological relevance.

  o Based on the rat crystal structure and the behaviour of the same construct in solution the C-terminal 10 kDa subdomain was proposed to be necessary and sufficient for self-association. Results here, however, demonstrate the same domain from *C. elegans* behaves as a monomer in solution and that the monomer most likely represents the biologically relevant unit of the crystal. Results here corroborate a recent assertion that the dimerisation mechanism based on the rat structure needs to be re-examined (section 3.3.2.1.).

- There is an inverse relationship between ceHsp70-CT thermal stability and pH, and, at pH 6.5, ceHsp70-CT unfolds via the accumulation of at least one "open" intermediate possibly suitable for domain-swap dimer formation.

  o Non biological domain-swaps are commonly observed in crystal structures and can provide insight into protein folding and flexibility (section 3.3.4.1.). Folding pathways of three-helix bundles have been proposed to be populated by open two-helix intermediates suitable for domain-swapped dimer formation. The thermal denaturation studies of ceHsp70-CT suggest that the rat domain-swapped structure may be a trapped folding intermediate on the three-helix bundle folding pathway.

- Work for this project has contributed to two papers; one published in Acta Crystallographica section F and one accepted for publication in Biochemical and Biophysical Research Communications.

  o Worrall, L., and Walkinshaw, M. D. (2006). Crystallization and X-ray data analysis of the 10 kDa C-terminal lid subdomain from Caenorhabditis elegans Hsp70. Acta Crystallograph Sect F Struct Biol Cryst Commun *62*, 938-943.
  o Worrall, L., and Walkinshaw, M. D. (2007). Crystal structure of the C-terminal three-helix subdomain from *C. elegans* Hsp70. Biochemical and Biophysical Research Communications *357*, 105-110.

7.1.3. Future work

- Although the sequence studied was selected based on the rat construct, examination of the *C. elegans* structure demonstrated that several recombinant tag residues

contributed to the hydrophobic packing of the three-helix bundle (section 3.3.5.). To exclude the possibility that these were inducing an artificial conformation it would be necessary to use an extended C-terminal construct incorporating the complete stable folding unit, specifically starting at residues Asp[535].

- A more detailed examination of the thermal stability and folding pathway of the *C. elegans* Hsp70 C-terminal subdomain is required to investigate the hypothesis that that domain-swapped form represents a folding intermediate. In particular, it would be interesting to determine whether the *C. elegans* subdomain is capable of forming domain-swapped dimers.

7.2. Biochemical and structural studies of two putative TPR domain containing co-chaperones

Work contributing to a previous doctoral degree identified two *C. elegans* TPR domain containing co-chaperones likely to interact with Hsp90. These proteins were found to be the *C. elegans* homologue for small glutamine-rich TPR containing protein (SGT) and a protein with homology to the C-terminal region of Hsp70/Hsp90 organising protein (Hop) but lacking the first 218 residues consisting of a TPR motif and DP-repeat region.

7.2.1. Project aims

- To clone, express and purify ceSGT and ceHop.

- To characterise the oligomeric state of both proteins.

- To characterise the interactions of the co-chaperones with Hsp70 and Hsp90.

- To crystallise and solve the structure of both proteins and/or subdomains thereof.

7.2.2. Major findings and conclusions

- Both ceSGT and ceHop were successfully cloned, expressed and purified to homogeneity (sections 4.2. and 5.2.).

- MALDI-TOF mass-spectrometry and glutaraldehyde cross-linking demonstrated that ceSGT is capable of forming dimers whilst gel-filtration shows these are elongated in shape and exist down to low nanomolar concentrations (section 4.3.2.).

  o SGT has been shown to oligomerise. Based on cross-linking results these were proposed to be dimers although, as evidenced here, these results alone

Structural and biochemical studies of the *C. elegans* Hsp70/Hsp90 chaperone system are not conclusive. Mass-spec results here confirm the ability of ceSGT to form dimers.

- Cross-linking and gel-filtration results are consistent with the dimerisation of ceHop (section 5.3.1.).

  o Hop too has been shown to exist as dimers and also to interact with Hsp90 as a dimer. Gel-filtration and cross-linking studies support this. Indirect evidence for the presence of a monomer-dimer equilibrium also comes from the shifting retention volume of free ceHop in gel-filtration studies of the interaction with Hsp90 (see section 5.3.2.)

- Gel-filtration analysis demonstrated both ceSGT and ceHop interact with Hsp90.

  o ceSGT interacts with the C-terminal domain of human Hsp90α with low affinity (section 4.3.3.1.).

  o ceHop interacts with the C-terminal domains of human and *C. elegans* Hsp90 with a dissociation constant of ~0.8 μM, in agreement with other studies of Hop-Hsp90 interactions (section 5.3.2.).

- The isolated TPR domain from ceSGT interacts with the C-terminal peptides from Hsp70 and Hsp90 with comparable affinities.

  o Studies of the interaction between Hsp70/Hsp90 with TPR co-chaperones have shown that many interactions can be reduced to the C-terminal Hsp peptides and the isolated TPR domains. The interaction of the ceSGT TPR domain and Hsp70/Hsp90 C-terminal peptides was investigated using ITC and CD, showing the ceSGT TPR domain interacted with both peptides with similar affinities of ~35 μM (section 4.3.3.2.).

- The isolated TPR domain is fully folded and peptide binding induces no change in secondary structure.

  o It has been suggested that a coupled binding-folding mechanism could be a means of regulating substrate binding by TPR domains. Results presented here disagree with this hypothesis with far-UV CD analysis of the ceSGT TPR domain showing no difference in the peptide free and bound secondary structure (section 4.3.3.2.).

- Crystals of full-length ceSGT and the isolated TPR domain were grown but were of insufficient quality for further studies.

    o   Full-length ceSGT was crystallised in 200 mM potassium thiocyanate, 5-20% PEG 3350, pH 7.0 (section 4.3.4.). Two crystal forms were obtained; small octahedral plates were grown at 4 °C using 20% PEG 3350 and long thin needles were grown at 20 °C using 5% PEG 3350. Both the small plates and fine needles, however, failed to diffract.

    o   Crystals of the isolated SGT TPR domain were grown from 100 mM HEPES pH 7.5, 800 mM sodium phosphate, 800 mM potassium phosphate. These crystals did diffract although only to 5 Å.

7.2.3. Future work

- Further work is necessary to fully characterise the interaction of both co-chaperones with Hsp70 and Hsp90. In light of the similar affinity of the ceSGT TPR domain for Hsp70 and Hsp90 peptides, studies of the interaction of full-length ceSGT with Hsp70 and Hsp90 are necessary to investigate whether regions outside the TPR domain or C-terminal peptides contribute to the interaction.

- The interaction of ceHop with Hsp70 was not studied. ceHop lacks domain TPR1, the major site of interaction with Hsp70, so it will be of interest to determine whether ceHop can interact with Hsp70. Recent results from yeast and human Hop homologues suggest that TPR1 and TPR2B have overlapping functions indicating the smaller *C. elegans* homologue will be able to support many of the functions of the full-length version.

- Structural studies are ongoing. Optimisation of crystallisation conditions of full-length ceSGT and the isolated TPR domains are necessary to obtain diffraction quality crystals. In the meantime small-angle X-ray scattering experiments are planned to determine a low resolution molecular envelope for the ceSGT dimer.

7.3. Prediction of the complete repertoire of *C. elegans* TPR co-chaperones

The group of TPR domain containing proteins that interact with Hsp70/Hsp90 represent a major class of co-chaperones. Although the number of members of this family is rapidly expanding, a thorough analysis of the published genomes is lacking. Hsp70/Hsp90 interacting TPR domains have a characteristic domain structure of three TPR repeats and conservation of key residues involved in the interaction most notably the five polar residues defined as the carboxylate-clamp. This well defined architecture makes the domain well

suited for modelling with profile hidden Markov models, probabilistic models of protein families.

7.3.1. Project aims

- To predict the complete repertoire of *C. elegans* TPR domain containing proteins capable of interacting with Hsp70 or Hsp90. A profile hidden Markov model (HMM) method was employed to search for Hsp70/Hsp90 interacting TPR domains in the *C. elegans* proteome and genome.

7.3.2. The *C. elegans* Hsp70/90 TPR co-chaperone family

- 12 proteins were identified with a characteristic TPR domain architecture and conservation of the carboxylate-clamp residues necessary for the interaction with Hsp70 or Hsp90.

  - These include *C. elegans* homologues for proteins already characterised as Hsp70/Hsp90 co-chaperones including SGT, Hop, Chip, PP5, FKB6, UNC45, CNS1, Tom70 and AIP (section 6.3.1. and 6.3.2.).

  - The remaining three proteins are uncharacterised and there is no published evidence of interactions with either Hsp70 or Hsp90 (section 6.3.3.). The most interesting of these is a WD-40/TPR repeat protein which, in Drosophila, has been shown to be involved in fat metabolism.

7.3.3. Future work

- The next step in this project is to clone all of the Hsp70/Hsp90 TPR co-chaperones and investigate their interaction with both Hsp70 and Hsp90. The most interesting starting point would be the three novel proteins.

- The analysis will also be extended to other sequenced genomes including *Drosophila melanogaster*, *Arabidopsis thaliana* and *Homo sapiens*.

A. Appendices

A.1. List of Hsp90 interacting proteins, curated by Cyril Picard (http://www.picard.ch/downloads/Hsp90interactors.pdf)

HSP90 INTERACTORS

Chaperones and relatives:

- Aha1 and its homolog Hch1
- Cdc37 (p50) and its relative Harc
- CS-containing p23 relatives SGT1, RAR1, Siah-1-interacting protein (SIP), Chp1
- Hsp70
- Human DnaJ homolog Hsj1b
- p23 (=Sba1)
- proteins with TPR motifs, including Hop (=Sti1), FKBP52 (and high MW plant homologs), FKBP51, FKBP8 (=FKBP38), cyclophilin-40 (Cpr6 and Cpr7), PP5 (and yeast Ppt1), Tom70, XAP-2 (=AIP=ARA9), Cns1 and its Drosophila relative Dpit47, CHIP, GCUNC-45 (also UNC-45 and She4), Tpr2 (=mDj11=CCRP), CRN, WISp39, Tah1, NASP, Toc64.
- S100A1
- Sse1
- valosin-containing protein (VCP)

Transcription factors:

- 12(S)-HETE receptor
- all vertebrate steroid receptors (GR, MR, ER, PR, AR)
- CAR
- cytoplasmic v-erbA
- EcR
- PPARα (PPARβ)
- PXR
- Hap1
- HSF-1
- IRF3
- Mal63
- p53
- PAS family members: Dioxin receptor (=AhR), Sim, HIF–1α, HIF-2α, HIF-3α
- Stat3 (also in caveolin-1 complexes in rafts)
- TonEBP/OREBP
- water mold *Achlya* steroid (antheridiol) receptor

Kinases:

- Akt/PKB
- ASK1
- Aurora B
- Bcr-Abl
- casein kinase IIα catalytic subunit
- Cdk2, Cdk4, Cdk6, Cdk9, Cdk11
- Chk1
- Death-associated kinases DAPK, DAPK2, DAPK3
- death domain kinase RIP
- eEF-2 kinase
- eIF2-α kinases HRI, Gcn2, Perk, PKR
- ErbB2 (and mutant EGF receptor)
- ERK5
- Flt3
- Fused
- GRK2
- IκB kinases α, β, γ, ε
- insulin receptor
- Integrin-linked kinase
- IRAK-1
- Ire1
- JAK1
- c-Kit mutant
- KSR
- Lkb1
- MEK
- MEKK1 and MEKK3
- Mik1
- MLK3
- MOK, MAK, MRK
- c-Mos
- NIK
- Nucleophosmin-Anaplastic Lymphoma Kinase
- PDK1
- Pim-1
- Pik1
- PKCλ
- Plk1
- pp60v-src, c-src
- src related tyrosine kinases: yes, fps, fes, fgr, and lck
- Raf-1, B-Raf, Ste11
- RET/PTC1
- Ron
- Slt2
- SSTK

- TAK1
- TBK1
- trkB
- VEGFR2
- Wee1, Swe1

Others:

- Annexin II
- ANP receptor
- Apaf-1
- apoB
- Bid
- calcineurin (Cna2; catalytic subunit)
- calmodulin
- calponin
- CFTR (nascent polypeptide)
- CIC-2 chloride channel
- Ctf13/Skp1 component of CBF3
- cytoskelettal proteins: actin, tubulin, myosin
- Dengue virus protein E
- DNA polymerase α
- eNOS, nNOS (?)
- ether-a-gogo-related cardiac potassium channel
- free $\beta\gamma$ subunit of G protein
- $G\alpha_0$, $G\alpha_{12}$
- GERp95 (= Argonaute-2)
- glutathione S-transferase subunit 3 (KS type)
- HDAC6
- Histones H1, H2A, H2B, H3 and H4
- knob complexes (in the membrane of Plasmodium-infected erythrocytes)
- macromolecular aminoacyl-tRNA synthetase complex
- Macrophage scavenger receptor
- Mdm2
- MMP2
- MRE11/Rad50/NBS1 (MRN) complex
- MTG8
- MUC1
- Na$^+$-K$^+$-Cl$^-$ cotransporter 1
- NB-LRR proteins RPM1 and RPS2
- Neuropeptide Y
- Nod1
- N-WASP
- P450 CYP2E1
- P2X$_7$ purinergic receptor
- PB2 subunit of influenza RNA pol.
- perilipin
- Mg^{2+}-dependent phosphatidate phosphohydrolase
- prolactin receptor

- proteasome
- Rab-αGDI
- Ral-binding protein 1
- reovirus protein σ1
- reverse transcriptase of hepatitis B virus
- ribosomal proteins S3 and S6
- R-protein I-2
- SIR2 (SIR2RP1 in Leishmania)
- SKP2 complexes
- SMYD3
- DNA helicase Ssl2
- survivin
- SV40 large T-antigen
- α-synuclein
- Tau protein
- telomerase
- thiopurine S-methyltransferase
- thrombin receptor (PAR-1)
- TLR4/MD-2 complex
- Vaccinia core protein 4a
- misfolded VHL
- Vimentin

A.2. Hsp70 C-terminal subdomain alignment (*C. elegans* residues 542-640)

```
                1        10        20        30        40        50
HSP70_THEPA-0   KLENYCYSMKNTLSEDQVKQKLGADEVD.NALNTITEALKWVETNQLAEHDEFED
HSP70_THEAN-0   KLENYCYSMKNTLSEDQVKQKLGADEVD.SALSTITDALKWVEANQLAEKEDEYED
HSP70_PYRSA-0   SLENYAYNIRNTVRDEKLKEKIQEEDKK.SIEEKVKEVLEFIETNEDLEKEEYEE
HSP7C_DICDI-0   KLENYAFTVKNSIKDEKVAAKISDSDKS.TIESETESVLKWLESNQTAEKDEYED
HSP70_PLACB-0   SLENYCYGVKSSLEDQKIKEKLQPSEIE.TCMKSITTILEWLEKNQLASKEEYES
HSP70_PLAFA-0   SLENYCYGVKSSLEDQKIKEKLQPAEIE.TCMKTITTILEWLEKNQLAGKDEYEA
HSP70_CHLRE-0   SLENYAYNMRNTIREDKVASQLSASDKE.SMEKALTAAMDWLEANQMAEVEEFEH
HSP70_DAUCA-0   ALENYAYNMRNTIKDDKIPGKLDAGDKE.KIETAVNEAIEWLEKNQLAEVDELED
HSP7E_SPIOL-0   ALENYAYNMRNIVKDEKIGAKLSEADKK.KIEEAIDASIQWLDGNQLAEADEFDD
HSP70_LUPPO-0   ALENYAYSMRNTITDDKITSKLPTEDKK.KIEDTVDGAISRLDGNQLPEVEEFED
HSP70_SOYBN-0   SLENYAYNMRNTIKDEKIGGKLSPDEKQ.KIEKAVEDAIQWLEGNQMAEVDEFED
HSP71_SOLLC-0   SLENYAYNMRNTVKDEKIGSKLSDDKK.KIEDAVDGAISWLSNQLAEVDEFED
HSP70_MAIZE-0   ALENYAYNMRNTIKDDKIASKLPAEDKK.KIEDAVDGAISWLDNQLAEVEEFED
HSP7C_PETHY-0   ALENYAYNMRNTIKDDKINSQLSAADKK.RIEDAIDEAIKWLDNNQLAEADEFED
HSP73_ARATH-0   ALENYAYNMRNTIKDEKIGEKLAGDDKK.KIEDSIEAAIEWLEANQLAECDEFED
HSP71_ARATH-0   ALENYAYNMRNTIQDEKIGEKLPAADKK.KIEDSIEQAIQWLEGNQLAEADEFED
HSP72_ARATH-0   ALENYAYNMRNTIRDEKIGEKLPAADKK.KVEDSIFEAIQWLDGNQLGEADEFED
HSP70_LEIBR-0   GLENYAYSMKNTVSDTNVSGKLEESDRS.ALNLAIDTALEWLNSNQEASKEEYEH
HSP70_LEIDO-0   GLENYAYSMKNTLGDSNVSGKLDDSDKA.TLNKEIDVTLEWLSSNQEATKEEYEH
HSP70_LEIAM-0   GLENYAYSMKNTLGDSNVSGKLDDTDKS.TLNKEIEAALEWLSSNQEATKEEYEH
HSP70_TRYCR-0   GLENYAFSMKNAVNDPNVAGKIEEADKK.TITSAVEEALEWLNNNQEASKEEYEH
HSP74_TRYBB-0   GLENYAFSMKNTINDPNVAGKLDDADKN.AVITAVEEALRWLNDNQEASLEEYNH
HSP72_CANAL-0   QLESYAYSLKNTLGEEQFKSKLDASEIE.EVTKAADETIAWLDSNQTATQEEFAD
HSP72_YEAST-0   QLESIAYSLKNTISEAG..DKLEQADKD.AVTKKAEETIAWLDSNTTATKEEFDD
HSP71_YEAST-0   QLESIAYSLKNTISEAG..DKLEQADKD.TVTKKAEETISWLDSNTTASKEEFDD
HSP71_PICAN-0   GLESYAYSLKQTASEKQFEEKVDASKRE.SLNKAIEETISWLDNNQSATTDEYED
HSP72_PICAN-0   GLESYAYSLKQTTSEKQFEEKVEASKRE.AFTKACDDTIAWLDENQTATAEEYDD
HSP74_YEAST-0   QLESYAFTLKNSVSENNFKEKVGEEDAR.KLEAAAQDAINWLDASQAASTEEYKE
HSP73_YEAST-0   QLESYAFTLKNTINEASFKEKVGEDDAK.RLETASQETIDWLDASQAASTDEYKD
HSP71_CANAL-0   QLESYAYSLKNTINDGEMKDKIGADDKE.KLTKAIDETISWLDASQAASTEEYED
HSP70_ACHKL-0   GLENYAYNLRNTLNDEKLQGKIDESDKK.KLDEKVTDIINWLDHNQSAEKEEFEA
HSP70_CLAHE-0   GLESYAYSLKNTVSDPKVEEKLSAEDKE.TLTGAIDKTVAWIDENQTATKEEYEA
HSP70_NEUCR-0   GLESYAYSLRNTLSDSKVDEKLDAADKE.KLKSEIDKIVAWLDENQQATREEYEA
HSP70_ALTAL-0   ALESYAYSLRNTLSDSKVDEKLDAGDKQ.KLTAEIDKTVQWLDNQTATKDEYES
HSP70_PARBR-0   GLESYAYSLRNTISDSKVDEKLDASDKE.KLKTEIDKTVSWLDENQTATKEEFEA
HSP70_AJECA-0   GLESYAYSLRNSLRHSKVDEKLEAGDKE.KLKSEIDKTVQWLDENQTATKEEYES
HSP70_TRIRU-0   GLESYAYSLKNTLSDSKVDEKLDAADKE.KLKSEIDKVVAWLDNQTATKEEYES
HSP70_BLAEM-0   GLESYAYNLRNTLNDDKVAGKMDAADKE.TLNKAIDETISWLDGNQEGAKDEYEH
HSP71_PUCGR-0   ALESYAYNLRNSLNDEKLADKFDAADKK.KLEDAVNSTISWLDHNTTAAKEEYEE
HSP71_SCHPO-0   HLESYAYSLRNSLDDPNLKDKVDASDKE.AIDKAVKETIEWLDHNTTAAKDEYED
HSP72_SCHPO-0   HLESYAYSLRNSLDDPNLKDKVDASDKE.TVDKAVKETIEWLDSNTTAAKDEFEA
HSP76_PIG-0     SLEAYVFHVKGSLHEESLRDKIPEEDRC.KVQDKCQEVLTWLEHNQLAEKEEYEH
HSP76_SAGOE-0   SLETHVFHVKGSLQEESLRDKIPKEDRH.KVQDKCQEVLAWLEHNQLADKEEYEH
HSP76_HUMAN-0   SLEAHVFHVKGSLQEESLRDKIPEEDRR.KMQDKCQEVLAWLEHNQLAEKEEYEH
HSP70_CERCA-0   NLESYVLAVKQAWT..TLVDKLSEREKS.EVTKACDDTIKWLDATRLADKEEYED
HSP74_PARLI-0   QLEGYAFNLKSAVDDAAAQSKLSPGDKE.TVTKAVNDVLQWLDSNSLADKEEFTY
HSP7A_DROME-0   QLESYCFQLRSTLDDEHLSSRFSPADRE.TIQQRSSETIAWLDANQLAERQEFEH
HSP70_SCHMA-0   SLESYVYTMKQQVEGE.LKEKIPESDRQ.VIISKCEDTISWLDVHQSAEKHEYES
HSP70_SCHJA-0   SLESYVYSMKQSVEGDEMKDKISESETIRWMDNNQLAAEKEEFEE
HSP74_ANOAL-0   QLEAYCFNLKQSLDGEG.ASKLSDADRK.TVQDRCEETLRWIDGNTMADKEEFEH
HSP71_ANOAL-0   QLEAYCFNLKQSLDGEG.SSKLSEADRR.TVQDRCDETLRWIDGNTMAEKEEYEH
HSP72_ANOAL-0   QLEAYCFNLKQSLDGEG.SSKLSDADRR.TVQDRCDETLRWIDGNTMAEKEEYEH
HSP70_DROME-0   ALESYVFNVKQAVEQAP.AGKLDEADKN.SVLDKCNDTIRWLDSNTTAEKEEFDH
HSP71_DROME-0   ALESYVFNVKQAVEQAP.AGKLDEADKN.SVLDKCNDTIRWLDSNTTAEKEEFDH
HSP71_DROSI-0   ALESYVFNVKQAV.DQAPAGKLDEADKN.SVLDKCNDTIRWLDSNTTAEKEEFDH
HSP73_DROME-0   ALESYVFNVKQSVEQAP.AGKLDEADKN.SVLDKCNETIRWLDSNTTAEKEEFDH
HSP72_DROME-0   ALESYVFNVKQSVEQAP.AGKLDEADKN.SVLDKCNETIRWLDSNTTAEKEEFDH
HSP74_DROME-0   ALESYVFNVKQSVEQAP.AGKLDEADKN.SVLDKCNETIRWLDSNTTAEKEEFDH
HSP75_DROME-0   ALESYVFNVKQSVEQAP.AGKLDEADKN.SVLDKCNETIRWLDSNTTAEKEEFDH
HSP72_DROSI-0   ALESYVFNVKQSVEQAP.AGKLDEADKN.SVLDKCNDTIRWLDSNTTAEKEEFDH
MAG29_DERFA-0   SLEGYAFQMKATLDEEAIKSKVSEEDRK.KILDKVDEVLKWLDANALAEKDEFEH
HSP71_RAT-0     ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDSNTLAEKEEFVH
HS70A_MOUSE-0   ALESYAFNMKSAVEDEGLKGKLSEADKK.KVLDKCQEVISWLDSNTLADKEEFVH
HS70B_MOUSE-0   ALESYAFNMKSAVEDEGLKGKLSEADKK.KVLDKCQEVISWLDSNTLADKEEFVH
HS70A_PIG-0     ALESYAFNMKSVVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HS70A_BOVIN-0   ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HS70B_BOVIN-0   ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HSP71_CERAE-0   ALESYALNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HSP71_HUMAN-0   ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HS70B_BOSMU-0   ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HS70B_PIG-0     ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HSP71_CANFA-0   ALESYAFNMKSAVEDEGLKGKISEADKK.KVLDKCQEVISWLDANTLAEKDEFEH
HS70L_HUMAN-0   ALESYAFNMKSVVSDEGLKGKISESDKK.KILDKCNELLSWLEVNQLAEKEEFDH
HS70L_RAT-0     ALESYAFNMKSAVGDEGLKDKISESDKK.KILDKCSEVLSWLEANQLAEKEEFDH
HS70L_MOUSE-0   ALESYAFNMKSAVGDEGLKDKISESDKK.KILDKCNELLSWLEANQLAEKEEFDH
HSP70_XENLA-0   ALESYAFNLKSMVEDENVKGKISDEDKR.TISEKCTQVISWLENNQLAEKEEYAF
HSP70_ECHGR-0   GLESYAFSMKSTVEDEKVKEKIGESDRR.RIMEKCEETVKWLDGNQQAEKEEYEH
HSP7C_ORYLA-0   GLESYAFNMKSTVEDEKLAGKISDEDKQ.KILDKCNEVISWLDKNQTAERDEYVH
HSP7D_MANSE-0   ALESYCFNMKSTMEDEKLKDKISDSDKQ.TILDKCNDTIKWLDSNQLADKEEYEH
HSP7D_DROME-0   GLESYCFNMKATLDEDNLKTKISDSDRT.TILDKCNETIKWLDANQLADKEEYEH
HSP7A_CAEEL-0   GLESYAFNLKQTIEDEKLKDKISPEDKK.KIEDKCDEILKWLDSNQTAEKEEFEH
HSP70_HYDMA-0   SLESYCYNMKQTVEDEKVKGKISEADKK.TIIKKCNETVEWVDKNQTAEKDQYEH
HSP70_BRUMA-0   ALESYAFNMKQTIEDEKLKDKISEADKK.KIQEKCDETVRWLDGNQTAEKDEFEH
HSP7C_BOVIN-0   SLKSYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIINWLDKNQTAEKEEFEH
HSP7C_MOUSE-0   SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIISWLDKNQTAEKEEFEH
HSP7C_CRIGR-0   SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIISWLDKNQTAEKEEFEH
HSP7C_RAT-0     SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIISWLDKNQTAEKEEFEH
HSP7C_PONPY-0   SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIINWLDKNQTAEKEEFEH
HSP7C_SAGOE-0   SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIINWLDKNQTAEKEEFEH
HSP7C_HUMAN-0   SLESYAFNMKATVEDEKLQGKINDEDKQ.KILDKCNEIINWLDKNQTAEKEEFEH
HSP7C_ICTPU-0   GLESYAFNMKSTVEDEKLKGKINDEDKQ.KILDKCNEVISWLDKNQTAEKDEYEH
HSP70_ONCMY-0   SLESYAFNMKSTVEDEKLQGKISDEDKT.KILEKCNEVIGWLDKNQTAEKEEYEH
HSP7C_BRARE-0   GLESYAFNMKSTVEDEKLKGKISDEDKQ.KILDKCNEVIGWLDKNQTAEREEFEH
HSP70_ONCTS-0   SLESYAFNMKSSVEDDNMKGKISQEDKK.KVVDRCDQTISWLENNQLGDKEEYEH
HSP70_PLEWA-0   TLESIAFNMKSTVEGDNLKDKISEDDRK.KIVDKCNQTISWMENNQMAEKEEYEH
```

```
                   60        70        80                  90
HSP70_THEPA-0   KLKHVEGVCNPLVTKLYQSGGAP....GAGPDMG...........AGFPGGA.
HSP70_THEAN-0   KLKHVEGVCNPLVTKLYQS.......GAAP..G.....GPDMS.GGFPGGA.
HSP70_PYRSA-0   KEKELKNFANPIISKLYQQ.......GSVP.......D.M....GNFSTGQ.
HSP7C_DICDI-0   KMKALEAVVNPIMSKLYQE.......GGMP.QG.....GGMPGGM.
HSP70_PLACB-0   KQKEAESVCAPIMSKIYQDVGGAA..GGMP..G....G.MP...GGMPGGM.
HSP70_PLAFA-0   KQKEAESVCAPIMSKIYQDAAGAA..GGMP..G....G.MP...GGMPGGM.
HSP70_CHLRE-0   HLKELEGLCNPIIITRLYQGGAGA..GGMP..........GGGAGAG.
HSP70_DAUCA-0   KLKELEGLCNPIIARLYQGRGDVP..IGGP..........GDMPGGGA.
HSP7E_SPIOL-0   KMKELESICNPIIAKMYQG.......GGDM..G.......GGMEDEG.
HSP70_LUPPO-0   KMKELESLCNPIIAKMYQGVLAQMVLVLLIM.......ADAPTGS.
HSP70_SOYBN-0   KQKELEGICNPIIAKMYQG.......AAGP..G.......GDVPMGA.
HSP71_SOLLC-0   KMKELEGICNPIIAKMYQG.......AGGD.........AGVPMDD.
HSP70_MAIZE-0   KMKELEGICNPIIAKMYXGEGAGMGAAAGM........D.....E.
HSP7C_PETHY-0   KMKELESICNPIIAKMYQ.......GGAG..GATMDED.GPSV..GGSAGSQ.
HSP73_ARATH-0   KMKELESICNPIIAKMYQGGE....AGGP..AA.....GGMDEDV.
HSP71_ARATH-0   KMKELESICNPIIAKMYQGAGGE....AGGP..GA....SGMDDDA.
HSP72_ARATH-0   KMKELESVCNPIIAKMYQGGAGGE...AGGP..GA....SGMDEDE.
HSP70_LEIBR-0   RQKELESTCNPIMTKMYQSM......GGGA.........GGMPGGGMP
HSP70_LEIDO-0   KQKELESVCNPIMTKMYQSMGGAG..GGMP..G....G.MPD.MSGMSGGA.
HSP70_LEIAM-0   KQKELENVCNPIMTKMYQSMGGGAGGMAGMP..DM........SSMSGAR.
HSP70_TRYCR-0   RQKELENLCTPIMTNMYQGMAG....AGMP.........GGMPGGM.
HSP74_TRYBB-0   RQKELEGVCAPILSKMYQGMGGGDGP.GGMP........EGMPGGMPGGMPGGM.
HSP72_CANAL-0   QQKELESKANPIMTKAYQAGATPSGAAGAAP.........GGFPGGA.
HSP72_YEAST-0   QLKELQEVANPIMSKLYQA.......GGAP..E....GAAP...GGFPGGA.
HSP71_YEAST-0   KLKELQDIANPIMSKLYQAGGAPGGAAGGAP.........GGFPGGA.
HSP71_PICAN-0   KRKELEGIANDALKDLYAA.......GGVP..G.....GAAP...GGFPGAG.
HSP72_PICAN-0   KRKELEQAGNEVLKDLYAE.......GGVP..G....G.AP...GGFPGAG.
HSP74_YEAST-0   RQKELEGVANPIMSKFYGAA......GGAP..GA.....GPVPGAG.
HSP73_YEAST-0   RQKELEGIANPIMTKFYGAGAGAGPGAGESG..G......FP....GSMPNSG.
HSP71_CANAL-0   KRKELESVANPIISGAYGAAGGAPGGAGGFP..G....AGGFP....GGAPGAG.
HSP70_ACHKL-0   KQKELEGIANPIMQKMYAAA......GGAP.........GGMPGGM.
HSP70_CLAHE-0   EQKQLESVANPVMMKIYGAE......GGAP.........GGMPGQG.
HSP70_NEUCR-0   RQKELEAIANPIMMKFYGAGGAPGGMPGAAP.........GGMPGGG.
HSP70_ALTAL-0   QQKELEGVANPIMMKFYGAGGE....GGMP.........GGMPGGGM
HSP70_PARBR-0   QQKELESVANPIMMKFYGA.......GGAP..GAGFPGAGGP...GGFPGAG.
HSP70_AJECA-0   QQKELEAVANPIMMKFYAGGEGAP..GGFP..G....AGGP...GGFPGGP.
HSP70_TRIRU-0   QQKELEGVANPIMMKFYGAG.GEGGAPGGFPGAG......AGGP...GGFPGAG.
HSP70_BLAEM-0   KQKELEGVANPIMTKLYSAA......GGAP.........GGMPGGFD
HSP71_PUCGR-0   HQKELEAVANPIMQKLYAGAGGAP..GGAP.........GGFPGGA.
HSP71_SCHPO-0   KQKELEGVANPIMAKIYQA.......GGAP..G....G.AP...GGMPGGA.
HSP72_SCHPO-0   KQKELESVANPIMAKIYQAGGAPGGMPGAAP.........GAAPGAA.
HSP76_PIG-0     QRKELEQICRPIFSRLYGA.......PGIP..G.......GSSCGAQ.
HSP76_SAGOE-0   QRKELEQICRPIFSRLYGG.......PGVP..G.......GSSCGAQ.
HSP76_HUMAN-0   QRKELEQICRPIFSRLYGG.......PGVP.........GSSCGTQ.
HSP70_CERCA-0   KMNTLTKLCTPIMTKLHS.......GGGA..G........Q...GASCGQQ.
HSP74_PARLI-0   KLEELQKTCSPIMAKMHAGTGGPR..GGGP.........QGFPSGG.
HSP7A_DROME-0   KQQELERICSPIITRLYQGA......GMAP.........PPTAGGS.
HSP70_SCHMA-0   KREELEKVCAPIITKVYQA.......GGMP.........GGMHEAS.
HSP70_SCHJA-0   KKSELEKVCMPIITAMNRAGGGVP..SGMP.........GGMPGAG.
HSP74_ANOAL-0   KMQELTKACSPIMTKLHQQA......AGGP..S.......PSSCAQQ.
HSP71_ANOAL-0   QMQELSRVCSPIMTKLHQQA......AGGP..Q.......PTSCGQQ.
HSP72_ANOAL-0   QMQELSRVCSPIMTKLHQQA......AGGP..Q.......PTSCGQQ.
HSP70_DROME-0   KLEELTRHCSPIMTKMHQQGAGAG..AGGP.........GANCGQQ.
HSP71_DROME-0   KLEELTRHCSPIMTKMHQQGAGAG..AGGP.........GANCGQQ.
HSP71_DROSI-0   KLEELTRHCSPIMTKMHQQGAGAG..AGAG.........GPGANCGQQ.
HSP73_DROME-0   KMEELTRHCSPIMTKMHQQGAGA...AGGP.........GANCGQQ.
HSP72_DROME-0   KMEELTRHCSPIMTKMHQQGAGA...AGGP.........GANCGQQ.
HSP74_DROME-0   KMEELTRHCSPIMTKMHQQGAGA...AGGP.........GANCGQQ.
HSP75_DROME-0   KMEELTRHCSPIMTKMHQQGAGA...AGGP.........GANCGQQ.
HSP72_DROSI-0   KMEELTRHCSPIMTKMHQQGAGAGA.AGGP.........GANCGQQ.
MAG29_DERFA-0   QRKELESVCNPIITKLYQQAGGAGA..GGMP.......GGFP....GGFPGTDG
HSP71_RAT-0     KREELERVCNPIISGLYQG.......AGAP..G......A...GGFGAQA.
HS70A_MOUSE-0   KREELERVCSPIISGLYQG.......AGAP..GA.....GGFGAQA.
HS70B_MOUSE-0   KREELERVCSPIISGLYQG.......AGAP..GA.....GGFGAQA.
HS70A_PIG-0     KRKELEQVCNPIISGLYQG.......AGGP..G....P.GGFGAPD.
HS70A_BOVIN-0   KRKELEQVCNPIISRLYQG.......AGGP..GA.....GGFGAQG.
HS70B_BOVIN-0   KRKELEQVCNPIISRLYQG.......AGGP..GA.....GGFGAQG.
HSP71_CERAE-0   KRKELEQVCNPIISGLYQG.......GGGP..G....P.GGFGAQG.
HSP71_HUMAN-0   KRKELEQVCNPIISGLYQG.......AGGP..G....P.GGFGAQG.
HS70B_BOSMU-0   KRKELEQVCNPIISRLYQG.......AGGP..GA.....GGFGAQA.
HS70B_PIG-0     KRKELEQVCNPIISGLYQG.......AGGP..GA.....GGFGAQA.
HSP71_CANFA-0   KRKELEQVCNPIITGLYQG.......AGGP..GA.....GGFGAQA.
HS70L_HUMAN-0   KRKELEQMCNPIITKLYQ.......GGCT..G....P.ACGTGYV.
HS70L_RAT-0     KRKELENMCNPIITKLYQ.......SGCT..G....P.TCAPGYT.
HS70L_MOUSE-0   KRKELENMCNPIITKLYQ.......SGCT..G....P.TCTPGYT.
HSP70_XENLA-0   QQKDLEKVCQPIITKLYQGGVP...GGVP.........GGMPGSS.
HSP70_ECHGR-0   RQKELESVCNPIIAKMYQEAGGVGGIPGGIP..G....GGMP...GGTPGGGI
HSP7C_ORYLA-0   QQKELEKVCNPIITKLYQSA......GGMP.........GGCQREC.
HSP7D_MANSE-0   KQKELEGICNPIITKLYQGAGGMP..GGMP.........GGMPGFPG
HSP7D_DROME-0   RQKELEGVCNPIITKLYQGAGFPP..GGMP..G....GP.GGMPGAA.
HSP7A_CAEEL-0   QQKDLEGLANPIISKLYQSA......GGAP.........GAAPGGA.
HSP70_HYDMA-0   KQKELEKVCNPIITKLYQAG......GGMP..G....G.MP...GGMPGGM.
HSP70_BRUMA-0   RQKELESVCNPIITKLYQSA......GGMP..G....G.MP...GGMPGGA.
HSP7C_BOVIN-0   QQKELEKVCNPIITKLYQSAGGGMP...GGMP.......GGMP...GGFPGGG.
HSP7C_MOUSE-0   QQKELEKVCNPIITKLYQSA......GGMP.......GGMP...GGFPGGG.
HSP7C_CRIGR-0   QQKELEKVCNPIITKLYQSA......GGMP..G....G.MP...GGFPGGG.
HSP7C_RAT-0     QQKELEKVCNPIITKLYQSA......GGMP..G....G.MP...GGFPGGG.
HSP7C_PONPY-0   QQKELEKVCNPIITKLYQSA......GGMP..G....G.MP...GGFPGGG.
HSP7C_SAGOE-0   QQKELEKVCNPIITKLYQSA......GGMP..G....G.MP...GGFPGGG.
HSP7C_HUMAN-0   QQKELEKVCNPIITKLYQSA......GGMP..G....G.MP...GGFPGGG.
HSP7C_ICTPU-0   QQKDLEKVCNPIITKLYQSDGGMP..GGMP.........DGMPGGFQ
HSP70_ONCMY-0   HQKELEKVCNPIITKLYQSA......GGMP..G....GGMPEGMAGGFPGAG.
HSP7C_BRARE-0   QQKELEKVCNPIITKLYQSAGGGMP...GGMP.......EGMP...GGFPGAG.
HSP70_ONCTS-0   QLKELEKVCQPIITKLYQQ.......GGMP.........TGCCGDQ.
HSP70_PLEWA-0   QQKELEKVCNSIITKLYQ........GGMP.........GGMPSGSS
```

```
                                          100
HSP70_THEPA-0    ....PPP......SS.SSGPTV..EEVD
HSP70_THEAN-0    ....APP......PQ.SSGPTV..EEVD
HSP70_PYRSA-0    ....NEQDDN...A..NMGPKI..EEVG
HSP7C_DICDI-0    ...........S..NDSPKSSNNKVD
HSP70_PLACB-0    ....PGG........MPGGGM..|....
HSP70_PLAFA-0    ....PSGMPG...G..MNFP...|....
HSP70_CHLRE-0    ....AAPSGG...S..GAGPKI..EEVD
HSP70_DAUCA-0    ....YGGSRG...SS.GAGPKI..EEVD
HSP7E_SPIOL-0    ....PTS..G...G..GAGPKI..EECR
HSP70_LUPPO-0    ...........G..GAGPKI..GEVD
HSP70_SOYBN-0    ..DMPAA......GAGPKI..EEVD
HSP71_SOLLC-0    .DAPPSG......GS.SAGPKI..EEVD
HSP70_MAIZE-0    ....DAPSGG....S.GAGPKI..EEVD
HSP7C_PETHY-0    ...........T..GAGPKI..EEVD
HSP73_ARATH-0    ....PPS......AG.GAGPKI..EEVD
HSP71_ARATH-0    ....PPA......SG.GAGPKI..EEVD
HSP72_ARATH-0    ....APPASG.....GAGPKI..EEVD
HSP70_LEIBR-0    DMSGMGGGQGPAAGA.SSGPKV..EEVD
HSP70_LEIDO-0    ....GPAGG...A..SSGPKV..EEVD
HSP70_LEIAM-0    ....PAG......GA.SSGPKV..EEVD
HSP70_TRYCR-0    ...PGGMPG...G..MPGGMP..|....
HSP74_TRYBB-0    ...GGGMGGA..AA.SSGPKV..EEVD
HSP72_CANAL-0    ...APE.......PS.NDGPTV..EEVD
HSP72_YEAST-0    ....PPA......PE.AEGPTV..EEVD
HSP71_YEAST-0    ....PPA......PE.AEGPTV..EEVD
HSP71_PICAN-0    .....GAAPG...A..DQGPSV..EEVD
HSP72_PICAN-0    ...GAPSTE...E..TQGPTV..EEVD
HSP74_YEAST-0    ..AGPTG......AP.DNGPTV..EEVD
HSP73_YEAST-0    ....ATG..G...GE.DTGPTV..EEVD
HSP71_CANAL-0    ..GPGGATG...GE.SSGPTV..EEVD
HSP70_ACHKL-0    ..PDMGGAGAPPPASHAQGPKI..EEVD
HSP70_CLAHE-0    AGAPPPG......AG.DDGPTV..EEVD
HSP70_NEUCR-0    ....PGS......ND.NEGPTV..EEVD
HSP70_ALTAL-0    PGGAPGGAAG.....DDGPTV..EEVD
HSP70_PARBR-0    ....VGGAHS...GG.DDGPTV..EEVD
HSP70_AJECA-0    ....GAGHASG..GG.DDGPTV..EEVD
HSP70_TRIRU-0    .....AGGAA...AH.DDGPTV..EEVD
HSP70_BLAEM-0    PSGAPPPA.....AD.TTGPTI..EEVD
HSP71_PUCGR-0    ....PGGFPGGAPAG.EDGPSV..EEVD
HSP71_SCHPO-0    ....PGGAPG...GA.DNGPEV..EEVD
HSP72_SCHPO-0    ....PGAAPG...G..DNGPEV..EEVD
HSP76_PIG-0      ...ARQG......AP.STGPVI..EEVD
HSP76_SAGOE-0    ..ARQG.......DR.STGPII..EEVD
HSP76_HUMAN-0    .....ARQG....DP.STGPII..EEVD
HSP70_CERCA-0    ...AGGFNG....G..HTGPTV..EEVD
HSP74_PARLI-0    ..............AG.GPTV..EEVD
HSP7A_DROME-0    ...NPGATG......GSGPTI..EEVD
HSP70_SCHMA-0    ....GAG......GGSGKGPTI..EEVD
HSP70_SCHJA-0    ....GGG........GKGPTI..EEVD
HSP74_ANOAL-0    ...AGGF.....GG.RTGPTV..EEVD
HSP71_ANOAL-0    ...AGGF.....GG.RTGPTV..EEVD
HSP72_ANOAL-0    ...AGGF.....GG.RTGPTV..EEVD
HSP70_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP71_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP71_DROSI-0    .....AGG....FGGYSGPTV..EEVD
HSP73_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP72_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP74_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP75_DROME-0    ...AGGF.....GG.YSGPTV..EEVD
HSP72_DROSI-0    ...AGGF.....GG.YSGPTV..EEVD
MAG29_DERFA-0    SGGGAAGGDG...G..KSGPTI..EEVD
HSP71_RAT-0      ....PKG......GS.GSGPTI..EEVD
HS70A_MOUSE-0    ....PKG......AS.GSGPTI..EEVD
HS70B_MOUSE-0    ...PPKG......AS.GSGPTI..EEVD
HS70A_PIG-0      ....LKG......GS.GSGPTI..EEVD
HS70A_BOVIN-0    ....PKG......GS.GSGPTI..EEVD
HS70B_BOVIN-0    ....PKG......GS.GSGPTI..EEVD
HSP71_CERAE-0    ....PKG..G...S..GSGPTI..EEVD
HSP71_HUMAN-0    ....PKG..G...S..GSGPTI..EEVD
HS70B_BOSMU-0    ....PKG......GS.GPGPTI..EEVD
HS70B_PIG-0      ....PKG......GS.GSGPTI..EEVD
HSP71_CANFA-0    ....PKG......GS.GSGPTI..EEVD
HS70L_HUMAN-0    ....PGR......P..ATGPTI..EEVD
HS70L_RAT-0      ....PGR......A..ATGPTI..EEVD
HS70L_MOUSE-0    ....PGR......A..ATGPTI..EEVD
HSP70_XENLA-0    ....CGAQARQ..GG.NSGPTI..EEVD
HSP70_ECHGR-0    PAGMAGGMSGDPSSG.GRGPTI..EEVD
HSP7C_ORYLA-0    ..........QE.VSLELVVLLAVA
HSP7D_MANSE-0    GAPGAGGAAP...GG.GAGPTI..EEVD
HSP7D_DROME-0    ...GAAGAAG...AG.GAGPTI..EEVD
HSP7A_CAEEL-0    ....AGG......AG.GAGPTI..EEVD
HSP70_HYDMA-0    ....PGSGSK...AS.SGGPTI..EEVD
HSP70_BRUMA-0    ....PGA..G...ST.GGGPTI..EEVD
HSP7C_BOVIN-0    ..APPSG......GA.SSGPTI..EEVD
HSP7C_MOUSE-0    ..APPSG......GA.SSGPTI..EEVD
HSP7C_CRIGR-0    ..APPSG..G...A..SSGPTI..EEVD
HSP7C_RAT-0      ..APPSG..G...A..SSGPTI..EEVD
HSP7C_PONPY-0    ..APPSG......GA.SSGPTI..EEVD
HSP7C_SAGOE-0    ..APPSG..G...A..SSGPTI..EEVD
HSP7C_HUMAN-0    ..APPSG..G...A..SSGPTI..EEVD
HSP7C_ICTPU-0    ELGAAPG......GG.SSGPTI..EEVD
HSP70_ONCMY-0    .GAAPGG......GG.SSGPTI..EEVD
HSP7C_BRARE-0    ....AAPGGG......SSGPTI..EEVD
HSP70_ONCTS-0    ....ARTSSG...DS.SQGPTI..EEID
HSP70_PLEWA-0    GAQARQG......SS.STGPTI..EEVD
```

A.3. Sequence alignment of Hsp70/Hsp90 interacting TPR domains

A.3.1. Hop TPR1 domain

```
                 1        10        20        30        40        50
gnl|tr|A0BUJ1  AQQFKDLGNQAFKENKFEDAAKFYSQAIELNPNDHILYSNRSGAYASLSKYEDAL
gnl|tr|A0E7Z9  AQQFKDLGNQAFKENKFEEAAKFYSQAIELNPNDHILYSNRSGSYASLSKYQEAL
gnl|tr|A1CG55  ADALKAEGNKAFSAKDYPTAIDKFTQAIQLDPSNYILYSNRSAVYAAQSEYQKAL
gnl|tr|A1D965  ADALKAEGNKAFSAKDYPTAIEKFTQAIELEPSNHILYSNRSAVYAAQSDYQKAL
gnl|sp|O35814  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPQNHVLYSNRSAAYAKKGDYQKAY
gnl|sp|O54981  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPQNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|O61650  VNELKEKGNQALSAEKFDEAVAAYTEAIALDDQNHVLYSNRSAAFAKAGKFQEAL
gnl|sp|P31948  VNELKEKGNKALSVGNIDDALQCYSEAIKLDPHNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|P90553  ATELKNKGNEEFSAGRYVEAVNYFSKAIQLDEQNSVLYSNRSACFAAMQKYKDAL
gnl|tr|P90647  ALEEKNKGNAAMSAGDFKAAVEHYTNAIQHDPQNHVLYSNRSAAYASLKDYDQAL
gnl|tr|Q017S6  ADEHKARGNAHFAKHEYAAAIDAFTSAIECDGTNHVFWSNRSAAYSGAEKWNEAL
gnl|tr|Q0CWW9  ADALKAEGNKAFSAKDYSTAVDKFTQAIAIEPENHILYSNRSAVYSAQGEYQKAL
gnl|tr|Q0JBE4  ADEAKAKGNAAFSAGRFEEAAAHFTDAIALAPDNHVLYSNRSAAYASLHRYPEAL
gnl|tr|Q0UEH1  ADALKAEGNKLFAEKKFTESIEKFSQAIELDPSNHVLYSNRSGAYASLKDWDKAL
gnl|tr|Q1DYZ0  ADALKAEGNKAFAAKDFNLAVEKFSAAIELDSSNHVLYSNRSGAYASLKNFDKAL
gnl|tr|Q22RN3  ATAFKNEGNKAFQENRFQDAIDAFTKAIEINPNDHVFYSNRSGAYASLNKLDEAL
gnl|tr|Q27U54  VNELKEKGNQALNAEKYQEAIEAYTEAILLDDKNHVLYSNRSAAYAKAGKFSEAL
gnl|tr|Q29NX2  VNELKEKGNTALNAEKFDEAVAAYTEAIALDSQNHVLFSNRSAAYAKAGKFAEAL
gnl|tr|Q2HEP5  ADELKALGNKAIAAKDFDDAIDKFTQAIALDGSNHILYSNRSAAYASKKDWDNAL
gnl|tr|Q2U285  ADALKAEGNKAFSAKDYPTAIDKFTQAIAIEPENHILYSNRSAVYSAQSEYEKAL
gnl|tr|Q3THQ5  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPQNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|Q3ZBZ8  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPQNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|Q3ZCU9  VNELKEKGNKALSVGNIDDALQCYSEAIKLDPHNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|Q45KR0  VSALKDQGNKALSAGNIDEAVRCYTEAVALDPSNHVLFSNRSAAYAKKGNYENAL
gnl|tr|Q4DYD9  ATELKNRGNQEFSSGRYKEAAEFFSQAINLDPSNHVLYSNRSACHAALHQYPNAL
gnl|tr|Q4JHN0  ATELKNKGNEEFSAGRYVEAVNYFSKAIQLDEQNSVLYSNRSACFAAMQKYKDAL
gnl|tr|Q4MZD6  MEDLKNLGNDAFKAGRFMDAVEFFTKAIELNPDDHVLYSNRSGAYASMYMYNEAL
gnl|tr|Q4QI58  ATELKNKGNEEFSAGRYVEAVNYFSKAIQLDEQNSVLYSNRSACFAAMQKYKDAL
gnl|tr|Q4R8N7  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPHNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|Q4SG24  VSALKDQGNKALSAGNIDEAVRCYTEAVALDPTNHVLFSNRSAAYAKKGSYEKAL
gnl|tr|Q4UBI5  MLDLKNLGNEAFKAGKFKEAAEFFTKAIELNPNDHVLYSNRSGAYASMYMYNEAL
gnl|tr|Q4XY67  AQRLKELGNKCFQEGKFDDSVKYFSDAIKNDPSDHVLYSNLSGAYSSLGRFYEAL
gnl|tr|Q4YNL4  AQRFKELGNKCFQEGKFEDSVKYFSDAIKNDPSDHVLYSNLSGAYSSLGRFYEAL
gnl|tr|Q51BN3  SEAAKARGTQAFKDQKFEEAIKEYTEAIKYDETNGVLYSNRSAASLEQFEKAL
gnl|tr|Q54DA8  ATEFKNQGNAAFSSKDYNSAVKCFDQAIELDPSNHILYSNRSASLLALDKNEDAL
gnl|tr|Q561A5  AVALKAEANKAFAAKDYTTAAKLYSDAIALDPSNHVLYSNRSATKAGLKDYEGAL
gnl|tr|Q57ZX0  ATELKNKGNQEFSSGRYREAAEFFSQAINLDPSNHVLYSNRSACFASLHQYAQAL
gnl|tr|Q59YX6  ADEYKAEGNKYFAAKDFEKAIEAFTKAIEASPENHVLYSNRSGSYASLKDFNNAL
gnl|tr|Q5ARF6  ADALKAEGNKAFAAKDYPTAVEKFTQAIELDSNNHVLYSNRSAVYAAQQEYEKAL
gnl|tr|Q5CCL7  VEQLKKKGNDALVNQNFDEAIKCYTEAIALDPTNHVLYSNRSAAHAKAENYEAAL
gnl|tr|Q5RKM3  VSQLKDQGNKALSAGNLEEAIRCYTEALTLDPSNHVLFSNRSAAYAKKGDYDNAL
gnl|tr|Q5XEP2  ADEAKAKGNAAFSSGDFNSAVNHFTDAINLTPTNHVLFSNRSAAHASLNHYDEAL
gnl|sp|Q60864  VNELKEKGNKALSAGNIDDALQCYSEAIKLDPQNHVLYSNRSAAYAKKGDYQKAY
gnl|tr|Q6H660  ADEAKAKGNAAFSAGRYEEAARHFTDAIALAPGNHVLYSNRSAALASVHRYSEAL
gnl|tr|Q7RJW7  AQRLKELGNKCFQEGKFEDSVKYFSDAIKNDPSDHVLYSNLSGAYSSLGRFYEAL
gnl|tr|Q7SET2  ADELKALGNKAIAEKNFDEAIDKFTQAIALDPSNHILYSNRSAAYASKKDWDNAL
gnl|tr|Q7ZWU1  ANALKEKGNKALSAGNLDEAVKCYTEAIKLDPKNHVLYSNRSAAYAKKKEFTKAL
gnl|tr|Q84TJ2  ADEAKAKGNAAFSSGDFNSAVNHFTDAINLTPTNHVLFSNRSAAHASLNHYDEAL
gnl|tr|Q8ILC1  AQRLKELGNKCFQEGKYEEAVKYFSDAITNDPLDHVLYSNLSGAFASLGRFYEAL
gnl|tr|Q8JHF9  ANALKEKGNKALSAGNLDEAVKCYTEAIKLDPKNHVLYSNRSAAYAKKKEFTKAL
gnl|tr|Q8L724  AEEAKSKGNAAFSSGDYATAITHFTEAINLSPTNHILYSNRSASYASLHRYEEAL
gnl|tr|Q9LNB6  AEEAKAKGNAAFSSGDFTTAINHFTEAIALAPTNHVLFSNRSAAHASLHQYAEAL
gnl|tr|Q9STH1  AEEAKSKGNAAFSSGDYATAITHFTEAINLSPTNHVLYSNRSASYASLHRYEEAL
gnl|sp|Q9USI5  AEELKAKGNAAFSKKDYKTAIDYFTQAIGLDERNHILYSNRSACYASEKDYADAL
gnl|tr|Q9VPN5  VNELKEKGNQALSAEKFDEAVAAYTEAIALDDQNHVLYSNRSAAFAKAGKFQEAL
```

```
                    60          70          80          90          100
gnl|tr|A0BUJ1   ADAEKCISLNSNFAKGYQRKGLALHYLGEFEKAIDAYQQGLAKDPNN
gnl|tr|A0E7Z9   TDADKCISINPNFAKGYQRKGLALHYLGEFEKAIEAYQQGLAKDPSN
gnl|tr|A1CG55   EDANKAVEIKPDWSKGWSRKGAASRGLGDLLGAHDAYEEALKLDPGN
gnl|tr|A1D965   DDANKAIEIKPDWSKGYSRKGAACRGLGDLLGAHDAYEEALKLDPSN
gnl|sp|O35814   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|sp|O54981   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|tr|O61650   EDAEKTIQLNPTWPKGYSRKGAAAAGLNDFMKAFEAYNEGLKYDPTN
gnl|sp|P31948   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|tr|P90553   DDADKCISIKPNWAKGYVRRGAALHGMRRYDDAIAAYEKGLKVDPSN
gnl|tr|P90647   ADGEKTVELKPDWSKGYSRKGAALCYLGRYADAKAAYAAGLEVEPTN
gnl|tr|Q017S6   RDAEKTIELKPEWGKGYGRKGAALFGMQKFDEARSAYALGLEKEPDN
gnl|tr|Q0CWW9   DDANKATEIKPDWSKGWSRKGAAYRGLGDLLAAHDAYEEALKIEPGN
gnl|tr|Q0JBE4   ADAERTVALRPDWAKGCSRLGAARLGLGDAAGAVAAYEKGLALEPSN
gnl|tr|Q0UEH1   ADASKTTELKPDWAKGWGRKGTALHGEGDLVGATEAFEEALKLDPNN
gnl|tr|Q1DYZ0   EDANKTTELKPDWPKGWGRKGAAMHGLGDLVGAHDAYEEALKLDPAN
gnl|tr|Q22RN3   ADAVQCISIKPDWAKGYQRKGHAEYELGKLSEAVATFKKGLEIEPNN
gnl|tr|Q27U54   EDAEKTIALNPTWAKGYSRKGAAAAGLHDYKKAFEAYNEGLKCDPKN
gnl|tr|Q29NX2   KDAEQTIALNPTWPKGYSRKGAAAAGLHDFMKAFEAFNEGLKYDPTN
gnl|tr|Q2HEP5   SDAEKTTELKADWPKGWGRKGTALYGKGDLLGAHDAYEQGLKIDPNN
gnl|tr|Q2U285   EDANKATEIKPDWSKGWQRKGAAYRGLGDLLAAHDAYEEALKIEPGN
gnl|tr|Q3THQ5   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|tr|Q3ZBZ8   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKQTYEEGLKHEANN
gnl|tr|Q3ZCU9   EDGCKTVDLKPDWGKGYSRKAAALELLNRFEEAKRTYEEGLKHEANN
gnl|tr|Q45KR0   QDACQTIKIKPDWGKGYSRKAAALEFLGRLEDAKATYHEGLRQEPNN
gnl|tr|Q4DYD9   QDAEKCVSIKPDWVKGYVRKGAALFLRRYEEAAAAYNKGLSLDPSS
gnl|tr|Q4JHN0   DDADKCISIKPNWAKGYVRRGAALHGMRRYDDAIAAYEKGLKVDPSN
gnl|tr|Q4MZD6   ADANKCIDLKPDWPKGYSRKGLCEYKLGNPEKAKETYNMGLAYDPNN
gnl|tr|Q4QI58   DDADKCISIKPNWAKGYVRRGAALHGMRRYDDAIAAYEKGLKVDPSN
gnl|tr|Q4R8N7   EDGCKTVELKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|tr|Q4SG24   EDACETIKLKPDWGKGYSRKAAALEFLSRLGEAKATYQEGLRQEPNN
gnl|tr|Q4UBI5   ADANKCIELKPDWPKGYSRKGLCEYKLGSPEKAKETYNLGLTYDPNN
gnl|tr|Q4XY67   ETANKCISIKNDWPKGYIRKACAEHGLRQLDNSEKTYLEGLKLDPNN
gnl|tr|Q4YNL4   ESANKCINIKNDWPKGYIRKACAEHGLRQLDNSEKTYLEGLKLDPNN
gnl|tr|Q51BN3   EDAEKTIEYKPDWSRGYSRKAFALHKLLERYEEAEEVCNSGLKIDPEN
gnl|tr|Q54DA8   TDAKKAIELKPDWSKGYLRETNALYKLGRFEEAEKSAEAGLKIDPTN
gnl|tr|Q561A5   EDAEKTIELDPSFSKGYARKGAALHGLRRFPDAVMAYESGLQAEPNN
gnl|tr|Q57ZX0   SDAEKCVSLKPDWVKGYVRHGAALHGLRRYDEAAAVYKKGLTVDPSS
gnl|tr|Q59YX6   KDAQECVKINPSWAKGYNRIAGAEFGLGNFDQAKSNYEKCLELDPNN
gnl|tr|Q5ARF6   EDAEKAVEIKPDWSKGHQRKGAAYRGIGDLLAAHDAYEEALKLEPGN
gnl|tr|Q5CCL7   EDAEKTVSLHPNWSKGYSRKGSVLAYLSRYEEAIEAYRTGLRLEPTN
gnl|tr|Q5RKM3   KDACQTIKIKPDWGKGYSRKAAALEFLGRLEDAKATYQEGLRQEPSN
gnl|tr|Q5XEP2   SDAKKTVELKPDWGKGYSRLGAAHLGLNQFDEAVEAYSKGLEIDPSN
gnl|sp|Q60864   EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANN
gnl|tr|Q6H660   ADAEKTVELKPDWAKGYSRLGAAHLGLGDAASAVAAYEKGLALDPTN
gnl|tr|Q7RJW7   ESANKCINIKNDWPKGYIRKACAEHGLRQLDNSEKTYLEGLKLDPNN
gnl|tr|Q7SET2   KDAEKTTEIKPDWPKGWGRKGTALFGKGDLLGANDAYEQGLKIDPNN
gnl|tr|Q7ZWU1   EDGSKTVELKADWGKGYSRKAAALEFLNRFEEAKKTYEEGLRHEPTN
gnl|tr|Q84TJ2   SDAKKTVELKPDWGKGYSRLGAAHLGLNQFDEAVEAYSKGLEIDPSN
gnl|tr|Q8ILC1   ESANKCISIKKDWPKGYIRKGCAEHGLRQLSNAEKTYLEGLKIDPNN
gnl|tr|Q8JHF9   EDGSKTVELKADWGKGYSRKAAALEFLNRFEEAKKTYEEGLRHEPTN
gnl|tr|Q8L724   SDAKKTIELKPDWSKGYSRLGAAFIGLSKFDEAVDSYKKGLEIDPSN
gnl|tr|Q9LNB6   SDAKETIKLKPYWPKGYSRLGAAHLGLNQFELAVTAYKKGLDVDPTN
gnl|tr|Q9STH1   SDAKKTIELKPDWSKGYSRLGAAFIGLSKFDEAVDSYKKGLEIDPSN
gnl|sp|Q9USI5   KDATKCTELKPDWAKGWSRKGAALHGLGDLDAARSAYEEGLKHDANN
gnl|tr|Q9VPN5   EDAEKTIQLNPTWPKGYSRKGAAAAGLNDFMKAFEAYNEGLKYDPTN
```

A.3.2. Hop TPR2A domain

```
                 1        10        20        30        40        50
gnl|tr|A0BUJ1   WEVQKNLGNEEYKNKNFENALQYYDAALQLNKEEALLYNNKAAAFIEQTKYDEAL
gnl|tr|A0E7Z9   WEVQKNLGNDEYKNKNFEKALQYYNAALELNKEEALLYNNKAAVFIEQKLYDQAL
gnl|tr|A1CG55   GDAEKKIGNDFYKKKQFDQAIEHYTKAWELN.KDITYLNNIGAAKFEKGDLQGAI
gnl|tr|A1D965   GDAEKKIGNDFYKKKQFDEAIEHYTKAWELN.KDITYLNNIGAAKFEKGDLQGAI
gnl|sp|O35814   ALKEKELGNDAYKKKDFDKALKHYDKAKELDPTNMTYITNQAAVHFEKGDYNKCR
gnl|sp|O54981   ALKEKEMGNEAYKKKDFDMALKHYDRAKELDPTNMTYITNQAAVHFEKGDYNKCR
gnl|tr|O61650   ARKEKELGNAAYKKKDFETALKHYHAAIEHDPTDITFYNNIAAVHFEKREYEECI
gnl|sp|P31948   ALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNQAAVYFEKGDYNKCR
gnl|tr|P90553   ALALKEEGNKLYLSKKFEEALTKYQEAQVKDPNNTLYILNVSAVYFEQGDYDKCI
gnl|tr|P90647   ALEEKELGNQAYKKKDFDTAIVHYKKAFELDPDNMTYLTNLAAVYMEQKNYEECV
gnl|tr|Q017S6   ALEAKDAGNAAYKKRDFDAAIAKYDEAIELDPEDISFLNNRAAANLEKGDFDACI
gnl|tr|Q0CWW9   GDAEKKIGNDFYKKKQFDEAIEHYTKAWELN.KDVTYLNNIGAAKFEKGDLQGAI
gnl|tr|Q0JBE4   AQEEKEAGNAAYK.KDFETAIQHYTKAMELDDEDVSYLTNRAAVYLEMGKYDECI
gnl|tr|Q0UEH1   ADELKKKGTEFYKKRQFDEAIENYTSAWETH.KDIAYKTNEGAARFEKGDYEGCI
gnl|tr|Q1DYZ0   ADKEKNLGTENYKKRQFDAAIEHYSKAWELH.KDITYLTNLSAAYFEKGEYQQTI
gnl|tr|Q22RN3   HEKVKNEGNEYYKSRNFDKALECYNKAIELQPTEILYYNNKAAVYIEQKNYDAAL
gnl|tr|Q27U54   GKKEKEAGNAAYKKKDFDNALNHYTKAMEYDPTDITFYNNIAAVYFERKQYDECI
gnl|tr|Q29NX2   AKLEKELGNAAYKKKEFDDALKHYNAAIEHDPTDITFYNNIAAVYFERKEYEECI
gnl|tr|Q2HEP5   ADKEKALGTENYKKRNFDEAIKHYQAAWDLH.KDITYLNNLGAAHFEKGDYQACI
gnl|tr|Q2U285   GDAEKKIGNEFYKKKQFDEAVQHYEKAWELN.KDITYLNNIGAAKFEKGDLQGAI
gnl|tr|Q3THQ5   ALKEKELGNDAYKKKDFDKAMKHYDRAKELDPTNMTYITNQAAVHFEKGDYNKCR
gnl|tr|Q3ZBZ8   ALREKELGNEAYKKKDFDTALKHYDKAKDLDPTNMTYITNQAAVYFEKGDYGQCR
gnl|tr|Q3ZCU9   ALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNQAAVYFEKGDYNKCR
gnl|tr|Q45KR0   ALKEKELGNTAYKNKDFETALKHYEEAVKHDPANMTYILNQAAVFFEKGELEKCR
gnl|tr|Q4DYD9   ALRKKEEGNALYKQRKFDEALQKYQEALAKDSTNTVYLLNITAVIFEKGEYAACV
gnl|tr|Q4JHN0   ALALKEEGNKLYLSKKFEEALTKYQEAQVKDPKNTLYILNVSAVYFEQGDYDKCI
gnl|tr|Q4MZD6   SNKYKEEGNNFYKQKKFTEALEMYNKAIELDPNNLLLENNKAAVYLEMGDYEKCI
gnl|tr|Q4QI58   ALALKEEGNKLYLSKKFEEALTKYQEAQVKDPNNTLYILNVSAVYFEQGDYDKCI
gnl|tr|Q4R8N7   ALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNQAAVYFEKGDYNKCR
gnl|tr|Q4SG24   ALKEKELGNSAYKTRDFESALKHYEAAIKHDPTNMSYISNKAAVYFEKGEFDKCR
gnl|tr|Q4UBI5   SKKYKEEGNNLYKQKKFAEALEMYNKAIELDPNNLLLENNKAAVYLEMGDYEKCI
gnl|tr|Q4XY67   GDEHKLKGNELYKQKKFEEALKEYDEAIKVNPNDIMYYYNKAAVYLEMKSYEKSI
gnl|tr|Q4YNL4   GEEHKLKGNDFYKQKKFEEALKEYDEAIKVNPNDIMYYYNKAAVYLEMKSYEKCI
gnl|tr|Q51BN3   AQQQKEKGNELYKQKKFNEAMECYDKAIELDPSDLTFKLNKSAVFLEMEKYDECI
gnl|tr|Q54DA8   SQKERDLGNKAYAKKEFEQAIVHYDKAVELDSSDILAMNNKAAVLIEQQKLDEAI
gnl|tr|Q561A5   AEEFKAQGNTSYKARKFDEAIEFYSKAWDLYPKDVTFLTNLSAVYFEQGEYQKCI
gnl|tr|Q57ZX0   ALRAKEEGNALYKQRKFDEALAKYDEASSLDPTNTVYLLNITAVFYEKGEYELCM
gnl|tr|Q59YX6   ADNAKAEGNALYKKRQFDEAIAAYNKAWELH.KDITYLNNRAAAEYEKGDYDAAI
gnl|tr|Q5ARF6   GDAEKKIGNDFYKKKQFDEAIEHYTKAWELN.KDVTYLNNIGAAKFEKGDLQGAI
gnl|tr|Q5CCL7   ALIQKDLGNDCYKKKEFDNAITHYEKAIEFDPTDITFYTNMAAVFFEQKEYEKCI
gnl|tr|Q5RKM3   ALKEKELGNAAYKKKDFATALKHYEEAIKHDPTNMTYLSNQAAVYFEKGDFDKCR
gnl|tr|Q5XEP2   AQKEKELGNAAYKKKDFETAIQHYSTAMEIDDEDISYITNRAAVHLEMGKYDECI
gnl|sp|Q60864   ALKEKELGNDAYKKKDFDKALKHYDRAKELDPTNMTYITNQAAVHFEKGDYNKCR
gnl|tr|Q6H660   ALKEKELGNAAYKKKDFETAIQHYSTAMEIDDEDISYLTNRAAVYIEMGKYDECI
gnl|tr|Q7RJW7   GEEHKLKGNDFYKQKKFEEALKEYDEAIKVNPNDIMYYYNKAAVYLEMKSYEKCI
gnl|tr|Q7SET2   ADKEKALGTECYKKRNFDEAIKHYQAAWDLY.KDIVYLNNLGAAYFEKGDYQACI
gnl|tr|Q7ZWU1   AQKEKELGNEAYKKKDFETALKHYGQARELDPANMTYITNQAAVYFEMGDYSKCR
gnl|tr|Q84TJ2   AQKEKELGNAAYKKKDFETAIQHYSTAMEIDDEDISYITNRAAVHLEMGKYDECI
gnl|tr|Q8ILC1   GDEHKLKGNEFYKQKKFDEALKEYEEAIQINPNDIMYHYNKAAVHIEMKNYDKAV
gnl|tr|Q8JHF9   AQKEKELGNEAYKKKDFETALKHYGQARELDPANMTYITNQAAVYFEMGDYSKCR
gnl|tr|Q8L724   ALKEKGEGNVAYKKKDFGRAVEHYTKAMELDDEDISYLTNRAAVYLEMGKYEECI
gnl|tr|Q9LNB6   AKKEKELGNAAYKKKDFETAIQHYSTAIEIDDEDISYLTNRAAVYLEMGKYNECI
gnl|tr|Q9STH1   ALKEKGEGNVAYKKKDFGRAVEHYTKAMELDDEDISYLTNRAAVYLEMGKYEECI
gnl|sp|Q9USI5   ADQEKQIGNENYKKRNFPVAIEQYKKAWDTY.KDITYLNNLAAAYFEADQLDDCI
gnl|tr|Q9VPN5   ARKEKELGNAAYKKKDFETALKHYHAAIEHDPTDITFYNNIAAVHFERKEYEECI
```

```
                        60         70         80         90        100
gnl|tr|A0BUJ1   EAIEEGLKV.LEVAKLLARKAKIYSLQNKFNEAIQFYEKSLVEDHVQ
gnl|tr|A0E7Z9   ESIEEGLKV.LEVAKLLARKAKVLSLQNKVDEAIQIYEKSLVEDHVQ
gnl|tr|A1CG55   ETCQKAVEEGRELAKAFARIGTAYEKLGDFTQAIEYYHKSLTEHRTP
gnl|tr|A1D965   EICQKAVEEGRELAKAFARIGTAYEKLGDFTQAIEYYHKSLTEHRTP
gnl|sp|O35814   ELCEKAIEVGRENAKAYARIGNSYFKEERYKDAIHFYNKSLAEHRTP
gnl|sp|O54981   ELCEKAIEVGRENAKAYARIGNSYFKEERYKDAIHFYNKSLAEHRTP
gnl|tr|O61650   KQCEKGIEVGRESAKSFARIGNTYRKLENYKQAKVYYEKAMSEHRTP
gnl|sp|P31948   ELCEKAIEVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|P90553   AECEHGIEHGRENAKLMTRNALCLQRQRKYEAAIDLYKRALVEWRNP
gnl|tr|P90647   NTCTEAIEVGRRVSRAFHRKGNAYMKMEKYAEAIDSYNRALTEHRNP
gnl|tr|Q017S6   GDCDAAIEKGRSIAKAMTRKGNALVKQGKLEEAVDQYQRSLTEHRTA
gnl|tr|Q0CWW9   ETCQKAVEEGRELAKAFARIGTAYEKLGDLKQAIEFYNKSLTEHRTP
gnl|tr|Q0JBE4   NDCDKAVERGRELSRALTRKGTALAKIKDYDVAIETYQKALTEHRNP
gnl|tr|Q0UEH1   KACQEAVDYGREVAKAFARIGTAYEKLGDLANAILFYQKAQTEHRTP
gnl|tr|Q1DYZ0   EACEKAISEGREMAKAFGRIGSSYEKLGDLPKAIVNYQKSLTEHRTP
gnl|tr|Q22RN3   ETVELALKVAQDNAKIFARKASILAKQEKYADSLYWYDKSMLEDNNP
gnl|tr|Q27U54   KMCEKGIEIGRENGKAFARIGNSYRKMEDYQQAKVYYEKAMSEHRTP
gnl|tr|Q29NX2   KQCEKGIEVGRENAKSFARIGNTYRKLENYKQAKIYFEKAMSEHRTP
gnl|tr|Q2HEP5   ETCTKAAEEGRSLAKSYARIGTAYEKQGDLAQAIDFYNKSLREHRTP
gnl|tr|Q2U285   EICQKAIEEGREVAKSYTRIGTAYEKLGDLTQAIEYYNKSLTEHRTP
gnl|tr|Q3THQ5   ELCEKAIEVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|Q3ZBZ8   ELCEKAIEVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|Q3ZCU9   ELCEKAIEVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|Q45KR0   ELCEKAIDVGRENAKALARIGNSYFKEEKYKEAVQYFNKSLTEHRTP
gnl|tr|Q4DYD9   EKCEEALEHGRENAKLMTREALCLQRLKRFDEAIALFKKALVEHRNA
gnl|tr|Q4JHN0   AECEHGIEHARENAKLMTRNALCLQKQRKYEAAIDLYKRALVEWRNP
gnl|tr|Q4MZD6   KTCNDAIDRRYDVSKIYNRLAACYTKMERYDDAILCYQKSLIENNTR
gnl|tr|Q4QI58   AECEHGIEHGRENAKLMTRNALCLQRQRKYEAAIDLYKRALVEWRNP
gnl|tr|Q4R8N7   ELCEKAIDVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|Q4SG24   ELCEEAIEVGRENAKALARIGNSYFKEEKYKEAIQYFNKSLTEHRTP
gnl|tr|Q4UBI5   KTCNDAIDRRYDVSKIYNRLAACYTKMEKYDDAISCYQKSLIENNTR
gnl|tr|Q4XY67   ETCIYAIENRYNFAKVYNRLAIGYINIKDYDKALEAYRKSLVEDNNR
gnl|tr|Q4YNL4   ETCIYAIENRYNFAKVYNRLAIGYINIKNYDKAIEAYRKSLVEDNNR
gnl|tr|Q51BN3   KLCNELLDEYKEQAKLFMRIGNAYFKQDKYTEALDFYKKSCTEKRTE
gnl|tr|Q54DA8   ETCKKALEKAQEISKVYTRLGNIYLKKNQLDDAYKAYSSAVLEDKNA
gnl|tr|Q561A5   ETCEKAVEEGRDLAKAYGRIGSSYSKLGDLAQAIKFFQKSLTEHRTP
gnl|tr|Q57ZX0   EKCENALEHGRENAKLMTRQALCLQKLKRFDEAIALFKKALVEHRNP
gnl|tr|Q59YX6   ATCEKAIDEGRDMAKSFARLGNIYLKKDELPEAVKNFEKSLTEHRTP
gnl|tr|Q5ARF6   ETCKNAIEGREHAKAFTRIGTAYEKLGDLDKAIENYNKSLTEHRTP
gnl|tr|Q5CCL7   KECEKAIEIGRENAKAFTRIGNAYKKMEQWKLAKTYFEKSMSEHRTP
gnl|tr|Q5RKM3   ELCEKAIDVGRENAKAYARIGNSYFKQEKYKEAVQFFNKSLTEHRTP
gnl|tr|Q5XEP2   KDCDKAVERGRELAKALTRKGTALGKMKDYEPVIQTYQKALTEHRNP
gnl|sp|Q60864   ELCEKAIEVGRENAKAYARIGNSYFKEEKYKDAIHFYNKSLAEHRTP
gnl|tr|Q6H660   KDCDKAVERGRELSRALTRKGTALAKLKDYDIAIETYQKALTEHRNP
gnl|tr|Q7RJW7   ETCIYAIENRYNFAKVYNRLAIGYINIKNYDKAIEAYRKSLVEDNNR
gnl|tr|Q7SET2   DTCKKAAEEGRARAKSLARIGSAYEKLGDLTNAIEYYNQSLREHRTP
gnl|tr|Q7ZWU1   ELCEKAIEVGRENAKAYARIGNSYFKEEKNKEAIQFFNKSLAEHRTP
gnl|tr|Q84TJ2   KDCDKAVERGRELAKALTRKGTALGKMKDYEPVIQTYQKAITEHRNP
gnl|tr|Q8ILC1   ETCLYAIENRYNFAKLYNRLAISYINMKKYDLAIEAYRKSLVEDNNR
gnl|tr|Q8JHF9   ELCEKAIEVGRENAKAYARIGNSYFKEEKNKEAIQFFNKSLAEHRTP
gnl|tr|Q8L724   EDCDKAVERGRELARALTRKGSALVKMKDFEPAIETFQKALTEHRNP
gnl|tr|Q9LNB6   EDCNKAVERGRELARALTRKGTALTKMKDYEPAIEAFQKALTEHRNP
gnl|tr|Q9STH1   EDCDKAVERGRELARALTRKGSALVKMKDFEPAIETFQKALTEHRNP
gnl|sp|Q9USI5   KTCEDAIEQGRELAKALGRLGTTYQKRGDLVKAIDYYQRSLTEHRTP
gnl|tr|Q9VPN5   KQCEKGIEVGRESAKSFARIGNTYRKLENYKQAKVYYEKAMSEHRTP
```

A.3.3. PP5 TPR domain

```
                                1        10        20        30        40        50
YGR123C/1-2443              ALERKNEGNVFVKEKHFLKAIEKYTEAIDLDSTQSIYFSNRAFAHFKVDNFQSAL
ENSXETP00000039599/1-2443  AEEELKEQANEYFRVKDYDRAVQYYTQAIGLSPDTAIYYGNRSLAYLRTECYGYAL
CG8402-PA/1-2443           AEQYKNQGNEMLKTKEFSKAIDMYTKAIELHPNSAIYYANRSLAHLRQESFGFAL
CG8402-PB/1-2443           AEQYKNQGNEMLKTKEFSKAIDMYTKAIELHPNSAIYYANRSLAHLRQESFGFAL
ENSCINP00000001412/1-2443  AEKFKEEANHLFKDKKYEEAIDLYTKAIEVNPKSAVYHANRSFANLRLENYGFAL
ENSCINP00000001419/1-2443  AEKFKEEANHLFKDKKYEEAIDLYTKAIEVNPKSAVYHANRSFANLRLENYGFAL
ENSMUSP00000003183/1-2443  AEELKTQANDYFKAKDYENAIKFYSQAIELNPGNAIYYGNRSLAYLRTECYGYAL
ENSGACP00000016905/1-2443  AELLKEKANTYFKEKDYDNAIKYYSEALEVNPTNAVYYSNRSLAYLRTECYGYAL
ENSCSAVP00000013266/1-2443 AEKFKEEANQLFKDKKYEEAVELYTRAIEANPKSSVYHANRSFAHLRLENYGFAL
ENSCSAVP00000013267/1-2443 AEKFKEEANQLFKDKKYEEAVELYTRAIEANPKSSVYHANRSFAHLRLENYGFAL
Y39B6A.2/1-2443            AGMIKDEANQFFKDQVYDVAADLYSVAIEIHPT.AVLYGNRAQAYLKKELYGSAL
ENSORLP00000001596/1-2443  AELLKEKANNYFKEKDYENAIKFYSEALELNPSNAIYYSNRSLAYLRTECYGYAL
ENSORLP00000001598/1-2443  AELLKEKANNYFKEKDYENAIKFYSEALELNPSNAIYYSNRSLAYLRTECYGYAL
AAEL005080-PA/1-2443       AEELKSQANEHFKNKDNDKAIQLYTEAIELDGSNAILYANRSFAYLRQEAFGYAL
ENSMMUP00000000309/1-2443  AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
ENSMMUP00000000308/1-2443  AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
ENSPTRP00000019194/1-2443  AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
ENSP00000012443/1-2443     AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
ENSRNOP00000023078/1-2443  AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
ENSDNOP00000005118/1-2443  AEELKTQANDYFKAKDYENAVKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
P53043/1-2443              ALERKNEGNVFVKEKHFLKAIEKYTEAIDLDSTQSIYFSNRAFAHFKVDNFQSAL
P53041/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q60676/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPGNAIYYGNRSLAYLRTECYGYAL
P53042/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q6BRL0/1-2443              AIKLKDEGNAYLKEHRHNYAIDSYTKAIELDPTNAVFYSNRAQVHIKLENYGLAI
Q2GW70/1-2443              ATDLKNQGNKAFAAHDWPTAIDLYTQAIELNSKEPTFWSNRAQAYLKTEAYGFAV
Q4I5W3/1-2443              AVELKNKGNKAFQSGDYPSAVDFYSQAIEKNDKEPTFFTNRAQAYIKTEAYGYAV
Q55WV5/1-2443              ALELKALANKAFKDKNFSKSIDFYTQAIALNPKEPTFWNNRAMSKAKMEEHGGAI
Q0V2H0/1-2443              ATALKNKGNDAFKNQDWPAALDFYTKAIELWDKEPSFYTNRAQANIKLESYGYAI
Q4WU01/1-2443              ATALKVQGNKAFAEHEWPTAVEFYTQAIDKYDREPSFFSNRAQAYIKLEAYGFAI
Q5AJP7/1-2443              ALEWKDKGNNLLKQHKYDEAIEAYTKAIEIDPNNAIFYSNRAQVQIKLENYGLAI
O43049/1-2443              ALELKNEANKFLKEGHIVQAIDLYTKAIELDSTNAILYSNRSLAHLKSEDYGLAI
Q1DUL9/1-2443              ATALKVAGNKAFAKHDWPEALGFYTKAIEKYDRDPSFWCNRAQANIKLEAYGYAI
Q4P3M8/1-2443              AQEHKLKGNEHFSAQRFDAAKHEYTLAIDLDPTIAAFYTNRAASENMLEQYNLAI
Q5KJE3/1-2443              ALELKALANKAFKDKNFSKSIDFYTQAIALNPKEPTFWNNRAMSKAKMEEHGGAI
Q6CFH3/1-2443              AEDLKNQGNKALLSGHYNDAVDLYTQAIELNPSNAVYYSNRAQAHIKNEAYGYAI
Q6CVM5/1-2443              ALELKNEGNKFVKEKLYAKAAEYYTKAIEHDPENTILYSNRAFTNLKLDNFQSSL
Q6FVW4/1-2443              ALEYKNEGNACVKVQDYAKAIEKYDKAIELDDTQSVYFSNRALCHLKLDNFQCAS
Q75EJ6/1-2443              ALDYKNEGNEWVKAKDYARAVEAYTRAIEADGTQSIFFSNRALANLKLDRFQSAL
Q5AJB8/1-2443              ALEWKDKGNNLLKQHKYDEAIEAYTKAIEIDPNNAIFYSNRAQVQIKLENYGLAI
O14428/1-2443              AIAFKNEGNKAFAAHDWPKAIEFYDKAIELNDKEPTFWSNRAQAHLKTEAYGYAI
Q2U919/1-2443              ATALKVQGNKAFGQHEWPTAVDFYTQAIAKYDREPSFFSNRAQAHIKLEAYGFAI
Q53FR0/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q53XV2/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q9BPW0/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q0V8L5/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q0V8M3/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q5R8T2/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q64538/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q68G16/1-2443              AEELKTQANDYFKAKDYENAIKFYSQAIELNPSNAIYYGNRSLAYLRTECYGYAL
Q58EP0/1-2443              AEKLKEKANDYFKDKDYENAIKYYTEALDLNPTNPIYYSNRSLSYLENYGYAL
O42205/1-2443              AEELKEQANEYFRVKDYDHAVQYYTQAIDLSPDTAIYYGNRSLAYLRTECYGYAL
Q68EP0/1-2443              AEELKEQANEYFRVKDYDRAVQYYTQAIGLSPDTAIYYGNRSLAYLRTECYGYAL
Q6GPS6/1-2443              AEELKEQANEYFRVKDYDHAVQYYTQAIDLSPDTAIYYGNRSLAYLRTECYGYAL
Q28EK7/1-2443              AEELKEQANEYFRVKDYDRAVQYYTQAIGLSPDTAIYYGNRSLAYLRTECYGYAL
Q4E5D0/1-2443              ADRLKNKGNEAFQEGKWHHAIELYTEALALH.KTPVILCNRAFAYLKTELAGAAL
Q9NES8/1-2443              AGMIKDEANQFFKDQVYDVAADLYSVAIEIHP.TAVLYGNRAQAYLKKELYGSAL
Q388N2/1-2443              ADKLKQLGNAAFSERKWHLAIDMYTKAIELTK.TPTLFCNRALAELRAELPGAAL
Q54RH6/1-2443              SDEYKAIANKHFSEQKYDLAAEVYTKAIKYHP.TAILYSNRSFSNFKNELYVNAL
Q5CJA8/1-2443              SEQYKIKGNESFKSGKYNEAIEYYTLAIKTSQASHIYYSNRALCHIRLENFGSAI
Q8IDE7/1-2443              CDALKNIGNKYFKENNYIISLRYYTEAIDLIKKSHIYYTNRSFCHIKLENYGTAI
Q4QE27/1-2443              SDRLKQEGNAYFQEKKFQHAVESYSQAIEAHK.THTLLCNRAFAYLKLELPGAAL
Q8WQR3/1-2443              CDALKNIGNKYFKENNYIISLRYYTEAIDLIKKSPIYYTNRSFCHIKLENYGTAI
Q4E1W0/1-2443              ADRLKNKGNEAFQEGKWHHAIELYTHVILCNRAFAYLKTELAGAAL
Q9GPZ6/1-2443              ADKLKQLGNAAFSERKWHLAIDMYTKAIELTK.TPTLFCNRALAELRAELPGAAL
Q22B29/1-2443              AEEFKQKGNDCFKHSKYQEASDFYTKAIDCHSTSPPYYSNRAFCQLKLENYGLAL
Q60TC7/1-2443              AGMIKDEANQFFKDQVYDVAADLYSVAIEIHPTAHVLYGNRAQAYLKKELYGSAL
Q962N7/1-2443              CDALKNIGNKYFKENNYIISLRYYTEAIDLIKKSHIYYTNRSFCHIKLENYGTAI
Q50XR4/1-2443              .MEYRELGNIAFSKQNYKEAIEQYSKAIQEDPLSCVNYTNRALCYLRINENELAL
Q7RFK9/1-2443              CNSLKNIGNKYFKENNYIISLKYYTAAIDFIKSFHIYYTNRSFCNMKLENYGLSI
Q9VH81/1-2443              AEQYKNQGNEMLKTKEFSKAIDMYTKAIELHPNSAIYYANRSLAHLRQESFGFAL
```

```
                                        60         70         80         90        100
                                        .          .          .          .          .
YGR123C/1-2443                  NDCDEAIKLDPKNIKAYHRRALSCMALLEFKKARKDLNVLLKAKPND
ENSXETP00000039599/1-2443       ADASRAIQLDAKYIKGYYRRAASNMALGKLKAALKDYETVVKVRPHD
CG8402-PA/1-2443                QDGVSAVKADPAYLKGYYRRAAAHMSLGKFKQALCDFEFVAKCRPND
CG8402-PB/1-2443                QDGVSAVKADPAYLKGYYRRAAAHMSLGKFKQALCDFEFVAKCRPND
ENSCINP00000001412/1-2443       EDATTAISCDKKYIKAYYRRASAYMSLGKFKLALRDLEAIVKVRPTD
ENSCINP00000001419/1-2443       EDATTAISCDKKYIKAYYRRASAYMSLGKFKLALRDLEAIVKVRPTD
ENSMUSP00000003183/1-2443       GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVRPND
ENSGACP00000016905/1-2443       ADATKALEIDKNYIKGYYRRATSNMALGKFKAALKDYETVVKVRPND
ENSCSAVP00000013266/1-2443      EDASTAIACDKKYIKAYYRRASAYMSLGKFKPALRDLEAVVKVRPTD
ENSCSAVP00000013267/1-2443      EDASTAIACDKKYIKAYYRRASAYMSLGKFKPALRDLEAVVKVRPTD
Y39B6A.2/1-2443                 EDADNAIAIDPSYVKGFYRRATANMALGRFKKALTDYQAVVKVCPND
ENSORLP00000001596/1-2443       ADATKALEVDKNYIKGYYRRATSNMALGKFKAALKDYETVVRVRPND
ENSORLP00000001598/1-2443       ADATKALEVDKNYIKGYYRRATSNMALGKFKAALKDYETVVRVRPND
AAEL005080-PA/1-2443            NDAVQAIKCNPNYLKGYYRRAGAHMALGKYKLALADLELVAKRCPND
ENSMMUP00000000309/1-2443       GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
ENSMMUP00000000308/1-2443       GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
ENSPTRP00000019194/1-2443       GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
ENSP00000012443/1-2443          GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
ENSRNOP00000023078/1-2443       GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
ENSDNOP00000005118/1-2443       ADATRAIEIDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
P53043/1-2443                   NDCDEAIKLDPKNIKAYHRRALSCMALLEFKKARKDLNVLLKAKPND
P53041/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q60676/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
P53042/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q6BRL0/1-2443                   SDCNEALKVDPNMMKAYYRRGISLMAILNYKEAQINFKEILKKMPND
Q2GW70/1-2443                   RDATKAIELKPSFVKAYYRRATAYAAILRPKEAVKDFKTCVKIDPGN
Q4I5W3/1-2443                   ADATKAIELNPKLVKAYYRRGLAKTAILRPKEAIDDFKTCVTLDPNN
Q55WV5/1-2443                   SDATKAVELNPSYAKAFYRRGLSQLAILRPTDAVSDFKKALAIEPGN
Q0V2H0/1-2443                   ADADKAIELDPNNVKAYYRRASANTSMLKHREALRDWKLVIKKAPND
Q4WU01/1-2443                   ADATKALELDPSYVKAYWRRALANTAILNYREALKDFKAVVKKEPNN
Q5AJP7/1-2443                   QDCDLVIKLDINFLKAYYRKGVSLMAILNHKQALENFKFILKKLPND
O43049/1-2443                   NDASKAIECDPEYAKAYFRRATAHIAIFQPKEAVGDFKKALALAPSD
Q1DUL9/1-2443                   ADATKAIELDPSYVKAYWRRAVANTAILNSREALKDFKTVVRKAPND
Q4P3M8/1-2443                   EDANQAIKLDPSYVKAYFRRATAYFKSNNLEAALQDFEHVLVHEPSN
Q5KJE3/1-2443                   SDATKAVELNPSYAKAFYRRGLSQLAILRPTDAVSDFKKALAIEPGN
Q6CFH3/1-2443                   EDSSKAIELDPTYIKAYFRRAVSNTAIIKHKDALVDFKKVVQLAPGD
Q6CVM5/1-2443                   DDAKRAIELDNNNLKAYHRRAMSYIGLLEFRKARKDLIIVLKFKPNE
Q6FVW4/1-2443                   QDCDKALALDPKNVKAYHRRGLACVGLLEFKKARADLTVVLKAKPSD
Q75EJ6/1-2443                   EDSARAIELDAGNVKAYHRRGLAHCGLLEWGPAKRDLEVVVRAKPGD
Q5AJB8/1-2443                   QDCDLVIKLDINFLKAYYRKGVSLMAILNHKQALENFKFILKKLPND
O14428/1-2443                   RDATKAIELNPGFVKAYYRRATAYAAILNPKEAVKDFKTCVKIAPDN
Q2U919/1-2443                   ADATKALELDPAYTKAYWRRALANTAILNYKDALRDFKVVAKREPNN
Q53FR0/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q53XV2/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q9BPW0/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q0V8L5/1-2443                   ADATRAVEMDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q0V8M3/1-2443                   ADATRAVEMDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q5R8T2/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPHD
Q64538/1-2443                   GDATRAIELDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPND
Q68G16/1-2443                   GDATRAIDLDKKYIKGYYRRAASNMALGKFRAALRDYETVVKVKPND
Q58EP0/1-2443                   ADATRALELDKNYLKGYYRRATSNMALGKFKAALKDYETVVRVRPND
O42205/1-2443                   ADASRAIQLDAKYIKGYYRRAASNMALGKLKAALKDYETVVKVRPHD
Q68EP0/1-2443                   ADASRAIQLDAKYIKGYYRRAASNMALGKLKAALKDYETVVKVRPHD
Q6GPS6/1-2443                   ADASRAIQLDAKYIKGYYRRAASNMALGKLKAALKDYETVVKVRPHD
Q28EK7/1-2443                   ADASRAIQLDAKYIKGYYRRAASNMALGKLKAALKDYETVVKVRPHD
Q4E5D0/1-2443                   TDADEALRLDPGYVKAYYRKASAHLYLGKHKEALKDFKTVVQLIPGD
Q9NES8/1-2443                   EDADNAIAIDPSYVKGFYRRATANMALGRFKKALTDYQAVVKVCPND
Q388N2/1-2443                   ADADAALGIEPTFAKAYYHKASAYLSLGKHKQALTNYKKVVDLAPQN
Q54RH6/1-2443                   QDAQTSHEMDPTYIKAYYRLGSAHLALRNFEEAKHFFKELLTKNPKE
Q5CJA8/1-2443                   EDSGESIKCCPSFSKAYYRRGIAYFNLLKYSLARKDFMMVLNLTQND
Q8IDE7/1-2443                   EDIDEAIKINPYYAKAYYRKGCSYLLLSDLKRASECFQKVLKL.TKD
Q4QE27/1-2443                   VDAQEAVEIDPGFVKAYYRKASAHLLLGKFKDAQKEFAAVLKLVPTE
Q8WQR3/1-2443                   EDIDEAIKINPYYAKAYYRKGCSYLLLSDLKRASECFQKVLKL.TKD
Q4E1W0/1-2443                   TDADEALRLDPGYVKAYYRKASAHLYLGKHKEALKDFKTVVQLIPGD
Q9GPZ6/1-2443                   ADADAALGIEPTFAKAYYHKASAYLSLGKHKQALTNYKKVVDLAPQN
Q22B29/1-2443                   EDSKTSIKLDPNFVKGYYREGSAYLALGKLEDARNSFKAAHKLQPKD
Q60TC7/1-2443                   EDADNAISIDPSYVKGFYRRATANMALGRFRKALADYQAVVKVVPND
Q962N7/1-2443                   EDIDEAIKINPYYAKAYYRKGCSYLLLSDLKRASECFQKVLKL.TKD
Q50XR4/1-2443                   ADSFNALQIDPSNPKAWFRHAVASANLKDNKTALLDLRRAMTLSS..
Q7RFK9/1-2443                   QDIDEAIKINPHYAKAYYRKGCSYLLLSDLKNASDCFQMVLKL.TKD
Q9VH81/1-2443                   QDGVSAVKADPAYLKGYYRRAAAHMSLGKFKQALCDFEFVAKCRPND
```

A.3.4. SGT TPR domain

```
                    1        10        20        30        40        50
gnl|tr|A1CQC9   SDKLKSEGNAAMARKEYSKAIDLYTQALSIAPSNPIYLSNRAAAYSASGEHEKAA
gnl|tr|A1D393   SDKLKSEGNAAMARKEYSKAIDLYTQALSIAPANPIYLSNRAAAYSASGQHEKAA
gnl|tr|O13797   AEKLKLEGNNAIAAKDYQKALDLYTKAIEIDPTSPVYYSNRAAAYNQLGQFENAV
gnl|sp|O43765   AERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYSKLGNYAGAV
gnl|sp|O70593   AERLKTEGNEQMKLENFEAAVHLYGKAIELNPANAVYFCNRAAAYSKLGNYVGAV
gnl|tr|Q0CQ57   SDKLKSEGNAAMARKEYTVAIERYTQALAIAPSNPIYLSNRAAAYSASGQHEKAA
gnl|tr|Q0UMT0   AERLKGLGNEAMKKKDYDSAIKHYTAALDIVPLNPIYLSNRAAAYSGQNKHQLAK
gnl|sp|Q12118   AEDLKMQGNKAMANKDYELAINKYTEAIKVLPTNAIYYANRAAAHSSLKEYDQAV
gnl|tr|Q1DHD7   SDRLKSEGNAAMARKDYIGAISFYTKALEIAPANPIYLSNRAAAFSASGNHARAV
gnl|tr|Q1HQM2   AENLKNEGNRLMKEEKYQEALNTYSKAISLDATNPVFYCNRAAAYSRLGDYQAAA
gnl|tr|Q21746   ANKLKEEGNDLMKASQFEAAVQKYNAAIKLN.RDPVYFCNRAAAYCRLEQYDLAI
gnl|tr|Q28H19   AERLKTEGNEQMKLENFESAISYYSKALELNPTNAVYYCNRAAAYSKLGNYAGAV
gnl|tr|Q29LB0   AESIKNEGNRLMKECKYNEALLQYNRAITFDPKNPIFYCNRAAAHIRLGDNERAV
gnl|tr|Q2GZN4   AEALKSKGNAAMAQRDYPAAIDLYTQALAVHPGNAIFLSNRAAAYSAARDHEAAR
gnl|tr|Q2U2H5   SDKLKSEGNAAMARKDYNSAIDLYTKALSIAPSNPIYLSNRAAAYSASGQHEKAA
gnl|tr|Q32LM2   AERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q3TN35   AERLKTEGNEQMKLENFEAAVHLYGKAIELNPANAVYFCNRAAAYSKLGNYVGAV
gnl|tr|Q4CLR4   AEEIKNKGNELMGMAKYKEAIAYYTKSIEMEPENHVFFANRAAAHTHLKDYDSAV
gnl|tr|Q4D5Z5   AEEIKNKGNELMGMAKYKEAIAYYTKSIEMEPENHVFFANRAAAHTHLKDYDSAV
gnl|tr|Q4IFB7   AEALKSKGNAAMAQKDYSAAINFYTQALAINASNAVYLSNRAAAHSANKDHASAR
gnl|tr|Q4P3F4   AEQLKAEGNKAMSAKDYGAAIEAYGKAIELNPNSPVYFSNRAAAFSQIGQHDSAI
gnl|tr|Q4PLZ5   AEKYKQEGNNMMKLEMYTAALECYTKAISLDGNNAVYYCNRAAAHSKLNNHADAI
gnl|tr|Q4Q720   AEQIKNKGNELMSQAKYKEAIAYYTKAIELQPDNAVFFANRAAAHTHLKDYNNAI
gnl|tr|Q4R6F4   AERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q4S298   AEQLKNEGNNHMKEENYRCAVECYTKAIELDLRNAVYYCNRAAAHSKLGNYMEAT
gnl|tr|Q4TAA5   AEALKNKGNDQMKMENFSAAVEFYSKAITVNPHNAVYFCNRAAAHSKLGNYAGAV
gnl|tr|Q4WTC0   SDKLKSEGNAAMARKEYSKAIDLYTQALSIAPANPIYLSNRAAAYSASGQHEKAA
gnl|tr|Q54VG4   AEKLKNEGNAKLNEGKHQEALSCYNKAILYDNTNAIYFANRAATYSALQNFEKSI
gnl|tr|Q560H9   AESLKTKGNQLMGQKLYDSAIEQYTEAIKLDP.NPVYYSNRAAAWGGAGQHEKAV
gnl|tr|Q585Z8   AEEIKNKGNELMGLANYKQAVAYYTKAIEMEPENHVFFANRAAAHTHLKDYRSAI
gnl|tr|Q5A0I8   ADELKVQGNRAMALKDYPEAIAKYTEAIGLDPSNVVYLSNRAAAHSSSQKHDKAV
gnl|tr|Q5BDU8   SDKLKSEGNAAMARKEYSVAIDLYTKALAIAPANPIYLSNRAAAYSASGQPQKAA
gnl|tr|Q5HZM2   AERLKTEGNEQMKVENFESAISYYTKALELNPANAVYYCNRAAAYSKLGNYAGAV
gnl|tr|Q5MAG3   AERLKTEGNEQMKVENFEAAVHFYGKAIELNPANAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q5ZHW6   AERLKTEGNEQMKAENFEAAVSFYGKAIELNPSNAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q5ZJ95   ADRLKDEGNNHMKEENYGAAVDCYTRAIELDPNNAVYYCNRAAAQSKLNKYSEAI
gnl|tr|Q622A6   ANKLKEEGNDLMKASQFDAAVQKYNAAIKLN.RDPVYFCNRAAAYCRLEQYDLAI
gnl|tr|Q6BJ86   ADALKAEGNRAMANKNFSEAIKKYTEAIELDGTNVVYLSNRAAAHSSSQHENAV
gnl|tr|Q6CCH5   ADKLKLEGNKALSQRNFEESIDLYTQAIDIDANNAVYYSNRAAAYSQLQLHDNAI
gnl|tr|Q6CSG7   AEELKLQGNKAMAAKKFDEAIEKYTAAIEVSPSNAVYYSNRAAAYSSLKQYEQAV
gnl|tr|Q6FVD6   AEALKLEGNKAMAGKDFELAIAKYSEAIEVLPTNAIYYANRAAAHSSLKQYDEAV
gnl|tr|Q6GM15   AEQLKDEGNGLMKEQNYEAAVDCYSQAIELDPNNAVYYCNRAAAQSQRGKHSEAI
gnl|tr|Q6GMI8   AEQLKNEGNNHMKEENYSSAVDCYTKAIELDQRNAVYYCNRAAAHSKLENYTEAM
gnl|tr|Q6NTZ8   AESLKTEGNQMKVENFESAVTYYTKALELNPRNAVYYCNRAAAYSKLGNYAGAV
gnl|tr|Q6NXA1   AERLKTDGNDQMKVENFSAAVEFYSKAIQLNPQNAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q6P2W1   AERLKTEGNEQMKLENFESAISYYSKALELNPTNAVYYCNRAAAYSKLGNYAGAV
gnl|tr|Q75CA7   AEALKLEGNRAMAAKDYETAIQKYTAAIEVLPTDAVYYANRAAAYSSLQQYEKAV
gnl|tr|Q7QDC4   AEGLKNEGNRLMKEEKYQEALNTYTKAINLDATNPVFYCNRAAAYSRLGDYVRAA
gnl|tr|Q7YW78   AEALKNQGNQCMKQEKFEEAVACYSKAIELSPYNAVYFCNRAAAHSRLDHHQDAI
gnl|tr|Q7ZUM6   AERLKTDGNDQMKVENFSAAVEFYSKAIQLNPQNAVYFCNRAAAYSKLGNYAGAV
gnl|tr|Q80W98   ADQLKDEGNNHMKEENYAAAVDCYTQAIELDPNNAVYYCNRAAAQSKLSHYTDAI
gnl|sp|Q8BJU0   AERLKTEGNEQMKLENFEAAVHLYGKAIELNPANAVYFCNRAAAYSKLGNYVGAV
gnl|sp|Q8VD33   ADQLKDEGNNHMKEENYAAAVDCYTQAIELDPNNAVYYCNRAAAQSKLSHYTDAI
gnl|sp|Q96EQ0   ADQLKDEGNNHMKEENYAAAVDCYTQAIELDPNNAVYYCNRAAAQSKLGHYTDAI
gnl|tr|Q9VJD4   AESIKNEGNRLMKENKYNEALLQYNRAIAFDPKNPIFYCNRAAAHIRLGENERAV
```

```
                    60         70         80         90        100
                    ·          ·          ·          ·          ·
gnl|tr|A1CQC9   EDAELATVVDPKYSKAWSRLGLARFDMADYHAAKEAYEKGIEAEGNG
gnl|tr|A1D393   EDAELATVVDPKYSKAWSRLGLARFDMADYKGAKEAYEKGIEAEGNG
gnl|tr|O13797   EDALTCLSLDPHHARAFGRLGRAKLSLGDAAAAADAYKKGLDFDPNN
gnl|sp|O43765   QDCERAICIDPAYSKAYGRMGLALSSLNKHVEAVAYYKKALELDPDN
gnl|sp|O70593   QDCERAIGIDPGYSKAYGRMGLALSSLNKHAEAVAYYKKALELDPDN
gnl|tr|Q0CQ57   EDAELATAVDPKYSKAWSRLGLARFDMADYHGAKEAYEKGIEAEGNG
gnl|tr|Q0UMT0   EDAEMAVAADPNYSKAWSRLGLANYVLGDAKGAMEAYKKGMDAEGGG
gnl|sp|Q12118   KDAESAISIDPSYFRGYSRLGFAKYAQGKPEEALEAYKKVLDIEGDN
gnl|tr|Q1DHD7   EDAEVAVAADPKYVKAWSRLGLAKFALGDAKGAAEAYEKGIEAEGNG
gnl|tr|Q1HQM2   DDCRMSLRYDPNYSKAYGRLGLAYSKMNKHEQALDAYQNALRIEPDN
gnl|tr|Q21746   QDCRTALALDPSYSKAWGRMGLAYSCQNRYEHAAEAYKKALELEPNQ
gnl|tr|Q28H19   RDCEEAITIDPNYSKAYGRMGLALSSLNKHAEAVGFYKQALILDPDN
gnl|tr|Q29LB0   TDCKSSLLYNNNYSKAYSRLGVAYSNMGKFNEAEQAYRKAIELEPEN
gnl|tr|Q2GZN4   ADAEAAVAVDPKYTKAWSRLGLARFALGDAKGSMEAYQKGIEYEGNG
gnl|tr|Q2U2H5   EDAELATAVDPKYSKAWSRLGLARFDLADFHGAKEAYEKGIEAEGNG
gnl|tr|Q32LM2   QDCERAICIDPSYSKAYGRMGLALSSLNKHTEAVAYYRKALELDPDN
gnl|tr|Q3TN35   QDCERAIGIDPGYSKAYGRMGLALSSLNKHAEAVAYYKKALELDPDN
gnl|tr|Q4CLR4   IDCERAIAINPNYSKAYSRLGTSLFYQEKYARAVDAFAKASELDPTN
gnl|tr|Q4D5Z5   IDCERAIAINPNYSKAYSRLGTSLFYQEKYARAVDAFAKASELDPTN
gnl|tr|Q4IFB7   SDAESAVSIDPAYTKAWSRLGFARFALGDARGAMEAYDRGIQHEGNG
gnl|tr|Q4P3F4   DDAKQASKIDPKFGKAYSRLGHALFSSGRYQEAVEAYQKGVEVDPSN
gnl|tr|Q4PLZ5   EDCQRALDIDPKYGKAYGRIGLAYASLNEHQRAKECYQKAVELDPEN
gnl|tr|Q4Q720   IDCERAIIINPEYSKSYSRLGTALFYQENYSRAVDAFTKACELDPDN
gnl|tr|Q4R6F4   QDCERAICIDPAYSKAYGRMGLALSSLNKHVEAVAYYKKALELDPDN
gnl|tr|Q4S298   CDCERAIGIDPTYSKAYGRMGLALTAMSKYPEAISYFKKALVLDPEN
gnl|tr|Q4TAA5   QDCERAIGIDPAYSKAYGRMGLALASVNKHSEAVGYYQKALELDPHN
gnl|tr|Q4WTC0   EDAELATVVDPKYSKAWSRLGLARFDMADYKGAKEAYEKGIEAEGNG
gnl|tr|Q54VG4   EDCLEAIKRNPNYGKAYTRMGSAYTSLGKFSEAMEAYNKAIELEPNN
gnl|tr|Q560H9   EDAEKALELDPKFTKAYSRLGHAHFSLGNYSDAVRAYENGLELDPTN
gnl|tr|Q585Z8   IDCERSISICPTYAKAYSRLGTTLFYQENYQRAVDAFSKACELDPTN
gnl|tr|Q5A0I8   EDAEKAIKLDPNFSKAYSRLGLAKYALGDAKGAMEAYKKGLEVEGET
gnl|tr|Q5BDU8   EDAELATSVDPKYSKAWSRLGLARFDLGDYHGAKEAYEHGIEAEGNG
gnl|tr|Q5HZM2   RDCEAAITIDPNYSKAYGRMGLALSSLNKHAEAVGFYKQALVLDPDN
gnl|tr|Q5MAG3   QDCERAICIDPAYSKAYGRMGLALSSLNKHVEAVAYYKKALELDPDN
gnl|tr|Q5ZHW6   RDCERAIGIDPNYSKAYGRMGLALSSLNKHTEAVVYYKKALELDPDN
gnl|tr|Q5ZJ95   KDCERAIAIDPKYSKAYGRMGLALTSVNKYEEAITSYQKALDLDPEN
gnl|tr|Q622A6   QDCRTALALDASYSKAWGRMGLAYSCQNRYEHAAEAYKKALELEPNQ
gnl|tr|Q6BJ86   KDAEKAIELNPKFSKSYSRLGLAKYALGDASAAMKAYEKGLEVEGDK
gnl|tr|Q6CCH5   ADADAAIRVDPNYSKAYASGDAQGALEAYEMGMKVEGDN
gnl|tr|Q6CSG7   KDAEQAIEVDPTYSKGFSRLGFAKYALNKPEEALDAYKKVLDIEGEK
gnl|tr|Q6FVD6   VDAEKAIEIDPAYSKGYSRLGFAKYAQNKPEEALEAYKKVMDLEGDK
gnl|tr|Q6GM15   TDCEKAISIDAKYSKAYGRMGRALVAMSRYKEAFESYQKALDLDPEN
gnl|tr|Q6GMI8   GDCERAIAIDPSYSKAYGRMGLALTSMSKYPEAISYFNKALVLDPEN
gnl|tr|Q6NTZ8   RDCEEAISIDPSYSKAYGRMGLALSSLNKHAESVGFYKQALVLDPEN
gnl|tr|Q6NXA1   QDCERAIGIDANYSKAYGRMGLALASLNKYSEAVSYYKKALELDPDN
gnl|tr|Q6P2W1   RDCEEAITIDPNYSKAYGRMGLALSSLNKHAEAVGFYKQALILDPDN
gnl|tr|Q75CA7   EDAEKATKVDSSYSKGFSRLGYAKYALGRHEEALEAYKRVLDIEGDN
gnl|tr|Q7QDC4   DDCRMALRHDPNYSKAWGRLGLAYSKMNEHKQAVTAYQNAIRLEPDN
gnl|tr|Q7YW78   NDCLKALEIDPYYSKAYGRMGIAYSSIGNHAKAVECYRKGLELDPNN
gnl|tr|Q7ZUM6   QDCERAIGIDANYSKAYGRMGLALASLNKYSEAVSYYKKALELDPDN
gnl|tr|Q80W98   KDCEKAIAIDSKYSKAYGRMGLALTAMNKFEEAVTSYQKALDLDPEN
gnl|sp|Q8BJU0   QDCERAIGIDPGYSKAYGRMGLALSSLNKHAEAVAYYKKALELDPDN
gnl|sp|Q8VD33   KDCEKAIAIDSKYSKAYGRMGLALTAMNKFEEAVTSYQKALDLDPEN
gnl|sp|Q96EQ0   KDCEKAIAIDSKYSKAYGRMGLALTALNKFEEAVTSYQKALDLDPEN
gnl|tr|Q9VJD4   TDCKSALVYNNNYSKAYCRLGVAYSNMGNFEKAEQAYAKAIELEPDN
```

A.4.   HMMer search results

Output for profile HMM search with Hsp70/Hsp90 interacting TPR model against *C. elegans* protein database. Table of hits with an E-value cutoff of 0.1 listed first followed by alignment of hits to model. Carboxylate-clamp residues highlighted in bold in alignments.

```
Scores for complete sequences (score includes all domains):
Sequence       Location       Score    E-value  N
--------       --------       -----    -------  ---
R09E12.3       V:773389       252.4    2.8e-72  2 KNNKR KNNKR * hop
R05F9.10       II:4902284     166.5      2e-46  1 KNNKR * sgt-1
Y39B6A.2       V:19190339     156.4    2.2e-43  1 KNNKR * pph-5
F30H5.1        III:491547     126.2    2.7e-34  1 RNNKR * unc-45
T09B4.10       I:6181318      109.7    2.6e-29  1 NKNKF * chn-1
C33H5.8        IV:7778004     103.8    1.5e-27  1 KNNKR * ?
F31D4.3        V:20841374      97.2    1.4e-25  1 KTNKR * fkb-6
C17G10.2       II:5594706      95.2    5.7e-25  1 KNNKR * cns-1
C34B2.5        I:10675171      84.7    8.6e-22  1 KNNKR * ttc1?
C56C10.10      II:6592449      76.2      3e-19  1 RNNKR * aip
T12D8.8        III:1361457     73.4    2.1e-18  1 RQKQF   hip
K04G7.3        III:7145552     67.3    1.4e-16  3 SNNCD INNDN NNNDN ogt-1
ZK370.8        III:8752004     48.1    8.5e-11  1 KNNKR * tom70
Y22D7AL.9      III:1606834     45.6    4.8e-10  1 HSNKR * ttc28?
C55B6.2        X:7194513       37.4    7.7e-08  1 LSRGQ   dnj-7
F52H3.5        II:10030120     24.2    8.8e-07  1 EVNKQ
Y73E7A.9       I:1610391       20.5    1.7e-06  1 RSNKR * adp
Y54E10BL.4     I:2996407       17.9    2.8e-06  1 YNRGQ   dnj-28
Y41G9A.1       X:2984269       11.0    9.9e-06  1 VNNQQ
F38B6.6        X:6684417        3.9    3.7e-05  1 YKNVN
C34C6.6a       II:8704748      -4.5    0.00017  1 NVRRN
Y110A7A.17a    I:5123919       -6.9    0.00027  1 CNLRG
F10C5.1.1      III:475452     -10.0    0.00047  1 CNLRG
T25F10.5       V:6762287      -12.8     0.0008  1 ATNTN
F32D1.3        V:4349598      -13.7    0.00094  1 KNNDG
T20B12.1       III:7386548    -16.7     0.0016  1 RHNEN
C18C4.10d.1    V:5575222      -21.0     0.0036  1 NVNKN
M7.2           IV:11084077    -21.9     0.0042  1 NINKN
* denotes conservation of carboxylate clamp and liklihood to interact
  with Hsp70 or Hsp90


Alignments of top-scoring domains:
R05F9.10: domain 1 of 1, from 105 to 205: score 166.5, E = 2e-46
                *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                   A++lKeeGN+++K+ ++e+A++kY+ Ai+l+  d+vy++NRAA+y +
     R05F9.10   105   ANKLKEEGNDLMKASQFEAAVQKYNAAIKLN-RDPVYFCNRAAAYCR 150

                   LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                   L++yd Ai+DC+ AL+ldp+++KA+ R+G+Ay ++++ye A e+++kale
     R05F9.10   151 LEQYDLAIQDCRTALALDPSYSKAWGRMGLAYSCQNRYEHAAEAYKKALE 200

                   ldPnn<-*
                   l+Pn+
     R05F9.10   201 LEPNQ     205


Y39B6A.2: domain 1 of 1, from 29 to 129: score 156.4, E = 2.2e-43
                *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                   A  +K+e N++FK++ y+ A ++Y+ Aie +p  av+y NRA++ylk
     Y39B6A.2    29   AGMIKDEANQFFKDQVYDVAADLYSVAIEIHP-TAVLYGNRAQAYLK 74

                   LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                   + y++A+eD++ A+++dp++vK++yRr++A++aLg++++Al d+q++++
     Y39B6A.2    75 KELYGSALEDADNAIAIDPSYVKGFYRRATANMALGRFKKALTDYQAVVK 124

                   ldPnn<-*
                   + Pn+
     Y39B6A.2   125 VCPND     129
```

```
R09E12.3: domain 2 of 2, from 140 to 241: score 142.3, E = 3.9e-39
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     A+e K++GNeyFK+++y  A ++Y++A++ +p++a++ysNRAAc  k
     R09E12.3    140   AQEEKNKGNEYFKKGDYPTAMRHYNEAVKRDPENAILYSNRAACLTK 186



                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     L ++++A++DC+  + ld++++K++ R++ ++ a+ ++ +A  +++ al+
     R09E12.3    187 LMEFQRALDDCDTCIRLDSKFIKGYIRKAACLVAMREWSKAQRAYEDALQ 236


                     ldPnn<-*
                     +dP n
     R09E12.3    237 VDPSN     241


F30H5.1: domain 1 of 1, from 8 to 114: score 126.2, E = 2.7e-34
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptd.....avyysNRA
                     Aee+++eGN++ K+++y +A e+Yt+A++l  ++++  ++v+y NRA
      F30H5.1      8   AEEIRDEGNAAVKDQDYIKADELYTEALQLTTDEdkalrPVLYRNRA 54


                     AcylkLgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledf
                     ++ lk +++ A  DCtkALe d  +vKAl+Rr +A+++Lg+   A +d
      F30H5.1     55 MARLKRDDFEGAQSDCTKALEFDGADVKALFRRSLAREQLGNVGPAFQDA 104


                     qkaleldPnn<-*
                     ++al l Pn+
      F30H5.1    105 KEALRLSPND     114


R09E12.3: domain 1 of 2, from 5 to 113: score 110.1, E = 1.9e-29
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     A + K+ GN+++K+k++e+A  +Y+KAiel+p+++++y+N+AA+y++
     R09E12.3      5   AIAEKDLGNAAYKQKDFEKAHVHYDKAIELDPSNITFYNNKAAVYFE 51


                     LgnydkAieDCtkALeldpn.......nvKAlyRrGqAylaLgkyeeAle
                     +++ ++++ C+kA+e++++++ + +  +KA  R G+A+++++++  A++
     R09E12.3     52 EKKFAECVQFCEKAVEVGREtradyklIAKAMSRAGNAFQKQNDLSLAVQ 101


                     dfqkaleldPnn<-*
                     +f ++l + +++
     R09E12.3    102 WFHRSLSEFRDP     113


T09B4.10.1: domain 1 of 1, from 5 to 103: score 109.7, E = 2.6e-29
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     Ae++   G +++ +k+y++A+++Y+KAi+++p  + yy NRA+cy++
    T09B4.10.1     5   AEQHNTNGKKCYMNKRYDDAVDHYSKAIKVNP-LPKYYQNRAMCYFQ 50


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     L+n +   eDC++ALel pn vK ly +G+ +l+ +ky eA+ ++ ka
    T09B4.10.1    51 LNNLKMTEEDCKRALELSPNEVKPLYFLGNVFLQSKKYSEAISCLSKA-- 98


                     ldPnn<-*
                     l+ n
    T09B4.10.1    99 LYHNA     103


C33H5.8: domain 1 of 1, from 8 to 108: score 103.8, E = 1.5e-27
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     A+ lKe+GNe+FK+kky +A   Y+K +e  p d++ +sNRA++ l
      C33H5.8      8   AQRLKEQGNEAFKKKKYHKAMTIYSKSLEHWP-DPIVFSNRAQAGLN 53


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                        A  DCt+AL ld++ +KA+yRr+qA+ aL+ ye A  d +++ +
      C33H5.8     54 ADLPLLAQIDCTAALNLDSTAAKAYYRRAQAFKALELYELAERDMKTCFK 103


                     ldPnn<-*
                     +  ++
      C33H5.8    104 YSNDP     108
```

```
F31D4.3.2: domain 1 of 1, from 254 to 370: score 97.2, E = 1.4e-25
               *->AeelKeeGNeyFKekkyeeAiekYtKAiellptd.............
                  A++ K++G+ y ++++ + A  kY++A e+l++++++++++  ++++
    F31D4.3.2   254    AKQAKDRGTMYLQKGNLKLAYNKYKRAEEVLEYEkstdpekmaeret 300


                  ..avyysNRAAcylkLgnydkAieDCtkALeldpnnvKAlyRrGqAylaL
                    +y+N++++ +k+++  ++i++C+k+Le +p nvKAlyR+++A+l +
    F31D4.3.2   301 ilNGAYLNLSLVCSKQNEQLECIKWCDKVLETKPGNVKALYRKATALLTM 350


                  gkyeeAledfqkaleldPnn<-*
                  ++  +A++ f+k++e++P+n
    F31D4.3.2   351 NEVRDAMKLFEKIVEVEPEN    370




C17G10.2: domain 1 of 1, from 95 to 200: score 95.2, E = 5.7e-25
               *->AeelKeeGNeyFKekkyeeAiekYtKAiellptd....avyysNRAA
                  Ae +KeeGN++FK kky  A ++Y+ +i+ ++ d++ +av+y NRAA
     C17G10.2    95    AEHHKEEGNKHFKFKKYRWATDCYSNGIKENSPDrklnAVLYFNRAA 141


                  cylkLgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfq
                  ++ +Lgn ++Ai+DC+    + dp++ K+  R ++++l+L+  ++Al++++
     C17G10.2   142 AQKHLGNLRSAIKDCSMGRKFDPTHLKGVIRGAECLLELEYAKDALNWIE 191


                  kaleldPnn<-*
                   + ++  +
     C17G10.2   192 SSKKIFAFT    200

C34B2.5: domain 1 of 1, from 18 to 125: score 84.7, E = 8.6e-22
               *->AeelKeeGNeyFKekkyeeAiekYtKAiellptd.....avyysNRA
                  + lK eGN++F ++ +e+A ekY++Ai+ +p  +++ +++++sN A
      C34B2.5    18    VDSLKKEGNNFFANGEFEKANEKYQEAIASCPPTstevqSILLSNSA 64


                  AcylkLgnydkAieDCtkALeldpnnvKAlyRrGqAylaL.gkyeeAled
                  A+ +kL ++++A+e+++k++e++++n+KAl Rr+ Ay ++++kye  +ed
      C34B2.5    65 AALIKLRKWESAVEAASKSIEIGATNEKALERRAFAYSNMsEKYENSIED 114


                  fqkaleldPnn<-*
                  +++  e  P+
      C34B2.5   115 YKQLQESLPKR    125


C56C10.10: domain 1 of 1, from 191 to 310: score 76.2, E = 3e-19
               *->AeelKeeGNeyFKekkyeeAiekYtKAi.................e
                  e+l+++GNe+F +k+y+eAi+ Y A+++ ++  +++++++  e
    C56C10.10   191    VEALRQKGNELFVQKDYKEAIDAYRDALtrldtlilrekpgepewvE 237


                  llptdavyysNRAAcylkLgnydkAieDCtkALeldpnnvKAlyRrGqAy
                  l+ +++ +y N ++cyl  g+  +A e  +++L+   +n+KAl+Rr++A+
    C56C10.10   238 LDRKNIPLYANMSQCYLNIGDLHEAEETSSEVLKREETNEKALFRRAKAR 287


                  laLgkyeeAledfqkaleldPnn<-*
                   a  k++eA ed++  l  +P
    C56C10.10   288 IAAWKLDEAEEDLKLLLRNHPAA    310


T12D8.8.1: domain 1 of 1, from 115 to 216: score 73.4, E = 2.1e-18
               *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                  A e + +  e+F +++++ A+ + t Aie +p  a ++  RA + lk
    T12D8.8.1   115    ASEERGKAQEAFSNGDFDTALTHFTAAIEANPGSAMLHAKRANVLLK 161


                  LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                  L+  +Ai+DC+kA+ ++p+ + ++  rG A+  Lgk+ eA+ d+ +a++
    T12D8.8.1   162 LKRPVAAIADCDKAISINPDSAQGYKFRGRANRLLGKWVEAKTDLATACK 211


                  ldPnn<-*
                  ld +
    T12D8.8.1   212 LDYDE    216
```

```
ZK370.8: domain 1 of 1, from 44 to 152: score 48.1, E = 8.5e-11
                 *->AeelKeeGNeyFKekkyeeAiekYtKAiel......lptdavyysNR
                    ee+K  GN  FKek+y+ A+e  tK++e  ++++++   a +y NR
     ZK370.8     44  LEEIKALGNLKFKEKQYDSALEAFTKGVEKagpnssDQIVAMLYQNR 90


                    AAcylkLgn.ydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAle
                    AAc  k g +    ++DC +AL++d +++KA+ R ++A+   gk  +Al
     ZK370.8     91  AACREKVGHsPFDILNDCMAALKVDKKYTKAYLRAAKALNDVGKKQDALA 140


                    dfqkaleldPnn<-*
                    ++ +a +ld
     ZK370.8    141  YLLAAFTLDSSL     152




Y22D7AL.9: domain 1 of 1, from 5 to 106: score 45.6, E = 4.8e-10
                 *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                    e++  e  ++ +++y+eA e+Y KA++ +p++ +++ N++A  lk
   Y22D7AL.9      5  LEKVVHEAGSAYSDGRYQEARELYEKALRDHPKNGILHANLSAILLK 51


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                    +   +A++ ++ +++l p  +KA+yR G+A+ aLg +++ +   +  ++
   Y22D7AL.9     52  IQLPPEALKHAEISVKLCPQWAKAYYRQGEAQRALGFLKKSIYSYCNGIR 101


                    ldPnn<-*
                    ldP
   Y22D7AL.9    102  LDPAG     106


C55B6.2: domain 1 of 1, from 26 to 127: score 37.4, E = 7.7e-08
                 *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                    +++ e G ++    ++ +A+ +Y Aiel+p+   +  RA yl
     C55B6.2     26  VAKHLELGSQFLARAQFADALTQYHAAIELDPKSYQAIYRRATTYLA 72


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                    +g  ++Ai D +++Lel+p++  A   rG+ +l++g++e A  df+ +l
     C55B6.2     73  MGRGKAAIVDLERVLELKPDFYGARIQRGNILLKQGELEAAEADFNIVLN 122


                    ldPnn<-*
                     d  n
     C55B6.2    123  HDSSN     127


K04G7.3a: domain 3 of 3, from 431 to 532: score 29.9, E = 3.1e-07
                 *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                    A+  +  N    ++k e+A ++Y KA+e  p+ a+++sN+A     +
   K04G7.3a    431  ADSQNNLANIKREQGKIEDATRLYLKALEIYPEFAAAHSNLASILQQ 477


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                    +g+   Ai  ++A+ + p+++ A+  +G+ + ++g+   A+ ++++a++
   K04G7.3a    478  QGKLNDAILHYKEAIRIAPTFADAYSNMGNTLKEMGDSSAAIACYNRAIQ 527


                    ldPnn<-*
                    ++P
   K04G7.3a    528  INPAF     532


K04G7.3a: domain 2 of 3, from 329 to 430: score 28.2, E = 4.2e-07
                 *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                    +++ + GN +  ++ ++ A+  Y +A+ l  + av + N+A +y +
   K04G7.3a    329  LDAYINLGNVLKEARIFDRAVSAYLRALNLSGNHAVVHGNLACVYYE 375


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                    +g d Ai+ +kA++l p + A+ +++A+ + g  eA + + kale
   K04G7.3a    376  QGLIDLAIDTYKKAIDLQPHFPDAYCNLANALKEKGSVVEAEQMYMKALE 425


                    ldPnn<-*
                    l P+
   K04G7.3a    426  LCPTH     430
```

```
F52H3.5: domain 1 of 1, from 43 to 148: score 24.2, E = 8.8e-07
                    *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                       +  +l  eG ++     +  +eAiek tKA+e++p+++++y+NRA++y
        F52H3.5   43    SLQLEREGVALAEGVRLDEAIEKFTKALEVCPKNPSAYNNRAQAYRL 89


                    LgnydkAieDCtkALeldpnnvK....AlyRrGqAylaLgkyeeAledfq
                    +++ +kA++D ++AL l    +K+  +A+  r++  y   g+ ++A  df
        F52H3.5   90  QNKPEKALDDLNEALSLAGPKTKtacqAYVQRASIYRLRGDDDKARTDFA 139


                    kaleldPnn<-*
                     a el
        F52H3.5  140  SAAELGSSF     148


Y73E7A.9: domain 1 of 1, from 282 to 396: score 20.5, E = 1.7e-06
                    *->AeelKeeGNeyFKekkyeeAiekYtKAiellpt..........davy
                         l+e+G  +  ek++ +Ai++Y+  i  ++++++++++   +v+
       Y73E7A.9  282    YPDLREIGSTAIREKHFAKAIDFYSDLIYRNDDreshqdhrafLSVC 328


                    ysNRAAcyl...kLgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgk
                    +sNRA + l +++ g+  +++ DC kALe+ + n+KAl R+ +++  ++
       Y73E7A.9  329  HSNRATALLlrrQRGDTYACVRDCIKALEIHRGNSKALLRLIKSFTTMEH 378


                    yeeAledfqkaleldPnn<-*
                     A ++ qk  e +Pn+
       Y73E7A.9  379  IGLARKCVQKFKEWYPND     396


Y54E10BL.4: domain 1 of 1, from 24 to 125: score 17.9, E = 2.8e-06
                    *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                       A+   e GN++F +++y +A+ +Y KAiel+pt  +   RA yl
      Y54E10BL.4   24   AQREYEAGNALFVNRQYSDALTHYHKAIELNPTMYQAIFRRATTYLA 70


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     g +  ++D + +L  +p+++ A   r++ +l++g +e A  df+  +
      Y54E10BL.4   71  FGRSKPGLADLDTVLSQKPDFAGARQQRASVLLKMGQLERAAADFRYLID 120


                    ldPnn<-*
                      +
      Y54E10BL.4  121  HSASQ     125


Y41G9A.1: domain 1 of 1, from 473 to 574: score 11.0, E = 9.9e-06
                    *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                       A +  ++GN ++ +++ ++A+  Y +A+  + +++ ++ N+ +
       Y41G9A.1  473    AHAQVNQGNIAYMNGDLDKALNNYREALNNDASCVQALFNIGLTAKA 519


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                    +gn ++A+e   k  +  nnv l  +++ y++L++  +A+e + +a
       Y41G9A.1  520  QGNLEQALEFFYKLHGILLNNVQVLVQLASIYESLEDSAQAIELYSQANS 569


                    ldPnn<-*
                    l Pn+
       Y41G9A.1  570  LVPND     574


K04G7.3a: domain 1 of 3, from 193 to 294: score 9.2, E = 1.4e-05
                    *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                       Ae++ + GN y  +++ ++A+e Y+ A++l+p+ + +y N+AA+
       K04G7.3a  193    AEAYSNLGNYYKEKGQLQDALENYKLAVKLKPEFIDAYINLAAALVS 239


                    LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     g+ ++A+ +   AL ++p+       +G+ + a+g++eeA+ ++ ka+e
       K04G7.3a  240  GGDLEQAVTAYFNALQINPDLYCVRSDLGNLLKAMGRLEEAKVCYLKAIE 289


                    ldPnn<-*
                     P+
       K04G7.3a  290  TQPQF     294
```

```
F38B6.6: domain 1 of 1, from 432 to 533: score 3.9, E = 3.7e-05
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     A+ + + G  +  ++ ++A + Y  Ai+l+p+   +++N+      k
      F38B6.6   432    AKIHYNLGKVLGDNGLTKDAEKNYWNAIKLDPSYEQALNNLGNLLEK 478


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     g+ + A   +A+ l p+++ A+  +G ++++L+ky eA + ++ +l
      F38B6.6   479 SGDSKTAESLLARAVTLRPSFAVAWMNLGISQMNLKKYYEAEKSLKNSLL 528


                     ldPnn<-*
                     + Pn
      F38B6.6   529 IRPNS    533


C34C6.6a: domain 1 of 1, from 360 to 461: score -4.5, E = 0.00017
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     +    G  y ++++  A++ + Ai+ +ptda +++ + A
    C34C6.6a   360    PDLQNALGVLYNLNRNFARAVDSLKLAISKNPTDARLWNRLGATLAN 406


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     +Ai + ++AL+l p++v A y +G ++++L  y+eAl+ f  ale
    C34C6.6a   407 GDHTAEAISAYREALKLYPTYVRARYNLGISCMQLSSYDEALKHFLSALE 456


                     ldPnn<-*
                     l
    C34C6.6a   457 LQKGG    461




Y110A7A.17a: domain 1 of 1, from 561 to 662: score -6.9, E = 0.00027
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     +     GN +  ++++ +Aie+  +Ai+l++  a +y+ +     +
   Y110A7A.17   561    PQSWCAAGNCFSLQRQHTQAIECMERAIQLDKRFAYAYTLLGHELIV 607


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     + + dkA    + AL l p+  A+y +G+ +l++     Al +qka+
   Y110A7A.17   608 QDELDKAAGSFRSALLLSPRDYRAWYGLGLVHLKKEQNLTALTNIQKAVN 657


                     ldPnn<-*
                     ++P+n
   Y110A7A.17   658 INPTN    662


F10C5.1.1: domain 1 of 1, from 400 to 501: score -10.0, E = 0.00047
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     e      N+   ++ e Ai++ ++A++l+p a+++ +   +++
   F10C5.1.1   400    WETCCIVANYHAIRRDSEHAIKFFQRALRLNPGLAALWVLIGHEFME 446


                     LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                     ++n +A   ++A+e+dp +  ++y +Gq y  ++   Al ++q+a +
   F10C5.1.1   447 MKNNAAACVSYRRAIEIDPADHRGWYGLGQMYDIMKMPAYALFYYQEAQK 496


                     ldPnn<-*
                     P +
   F10C5.1.1   497 CKPHD    501


T25F10.5: domain 1 of 1, from 318 to 421: score -12.8, E = 0.0008
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     e++    + y+  +k e A ++Y ++++   + + +++N+ +c +
    T25F10.5   318    IEAIACVATTYYYGGKPELAMRYYRRILQMGVSSPELFLNIGLCCMA 364


                     LgnydkAieDCtkALeldpnnvKA..lyRrGqAylaLgkyeeAledfqka
                     +++d A+   +A   ++v A+ +y +Gq + +g++  A  f+ a
    T25F10.5   365 AQQFDFALSSILRAQSTMTDDVAAdvWYNIGQILVDIGDLVSAARSFRIA 414


                     leldPnn<-*
                     l  dP+
    T25F10.5   415 LSHDPDH   421
```

```
F32D1.3: domain 1 of 1, from 625 to 720: score -13.7, E = 0.00094
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     +  lK + N+ F   +            KA  l p+ ++ + N+A++ ++
      F32D1.3   625     MAHLKIRQNRSFEVENLLR------KAMTLAPESVTVLQNIALAEFH 665


                  LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                  ++ny + +    +kAL ldp++   l  +++ +++  ++  e   +++k++e
      F32D1.3   666 MQNYNRSLLFYRKALHLDPTHLDSLQGIANLLQQTQNHVESETFYRKVME 715


                  ldPnn<-*
                   Pn
      F32D1.3   716 AQPNS     720


T20B12.1: domain 1 of 1, from 467 to 568: score -16.7, E = 0.0016
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiellptdavyysNRAAcylk
                     A +++  G  +  +kk+eeA ++  + +el p +   + N   c  k
     T20B12.1   467    ARAHRSLGHLLLMDKKFEEAYKHLRRSLELQPIQLGTWFNAGYCAWK 513


                  LgnydkAieDCtkALeldpnnvKAlyRrGqAylaLgkyeeAledfqkale
                  L+n+++  ++  + + l p++  A+ +  Ay   g    +A + +q+al+
     T20B12.1   514 LENFKESTQCYHRCVSLQPDHFEAWNNLSAAYIRHGQKPKAWKLLQEALK 563


                  ldPnn<-*
                  ++ +
     T20B12.1   564 YNYEH     568


C18C4.10d.1: domain 1 of 1, from 290 to 407: score -21.0, E = 0.0036
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiel........lptdavyys
                     A+  l +   + K +k+++A + ++A+e +++  ++++p+  a  ++
     C18C4.10d.   290    AATLNNLAVLFGKRGKFKDAEPLCKRALEIrekvlgddHPDVAKQLN 336


                  NRAAcylkLgnydkAieDCtkALel........dpnnvKAlyRrGqAyla
                  N+A+     +g+y+++ +  ++ALe+ +++  +++dpn +K    + +Ayl+
     C18C4.10d.   337 NLALLCQNQGKYEEVEKYYKRALEIyesklgpdDPNVAKTKNNLSSAYLK 386


                  LgkyeeAledfqkaleldPnn<-*
                  +gky+eA e ++++l+ + +
     C18C4.10d.   387 QGKYKEAEELYKQILTRAHER     407


M7.2: domain 1 of 1, from 243 to 360: score -21.9, E = 0.0042
                  *->AeelKeeGNeyFKekkyeeAiekYtKAiel........lptdavyys
                     A+  l      y  +  ++++A ++  KA++ + +  ++++ + a+ ++
         M7.2   243    ATMLNVLAIVYRNQENFKDAAIYLEKALSIrvqccgenHHSVAATLN 289


                  NRAAcylkLgnydkAieDCtkALeld........pnnvKAlyRrGqAyla
                  N+A +y k g+y++     C++ALe+ ++   ++++p+ +K l  +G   ++
         M7.2   290 NLAIAYGKRGKYKESEPLCKRALEIRknllgpnhPDVAKQLTNLGIVTQQ 339


                  LgkyeeAledfqkaleldPnn<-*
                  L+kyee  ++f++al ++
         M7.2   340 LEKYEETENYFKQALSIYNRA     360
```